Modern radio engineering and telecommunication systems

Современные радиотехнические и телекоммуникационные системы

UDC 621.391.8
https://doi.org/10.32362/2500-316X-2026-14-4-82-95
EDN VLPBLV

CC BY

RESEARCH ARTICLE

Hybrid neural network architecture ResBiLSTM-BiGRU for efficient noise suppression in radio channels with digitally modulated signals

Nguyen Van Dung @**,**
Phan Trong Hanh

Le Quy Don Technical University, Hanoi, Vietnam
@ *Corresponding author, e-mail: nguyenvandungvtdt@lqdtu.edu.vn*

• **Submitted:** 24.11.2025 • **Revised:** 24.01.2026 • **Accepted:** 25.05.2026

Abstract
Objectives. The work sets out to develop an effective method for suppressing additive Gaussian noise in radio channels to ensure the high-quality reconstruction of digitally modulated signals at low signal-to-noise ratio (SNR) levels.
Methods. To address this problem, a hybrid architecture is proposed that combines Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) recurrent neural networks. The ResBiLSTM-BiGRU model enables more stable training and improved reconstruction quality, as well as more accurate modeling of the temporal structure of signals. The performance of the model was evaluated using a multi-modulation dataset with various input SNR levels.
Results. As confirmed by the experimental results, the proposed model achieves an average improvement of 2–3 dB in output SNR compared with the Relativistic Average Generative Adversarial Network (RaGAN) under identical simulation conditions, as well as a more than a 10-dB improvement as compared with the wavelet-transform-based method. The model is characterized by lower mean squared error values and higher correlation coefficients across the entire input SNR range from −10 to 10 dB. The peak SNR values also indicate high signal reconstruction quality, especially at low input SNR levels. Analysis of architectural parameters shows that the best performance is achieved with 128 hidden units and a deep ResBiLSTM-BiGRU architecture.
Conclusions. The obtained results confirm the effectiveness and practical applicability of the proposed ResBiLSTM-BiGRU hybrid architecture for suppressing additive Gaussian noise in radio channels and reconstructing digitally modulated signals, particularly under low-SNR conditions.

Keywords: cognitive radio networks, automatic modulation classification, digitally modulated signals, deep learning, hybrid ResBiLSTM-BiGRU network, noise suppression, signal-to-noise ratio, mean squared error, correlation coefficient, peak signal-to-noise ratio

© Nguyen Van Dung, Phan Trong Hanh, 2026

For citation: Nguyen V.D., Phan T.H. Hybrid neural network architecture ResBiLSTM-BiGRU for efficient noise suppression in radio channels with digitally modulated signals. *Russian Technological Journal.* 2026;14(4):82–95. https://doi.org/10.32362/2500-316X-2026-14-4-82-95, https://elibrary.ru/VLPBLV

Financial disclosure: The authors have no financial or proprietary interest in any material or method mentioned.

The authors declare no conflicts of interest.

НАУЧНАЯ СТАТЬЯ

Гибридная архитектура нейронной сети ResBiLSTM-BiGRU для эффективного подавления шума в радиоканале с цифровыми модулированными сигналами

В.З. Нгуен @,
Ч.Х. Фан

Вьетнамский государственный технический университет им. Ле Куй Дона, Ханой, Вьетнам
@ Автор для переписки, e-mail: nguyenvandungvtdt@lqdtu.edu.vn

• **Поступила:** 24.11.2025 • **Доработана:** 24.01.2026 • **Принята к опубликованию:** 25.05.2026

Резюме

Цели. Целью данной работы является разработка эффективного метода подавления аддитивного гауссовского шума в радиоканале, обеспечивающего высокое качество восстановления цифровых модулированных сигналов при низких значениях отношения сигнал/шум (signal-to-noise ratio, SNR).

Методы. Для решения поставленной задачи предложена гибридная архитектура, объединяющая рекуррентные нейронные сети типов долгой краткосрочной памяти (long short-term memory, LSTM) и управляемых рекуррентных блоков (gated recurrent unit, GRU). Модель ResBiLSTM-BiGRU обеспечивает более стабильное обучение, повышенное качество восстановления и более точное моделирование временной структуры сигналов. Эффективность модели оценивалась на многомодуляционном наборе данных при различных уровнях входного SNR.

Результаты. Экспериментальные исследования показали, что предложенная модель в среднем обеспечивает увеличение выходного SNR на 2–3 дБ по сравнению с релятивистской усредненной генеративно-состязательной сетью (relativistic average generative adversarial networks, RaGAN) и более чем на 10 дБ по сравнению с методом на основе вейвлет-преобразования при одинаковых условиях моделирования. Модель характеризуется меньшими значениями среднеквадратической ошибки и более высокими коэффициентами корреляции во всем диапазоне входных SNR от −10 до 10 дБ. Значения пикового SNR также свидетельствуют о высоком качестве восстановления сигналов, особенно при низких уровнях входного SNR. Анализ архитектурных параметров показал, что наилучшие результаты достигаются при использовании 128 скрытых узлов и глубокой архитектуры ResBiLSTM-BiGRU.

Выводы. Полученные результаты подтверждают эффективность и практическую применимость гибридной архитектуры ResBiLSTM-BiGRU для подавления аддитивного гауссовского шума в радиоканале и восстановления цифровых модулированных сигналов, особенно в условиях низкого SNR.

Ключевые слова: когнитивные радиосети, автоматическая классификация модуляции, сигналы цифровой модуляции, глубокое обучение, гибридная сеть ResBiLSTM-BiGRU, подавление шума, отношение сигнал/шум, среднеквадратическая ошибка, коэффициент корреляции, пиковое отношение сигнал/шум

Для цитирования: Нгуен В.З., Фан Ч.Х. Гибридная архитектура нейронной сети ResBiLSTM-BiGRU для эффективного подавления шума в радиоканале с цифровыми модулированными сигналами. *Russian Technological Journal*. 2026;14(4):82–95. https://doi.org/10.32362/2500-316X-2026-14-4-82-95, https://elibrary.ru/VLPBLV

Прозрачность финансовой деятельности: Авторы не имеют финансовой заинтересованности в представленных материалах или методах.

Авторы заявляют об отсутствии конфликта интересов.

INTRODUCTION

The increased use of modern communication technologies such as 5G/6G networks and the Internet of Things (IoT) has led to greater demands for high-quality signal transmission through complex, noisy communication channels. This has given rise to a significant challenge in the form of additive white Gaussian noise (AWGN). By significantly reducing signal quality, AWGN complicates various applications, including automatic modulation classification, spectral sensing and resource management in cognitive radio networks.

Traditional noise suppression methods, such as linear filtering and signal transformations in the time and frequency domains, have been used for several decades. While these methods demonstrate satisfactory performance in linear channels with medium to high signal-to-noise ratios (SNRs), they become less effective in the presence of nonlinear signals and low SNR conditions. This limits their practical applicability. However, they are ineffective in the presence of nonlinear signals and low SNR conditions.

