electronics

MDPI

Article

A Hybrid Self-ONN and Vision Mamba Architecture for Robust Radio Interference Recognition in GNSS Applications

Nursultan Meirambekuly [1,2,3], Margulan Ibraimov [4,*], Bakyt Khaniyev [4,5,*], Beibit Karibayev [4], Alisher Skabylov [4], Nursultan Uzbekov [1,6], Sungat Koishybay [4], Timur Dautov [4], Ainur Khaniyeva [4] and Bagdat Kozhakhmetova [7]

1 Space Technology Center, Al-Farabi Kazakh National University, Almaty 050040, Kazakhstan; nursultan.meirambekuly@kaznu.edu.kz (N.M.); uzbekov.nursultan@gmail.com (N.U.)
2 Department of Science and Innovation, Kazakh-British Technical University, Almaty 050010, Kazakhstan
3 Ecology Research Institute, Khoja Akhmet Yassawi International Kazakh-Turkish University, Turkistan 161200, Kazakhstan
4 Faculty of Physics and Technology, Al-Farabi Kazakh National University, Almaty 050040, Kazakhstan; beybit.karybaev@kaznu.edu.kz (B.K.); alisher.skabylov@kaznu.edu.kz (A.S.); sungatkoishybai@gmail.com (S.K.); dautov14@list.ru (T.D.); ainur.khaniyeva@kaznu.edu.kz (A.K.)
5 Research Institute of Experimental and Theoretical of Physics, Al-Farabi Kazakh National University, Almaty 050040, Kazakhstan
6 Orion System LLP, Almaty 050004, Kazakhstan
7 Department of Telecommunications Engineering, Almaty University of Power Engineering and Telecommunications named after Gumarbek Daukeyev, Almaty 050013, Kazakhstan; kozhahmetova.ba@gmail.com
* Correspondence: margulan.ibraimov@kaznu.edu.kz (M.I.); khaniyev.bakyt@kaznu.kz (B.K.)

Abstract

Radio-frequency interference (RFI) poses a critical challenge for modern high-precision Global Navigation Satellite System (GNSS) applications, as both intentional and unintentional interference can significantly degrade positioning accuracy and reliability. With increasingly sophisticated interference sources, robust and computationally efficient automatic recognition methods are required for next-generation GNSS receivers. Although deep learning approaches show strong potential for interference detection, their high computational cost often limits deployment in resource-constrained navigation hardware. This paper proposes a hybrid deep learning architecture for radio interference recognition in high-precision GNSSs. The framework employs a dual-branch design integrating complementary signal representations. A Self-Organizing Operational Neural Network (Self-ONN) extracts nonlinear temporal features from raw one-dimensional signals, while a Vision Mamba state-space model processes two-dimensional time-frequency spectrograms obtained via Short-Time Fourier Transform (STFT). The fused features enable accurate classification of diverse interference types with high computational efficiency. Experiments on a synthetic dataset demonstrate that the proposed model achieves 99.83% accuracy and F1-score, outperforming ResNet18, VGG16, and Vision Transformer while reducing computational complexity by up to 42% and improving inference speed by up to 35%, supporting its applicability for intelligent interference monitoring in GNSS receivers.

Keywords: GNSS applications; radio interference recognition; deep learning; hybrid neural network; Self-ONN; Vision Mamba; real-time signal processing

Check for updates

Academic Editor: Massimiliano Pieraccini

Received: 8 March 2026
Revised: 26 March 2026
Accepted: 29 March 2026
Published: 3 April 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

GNSS represents a critical infrastructure component supporting positioning, navigation, and timing services in civil, industrial, and scientific domains. Modern applications

such as autonomous transportation, precision agriculture, geodetic monitoring, and safety-critical navigation rely on the continuous availability and integrity of GNSS signals [1–5]. However, GNSS signals received at the Earth's surface exhibit extremely low power levels, making them susceptible to radio-frequency interference (RFI) [6,7]. Both intentional interference (e.g., jamming and spoofing) and unintentional electromagnetic emissions can degrade positioning accuracy or disrupt navigation solutions entirely [8]. Ensuring robust interference recognition has therefore become a fundamental requirement for next-generation high-precision GNSS receivers [9].

While interference detection has been extensively studied, the practical challenge extends beyond binary detection toward reliable classification of interference types. Accurate identification of interference structure enables adaptive mitigation strategies, receiver reconfiguration, and post-processing integrity analysis [10]. Formally, the problem can be defined as a supervised classification task in which a received signal is assigned to one of several predefined interference categories under varying SNR conditions. The classifier is required to map the input signal to a discrete class label representing a specific type of interference, while simultaneously maintaining high discriminative accuracy and meeting the real-time computational constraints imposed by embedded GNSS receiver hardware [11].

Early approaches to interference recognition relied on manually engineered features derived from amplitude statistics, temporal characteristics, or spectral parameters. Dovis and Musumeci proposed an interference detection and mitigation technique based on the Karhunen-Loève transform, enabling the separation of interference components in a transformed signal domain [12]. Sharifi-Tehrani et al. introduced a GNSS jamming detection method based on random matrix theory, where statistical properties of the received signal were exploited to identify interference presence [13]. Although such signal processing techniques can be effective in controlled scenarios, they strongly depend on expert-designed feature extraction pipelines and exhibit limited robustness to non-stationary or previously unseen interference patterns.

With the advancement of parallel computing and GPU-accelerated processing, deep learning methods have emerged as powerful tools for automatic feature extraction from complex radio signals [14]. Blais et al. proposed a convolutional neural network (CNN) framework that processes delay Doppler correlation maps of GNSS signals for multipath detection, achieving detection accuracy above 93% even under degraded reception conditions with carrier-to-noise ratios as low as 36 dB-Hz [15]. Cai et al. developed a jamming recognition approach combining multistep singular spectrum analysis with a channel-spatial attention CNN, achieving classification accuracy exceeding 90% at JNR = −2 dB under noisy interference conditions [16]. However, many high-capacity architectures rely on quadratic-complexity attention mechanisms or deep convolutional stacks, resulting in substantial memory and computational requirements. Such models are difficult to deploy in embedded GNSS receivers operating under strict power and latency constraints.

