Computer Modeling in
Engineering & Sciences

Tech Science Press

https://doi.org/10.32604/cmes.2026.083624

ARTICLE

A Lightweight Manifold-Aware State Space Model for Efficient Seismic Signal Classification at the Edge

Pingan Peng[1,2], Qi Zhang[1,*], Ya Liu[1], Yue Han[1], Zhida Jiang[1], Linli Chen[1] and Xuefeng Huo[1]

[1]School of Resources and Safety Engineering, Central South University, Changsha, China
[2]State Key Laboratory for Fine Exploration and Intelligent Development of Coal Resources, China University of Mining and Technology, Xuzhou, China

*Corresponding Author: Qi Zhang. Email: 255512048@csu.edu.cn

Received: 07 April 2026; Accepted: 30 July 2026; Published: 28 August 2026

ABSTRACT: Unfilled goafs located beneath urban areas pose a significant threat to surface safety, and microseismic monitoring is an important tool for capturing rock-mass microfractures and supporting early warning. However, under field conditions, existing sequential models face challenges such as high computational complexity and deep feature degradation when deployed on edge devices, primarily due to the non-stationary characteristics of microseismic signals and the presence of complex ambient background noise. To address these issues, this paper proposes the Hierarchical Network with Manifold Hybrid Connection (H-NET-mHC), a lightweight manifold-aware state space model designed for edge computing. The model introduces a content-aware dynamic chunking strategy to adaptively identify informative local segments, effectively suppressing noise and extracting key phases while compressing long sequences. A manifold hybrid connection (mHC) module concurrently constrains feature-flow propagation using doubly stochastic matrices, thereby mitigating nonlinear degradation of high-dimensional features and stabilizing network training dynamics. Multiclass classification and ablation experiments using real-world data from a gypsum mine in Changde demonstrate that, with only 0.35 million parameters, H-NET-mHC achieves an accuracy of 97.77% and a microseismic-event recall of 96.11%. Under severe class imbalance, the two proposed mechanisms produce a complementary recall-oriented effect rather than an across-the-board improvement in all aggregate metrics. The resulting architecture balances computational cost and classification performance, supporting the engineering requirements for intelligent perception and early warning of mine microseismic signals.

KEYWORDS: Microseismic monitoring; lightweight; edge computing; state space model; dynamic chunking; manifold hybrid connection

1 Introduction

The extraction of mineral resources has promoted economic development. Due to the limitations of early mining technologies, some mining areas were not backfilled after the completion of extraction [1]. With the expansion of urban areas, unfilled goafs located beneath urban built-up areas pose significant risks to urban safety [2]. Because gypsum rock masses are rheological and water-soluble, these goafs may subside [3]. Such subsidence can damage urban road networks, underground utility corridors, and surface buildings, thereby threatening public safety [4].

To support early hazard warning, microseismic monitoring is used to capture elastic waves released by rock-mass micro-fractures. Previous field studies have demonstrated that microseismic monitoring can support rock-mass hazard assessment and stability evaluation by characterizing damage initiation, damage evolution, and energy release processes. For example, Ma et al. combined mechanical analysis, numerical simulation, and microseismic monitoring to evaluate coal-floor water-inrush hazards, while another study used microseismic energy density and magnitude-frequency analysis to assess the stability of a high-steep rock slope and the effectiveness of its reinforcement [5,6]. In urban mining environments, however, true events must be separated from traffic, machinery, and other ambient disturbances. Beyond fracture monitoring, microseismic observations have also been used to investigate subsurface migration and leakage processes [7]. Statistical phase detectors provide interpretable arrival decisions [8], while recurrent classifiers model temporal dependencies [9]. End-to-end detection and picking frameworks reduce manual feature engineering [10,11], and joint detection-picking methods further integrate related seismic tasks [12]. Convolutional and multi-scale networks capture local waveform patterns [13,14]. Transformer-based models extend the contextual receptive field [15,16], building on the general self-attention architecture [17]. Attention refinements emphasize informative phases [18,19]. Generative and lightweight approaches have also been explored for robustness and efficient recognition [20,21]. Despite these advances, recurrent models remain constrained by serial computation [22], convolutional models have limited long-range context, and Transformer or generative architectures can be expensive for resource-constrained nodes.

Microseismic and seismic records are also strongly non-stationary: their frequency content and transient structure vary over time. Time-frequency and decomposition analyses have been used to characterize this variability [23,24]. Related studies investigate transient spectral representations [25,26] and noise-robust feature extraction [27,28]. Traditional methods based on characteristic parameters [29] or wavelets [30] remain useful but require handcrafted representations. Deep networks may exhibit optimization instability during nonlinear stacking [31,32], while repeated transformations can also degrade useful features [33,34]. Structured state space models (SSMs) provide an efficient foundation for long-sequence modeling [35]. Mamba introduces input-dependent selective state transitions with linear sequence complexity [36]. Vision Mamba further demonstrates the adaptability of selective SSMs to structured signals [37]. These properties offer a promising basis for edge-oriented one-dimensional (1D) signal analysis. Table 1 summarizes the technical routes most directly related to the present study.