In recent years, deep learning has opened up new possibilities for radio signal processing. Models such as denoising autoencoders (DAEs) [9, 10], generative adversarial networks (GANs) [11–13], diffusion models [14, 15], and transformers have been shown to be effective at suppressing noise. However, each of these models has its limitations. For instance, DAEs become less efficient at low SNRs. GANs and their variations are difficult to train and require a lot of computational resources. Although diffusion models and transformers perform well in high-noise scenarios, they also have high computational requirements and long output delays. Long short-term memory (LSTM) and gated recurrent unit (GRU) networks do not exploit their complementary capabilities fully when used separately.

Based on these findings, the present work proposes a ResBiLSTM-BiGRU hybrid architecture that combines the strengths of LSTMs and GRUs. While LSTMs facilitate the learning of long-term dependencies, GRUs offer compactness and rapid convergence. The study aims to develop an efficient and practical approach to AWGN suppression in digital radio channels for next-generation communication systems.

Specifically, the study:

- develops a specialized ResBiLSTM-BiGRU hybrid architecture for noise reduction in digital radio signals;
- conducts a thorough analysis of the results across a wide SNR range of from −10 dB to 10 dB and various modulation techniques; and
- demonstrates the efficacy of the proposed approach compared to traditional noise reduction methods using metrics such as SNR, mean square error (MSE), root mean square error (RMSE) and peak signal-to-noise ratio (PSNR).

1. RELATED RESEARCH

Over several decades, both conventional techniques and modern deep learning-based approaches have been used to address AWGN. While conventional techniques remain popular due to their simplicity and reliability, they are less effective under more challenging scenarios, such as those involving low SNR or nonlinear channels. Table 1 summarizes the key features, advantages, and limitations of some of the most commonly used conventional noise reduction methods.

Contemporary approaches based on deep learning offer high flexibility and efficacy in addressing AWGN, especially in low SNR settings. However, there are still several challenges to overcome when implementing deep learning models for noise reduction in digital communications. Table 2 outlines some recent research efforts to address AWGN in digital modulation systems.

The experimental phase of the research involves a comparative analysis of two noise reduction methods. One of these techniques is based on wavelet transformations (WT), while the other uses relativistic average generative adversarial networks (RaGAN). The results suggest that a hybrid architecture combining ResBiLSTM and BiGRU networks achieves more consistent signal recovery accuracy when used in challenging communication channels with low SNR for AWGN suppression across various modulation techniques.

Table 1. Characteristics of conventional noise reduction methods

Method group	Primary methods	Key features	Limitations
Linear filtering [1, 2]	Wiener filtering, Kalman filtering, adaptive least mean squares (LMS) algorithm, and recursive least squares (RLS) algorithm	• Wiener filter optimizes MSE in a linear environment; • Kalman filtering is effective for linear dynamical systems; • LMS/RLS provides adaptive filtering for a changing channel	• linear properties of signal and noise are required; • reduced efficiency with low SNR; • limitations when working in non-linear channels or with non-stationary noise
Noise reduction in the transformed area [3–6]	WT, Fourier transform, and discrete cosine transform (DCT)	• WT provides signal analysis in both time and frequency domains simultaneously; • the Fourier transform and DCT represent a signal exclusively in the frequency domain; • these methods are most effective for signals with a pronounced and stable frequency structure	• the effectiveness of WT depends on the choice of the basic function and the threshold; • the Fourier transform and DCT are ineffective with AWGN; • partial loss of local information is possible
Statistical methods [7, 8]	Minimum mean square error (MMSE), maximum *a posteriori* probability (MAP) method, principal component analysis (PCA), singular value decomposition (SVD)	• MMSE/MAP uses probabilistic signal and noise models; • PCA/SVD allows you to separate signal and noise components; • effective with a complete statistical description of the environment	• precise knowledge of signal and noise distributions is required; • high computational complexity; • efficiency is significantly reduced with an inaccurate model

Table 2. Characteristics of deep learning methods

Author(s), Year, Study	Architecture	Data / Channel	Key concepts and benefits	Limitations
Ruifeng Duan, Ziyu Chen, Haiyan Zhang, Xu Wang, Wei Meng, Guodong Sun, 2023 [10]	Autoencoder with skip connections and channel-wise attention mechanism	Digital modulated signals; AWGN	Combination of two residual encoder-decoder network connections and an advanced SE-ResNet[1] module to enhance feature extraction and reduce computational complexity	Comparison only with RLS, LMS, and PCA
Shoushuai He, Lei Zhu, Changhua Yao, Lei Wang, Zhen Qin, 2022 [11]	Modified GAN architecture	Digital modulated signals; AWGN + interference	Restoring parts is better than traditional filters	The difficulty of learning, the difficulty of setting up
Liang Peng, Shengliang Fang, Youchen Fan, Mengtao Wang, Zhao Ma, 2023 [12]	RaGAN	Digital modulated signals; AWGN	Using GAN to generate a clean signal from a noisy one, improving the quality of recovery compared to traditional filters	Learning complexity, high computational load
Tong Wu, Zhiyong Chen, Dazhi He, Liang Qian, Yin Xu, Meixia Tao, Wenjun Zhang, 2023 [14]	The diffusion model is used to predict and suppress noise after channel alignment	Relay and AWGN communication channel	Determines the diffusion process according to the conditional distribution of the received signal, and can be integrated as a noise reduction module to improve signal quality	High computational complexity
Zelin Ji, Shuo Wang, Kuojun Yang, Qinchuan Zhang, Peng Ye, 2024 [16]	GRU-noise reduction module with classifier	Cyclic signal; digital modulated signal; AWGN channel	Pre-training of the noise canceller on a cyclic signal followed by fine-tuning on a digitally modulated signal; increased accuracy by more than 20% at low SNR values	Limited set of modulations; no evaluation using clean noise reduction metrics (MSE/SNR)

[1] The Squeeze-and-Excitation Residual Network (SE-ResNet) is a convolutional neural network architecture combining the skip connections of ResNet with a channel-wise attention mechanism based on Squeeze-and-Excitation blocks. This enables the adaptive recalibration of feature channels by modeling their mutual dependencies.

2. PROPOSED METHOD

2.1. ResBiLSTM-BiGRU network architecture for suppressing additive Gaussian noise and compensating for an imperfect transmission channel of basic I/Q[2] data

The paper assumes that the received data is basic I/Q data transmitted over a communication channel and corrupted by AWGN. In order to provide an effective noise suppression model for these signals, a ResNet[3]-based architecture called ResBiLSTM-BiGRU is proposed. The learning branch of this architecture is a hybrid network combining LSTM and GRU in cascade.

Before discussing this hybrid network architecture, a comparison of the key characteristics of LSTMs and GRUs is presented [17]. Table 3 summarizes their main differences.

Figure 1 depicts the overall architecture of the ResBiLSTM-BiGRU network. The network follows a similar structure to ResNet, comprising two parallel branches. The right branch serves as the primary data transmission path to ensure stable information flow and prevent the occurrence of gradient attenuation or explosion within the network. In contrast, the trainable left branch incorporates a hybrid LSTM-GRU structure alongside auxiliary layers, which is referred to as a "Hybrid LSTM-GRU."

The model receives as input a basic I/Q signal contaminated by AWGN, represented as a 3D tensor with dimensions of [B, T, 2]. Here, B represents the batch size, T denotes the length of the time sequence, and 2 is the number of channels (I and Q). Due to the direct connection provided by the right branch, we will focus on data processing within the Hybrid LSTM-GRU branch for further analysis.