A widely adopted strategy in modern interference recognition frameworks involves transforming one-dimensional temporal signals into two-dimensional time-frequency representations using techniques such as the STFT, enabling the application of image-based neural architectures originally developed for computer vision tasks. Wu et al. proposed a jamming recognition method based on wavelet packet decomposition (WPD) combined with a deep learning network, where time-frequency spectrogram representations were used as inputs to improve feature extraction and classification accuracy while maintaining favorable computational efficiency [17]. Chen et al. developed a jamming recognition framework that generates STFT-based time–frequency distribution maps and employs a lightweight CNN to achieve efficient interference classification suitable for real-time signal processing scenarios [18]. Conversely, models operating directly on one-dimensional sig-

nals may fail to capture global spectral-temporal structures that are more clearly expressed in the time-frequency domain. Recent studies suggest that combining complementary representations can improve robustness under low SNR and structurally similar interference scenarios [19,20]. At the same time, emerging state-space neural architectures offer linear-complexity sequence modeling, providing an attractive alternative to transformer-based approaches for real-time signal processing [21].

Despite these developments, existing GNSS interference recognition frameworks typically rely on a single-domain representation (either temporal or time-frequency) or employ computationally intensive architectures unsuitable for real-time deployment. The absence of computationally efficient multimodal fusion strategies tailored specifically to GNSS interference classification motivates further research.

Unlike existing multimodal RF classification approaches that typically rely on convolutional feature extractors or transformer-based architectures, the proposed framework integrates Self-ONN and Vision Mamba within a unified design. This combination enables simultaneous nonlinear temporal modeling and efficient long-range dependency capture with linear computational complexity, making the approach particularly suitable for resource-constrained GNSS receiver applications. This distinguishes the proposed method from existing multimodal approaches by explicitly integrating data-efficient nonlinear temporal modeling with linear-complexity state-space sequence processing, enabling an improved balance between classification performance and computational efficiency.

The objective of this study is to design and experimentally validate a multimodal deep learning framework for GNSS radio-frequency interference classification that exploits complementary temporal and spectral representations to maintain high accuracy under noisy conditions while significantly reducing computational overhead compared to transformer-based models.

To achieve this goal, we propose a dual-branch architecture integrating a Self-ONN for nonlinear temporal feature extraction from raw one-dimensional signals and a Vision Mamba state-space model for efficient processing of two-dimensional spectrograms.

The outputs of both branches are fused at the feature level to construct a unified representation for final classification.

The main contributions of this work are:

1. Formulation of GNSS interference recognition as a computationally constrained multimodal classification problem.
2. Development of a hybrid Self-ONN and Vision Mamba architecture that combines nonlinear temporal modeling with linear-complexity state-space processing for efficient multimodal feature extraction.
3. A feature-level fusion strategy enabling complementary representation learning with reduced computational complexity.
4. Comprehensive experimental evaluation including accuracy, macro-F1 score, and robustness analysis under varying SNR conditions.
5. Analysis of computational efficiency in comparison with convolutional and transformer-based baselines.

Although the study is conducted using controlled synthetic interference signals to ensure reproducibility and parameter transparency, the modeling framework is designed to approximate representative GNSS interference structures. The limitations of simulation-based evaluation and directions for real-world validation are discussed in subsequent sections.

2. Methodology

The proposed methodology focuses on a computationally efficient deep learning framework for the classification of complex RFI. The architecture is designed to meet the stringent real-time requirements of next-generation high-precision GNSS receivers by balancing recognition accuracy with computational overhead.

2.1. Dataset Description and Signal Modeling

To validate the framework's robustness in GNSS-relevant scenarios, we utilized a representative dataset of synthetic radio-frequency interference waveforms. These signals represent typical interference patterns relevant to GNSS and radio navigation environments. The dataset comprises 18,000 samples across six sophisticated interference classes: Deceptive Jamming: Digital Deceptive Jamming (DDJ), Velocity Deception Jamming (VDJ), and Distance Frequency Time Jamming (DFTJ). Interrupted Sampling Interference: Interrupted Sampling Deceptive Jamming (ISDJ), Interrupted Sampling Repeater Jamming (ISRJ), and Interrupted Sampling Linear Jamming (ISLJ). No additional data augmentation was applied in order to preserve the controlled characteristics of the simulated interference signals.

The signals were modeled based on Linear Frequency Modulation (LFM) with a bandwidth of 15 MHz and a sampling rate of 25 MHz. The dataset was divided into training (70%), validation (15%), and testing (15%) subsets using stratified sampling to ensure balanced class distribution. Table 1 summarizes the technical parameters used for signal synthesis.

Table 1. Parameters used for signal generation.

Signal	Parameter	Value Range
LFM	Time width Bandwidth Sampling frequency	25 μs 15 MHz 25 MHz
DDJ	False target delay	10–30 μs
VDJ	Doppler shift	−5–5 MHz
DFTJ	Number of false targets False target delay	3–4 10–30 μs
ISDJ	Number of direct targets Slice cycles	1 2–3 μs
ISRJ	Number of direct targets Slice cycles	2–3 2–3 μs
ISLJ	Number of direct targets Slice cycles	3–4 2–3 μs

All signals are based on LFM waveforms [22]:

$$s(t) = A\exp\left(j\left(2\pi f_0 t + \pi\frac{B}{T}t^2\right)\right) \quad (1)$$

where $s(t)$ is the transmitted LFM signal, A denotes the signal amplitude, f_0 is the initial frequency, B represents the bandwidth, T is the pulse duration, and t is the time variable.

Although GNSS signals are based on spread-spectrum modulation rather than pure LFM, LFM-based modeling is widely used to simulate structured interference distortions due to its controllable spectral properties. The use of synthetically generated signals enables controlled evaluation of classification performance under diverse interference conditions representative of GNSS and radio navigation environments. This simulation-

based approach allows systematic variation in signal parameters, ensures experimental reproducibility, and provides a consistent benchmark for comparing deep learning architectures for interference recognition. These interference patterns reflect signal distortions that may affect high-precision GNSS receivers operating in contested or electromagnetically congested environments.