Table 1: Technical context and research gap relevant to H-NET-mHC.

Technical Route	Representative Studies	Main Contribution	Limitation for This Task
Statistical and wavelet processing	P-phase arrival identification scheme (PAI-S/K) [8]; characteristic parameters [29]; Haar wavelets [30]	Interpretable event and arrival features	Relies on handcrafted representations and thresholds
Convolutional /recurrent neural network event recognition	Long short-term memory [9]; convolutional classifiers [13,14]	Learns temporal or local multi-scale patterns	Serial recurrence or limited long-range context
Attention and end-to-end models	End-to-end picking [10,11]; joint tasks [12]; Transformer models [15,16]; attention refinements [18,19]	Strong contextual and multi-task representation	Higher compute and memory cost for long sequences

(Continued)

Table 1 (continued)

Technical Route	Representative Studies	Main Contribution	Limitation for This Task
Lightweight neural networks	Kernel evaluation [21]; MobileNet V3 [38]; ShuffleNet V2 [39]	Compact operators and mobile-oriented design	Primarily developed for static image features
Structured state space models	Mamba and Vision Mamba [36,37]	Input-dependent long-range modeling with linear complexity	Needs task-specific temporal compression and stable deep fusion

In response to these challenges, this study proposes H-NET-mHC, a lightweight, manifold-aware, hierarchical 1D SSM for single-channel microseismic classification in an underground gypsum mine in Changde, China. The model is designed around the task-specific coupling of content-aware temporal compression and mHC-stabilized feature propagation. Dynamic chunking is integrated between hierarchical SSM stages to reduce sequence length while preserving transient phase regions under complex ambient noise, whereas the mHC unit is inserted after each normalized Mamba update to regulate high-variance feature propagation through doubly stochastic mixing constraints. By combining phase-aware compression with stable deep feature evolution, H-NET-mHC provides a compact architecture for edge-oriented microseismic signal classification. The 0.35M-parameter model was evaluated on a severely imbalanced real-world dataset using overall accuracy, class-wise metrics, ablation analysis, and central processing unit (CPU)-only inference measurements. The results indicate its potential for efficient signal recognition in resource-constrained monitoring scenarios.

2 Methodology

2.1 Preliminaries

The core architecture of the hierarchical network (H-NET) is built on structured SSMs. Inspired by classical control theory, SSMs aim to map a one-dimensional input signal $x(t) \in \mathrm{R}$ to an output response $y(t) \in \mathrm{R}$ through an implicit latent state $h(t) \in \mathrm{R}^{N}$. In the continuous-time domain, the evolution process of a linear time-invariant system is described by the following ordinary differential equation:

$$\dot{h}(t) = Ah(t) + Bx(t), \quad (1)$$

$$y(t) = Ch(t), \quad (2)$$

where $A \in \mathrm{R}^{N\times N}$ is the state evolution matrix, typically initialized as a high-order polynomial projection operator (HiPPO) matrix to effectively capture long-range historical information [40]; $B \in \mathrm{R}^{N\times 1}$ and $C \in \mathrm{R}^{1\times N}$ are the input and output projection matrices, respectively. However, the signals acquired by the microseismic monitoring system form a discrete sequence after analog-to-digital conversion. To process this signal on a digital computer, the aforementioned continuous system must be discretized using a time step Δ. By adopting the zero-order hold (ZOH) principle, the discretized parameters $\overline{A}$ and $\overline{B}$ can be derived from the following equations.

$$\overline{A} = \exp\left(\Delta A\right), \tag{3}$$

$$\overline{B} = \left(\Delta A\right)^{-1}\left(\exp\left(\Delta A\right) - I\right) \cdot \Delta B. \tag{4}$$

The discretized SSM supports two equivalent computational forms for training and inference.

Linear Recurrent Form

During inference on edge devices, the model operates sequentially across time steps, incrementally processing the arriving microseismic data points with a time complexity of $O(1)$. This satisfies the low-latency requirements of real-time monitoring.

$$h_t = \overline{A}h_{t-1} + \overline{B}x_t, \tag{5}$$

$$y_t = Ch_t. \tag{6}$$

Convolutional Parallel Form

During model training, recurrent computation struggles to exploit the parallel computing power of graphics processing units (GPUs). By expanding the recursive formula, the output y can be represented as a convolution operation between the input sequence x and a global convolution kernel $\overline{K}$.

$$y = x * \overline{K}, \tag{7}$$

$$\overline{K} = C\overline{B}, C\overline{A}\overline{B}, \ldots, C\overline{A}^{L-1}\overline{B}, \tag{8}$$

where L is the input sequence length. This form allows for parallel acceleration using the fast Fourier transform, effectively bypassing the serial computation bottleneck of recurrent neural networks (RNNs). In traditional SSMs, such as the Structured State Space Sequence model (S4), the matrices A, B, and C are time-invariant and independent of the input. However, the Mamba architecture adopted by H-NET introduces an input-dependent selection mechanism, making the parameters Δ, B, and C functions of the input x_t. Consequently, the model can dynamically adjust the state compression rate based on the current signal content, thereby achieving content-based information filtering.