At the initial stage of processing, the data is passed through bidirectional long short-term memory (BiLSTM) layers. These layers extract time dependencies in both the forward and backward directions to capture long-term correlations within the data. The output of these layers is then fed into bidirectional gated recurrent units (BiGRU), which refine the features and capture local dependencies with higher frequency and lower computational cost due to their simplified gating structure.

Following the BiGRU processing, the output is normalized using layer normalization for stability in learning. The normalized data is then fed into a linear layer to reduce the dimension of the input. Finally, the outputs the direct and indirect paths are combined to generate the final output of the model.

This structure is intended to suppress AWGN and compensate for distortions caused by imperfections in the transmission channel that affect the received signal.

Table 3. Comparative characteristics of LSTM and GRU architectures

Characteristic	LSTM	GRU
Number of gates	Three: entry, forgetting, and exit	Two: update, reset
Number of states	Two: cell state and hidden state	One: hidden state
Memory mechanism	Linear transfer of information through the cell state resistant to long-term dependencies	Merging new and previous information in a hidden state, effective for medium and short dependencies
Model complexity	Higher	Lower
Number of parameters	Significantly more	Less (by 25–30% compared to LSTM)
Learning rate	Lower	Higher
Modeling long-term dependencies	Very high	High, but inferior to LSTM
Modeling short-term dynamics	Less sensitive	More sensitive
Overfitting tendency	Higher	Lower
Memory costs	Higher	Lower
Convergence behavior	Slower, hyperparameter-sensitive	Faster and more stable
Typical applications	Long sequences, signals with trend or drift	Noisy signals, real-time processing, limited resources

[2] I stands for the in-phase component, Q stands for the quadrature component, and the quadrature component refers to two orthogonal components of a complex signal.

[3] A residual network is an architecture for convolutional neural networks that uses skip connections.

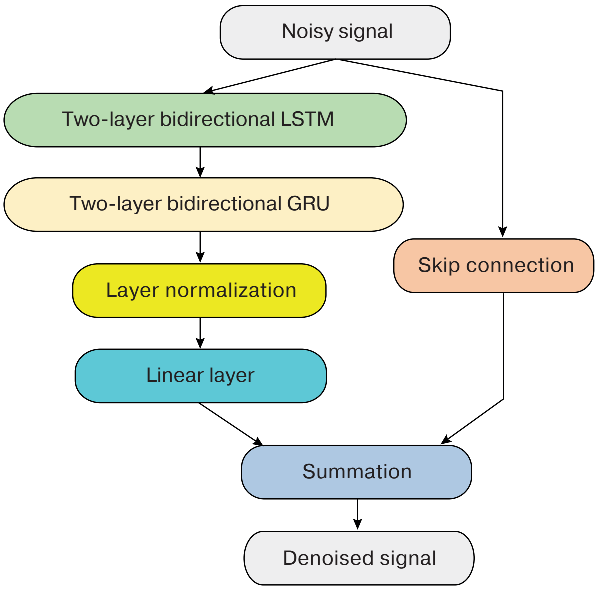

Fig. 1. Architecture of the ResBiLSTM-BiGRU hybrid network for noise suppression mixed with a digital modulation signal

2.2. Justification of the feasibility of the Hybrid LSTM-GRU structure

As illustrated in Fig. 1, a key feature of the proposed model is its hybrid architecture, which combines LSTM and GRU (Hybrid LSTM-GRU). Figure 2 provides a more detailed representation of this architecture.

Figure 2 depicts a network that operates on basic I/Q data. This modulated signal, which has been reduced to the base frequency band, contains information about amplitude, frequency and phase modulation. When transmitting over a channel, this signal is subject to additive noise, and can be modelled as follows:

$$x(n) = A(n)e^{j\phi} + \eta(n), \qquad (1)$$

where $A(n)$ and $\phi(n)$ are slowly varying components carrying information about modulation, the relationship between symbols and details of the transmission medium (including multipath effects and fading), and where $\eta(n)$ represents additive noise typically modelled as a white Gaussian process.

In this context, processing I/Q data involves suppressing additive noise, $\eta(n)$, and identifying signal patterns, as well as accounting for the effects of multipath/fading in $A(n)$ and $\phi(n)$. In the time domain, AWGN $\eta(n)$ varies rapidly and carries short-term information, whereas the signal patterns and multipath/fading effects are considered to be more slowly varying and to carry long-term information. In other words, the signal $x(n)$ in equation (1) exhibits multiscale temporal behavior.

The mathematical model of the I/Q data presented in equation (1) and the above analysis determine the network architecture shown in Fig. 2 directly.

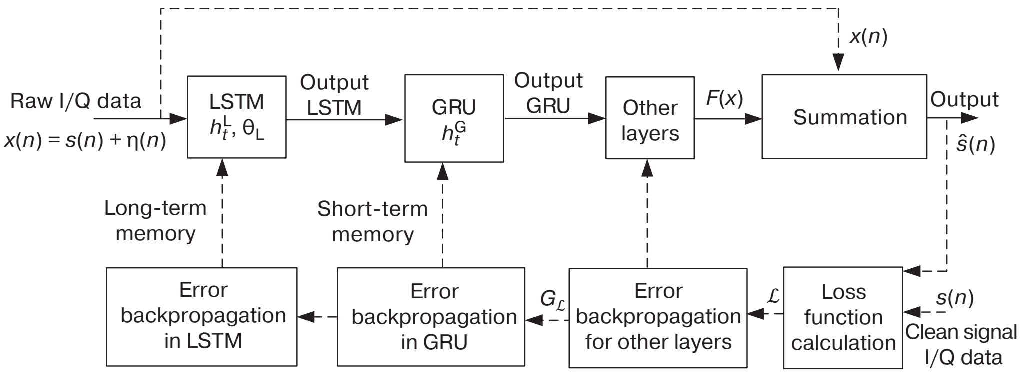

Fig. 2. Sequential Hybrid LSTM-GRU architecture: $x(n)$ is the noisy signal, $s(n)$ is the clean signal, $\hat{s}(n)$ is the reconstructed signal, h_t^L, h_t^G are the hidden states of the LSTM and GRU networks, respectively, θ_L is the set of LSTM parameters, and $F(x)$ represents the transformation of the input signal through the network. The final output is produced using skip connection: $\hat{s}(n) = x(n) + F(x)$; $\mathcal{L}$ is the loss function, $G_{\mathcal{L}}$ is the loss gradient, n is the count, and t is the time step

The underlying I/Q signals exhibit multiscale temporal patterns and can also be considered statistically independent, since the white Gaussian noise component is uncorrelated with the others. During the training process, the network architecture (Fig. 2) enables the LSTM unit to capture primarily long-term signal components, while the GRU specializes in modeling short-term dynamics. From a physical perspective, the LSTM unit focuses on extracting characteristic signal patterns and multipath or fading effects, whereas the GRU unit mainly extracts information relating to the additive white noise component.

Therefore, the multiscale time structure and statistical independence of the I/Q signal components are necessary for the hybrid neural network to automatically allocate roles to the LSTM and GRU components without imposing any prior constraints.