2.2. Data Preprocessing

The preprocessing stage is designed to ensure a robust multi-domain feature representation by preparing the input data for both temporal and spectral analysis. For the 1D temporal stream, raw interference signals undergo amplitude normalization to a standardized range of $[-1, 1]$, followed by a signal alignment process where sequences are fixed to a length of 4096 samples using zero-padding or truncation as required. In practical scenarios, variable-length signals can be processed using sliding windows or streaming-based inference. Concurrently, for the 2D time-frequency stream, the STFT is employed to extract spectral dynamics using a Hamming window with a size of 256 and a 50% overlap. The STFT is defined as [23]:

$$X(t, f) = \sum_{n=-\infty}^{\infty} x[n]w[n - t]e^{-j2\pi fn} \tag{2}$$

where $X(t, f)$ is the time-frequency representation of the signal, $x[n]$ denotes the discrete-time input signal, $w[n - t]$ is the window function centered at time t, and f represents the frequency variable.

The resulting complex time-frequency representations are converted into RGB spectrogram images with a spatial resolution of 224×224 pixels. To handle the complex-valued output of the STFT, the magnitude information was extracted and mapped to the jet color scale to form the input for the Vision Mamba stream. To facilitate rapid convergence and ensure numerical stability during the training phase, these images are further normalized based on the global mean and standard deviation of the training dataset. This dual-path preparation allows the framework to simultaneously exploit high-resolution waveform morphology and macroscopic spectral patterns.

2.3. Proposed Hybrid Architecture

To address the above challenges, a hybrid deep learning architecture was developed that integrates complementary signal representations within a unified classification framework. The proposed hybrid neural architecture consists of two parallel processing branches designed to extract complementary features from the same input signal. One branch operates directly in the temporal domain, while the second branch processes the time-frequency representation of the signal. The extracted features are subsequently fused and used for final classification. This dual-stream design enables the model to capture both fine-grained waveform characteristics and global spectral-temporal patterns associated with various interference types. Such integration is particularly relevant for GNSS and radio navigation applications, where interference signals may exhibit subtle variations in amplitude, phase, or spectral structure. The overall architecture of the proposed framework is illustrated in Figure 1.

The first branch processes raw one-dimensional signals in the temporal domain using a Self-ONN. Unlike conventional convolutional neural networks, Self-ONN replaces fixed convolutional kernels with learnable operational neurons that adapt their functional form during training. This design enables the network to model nonlinear transformations of the input signal more flexibly and to capture subtle variations in waveform morphology. Such properties are particularly beneficial for interference recognition tasks where differences

between signal classes may be characterized by small phase shifts, amplitude variations, or temporal distortions. The temporal branch receives normalized one-dimensional input signals and processes them through a sequence of operational layers, pooling operations, and nonlinear activations. The resulting feature vector encodes temporal dependencies and morphological characteristics relevant to interference classification.

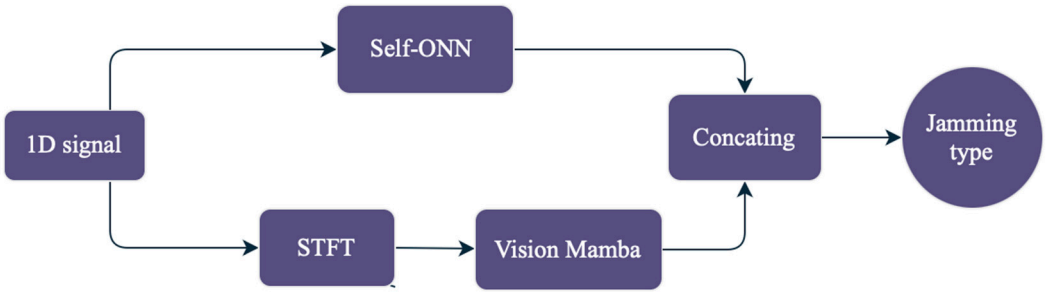


Figure 1. Overall architecture of the proposed hybrid interference recognition framework.

Unlike standard CNNs that rely on linear operations, the Self-ONN branch utilizes nonlinear operational neurons. The output of an operational neuron y can be represented as a nodal transformation [24]:

$$y = \varphi(x, \omega) + b, \tag{3}$$

where φ is a learnable nonlinear function, ω represents the weights, and b is the bias. This allows the network to approximate the complex nonlinear distortions of the interference signal with significantly fewer parameters, achieving high discriminative power in the temporal domain.

The second branch operates in the time-frequency domain. Each input signal is transformed into a two-dimensional spectrogram using STFT. This representation highlights spectral dynamics and transient patterns that are often indicative of specific interference types affecting GNSS and radio navigation signals. The generated spectrograms are processed by a Vision Mamba model based on state-space neural architectures. State-space models (SSMs) provide an efficient mechanism for capturing long-range dependencies in sequential data while maintaining lower computational complexity compared with conventional transformer-based architectures. The Vision Mamba encoder extracts discriminative spatial-temporal features from the spectrogram representation, enabling effective characterization of complex interference patterns such as interrupted sampling, repeater-based interference, and frequency-modulated signals.

The spatial stream leverages the Vision Mamba architecture, which replaces the attention mechanism of Transformers with a Selective Scan Mechanism (S6). The state-space model is formulated as [25]:

$$h(t) = \mathbf{A}h(t-1) + \mathbf{B}x(t) \tag{4}$$

$$y(t) = \mathbf{C}h(t) \tag{5}$$

where $h(t)$ denotes the hidden state vector at time step t, $x(t)$ is the input signal, $y(t)$ represents the output, and $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are learnable state-space matrices that define the system dynamics and observation model.

In the S6 mechanism, the matrices $\mathbf{B}$ and $\mathbf{C}$ are made input-dependent, allowing the model to selectively propagate or forget information across the spectrogram. This transition from quadratic $O(L^2)$ complexity of Transformers to linear $O(L)$ complexity is a key factor enabling a significant reduction in computational overhead compared with conventional transformer-based architectures.

The resulting feature embeddings represent global spectral-temporal information complementary to the temporal features extracted by the Self-ONN branch. A detailed structure of the proposed hybrid architecture is presented in Figure 2.