2.2 Overall Architecture

As shown in Fig. 1, the H-NET-mHC constructed in this paper is a lightweight hierarchical network designed for the computational constraints of edge devices. To address the characteristics of microseismic signals, such as long sequences, strong non-stationarity, and complex ambient background noise, the architecture is built on SSMs and uses a multi-scale processing framework that balances local features with global long-range dependencies. The main network components and their functions are described below.

Stem Layer for Shallow Feature Mapping: The stem performs initial dimensionality reduction and extracts features from the one-dimensional microseismic waveform. Through convolution operations, the original time-domain signal is projected into a high-dimensional feature space. While extracting initial local high-frequency variations, this operation reduces the sequence length and lowers the computational overhead for subsequent state space modeling.

Hierarchical Stages 1–3 for Temporal Encoding: These stages capture multiscale features of signal evolution over time. The network expands the channel dimensions and compresses the temporal dimensions through three progressive stages by stacking H-NET-mHC blocks. This pyramid-like structural design allows the model to extract transient local features in the shallow stages and correlate phase information in the deep stages, thereby forming a global representation of the microseismic event evolution process.

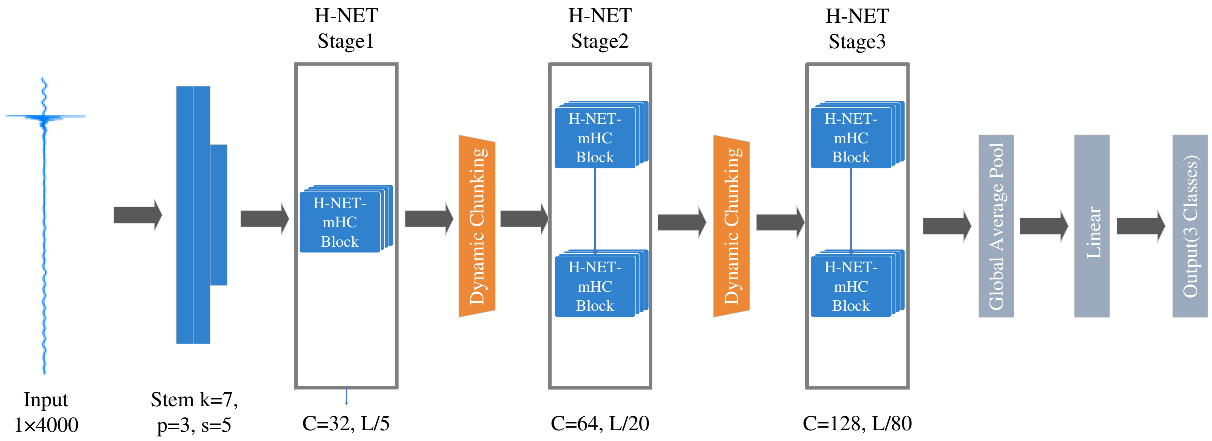

Figure 1: Overall architecture of H-NET-mHC.

Content-Aware Dynamic Chunking: This module downsamples features between encoding stages. Traditional static pooling poses a risk of losing critical information in high-noise environments; instead, this module dynamically assigns weights to informative waveform peaks based on the energy distribution of the signal content. Its role is to compress the sequence length while reducing the dilution of weak signals by background noise.

Global Classification Head: This head aggregates the final features and performs event classification. Temporal features are mapped to a single global context vector through a global average pooling layer to eliminate time-step differences. Finally, a linear classifier produces a three-class logit vector for the current signal segment, corresponding to a microseismic event, deterministic vibration, and noise.

For reproducibility, the main architectural settings are specified as follows. Each Mamba block uses d_state = 16, d_conv = 4, an expansion factor of 2, and dt_rank = ceil (d_model/16). The three hierarchical stages have channel dimensions of 32, 64, and 128, with depths of 1, 2, and 2, respectively. The stem layer is implemented as a one-dimensional convolution with a kernel size of 7, padding of 3, and stride of 5. Within each of the five Mamba blocks, the feature flow is processed in the order root mean square layer normalization (RMSNorm), Mamba update, and mHC fusion with the original feature flow.

2.3 Content-Aware Dynamic Chunking

In hierarchical sequence modeling, progressive downsampling along the temporal dimension as network depth increases is a common strategy used to reduce computational complexity and expand the model's effective receptive field. However, traditional convolutional networks typically employ static pooling or fixed-stride convolutions to achieve sequence compression. For the microseismic monitoring task, such content-agnostic operations may not accommodate pronounced signal nonstationarity. Under complex ambient noise, fixed-rule temporal compression may attenuate key phase segments while propagating high-energy noise into deeper features, thereby degrading classification and representation.