2.3. Dataset

The study is based on the RadioML 2018.01A open DeepSig[4] dataset. This dataset, which is widely used for radio signal modulation recognition tasks, includes 24 different modulation types, presented at 26 SNR levels ranging from −20 to +30 dB in 2 dB increments. A total of 4096 individual frames are included for each "modulation–SNR" pair. Each frame contains 1024 complex signal samples, which are represented as I/Q components with dimensions of (1024, 2). The total dataset comprises 2555904 frames.

To tailor the dataset for noise suppression, signals with an SNR of 30 dB are isolated and used as a reference for "clean" signals with minimal noise. Based on these, a noisy set is formed in which AWGN is added to each signal at SNR levels ranging from −10 to +10 dB in 2 dB increments according to a Gaussian distribution. Thus, two subsets are formed: clean modulated signals and noisy modulated signals.

The main parameters of the dataset are given in Table 4.

The dataset is divided into three parts for training, validation, and testing according to the proportions 80%, 10%, and 10%, respectively. Two models are trained using the training and validation sets: a Hybrid LSTM-GRU and a RaGAN. The test set is then used to compare the efficacy of these two models and evaluate the performance of the WT-based noise suppression methods.

Table 4. Dataset main parameters

Parameter	Value
Modulation types	8 types: BPSK[5], QPSK[6], OQPSK[7], 8PSK[8], 16QAM[9], 64QAM[10], GMSK[11], and OOK[12]
SNR ranges	11 values ranging from −10 to 10 dB in 2 dB increments
Number of samples in one frame	1024
Number of frames per "modulation–SNR" pair	4096
Data format in the frame	1024 complex I/Q samples (in-phase and quadrature components), size (1024, 2)
Total number of frames	360448 frames (8 × 11 × 4096)
Communication channel	AWGN

3. EXPERIMENTAL RESULTS

To evaluate the effectiveness of the proposed ResBiLSTM-BiGRU hybrid architecture, experiments have been conducted using a test dataset. The performance of the model is compared to two representative methods, namely a deep RaGAN model and a conventional WT-based noise suppression method.

The RaGAN model uses a generative adversarial network with a relativistic discriminator based on bidirectional LSTM (BiLSTM) networks. We denote the original (real) signal as s_r, while the synthesized (fake) signal is denoted as s_f. The discriminator calculates the probability that the original data is more realistic than the synthesized data. The loss functions for the discriminator (D) and the generator (G) can be expressed as follows:

$$L_\mathrm{D}^\mathrm{Ra} = -E\left[\log\left(\mathrm{D}_\mathrm{RaGAN}\left(s_\mathrm{r}, s_\mathrm{f}\right)\right)\right] - \\ -E\left[\log\left(1-\mathrm{D}_\mathrm{RaGAN}\left(s_\mathrm{f}, s_\mathrm{r}\right)\right)\right],$$

$$L_\mathrm{G}^\mathrm{Ra} = -E\left[\log\left(1-\mathrm{D}_\mathrm{RaGAN}\left(s_\mathrm{r}, s_\mathrm{f}\right)\right)\right] - \\ -E\left[\log\left(\mathrm{D}_\mathrm{RaGAN}\left(s_\mathrm{f}, s_\mathrm{r}\right)\right)\right],$$

where $E[.]$ denotes the expectation operator, and $\mathrm{D}_\mathrm{RaGAN}$ is the relativistic discriminator used in RaGAN.

[4] https://www.deepsig.ai/datasets/. Accessed November 11, 2025.
[5] Binary Phase Shift Keying.
[6] Quadrature Phase Shift Keying.
[7] Offset Quadrature Phase Shift Keying.
[8] 8-Phase Shift Keying.
[9] 16-Quadrature Amplitude Modulation.
[10] 64-Quadrature Amplitude Modulation.
[11] Gaussian Minimum Shift Keying.
[12] On-Off Keying.

To improve the quality of clean signal recovery, the overall loss function of the generator is formulated as a weighted sum of content loss and adversarial loss as follows:

$$L_{\mathrm{G}} = L_{\mathrm{con}} + \lambda L_{\mathrm{G}}^{\mathrm{Ra}},$$

where λ is the balancing coefficient. Content loss, which is based on MSE and absolute difference, is defined as:

$$L_{\mathrm{MSE}} = \frac{1}{n}\sum_{i=1}^{n}\left(s_{\mathrm{r}_i} - \mathrm{G}\left(s_{\mathrm{f}}\right)_i\right)^2, \quad L_1 = \frac{1}{n}\sum_{i=1}^{n}\left|s_{\mathrm{r}_i} - \mathrm{G}\left(s_{\mathrm{f}}\right)_i\right|,$$

$$L_{\mathrm{con}} = \frac{L_{\mathrm{MSE}} + L_1}{2}.$$

This combination of loss functions allows the generator to accurately reconstruct the noisy signal while considering both the quality of content and competitive component.

The classical WT method operates in the frequency domain. The noisy signal $y(t)$ is decomposed into a set of wavelet coefficients:

$$c_{j,k} = \int y(t)\psi_{j,k}(t)dt,$$

where $\psi_{j,k}(t)$ are the basic wavelet functions at scale j and position k. Threshold filtering of coefficients is then applied:

$$\tilde{c}_{j,k} = \begin{cases} \mathrm{sign}\left(\mathrm{c}_{j,k}\right)\left(\left|\mathrm{c}_{j,k}\right| - \lambda\right), & \left|\mathrm{c}_{j,k}\right| > \lambda, \\ 0, & \left|\mathrm{c}_{j,k}\right| \le \lambda. \end{cases}$$

Then, the signal is reconstructed using the inverse wavelet transform, as follows:

$$\hat{x}(t) = \sum_{j,k}\tilde{c}_{j,k}\psi_{j,k}(t).$$

This approach enables the removal of high-frequency noise components by setting the corresponding coefficients to zero.

The efficacy of noise suppression was evaluated using standard quality recovery metrics, such as MSE, SNR, correlation coefficient (CC) and PSNR. The formulas for these metrics can be found in Table 5.

The model is trained using MSE loss function and the Adam [13] optimizer for 150 epochs. During training, the adaptive reduction mechanism called Reduce learning rate on plateau or ReduceLROnPlateau, is employed, whereby the learning rate decreases if there is no improvement in the metrics over five consecutive epochs. To prevent overfitting and reduce calculation time, the Early stopping[14] strategy is also employed, which is triggered when the loss function stops decreasing. As shown in Fig. 3, the learning outcomes demonstrate a stable decrease in the loss function for both the training and test samples, as well as the model achieving stable and optimal error values.