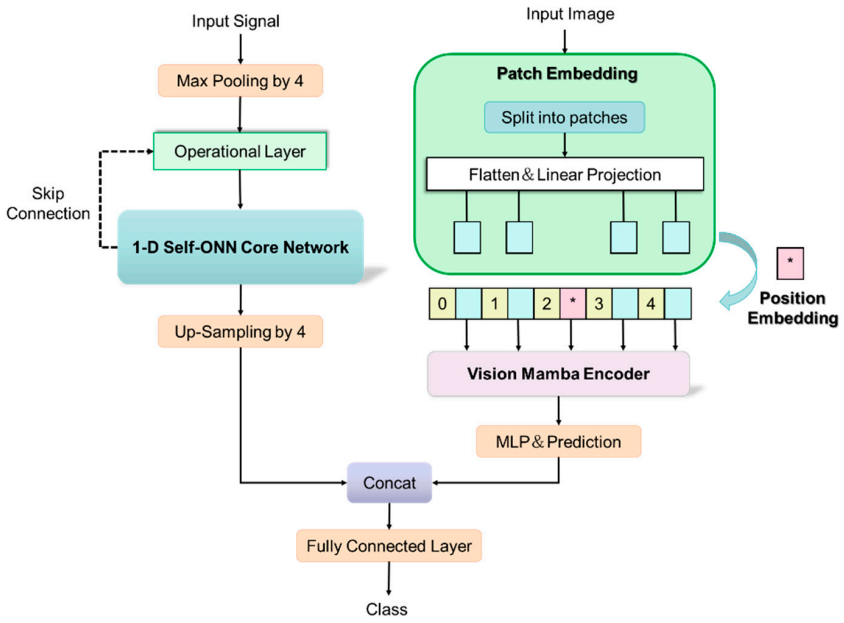


Figure 2. Detailed architecture of the hybrid Self-ONN and Vision Mamba model. * denotes a learnable position embedding vector that is added to each patch embedding to encode its spatial position.

The proposed classifier is defined as:

$$f_\theta(x) = g_\phi(f_T(x), f_S(S(x))), \tag{6}$$

where f_T—temporal encoder (Self-ONN); f_S—spectral encoder (Vision Mamba); and g_ϕ—fusion classifier.

After feature extraction in both branches, the resulting feature vectors are concatenated to form a unified representation. This fusion strategy combines temporal waveform characteristics with spectral-temporal features, enabling the model to exploit complementary information from both domains.

The fused feature vector is passed to a fully connected classification module consisting of dense layers with nonlinear activation functions and a final SoftMax layer. The output layer produces class probabilities corresponding to different interference types. This multimodal fusion strategy enhances classification accuracy and robustness by leveraging diverse signal representations while maintaining computational efficiency suitable for real-time deployment in GNSS and radio navigation systems.

2.4. Experimental Setup and Training Protocol

Model training was conducted using an NVIDIA GeForce RTX 2080 Ti GPU in a PyTorch v2.1.1-cu118 environment with Python 3.10.13. The AdamW optimizer was used with weight decay regularization and cosine annealing learning rate scheduling. The batch size was set to 48 and the models were trained for 10 epochs. Input spectrograms were resized to 224 × 224 pixels.

The Vision Mamba encoder used 24 layers with an embedding dimension of 192, while the Self-ONN branch processed one-dimensional input sequences with adaptive operational layers. This configuration ensures a balance between classification performance and computational efficiency, making the proposed framework suitable for real-time interference monitoring in GNSS-related applications. The model was optimized using Cross-

Entropy loss to minimize the discrepancy between the predicted probability distribution and the actual class labels of the GNSS interference signals. The described methodological framework provides a balanced trade-off between recognition accuracy and computational efficiency. Such characteristics are particularly important for real-time interference monitoring modules integrated into next-generation GNSS receivers and navigation systems.

3. Results and Discussion

3.1. Visualization of Interference Signals

To illustrate the characteristics of the considered interference signals, Figure 3 presents both one-dimensional waveforms and corresponding time-frequency spectrograms with additive Gaussian noise for all six interference types used during training and testing.

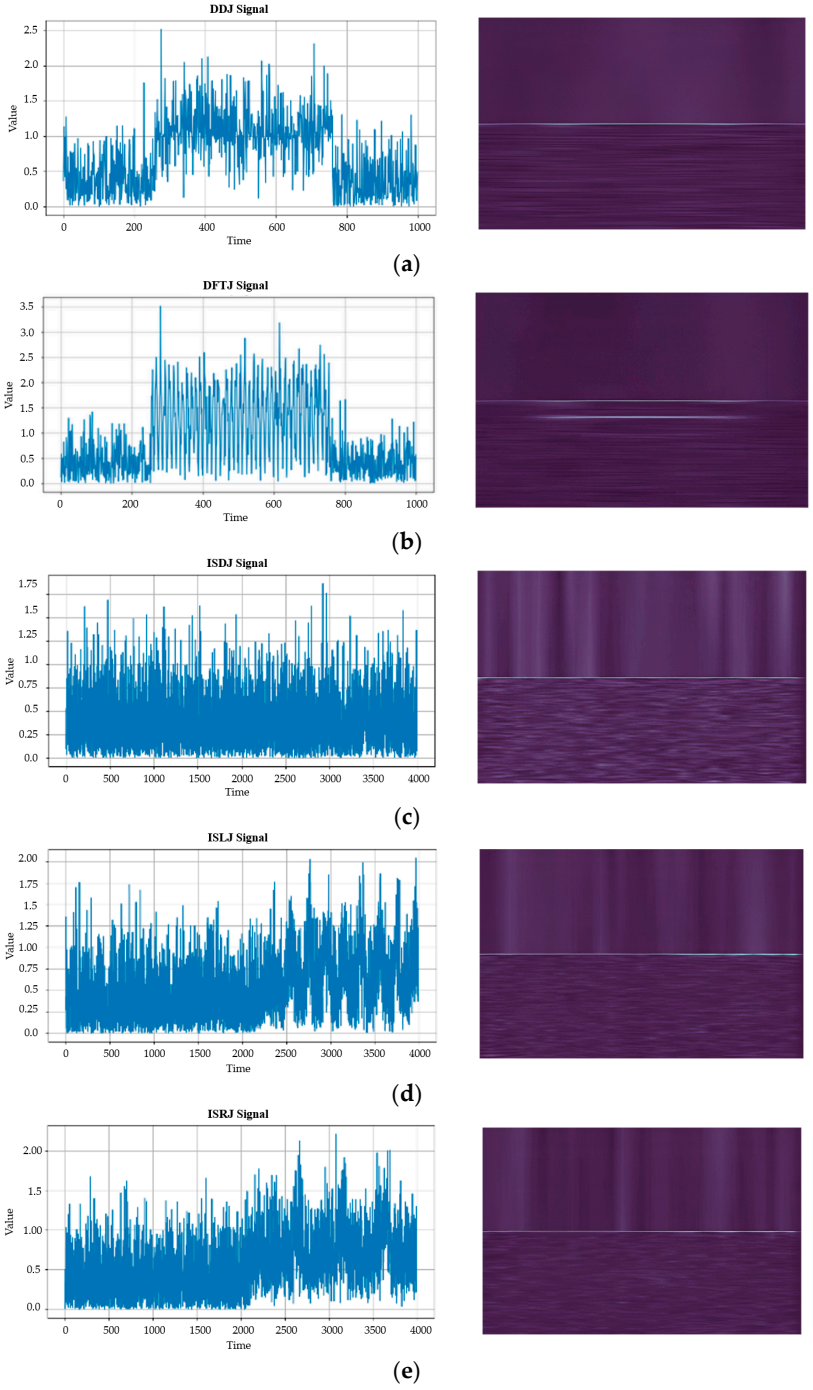
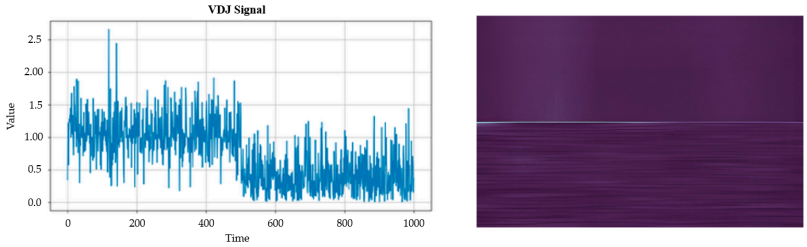