To improve the representation of transient phases, this paper adopts and adapts the lightweight content-aware dynamic chunking concept introduced in [41], as shown in Fig. 2. Instead of using fixed-stride downsampling, this module introduces a data-driven local importance evaluation mechanism: it first estimates the importance of local temporal segments, and then performs feature aggregation and downsampling based on these weights to achieve adaptive temporal aggregation.

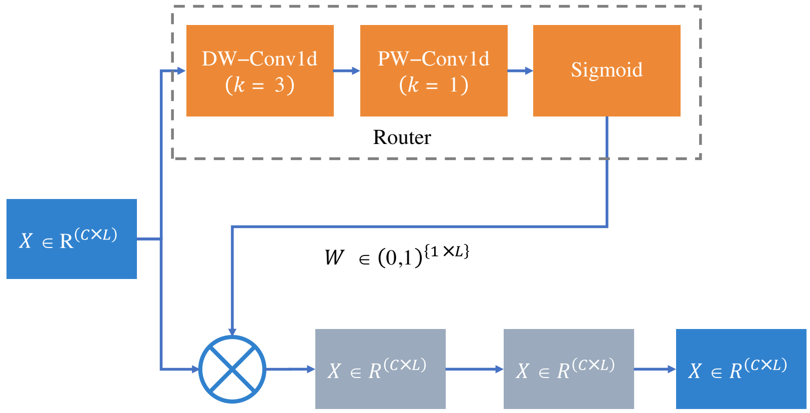

Figure 2: Algorithm structure diagram of the content-aware dynamic chunking layer.

Specifically, given an input feature sequence $X \in R^{C\times L}$, where C is the number of channels and L is the sequence length, a lightweight router calculates an importance weight for each time step. The calculation process of this router is as follows.

$$W = \sigma\left(\mathrm{Conv}_{pw}\left(\mathrm{Conv}_{dw}\left(X\right)\right)\right), \tag{9}$$

where Conv_{dw} denotes a depthwise convolution with a kernel size of 3, which captures local temporal context with few parameters; Conv_{pw} is a pointwise convolution with a kernel size of 1, which fuses channel information and maps it to a scalar response; and σ denotes the sigmoid activation function. The values in the resulting weight matrix $W \in R^{1\times L}$ lie in the range (0, 1) and characterize the relative contribution of different time steps to the classification task.

After obtaining the weight response, element-wise weighting is performed on the original features:

$$X_{weighted} = X \odot W, \tag{10}$$

where ⊙ denotes element-wise multiplication. This mechanism can reduce the impact of noise segments on deep representations at the feature level and enhance the response intensity of potential phase regions, thereby improving the model's representation capability for weak microseismic events. Finally, structured chunking-based downsampling and normalization are applied to the weighted features $X_{weighted}$.

$$X_{down} = \mathrm{LayerNorm}\left(\mathrm{Conv}_{down}\left(X_{weighted}\right)\right), \tag{11}$$

where Conv_{down} is a downsampling convolution operation with a stride of s, which compresses the sequence length to L/s while simultaneously completing the cross-stage mapping of feature dimensions. Subsequently, layer normalization is applied to provide a stable input for the next stage. Overall, this module transforms the fixed temporal compression process into an adaptive aggregation process based on local content responses, improving the model's feature retention capability and robustness under strong background noise while adding little computational overhead.

In the implementation, both content-aware dynamic chunking layers use a depthwise router kernel of 3 followed by a pointwise 1 × 1 projection. The subsequent downsampling convolution uses a kernel size of 4 and a stride of 4, thereby reducing the temporal length between hierarchical stages while completing the channel transformation.

2.4 Manifold Hybrid Connection

In the design of deep sequence models, the residual connection is typically formulated as $x_{out} = x + f(x)$. However, when processing non-stationary microseismic signals, deep models face risks of training instability and feature degradation. Traditional linear superposition leads to the accumulation of feature variance as the network depth increases, which may trigger anomalies in the state flow. To alleviate this problem and maintain the stable evolution of the hidden states in state space models, this paper introduces the mHC module.

As shown in Fig. 3, mHC constrains feature fusion using channel-wise mixing matrices projected onto the Birkhoff polytope, the convex set of doubly stochastic matrices. For a given input feature flow $x \in R^{B\times L\times D}$ and the update flow $\Delta \in R^{B\times L\times D}$ processed by the Mamba Block, mHC combines the two flows using channel-wise dynamic weights. To generate mixing weights that satisfy the doubly stochastic constraints, this paper introduces the Sinkhorn-Knopp matrix scaling algorithm [42].