Table 5. Metrics for evaluating the quality of noise reduction

Metric	Formula
Signal-to-noise ratio (SNR)	$\mathrm{SNR} = 10\cdot\lg\left(\frac{\sum_{i=1}^{N}x_i^2}{\sum_{i=1}^{N}\left(x_i - \hat{x}_i\right)^2}\right)$ dB, where x is the original clean signal, $\hat{x}$ is reconstructed signal, N is the number of samples
Mean squared error (MSE)	$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}\left(x_i - \hat{x}_i\right)^2$
Correlation coefficient (CC)	$\mathrm{CC} = \frac{\sum_{i=1}^{N}\left(x_i - \bar{x}\right)\left(\hat{x}_i - \bar{\hat{x}}\right)}{\sqrt{\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2}\sqrt{\sum_{i=1}^{N}\left(\hat{x}_i - \bar{\hat{x}}\right)^2}}$, where $\bar{x}$ are the average values of the original clean signal, $\bar{\hat{x}}$ are average values of the reconstructed signal
Peak signal-to-noise ratio (PSNR)	$\mathrm{PSNR} = 10\cdot\lg\frac{x_{\max}^2}{\mathrm{MSE}}$. The $x_{\max}$ parameter is defined as the maximum amplitude value of a clean one-dimensional signal, calculated from its discrete time realization in a test sample

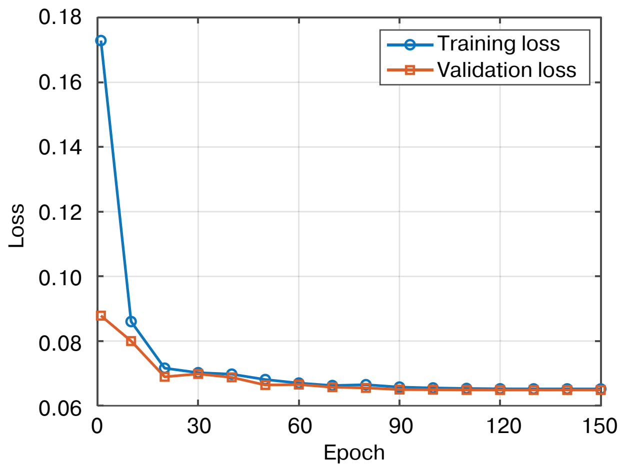

Fig. 3. Values of training and validation loss during the training process

[13] Adaptive moment estimation is an optimization algorithm used for training neural networks that combines the concepts of the momentum method and adaptive gradient scaling.

[14] Early stopping is a regularization method in which the learning process of a neural network is interrupted when the error on the validation data stops decreasing (and usually begins to increase), which prevents overfitting.

Figure 4a shows the results of comparing different noise suppression methods using the SNR metric. The proposed model delivers optimal performance across the entire input SNR range, spanning from −10 to +10 dB. Compared to RaGAN, the average improvement is approximately 2.27 dB. At an extremely low SNR level of −10 dB, the improvement is just 0.41 dB; however, it then gradually increases to stabilize within the range of 3.0–3.5 dB for high SNR values. Compared to WT, the advantage is even more pronounced, providing an average increase of 10.5 dB. In the low SNR range of −10 to 0 dB, the increase ranges from 6.1 to 12.3 dB. In the high SNR range (2 to 10 dB), the noise-suppression improvement remains between 11 and 13.5 dB.

Figure 4b shows the results of comparing the average MSE values. The proposed model consistently provides the lowest MSE, indicating more accurate signal reconstruction than RaGAN and WT. A particularly significant improvement can be observed compared to the WT; here, the MSE is 0.2556 at an SNR of −10 dB, whereas for WT it is 1.077. At an SNR of 10 dB, the MSE decreases to 0.0024, whereas WT remains at 0.0233 (almost ten times higher). Compared to RaGAN, the difference is less pronounced but still evident across the entire SNR range. For instance, at an SNR of −10 dB, the model's MSE is 0.255 *vs* 0.279 for RaGAN. At an SNR of 10 dB, the respective values are 0.0024 and 0.0042, indicating approximately 1.7 times greater accuracy.

Figure 4c shows a comparison of the correlation coefficient between the original and reconstructed signals. The proposed model achieves the highest values across the entire SNR range, confirming its superior ability to preserve the temporal structure of the signal. At an SNR of −10 dB, the correlation coefficient is 0.726, which is higher than that of RaGAN (0.718) and almost double that of WT (0.366). As the SNR increases, the performance of all models improves; however, the proposed method consistently outperforms the others. At an SNR of 0 dB, the correlation coefficient reaches 0.980, surpassing the values obtained for RaGAN (0.971) and WT (0.822). At high SNR levels (>6 dB), all techniques approach the saturation region (>0.99 for the proposed method and RaGAN). Nevertheless, the proposed method continues to outperform WT (0.979 at 10 dB *vs* 0.998 for the former).

Figure 4d presents the PSNR comparison results. The proposed model exhibits the highest PSNR values over the entire SNR range, confirming its superior signal reconstruction capabilities. At an SNR of −10 dB, the PSNR is 11.72 dB, which exceeds the corresponding values for RaGAN (11.31 dB) and WT (3.47 dB). As the SNR increases, the gap between the proposed model and

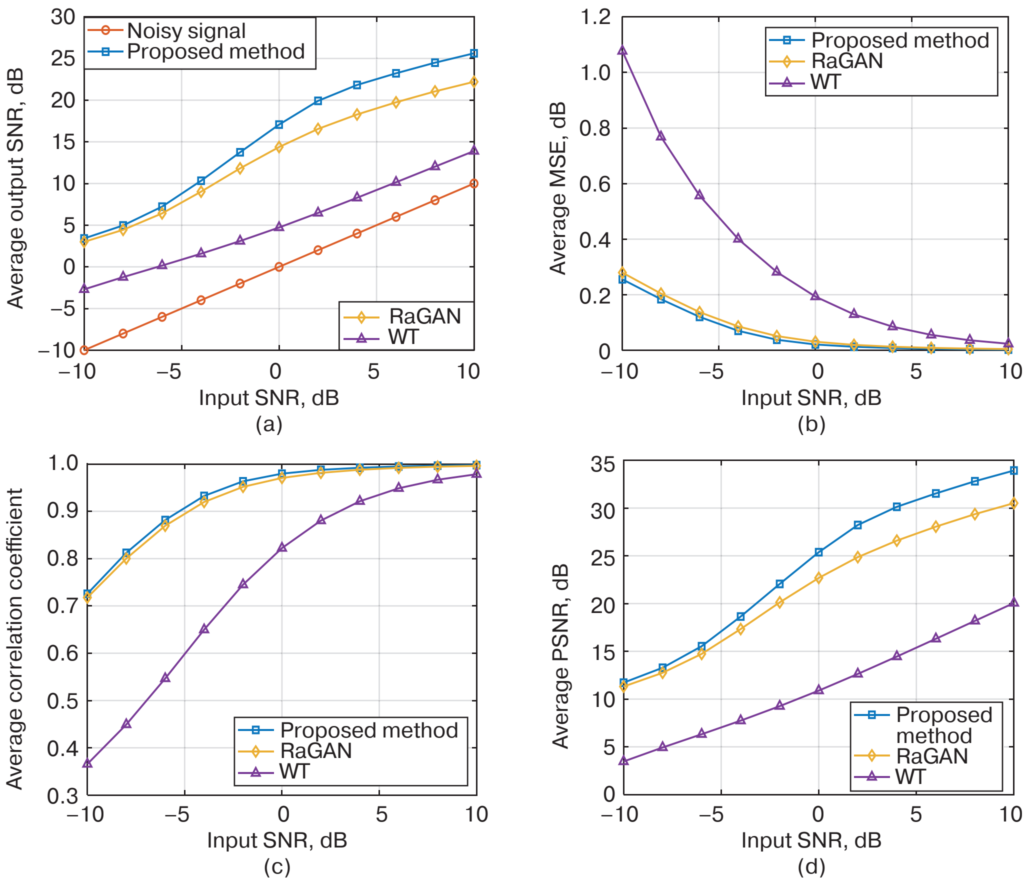

Fig. 4. Results of comparing noise suppression methods: (a) output SNR, (b) MSE, (c) correlation coefficient, and (d) PSNR