Figure 3. *Cont.*



(**f**)

Figure 3. Noisy interference signals and corresponding spectrograms for all considered classes: (**a**) DDJ, (**b**) DFTJ, (**c**) ISDJ, (**d**) ISLJ, (**e**) ISRJ, (**f**) VDJ.

These visualizations highlight distinctive temporal and spectral patterns associated with each interference class and demonstrate the complexity of the classification task. The presence of noise and structural similarities between certain interference types confirms the necessity of robust feature extraction methods capable of capturing both temporal waveform morphology and time-frequency characteristics. Such conditions closely reflect real-world GNSS interference scenarios, where received signals are often affected by noise, multipath propagation, and various distortions. The visual analysis demonstrates that several interference classes exhibit similar spectral signatures, making reliable classification challenging when relying on a single-domain representation. To quantitatively evaluate the performance of the proposed hybrid architecture and baseline models, we used accuracy A and the macro-averaged F1-score, as these metrics provide a comprehensive assessment of classification reliability across balanced datasets. These metrics are formally defined as follows:

Accuracy represents the proportion of correctly classified interference samples out of the total number of samples [26]:

$$\mathrm{A} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \tag{7}$$

The macro-averaged F1-score is defined as [26]:

$$\mathrm{F1}_{\mathrm{macro}} = \frac{2\mathrm{TP}}{2\mathrm{TP} + \mathrm{FP} + \mathrm{FN}}, \tag{8}$$

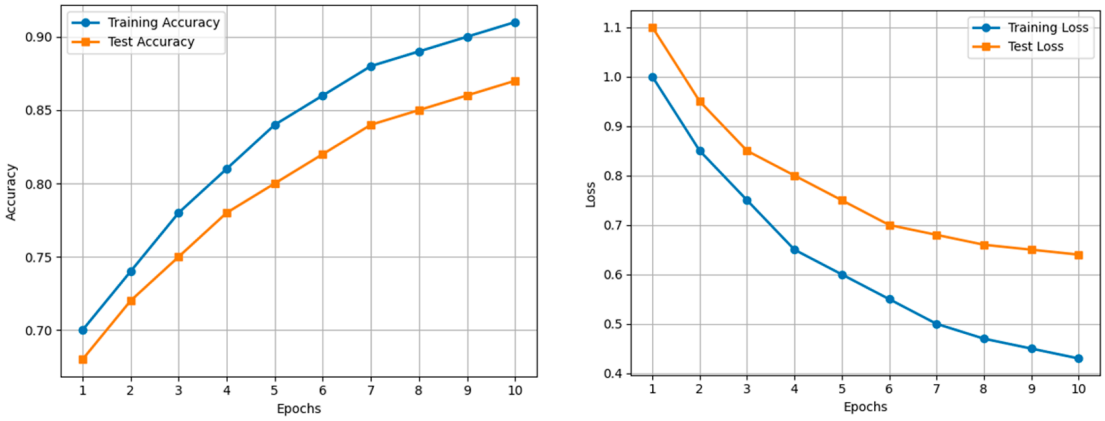
where TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively.

3.2. Baseline Performance of Vision Mamba

To establish a baseline for time-frequency domain classification, the Vision Mamba model was evaluated independently using spectrogram-based inputs. The model demonstrated stable convergence during training and achieved a classification accuracy of 92.5% and an unweighted F1-score (macro-averaged F1-score) of ~92.4% on the test dataset.

These results confirm that Vision Mamba effectively captures global spectral-temporal structures and transient interference patterns. However, its performance remains lower than that of deep convolutional architectures such as ResNet18 and VGG16, which achieved classification accuracies of 99.75% and 97.67%, respectively. This performance gap indicates that spectrogram-based representations alone may be insufficient for reliable discrimination between structurally similar interference types. In particular, interference classes such as ISDJ and ISRJ exhibit highly similar time-frequency patterns, resulting in increased classification ambiguity when only spectral features are considered. These findings highlight the limitations of single-modality approaches and motivate the integration of complementary temporal-domain representations capable of capturing subtle waveform variations. The

training and validation accuracy and loss curves for the Vision Mamba model are shown in Figure 4.