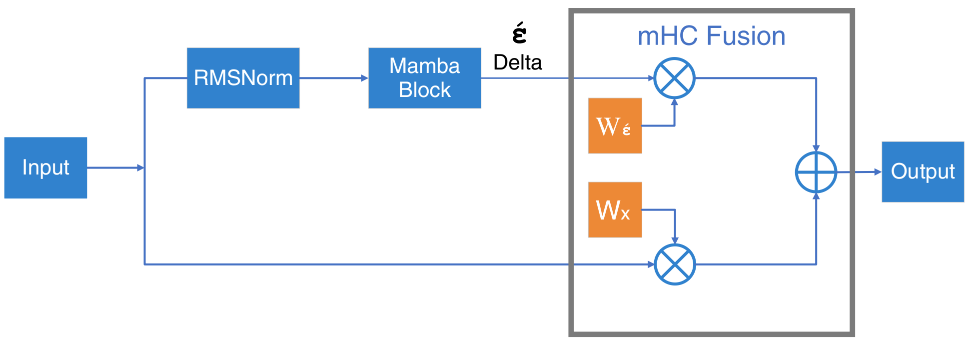

Figure 3: Algorithm structure diagram of the manifold hybrid connection.

Specifically, a learnable parameter tensor $\alpha \in R^{D\times 2\times 2}$ is initialized to represent the interaction between the original flow and the update flow across the feature channels. To ensure the non-negativity of the weights, an exponential mapping is first applied.

$$M = \exp(\alpha). \tag{12}$$

Subsequently, row normalization and column normalization are alternately performed on matrix M. In practical computations, this iterative process is limited to K iterations:

$$M \leftarrow \frac{M}{\sum_j M_{ij} + \varepsilon}, \tag{13}$$

$$M \leftarrow \frac{M}{\sum_i M_{ij} + \epsilon}, \tag{14}$$

where ϵ is a constant set to 10^{-6} to avoid division by zero. After K iterations, this matrix converges toward a doubly stochastic matrix within the Birkhoff polytope $P \in \mathrm{R}^{D\times 2\times 2}$, meaning that the slices for each channel simultaneously satisfy the condition that both row sums and column sums equal 1, i.e., $\sum_i P_{c,ij} = \sum_j P_{c,ij} = 1$.

After obtaining the transition matrix $P = \text{Sinkhorn}(\alpha)$ under the doubly stochastic constraints, specific elements are extracted as the feature mixing weights.

$$w_x = P_{:,0,0}, \tag{15}$$

$$w_\Delta = P_{:,0,1}. \tag{16}$$

The final output feature x_{out} is obtained by the element-wise multiplication and addition of the original flow and the update flow according to these weights.

$$x_{out} = w_x \odot x + w_\Delta \odot \Delta, \tag{17}$$

By defining feature fusion within the Birkhoff polytope of doubly stochastic matrices, the mHC mechanism helps limit unbounded growth or severe attenuation of the feature flow's norm during deep propagation. This method effectively mitigates degradation of high-dimensional sequential features during nonlinear stacking, ensures the stable evolution of the deep state flow, and thereby improves the stability of deep network training and inference. Moreover, mHC introduces additional parameters only at the $O(D)$ level, satisfying the low-compute deployment requirements of H-NET-mHC on edge computing devices. The Sinkhorn-Knopp normalization is performed with K = 5 alternating row and column normalization iterations, and epsilon = 1×10^{-6} is used for numerical stability.

3 Experiments

3.1 Setup

3.1.1 Dataset

The data used in this experiment originated from single-channel (1D) microseismic and ambient background noise signals collected throughout 2025 from an underground mine with an unfilled goaf in Changde City, Hunan Province, China, as shown in Fig. 4. The monitoring system included 23 monitoring points and 23 channels. Each monitoring point was recorded as a single-axis channel, with a sensor sensitivity of 83.0000, hardware gain of 2, and installation direction recorded as 2 in the raw file header. When vibration was detected, the monitoring system automatically saved an individual waveform sample containing 4000 samples over 2 s, corresponding to a sampling frequency of 2 kHz and a sampling interval of 500 microseconds (0.5 ms). The dataset contains 322,665 signal samples, which were divided into three categories using expert annotation: 9689 microseismic events, 38,436 deterministic vibration samples, and 274,540 pure noise samples. In field monitoring, noise typically dominates the other categories. To preserve this realistic imbalance, the dataset was not artificially balanced, making minority-class signals more difficult to detect. During the data preprocessing stage, Z-score standardization was applied to each waveform individually to eliminate scale differences. The random seed was fixed at 42, and the dataset was partitioned at the individual 4000-point waveform-sample level into training, validation, and test sets at a ratio of 70%:15%:15%.