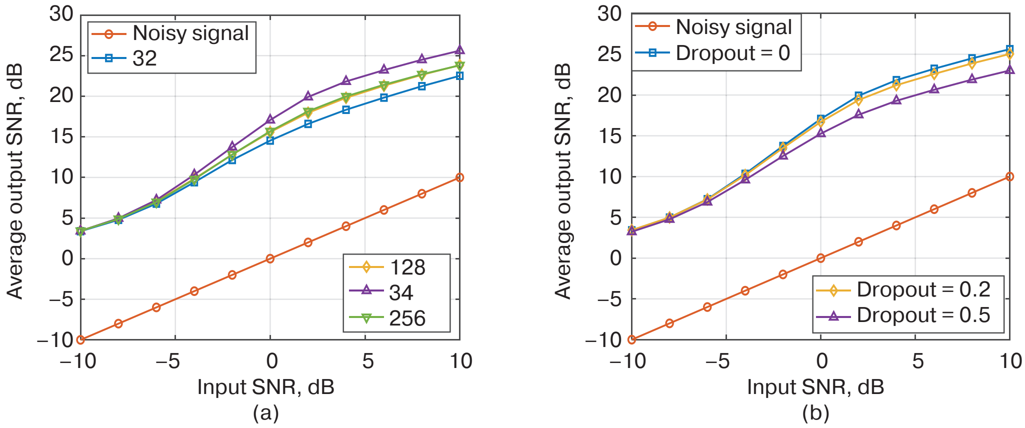

Fig. 5. Results of comparing noise suppression methods: (a) by the number of hidden nodes and (b) by the dropout value

RaGAN remains consistent at around 2–3 dB. The benefit is particularly notable in comparison to WT: at 0 dB, the PSNR is 25.38 dB for the model *vs* 10.89 dB for WT; at 10 dB, the model's PSNR reaches 33.93 dB, which is nearly 14 dB greater than that of WT (20.06 dB).

The influence of the number of hidden layers in the ResBiLSTM-BiGRU hybrid model is illustrated in Fig. 5a. A significant improvement in signal reconstruction accuracy is achieved by increasing the number of layers from 32 to 64. For example, at an SNR of 0 dB, the performance metric increases from 14.54 dB to 15.59 dB. Further increasing the number of layers to 128 results in optimal performance, with an SNR of 0 dB yielding a value of 17.07 dB. This is 1.5 dB higher than the 64-layer configuration and 2.5 dB higher than the 32-layer configuration. However, no further improvement is observed with 256 layers, at which configuration the performance remains comparable to that of the 64-layer model, and is lower than that of the 128-layer variant over the entire SNR range. These results suggest that 128 layers provide the best balance between reconstruction accuracy and model complexity.

Figure 5b shows the impact of the dropout rate[15] on model performance. Without dropout, the model exhibits maximum efficiency. Setting the dropout value to 0.2 results in only a minor decrease in reconstruction accuracy while maintaining overall stability. However, increasing the dropout rate to 0.5 results in a significant reduction in performance across the entire SNR range.

Figure 6 depicts the results of comparing various combinations of the LSTM and GRU blocks. Using LSTM alone provides limited performance, for

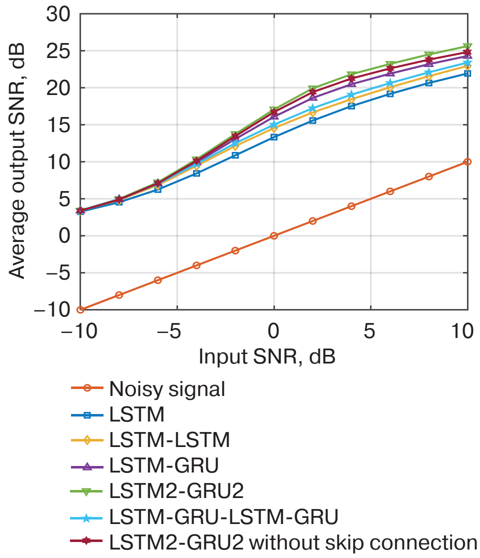

Fig. 6. Results of comparing performance for different versions of the hybrid LSTM-GRU architecture

[15] A hyperparameter of the neural network regularization method that determines the probability of unintentional pruning of neurons during the training process in order to prevent overfitting.

example, 13.33 dB at an SNR of 0 dB. Adding the GRU component (the LSTM-GRU configuration) significantly improves signal reconstruction accuracy, achieving 16.08 dB at 0 dB. The best performance is demonstrated by deeper architectures, such as LSTM2-GRU2 with a skip connection (ResBiLSTM-BiGRU), which achieve maximum values of 17.07 dB at 0 dB and 25.62 dB, respectively, for all configurations considered. Although variants of the LSTM-GRU-LSTM-GRU and LSTM2-GRU2 configurations without a skip connection also perform well, they are inferior to the LSTM2-GRU2-Res architecture in terms of stability and absolute metric values across the entire SNR range.

Figure 7 shows the results of the proposed method for the following eight types of modulation: BPSK, QPSK, OQPSK, 8PSK, 16QAM, 64QAM, GMSK, and OOK. At low input SNRs, 64QAM, 16QAM, and 8PSK demonstrate the lowest efficiency, exhibiting a lower output SNR, PSNR, and correlation coefficient, while the MSE is higher. BPSK, GMSK, and OOK show the best results under the same conditions. As the SNR increases, the difference between the modulations decreases; however, higher orders of QAM and 8PSK continue to demonstrate slightly lower efficiency.

Figure 8 shows some examples of time domain signal comparisons at different SNR levels after noise suppression methods have been applied.

CONCLUSIONS

A novel hybrid ResBiLSTM-BiGRU neural network architecture for noise suppression in digital modulation signals transmitted over an AWGN channel is proposed. Experimental results confirm that the proposed model achieves an average SNR reconstruction improvement of 2–3 dB compared to RaGAN, and an improvement of over 10 dB as compared to WT. The model also demonstrates the lowest MSE values and highest correlation coefficients across the entire input SNR range from −10 to 10 dB. The PSNR values additionally confirm the model's high signal reconstruction accuracy, particularly at low SNR levels.