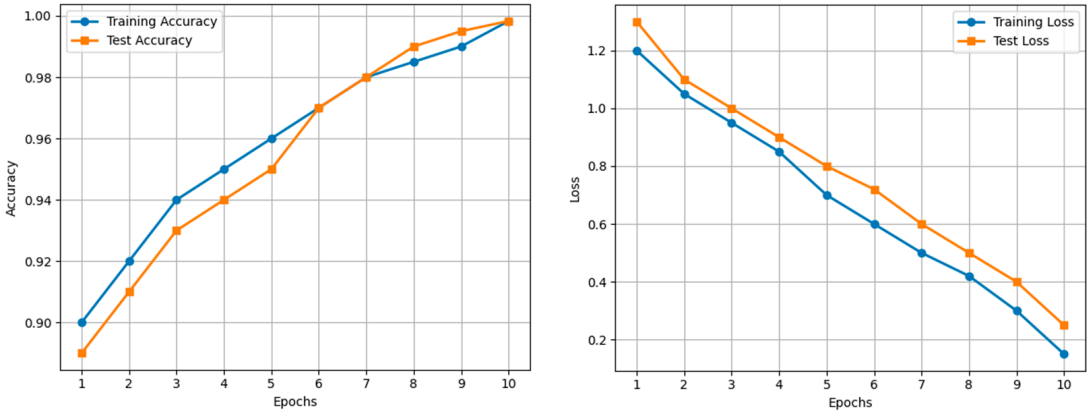
(**a**) (**b**)

Figure 4. Training and validation curves for the Vision Mamba model: (**a**) accuracy curve of Vision Mamba; (**b**) loss curve of Vision Mamba.

3.3. Performance of the Proposed Hybrid Architecture

The proposed hybrid architecture combining Self-ONN and Vision Mamba demonstrated superior performance across all evaluation metrics. The model achieved a classification accuracy of 99.83% and a macro-averaged F1-score of 99.83% on the test dataset, outperforming all baseline models. The observed improvement confirms that the fusion of temporal and time-frequency representations provides a more informative feature space for interference recognition.

The Self-ONN branch effectively captures fine-grained temporal waveform distortions and nonlinear signal dynamics, while the Vision Mamba branch extracts global spectral-temporal patterns. The combination of these complementary features enables more reliable discrimination between structurally similar interference classes. Training and validation curves for the hybrid model, presented in Figure 5, demonstrate stable convergence and minimal overfitting. Results were consistent across three independent training runs with variance below 0.05%.



(**a**) (**b**)

Figure 5. Training and validation curves for the proposed hybrid architecture: (**a**) accuracy curve; (**b**) loss curve.

3.4. Confusion Matrix Analysis

The confusion matrix of the hybrid model is presented in Figure 6. The results demonstrate near-perfect classification performance across all six interference categories. Only two samples from the ISRJ class were misclassified, most likely due to structural similarities to ISDJ signals. The confusion matrix confirms that the model achieves balanced performance across all classes without bias toward any particular interference type. This balanced classification performance is particularly important for GNSS interference monitoring applications, where reliable identification of diverse interference types is required for adaptive mitigation strategies. The minor confusion between certain classes arises from the subtle phase and frequency modulation overlaps in simulated GNSS environments, where deceptive jamming echoes resemble authentic signal reflections.

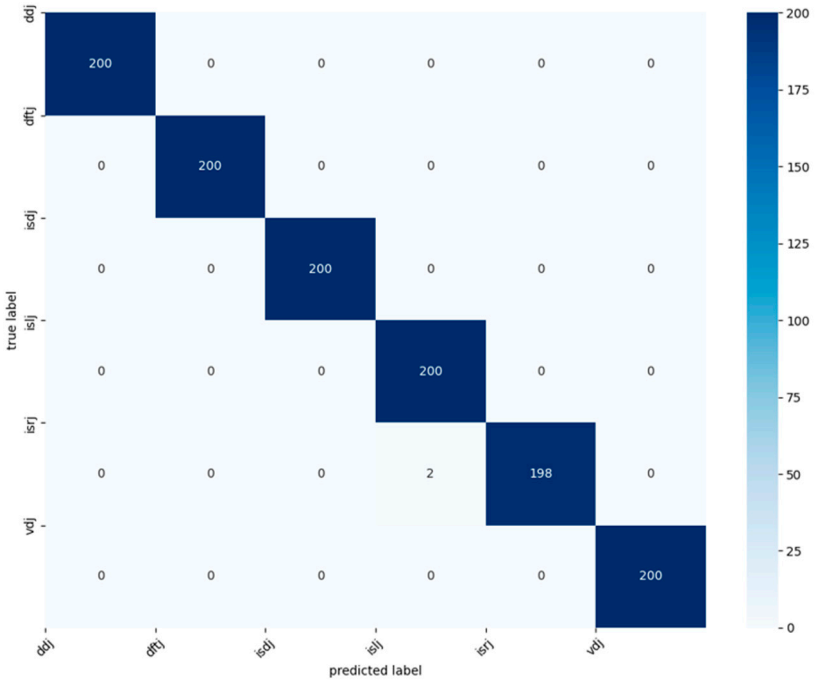


Figure 6. Confusion matrix of the proposed hybrid model.

3.5. Comparative Analysis with Existing Models

A comparative evaluation of all considered models is summarized in Table 2. The results clearly demonstrate that the proposed hybrid architecture achieves the highest classification accuracy and F1-score among all tested approaches. While ResNet18 and the standalone Self-ONN model individually achieved strong performance (~99.75%), the hybrid model provided further improvement due to the integration of complementary signal representations. Vision Transformer exhibited significantly lower performance on this dataset, indicating that purely attention-based architectures may be less effective for interference classification tasks where fine-grained temporal patterns play a critical role. The relatively low performance of the Vision Transformer can be attributed to the limited dataset size and the absence of large-scale pretraining, which is typically required for transformer-based architectures to achieve optimal performance. The Vision Mamba model demonstrated competitive performance with reduced computational complexity but benefited significantly from integration with the temporal Self-ONN branch. The hybrid approach therefore combines high classification accuracy with improved computational efficiency, making it suitable for real-time deployment.

Table 2. Comparative performance of evaluated models.

Model	Accuracy	Macro-Averaged F1-Score
VGG16	0.9767	0.9766
ResNet18	0.9975	0.9975
Vision transformer	0.4267	0.3369
ONN	0.9975	0.9975
VisionMamba	0.9250	0.9240
Proposed hybrid model	0.9983	0.9983

The superior performance of the hybrid architecture can be explained by the complementary nature of temporal and time-frequency representations. While spectrogram-based models capture global spectral patterns, temporal-domain processing preserves fine waveform distortions that may be partially lost during transformation. The integration of these representations enables the model to construct a more discriminative feature space, improving robustness to noise and structural similarity between interference types. This multimodal representation is particularly relevant for GNSS interference scenarios, where signal distortions may manifest simultaneously in both temporal and spectral domains.