Based on the physical mechanisms, waveform morphology, and monitoring context of the recorded signals, the samples were classified into three categories to evaluate the model's ability to recognize signals from an underground mine. Microseismic events, as shown in Fig. 5a, are characterized by an impulsive first arrival followed by subsequent energy attenuation, reflecting transient micro-fracturing processes within the rock mass. Noise records, as shown in Fig. 5b, mainly represent ambient background disturbances; their waveforms generally lack clear structural patterns and are dominated by low-amplitude random fluctuations. Because the unfilled goaf is located beneath an urban built-up area, deterministic vibration records, as shown in Fig. 5c, are mainly associated with non-event disturbances transmitted downward from surface activities, such as urban traffic and ground mechanical vibrations. These signals typically exhibit regular waveform patterns, including sustained oscillations and periodic repetitions. The annotation procedure consisted of initial labeling by one author, label review by the other authors, and an additional spot-check. Ambiguous samples were discussed by multiple authors until a final category was agreed upon.

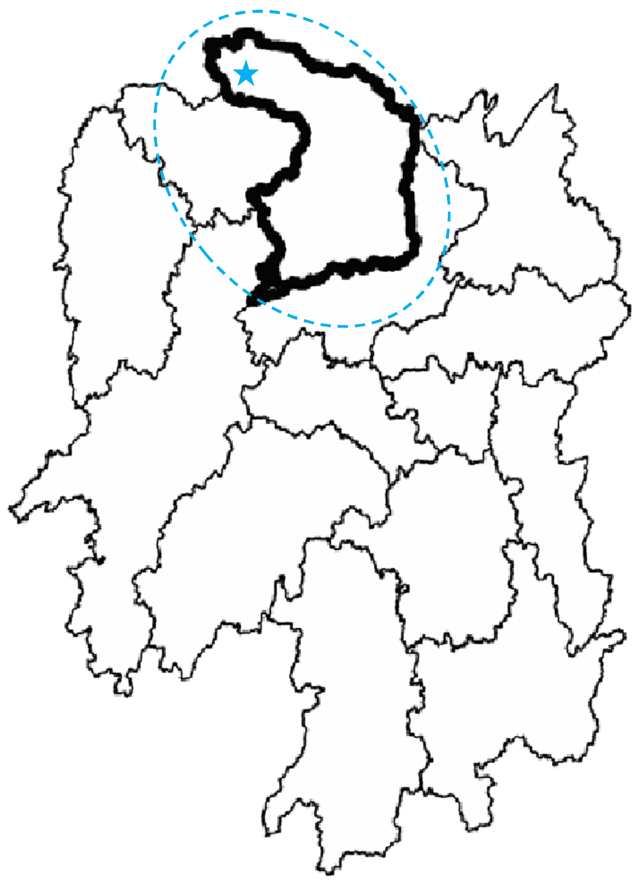

Figure 4: Geographical location map of Changde City, China.

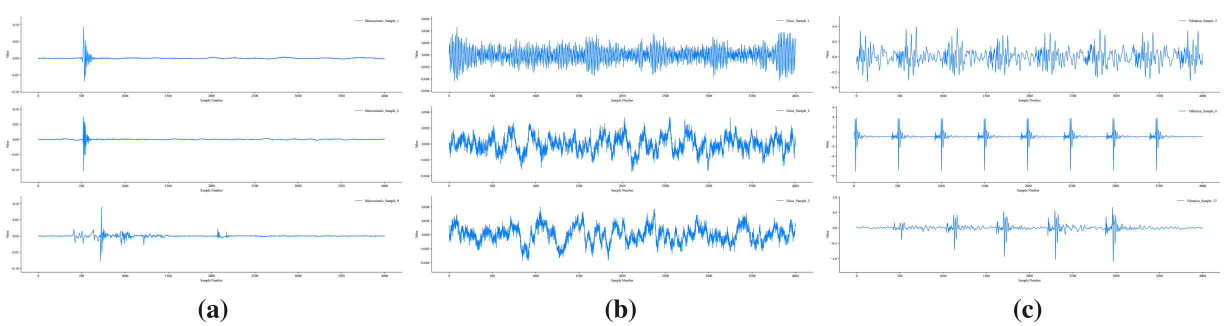

Figure 5: Signal waveforms of different data types. (**a**) Microseismic event; (**b**) Noise; (**c**) Deterministic vibration.