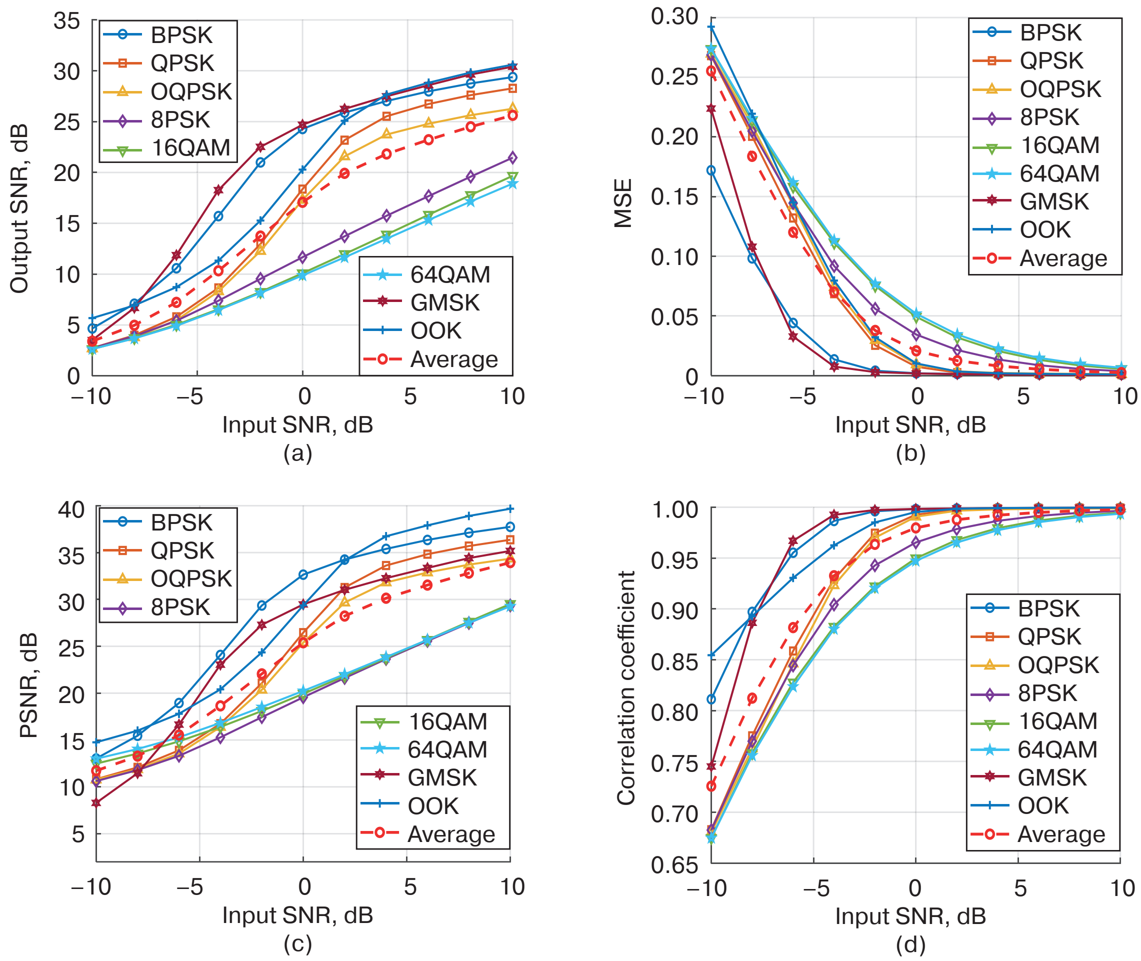

Fig. 7. Results of evaluating the accuracy of the proposed method for eight types of modulation: BPSK, QPSK, OQPSK, 8PSK, 16QAM, 64QAM, GMSK, and OOK for different SNR input levels: (a) output SNR, (b) MSE, (c) PSNR, and (d) correlation coefficient

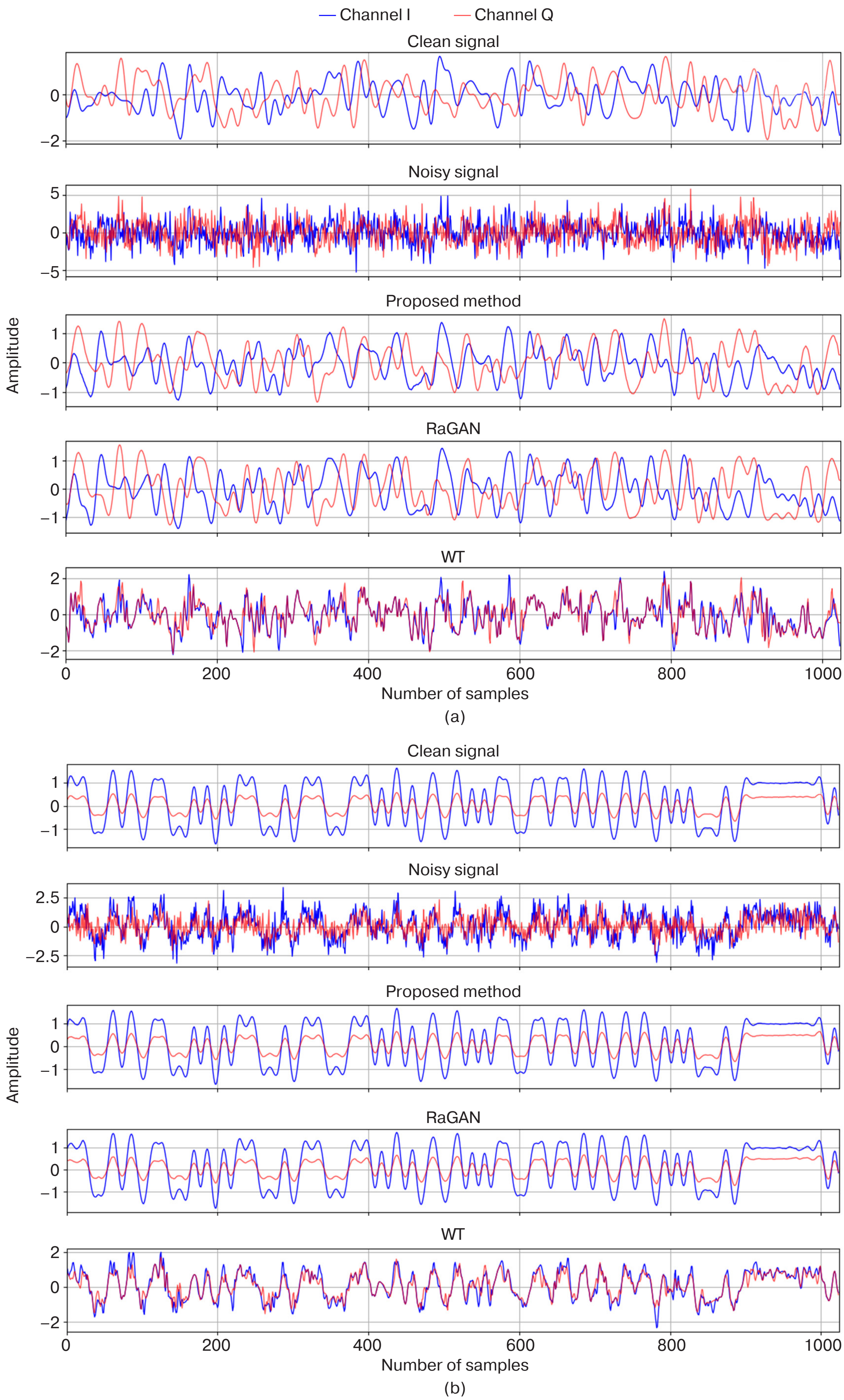

Fig. 8. Signals after noise suppression by various models for the modulated signal: (a) 16QAM at SNR = −6 dB and (b) BPSK at SNR = 0 dB

Analysis of the parameters shows that the optimal configuration has 128 hidden nodes, while additional nodes increase the dropout rate leading to a significant reduction in accuracy. Among the studied architectures, the deep ResBiLSTM-BiGRU model demonstrates the highest overall efficiency. These results confirm the effectiveness and promising potential of the proposed ResBiLSTM-BiGRU hybrid model for suppressing additive Gaussian noise.