3.6. Robustness to Noise and Discussion

Additional experiments with additive white Gaussian noise (AWGN) were conducted to evaluate model robustness under low signal-to-noise ratio conditions. The hybrid model maintained classification accuracy above 97% even at reduced SNR levels, while single-domain models exhibited more significant performance degradation. This demonstrates that multimodal feature fusion provides redundancy that improves resilience to signal distortions. From a practical perspective, the proposed hybrid framework offers a favorable balance between recognition accuracy and computational efficiency. The reduced parameter counts and lower computational complexity compared with transformer-based architectures make the model suitable for real-time interference monitoring modules in next-generation GNSS receiver systems. These results confirm the feasibility of integrating the proposed hybrid interference recognition framework into high-precision GNSS receivers for real-time signal integrity monitoring and interference mitigation. Despite the high classification accuracy, several limitations should be noted. The evaluation was conducted using synthetically generated signals, which may not fully represent all distortions present in real GNSS environments. In addition, the use of STFT introduces sensitivity to preprocessing parameters such as window size and overlap. Future work will focus on validation using real GNSS signal recordings, evaluation under varying channel conditions, and the development of adaptive cross-domain fusion mechanisms. Although the dataset is synthetically generated, the modeled interference patterns reflect distortions commonly observed in GNSS receivers operating in contested electromagnetic environments, including repeater-based interference, spoofing-like structures, and broadband jamming scenarios.

The reported performance should be interpreted as an upper-bound under controlled simulation conditions. In real GNSS environments, factors such as RF front-end distortions, multipath propagation, oscillator instability, and non-Gaussian interference may affect signal characteristics and reduce classification accuracy. Therefore, further validation using real GNSS recordings and publicly available RF datasets is required to assess practical applicability. From a computational perspective, the proposed architecture balances classification performance and efficiency. Vision Mamba enables linear-complexity sequence modeling, while the Self-ONN branch provides expressive nonlinear feature extraction with a relatively low parameter count. This results in reduced computational overhead compared with transformer-based approaches, supporting its suitability for deployment

in resource-constrained GNSS receivers. However, direct embedded latency and memory evaluations remain part of future work.

Although the dataset size is moderate, overfitting was mitigated through stratified data splitting, consistent results across multiple runs, and stable convergence with minimal divergence between training and validation curves. Nevertheless, larger datasets and data augmentation may further improve robustness. From a system integration perspective, the framework can be incorporated into GNSS receivers as an interference monitoring module alongside the signal tracking stage. The dual-branch design enables efficient processing of temporal and time-frequency representations, supporting real-time operation and enabling adaptive interference mitigation to improve positioning reliability in challenging environments.

The choice of STFT parameters may influence the resulting time-frequency representation, and a systematic sensitivity analysis of window size and overlap will be considered in future work. The improved classification performance is achieved at the cost of increased architectural complexity due to the dual-branch design and the need for preprocessing in both temporal and time-frequency domains. This trade-off reflects a balance between accuracy and computational simplicity, which is important for practical deployment considerations. Future work will include validation using real GNSS interference data collected in practical electromagnetic environments to further assess model generalization and deployment performance.

4. Conclusions

This study presented a hybrid deep learning framework for robust radio-frequency interference recognition in high-precision GNSS applications. The proposed architecture combines temporal-domain processing using Self-ONN with time-frequency feature extraction based on a Vision Mamba state-space model. By integrating complementary signal representations within a multimodal framework, the approach achieves high classification accuracy while maintaining computational efficiency suitable for real-time deployment.

Experimental results show that the proposed architecture outperforms conventional deep learning models, including ResNet18, VGG16, Vision Transformer, and standalone Vision Mamba. The model achieved a classification accuracy and macro-averaged F1-score of 99.83%, demonstrating the effectiveness of multimodal feature fusion for distinguishing structurally similar interference types under noisy conditions. Confusion matrix analysis confirmed balanced performance across all interference classes with minimal misclassification.

The results indicate that combining temporal waveform analysis with time–frequency representations provides a more discriminative feature space for interference recognition while maintaining a favorable balance between accuracy and computational complexity. Although the evaluation was conducted on synthetically generated signals, future work will focus on validation using real GNSS recordings, testing under diverse channel conditions, and further optimization for embedded deployment in intelligent GNSS receivers.

Author Contributions: Conceptualization, M.I. and B.K. (Bakyt Khaniyev); methodology, N.M. and B.K. (Beibit Karibayev); software, T.D., N.U. and A.S.; validation, T.D., N.U. and A.S.; formal analysis, B.K. (Beibit Karibayev); writing—original draft preparation, S.K. and B.K. (Bagdat Kozhakhmetova); writing—review and editing, B.K. (Bakyt Khaniyev) and A.K.; visualization, B.K. (Bagdat Kozhakhmetova), A.K. and T.D.; supervision, N.M. and M.I.; project administration, N.M.; funding acquisition, N.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Committee of Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan, grant number AP23487878.

Data Availability Statement: Data is contained within the article.

Conflicts of Interest: Author Nursultan Uzbekov was employed by the company Orion System LLP. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

RFI	Radio-frequency Interference
GNSS	Global Navigation Satellite System
Self-ONN	Self-Organizing Operational Neural Network
STFT	Short-Time Fourier Transform
SNR	Signal-to-Noise Ratio
CNN	Convolutional Neural Network
JNR	Jammer-to-Noise Ratio
GPU	Graphics Processing Unit
WPD	Wavelet Packet Decomposition
DDJ	Digital Deceptive Jamming
VDJ	Velocity Deception Jamming
DFTJ	Distance Frequency Time Jamming
ISDJ	Interrupted Sampling Deceptive Jamming
ISRJ	Interrupted Sampling Repeater Jamming
ISLJ	Interrupted Sampling Linear Jamming
LFM	Linear Frequency Modulation
SSM	State-space Model
AWGN	Additive White Gaussian Noise