3.1.2 Evaluation Indicators

To comprehensively evaluate the multiclass classification performance of the proposed H-NET-mHC model in resource-constrained edge environments, this paper uses accuracy, precision, recall, F1 score, and the normalized confusion matrix as the primary quantitative metrics [43]. To address the severe class imbalance created by abundant ambient background noise and sparse microseismic signals in mine monitoring, this paper further uses receiver operating characteristic (ROC) curves and the area under the curve (AUC) to assess discriminative ability across classification thresholds [44]. For each one-vs.-rest class, an AUC closer to 1 indicates stronger discrimination; the macro-average AUC summarizes the three class-wise AUC values with equal weighting. The formulas for the core evaluation indicators are as follows.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \tag{18}$$

$$Precision = \frac{TP}{TP + FP}, \tag{19}$$

$$Recall = \frac{TP}{TP + FN}, \tag{20}$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}, \tag{21}$$

$$FPR = \frac{FP}{FP + TN}, \tag{22}$$

$$C_{ij}^{norm} = \frac{C_{ij}}{\sum_k C_{ik}}. \tag{23}$$

Because this is a three-class task, true positive (TP), true negative (TN), false positive (FP), and false negative (FN) are defined separately for each class using a one-vs.-rest strategy. For class c, TP_c is the number of samples of class c predicted as c; FN_c is the number of samples of class c predicted as another class; FP_c is the number of samples from other classes predicted as c; and TN_c contains all remaining samples. Precision, recall, and F1 are calculated for each class and then macro-averaged with equal 1:1:1 class weights. Overall accuracy is the number of correct predictions divided by the total number of samples. The class weights [3.5, 2.0, 1.0] are used only in the weighted cross-entropy loss during training. For the row-normalized confusion matrix, element C_ij is divided by the number of samples whose true class is i; consequently, each row sums to one and each diagonal element represents class-wise recall.

3.2 Parameters and Network Settings

The model was implemented in PyTorch. Training used the AdamW optimizer (Adam with decoupled weight decay), with an initial learning rate of 0.0005, weight decay of 0.01, a batch size of 128, and 20 epochs. Weighted cross-entropy with class weights [3.5, 2.0, 1.0] for microseismic events, deterministic vibration, and noise, respectively, was used to reduce the effect of severe class imbalance during model optimization. These weights were applied only to the training loss, whereas the reported precision, recall, and F1-score were calculated as equal-weight macro averages. ReduceLROnPlateau (factor = 0.5, patience = 3) adjusted the learning rate according to validation loss. Gradients were clipped to a maximum norm of 1.0. The random seed was fixed at 42 and deterministic behavior in the NVIDIA CUDA (Compute Unified Device Architecture) Deep Neural Network library (cuDNN) was enabled. The dataset was divided into training, validation, and test subsets at a ratio of 70%/15%/15% using a seed-controlled waveform-sample-level random split. The same dataset ordering, seed, and splitting procedure were used for all models. The complete optimization settings are summarized in Table 2.

H-NET-mHC uses a compact hierarchical structure. As detailed in Table 3, the 4000-point input is reduced by the stem from 4000 to 800 time steps, followed by three Mamba stages with temporal lengths 800, 200, and 50. Content-aware dynamic chunking is applied between adjacent stages. Global average pooling, layer normalization (LayerNorm), and a linear layer produce the three-class output. Including the 64 trainable affine parameters of the stem batch normalization (BatchNorm) layer, the complete model contains

353,221 trainable parameters. Input and output dimensions in Table 3 exclude the batch dimension and are reported as channels × sequence length; the classifier output is the three-class logit vector.

Table 2: Core hyperparameter settings for the experiment.

Parameter	Value	Description
Optimizer	AdamW	Adam with decoupled weight decay
Initial Learning Rate	0.0005	Initial learning rate
Batch Size	128	Number of samples per training batch
Epochs	20	Total number of training epochs
Loss Function	Weighted Cross-Entropy	Class weights [3.5, 2.0, 1.0]
Learning-Rate Scheduler	ReduceLROnPlateau	Dynamic decay based on validation loss (patience = 3, factor = 0.5)
Gradient Clipping	Max norm 1.0	Maximum norm for gradient clipping
Random Seed	Fixed (42)	Global random seed to ensure experimental reproducibility
Weight Decay	0.01	AdamW weight-decay regularization

Table 3: Details of H-NET-mHC model parameters.

Module	Main Components	Input	Output	Parameters
Stem Layer	1D convolution, BatchNorm, sigmoid linear unit (SiLU)	1 × 4000	32 × 800	320
Stage 1	MambaBlock, mHC, RMSNorm	32 × 800	32 × 800	10,080
Dynamic Chunking 1	Router (depthwise + pointwise), downsampling 1D convolution	32 × 800	64 × 200	8545
Stage 2 (Depth = 2)	MambaBlock, mHC, RMSNorm	64 × 200	64 × 200	65,920
Dynamic Chunking 2	Router (depthwise + pointwise), downsampling 1D convolution	64 × 200	128 × 50	33,473
Stage 3 (Depth = 2)	MambaBlock, mHC, RMSNorm	128 × 50	128 × 50	234,240
Classifier	Global average pooling, LayerNorm, linear	128 × 50	3	643
Total		1 × 4000	3	353,221

3.3 Training and Testing Results

To ensure the reliability and reproducibility of the experimental results, all training, validation, and testing phases of the models in this study were conducted on the same hardware and software platform. The specific configurations of the experimental environment are detailed in Table 4. The experimental platform comprised an Intel Core i5-14600KF processor and 32 gigabytes (GB) of memory, and one NVIDIA GeForce RTX 5060 GPU for tensor computation. At the software level, Python 3.13.7, PyTorch 2.9.0, and CUDA 13.1 were used for GPU-accelerated model training.