Future research will focus on integrating the noise reduction module with the automatic modulation classification system in order to improve the overall performance of a radio communication system.

Authors' contribution

All authors contributed equally to the research work.

REFERENCES

1. Wang W. Evaluation of Adaptive Filtering Algorithms in Signal Denoising: Insights from Simulation and Algorithmic Enhancement. *Theoretical and Natural Science*. 2025;101(1):61–69. https://doi.org/10.54254/2753-8818/2025.CH22293
2. Lei J., Zhou J., Jian Z., Liu H., Zhang L., Feng Y., Wu Z. A Wiener Filter Denoising Based Intelligent Modulation Recognition. In: *2022 IEEE/CIC International Conference on Communications in China* (*ICCC*). IEEE. 2022. P. 88–93. https://doi.org/10.1109/ICCC55456.2022.9880789
3. Guo T., Zhang T., Lim E., López-Benítez M., Ma F., Yu L. A Review of Wavelet Analysis and Its Applications: Challenges and Opportunities. *IEEE Access*. 2022;10:58869–58903. https://doi.org/10.1109/ACCESS.2022.3179517
4. Akkaya S. Wavelet-Based Denoising Strategies for Non-Stationary Signals in Electrical Power Systems: An Optimization Perspective. *Electronics*. 2025;14(16):3190. https://doi.org/10.3390/electronics14163190
5. Facciolo G., Pierazzo N., Morel J.-M. Conservative Scale Recomposition for Multiscale Denoising. (The Devil is in the High Frequency Detail). *SIAM J. Imaging Sci*. 2017;10(3):1603–1626. https://doi.org/10.1137/17M1111826
6. Wahab M.F., Gritti F., C. O'Haver T. Discrete Fourier transform techniques for noise reduction and digital enhancement of analytical signals. *TrAC Trends in Analytical Chemistry.* 2021;143(4):116354. https://doi.org/10.1016/j.trac.2021.116354
7. Zhao S., Ma L., Xu L., Liu M., Chen X. A Study of Fault Signal Noise Reduction Based on Improved CEEMDAN-SVD. *Appl. Sci.* 2023;13(19):10713. https://doi.org/10.3390/app131910713
8. Wang H., Yao Y., Zhang G., Pan J., Gao L., Jin H., Wang C. Enhanced Noise Suppression in Partial Discharge Signals via SVD and VMD with Wavelet Thresholding. *Modelling and Simulation in Engineering*. 2024;2024:5676986. https://doi.org/10.1155/2024/5676986
9. Almazrouei E., Gianini G., Almoosa N., Damiani E. A Deep Learning Approach to Radio Signal Denoising. In: *2019 IEEE Wireless Communications and Networking Conference Workshop* (*WCNCW*). IEEE. 2019. https://doi.org/10.1109/WCNCW.2019.8902756
10. Duan R., Chen Z., Zhang H., Wang X., Meng W., Sun G. Dual Residual Denoising Autoencoder with Channel Attention Mechanism for Modulation of Signals. *Sensors*. 2023;23(2):1023. https://doi.org/10.3390/s23021023
11. He S., Zhu L., Yao C., Wang L., Qin Z. A Novel Approach Based on Generative Adversarial Network for Signal Enhancement in Wireless Communications. *Wireless Communications and Mobile Computing*. 2022. https://doi.org/10.1155/2022/7050573
12. Peng L., Fang S., Fan Y., Wang M., Ma Z. A Method of Noise Reduction for Radio Communication Signal Based on RaGAN. *Sensors*. 2023;23(1):475. https://doi.org/10.3390/s23010475
13. Yao H., Zhou Q., Chen Z., Huang L., Xu D., Xuan Q. Denoising of Radio Modulation Signal Based on Deep Learning. In: Xuan Q., Xiang Y., Xu D. (Eds.). *Deep Learning Applications*. 2023. Ch. 7. P. 125–143. https://doi.org/10.1142/9789811266911_0007
14. Wu T., Chen Z., He D., Qian L., Xu Y., Tao M. CDDM: Channel Denoising Diffusion Models for Wireless Semantic Communications. *IEEE Trans. on Wireless Commun*. 2024;23(9):11168–11183. https://doi.org/10.1109/TWC.2024.3379244
15. Letafati M., Ali S., Latva-Aho M. Conditional Denoising Diffusion Probabilistic Models for Data Reconstruction Enhancement in Wireless Communications. *IEEE Trans. on Machine Learning in Communications and Networking*. 2025;3:133–146. https://doi.org/10.1109/TMLCN.2024.3522872
16. Ji Z., Wang S., Yang K., Zhang Q., Ye P. Transfer Learning Guided Noise Reduction for Automatic Modulation Classification. In: *2025 IEEE 102nd Vehicular Technology Conference* (*VTC2025-Fall*). IEEE. 2025. https://doi.org/10.1109/VTC2025-Fall65116.2025.11310099
17. Chung J., Gulcehre C., Cho K.H., Bengio Y. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. *arXiv* preprint. arXiv:1412.3555. 2014. http://arxiv.org/abs/1412.3555
18. O'Shea T.J., Roy T., Clancy T.C. Over-the-Air Deep Learning Based Radio Signal Classification. *IEEE J. Selected Topics in Signal Process*. 2018;12(1):168–179. https://doi.org/10.1109/JSTSP.2018.2797022

About the Authors

Nguyen Van Dung, Cand. Sci. (Eng.), Lecturer, Department of Circuit Theory – Measurement, Faculty of Radio-Electronic Engineering, Le Quy Don Technical University (236, Hoang Quoc Viet, Ha Noi, Viet Nam). E-mail: nguyenvandungvtdt@lqdtu.edu.vn. https://orcid.org/0000-0002-2810-1204

Phan Trong Hanh, Cand. Sci. (Eng.), Lecturer, Department of Circuit Theory – Measurement, Faculty of Radio-Electronic Engineering, Le Quy Don Technical University (236, Hoang Quoc Viet, Ha Noi, Viet Nam). E-mail: maitrongphan@gmail.com. https://orcid.org/0009-0006-2093-7150

Об авторах

Нгуен Ван Зунг, к.т.н., преподаватель, кафедра теории цепей – измерения, Факультет радиотехники и электроники, Вьетнамский государственный технический университет им. Ле Куй Дона (Вьетнам, Ханой, ул. Хоанг Куок Вьет, д. 236). E-mail: nguyenvandungvtdt@lqdtu.edu.vn. https://orcid.org/0000-0002-2810-1204

Фан Чонг Хань, к.т.н., преподаватель, кафедра теории цепей – измерения, Факультет радиотехники и электроники, Вьетнамский государственный технический университет им. Ле Куй Дона (Вьетнам, Ханой, ул. Хоанг Куок Вьет, д. 236). E-mail: maitrongphan@gmail.com. https://orcid.org/0009-0006-2093-7150

Translated from Russian into English by K. Nazarov
Edited for English language and spelling by Thomas A. Beavitt