References

1. Gao, N.; Yan, Z.; Wang, Y. Performance enhancement of GNSS receiver using an INS-aided tracking loop for spinning vehicle. *Aerosp. Sci. Technol.* **2026**, *175*, 112036. [CrossRef]
2. Huang, J.; Yang, R.; Zhan, X. Constraint navigation filter for space vehicle autonomous positioning with deficient GNSS measurements. *Aerosp. Sci. Technol.* **2022**, *120*, 107291. [CrossRef]
3. Saurova, K.; Alipbayev, K.; Nysanbaeva, S.; Karibayev, B.; Khaniyev, B.; Uzbekov, N.; Orynbassar, S.; Geroyeva, Z.; Mukhamedgali, A.; Meirambekuly, N. Research of methods to improve the accuracy of the star sensor for global navigation satellite system technology. *Eng. Sci.* **2024**, *33*, 1351. [CrossRef]
4. Jiang, W.; Chen, Y.; Chen, Q.; Chen, H.; Pan, Y.; Liu, X.; Liu, T. High precision deformation monitoring with integrated GNSS and ground range observations in harsh environment. *Measurement* **2022**, *204*, 112179. [CrossRef]
5. Sukhenko, A.; Meirambekuly, N.; Syzdykov, A.; Mukhamedgali, A.; Mellatova, Y. GNSS for High-Precision and Reliable Positioning: A Review of Correction Techniques and System Architectures. *Appl. Sci.* **2025**, *15*, 12304. [CrossRef]
6. Humphreys, T.E. Detection strategy for cryptographic GNSS anti-spoofing. *IEEE Trans. Aerosp. Electron. Syst.* **2013**, *49*, 1073–1090. [CrossRef]
7. Psiaki, M.L.; Humphreys, T.E. GNSS spoofing and detection. *Proc. IEEE* **2016**, *104*, 1258–1270. [CrossRef]
8. Broumandan, A.; Jafarnia-Jahromi, A.; Dehghanian, V.; Nielsen, J.; Lachapelle, G. GNSS spoofing detection in handheld receivers based on signal spatial correlation. In Proceedings of the 2012 IEEE/ION Position, Location and Navigation Symposium, Myrtle Beach, SC, USA, 23–26 April 2012; pp. 479–487.
9. Ivanov, V.; Scaramuzza, M.; Wilson, R.C. Deep temporal semi-supervised one-class classification for GNSS radio frequency interference detection. *J. Navig.* **2024**, *77*, 59–81. [CrossRef]
10. Dovis, F.; Musumeci, L. Use of the Karhunen–Loève transform for interference detection and mitigation in GNSS. *ICT Express* **2016**, *2*, 33–36. [CrossRef]
11. Sharifi-Tehrani, O.; Sabahi, M.F.; Danaee, M.R. GNSS jamming detection of UAV ground control station using random matrix theory. *ICT Express* **2021**, *7*, 239–243. [CrossRef]
12. Dutta, P.; Halder, T.; Banerjee, S.; Basak, A.; Nanda, S.; Chakravarty, D. Analysis of jamming and anti jamming techniques for Galileo GNSS. *Mater. Today Proc.* **2022**, *58*, 489–495. [CrossRef]

13. Wu, Q.; Li, G.; Wang, X.; Luo, H.; Li, L.; Chen, Q.; Jing, X. Jamming recognition method based on wavelet packet decomposition and improved deep learning. *Digit. Commun. Netw.* **2025**, *14*, 1469–1478. [CrossRef]
14. Skabylov, A.; Zhexebay, D.; Khokhlov, S.; Agishev, A.; Abdizhalilova, L.; Zhakipova, M.; Azamat, R.; Orazakova, A.; Qin, Y.; Ibraimov, M. Seismic P-Wave Detection Using CWT and Deep Image Classification With YOLO. *IEEE Access* **2025**, *13*, 181267–181285. [CrossRef]
15. Cai, X.; Li, P.; Liu, M.; Chen, Y.; Lu, J. Radio jamming recognition algorithm based on MS-SSA and the CSA-CNN. *Digit. Signal Process.* **2025**, *159*, 105019. [CrossRef]
16. Blais, A.; Couellan, N.; Munin, E. A novel image representation of GNSS correlation for deep learning multipath detection. *Array* **2022**, *14*, 100167. [CrossRef]
17. Gyagenda, N.; Hatilima, J.V.; Roth, H.; Zhmud, V. A review of GNSS-independent UAV navigation techniques. *Robot. Auton. Syst.* **2022**, *152*, 104069. [CrossRef]
18. Chen, X.Y.; Liu, Y.; Wang, C. DRFM-based jamming signal recognition method guided by target detection. *Procedia Comput. Sci.* **2023**, *221*, 1013–1020. [CrossRef]
19. Elyousseph, H.; Altamimi, M.L. Deep learning radio frequency signal classification with hybrid images. In Proceedings of the 2021 IEEE International Conference on Signal and Image Processing Applications (ICSIPA), Kuala Terengganu, Malaysia, 13–15 September 2021; pp. 7–11. [CrossRef]
20. Wang, P.; Ma, K.; Bai, Y.; Sun, C.; Wang, Z.; Chen, S. Wireless interference recognition with multimodal learning. *IEEE Trans. Wirel. Commun.* **2024**, *23*, 18576–18591. [CrossRef]
21. Gu, A.; Goel, K.; Ré, C. Efficiently modeling long sequences with structured state spaces. *Int. Conf. Learn. Represent. (ICLR)* **2021**, *32*. [CrossRef]
22. Shang, W.; Guo, Y.; Yang, L. Multi-component LFM signal representation method under impulsive noise: Principle, method and application. *Digit. Signal Process.* **2026**, *173*, 105868. [CrossRef]
23. Goyal, D.; Pabla, B.S. Condition based maintenance of machine tools—A review. *CIRP J. Manuf. Sci. Technol.* **2015**, *10*, 24–35. [CrossRef]
24. Mohammed, H.H.; Malik, J.; Al-Maadeed, S.; Kiranyaz, S. 2D self-organized ONN model for handwritten text recognition. *Appl. Soft Comput.* **2022**, *127*, 109311. [CrossRef]
25. Cai, C.; Zhang, L.; Zhou, J.; Zhou, L. Sky images based photovoltaic power forecasting: A novel approach with optimized VMD and Vision Mamba. *Results Eng.* **2024**, *24*, 103022. [CrossRef]
26. Göger, S.; Darol, E.S.; Uzun, S. Multiple screen addiction and neurological complaints in adolescents: A machine learning-based classification model. *Acta Psychol.* **2026**, *264*, 106491. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.