To provide direct computational evidence beyond parameter count, CPU-only inference of H-NET-mHC was measured using its trained weights on the Intel Core i5-14600KF listed in Table 4. PyTorch 2.9.0 (CPU build), batch size 1, one CPU thread, and an input shape of $1 \times 1 \times 4000$ were used. The model was evaluated in five repetitions, each with 20 warm-up runs and 100 timed runs, giving 500 timed inferences. H-NET-mHC requires approximately 1.35 mebibytes (MiB) of 32-bit floating-point (FP32) parameter memory, and its serialized weight file occupies 1.38 MiB. Its mean, median, and 95th-percentile latencies were 41.11, 40.99, and 42.59 ms, respectively, corresponding to 24.32 samples/s. These measurements provide a reproducible CPU runtime and memory-footprint reference for the proposed model.

Table 4: Hardware, Software, and CPU benchmark specifications.

Hardware/Software	Specifications/Versions
GPU	NVIDIA GeForce RTX 5060
CPU	Intel(R) Core(TM) i5-14600KF
Random-Access Memory (RAM)	32 GB
Language	Python 3.13.7
Framework	PyTorch 2.9.0
CUDA	13.1
CPU Benchmark Protocol	PyTorch 2.9.0 CPU; batch 1; one thread; input $1 \times 1 \times 4000$; 5 repetitions; each with 20 warm-ups + 100 timed runs per model

As shown in the confusion matrix in Fig. 6a, where the vertical axis represents the true labels and the horizontal axis represents the predicted labels, the overall prediction results are primarily distributed along the main diagonal. Under the condition of an imbalanced test set, the model misclassified 80 noise samples and 131 deterministic-vibration samples as microseismic events, demonstrating a low false-positive rate (FPR). Simultaneously, out of 1439 true microseismic events, the model correctly identified 1383, with 24 samples falsely predicted as noise. This indicates that the H-NET-mHC model can effectively extract the phase features of sparse microseismic signals against a background of non-event noise.

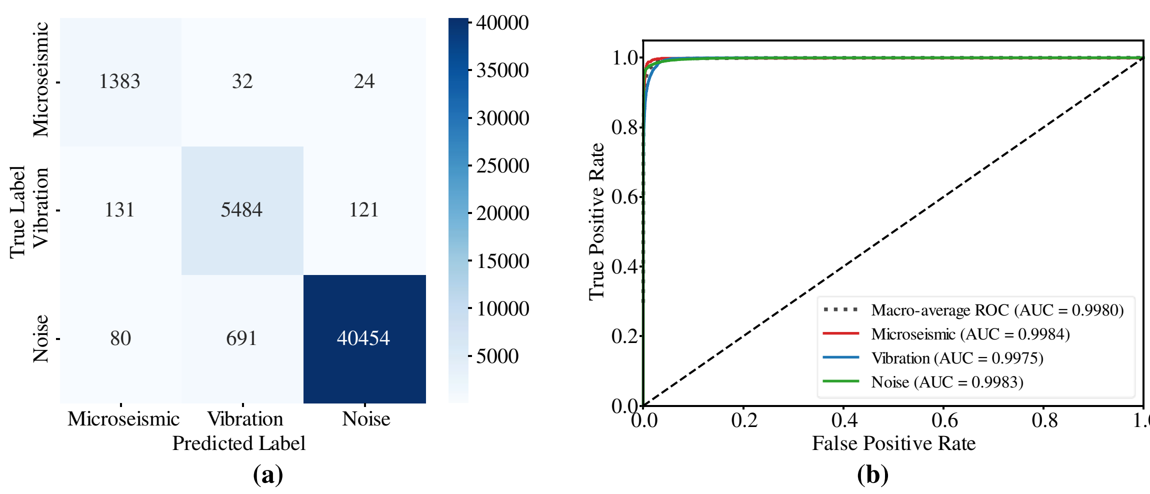

Figure 6: Classification performance of H-NET-mHC. (**a**) Confusion matrix; (**b**) Class-wise ROC curves.

To further quantify the model's discriminative ability across different classification thresholds, one-vs.-rest ROC curves were generated. As shown in Fig. 6b, the AUC values for microseismic events, deterministic vibration, and noise are 0.9984, 0.9975, and 0.9983, respectively, with a macro-average AUC of 0.9980. These results indicate that H-NET-mHC maintains a high true-positive rate and a low false-positive rate under complex background noise and class imbalance, demonstrating strong inter-class discriminative capability.

3.4 Main Results

To evaluate the performance of H-NET-mHC in resource-constrained edge-oriented scenarios, representative baseline models from different architectural families were selected for quantitative comparison. These baselines include 18-layer residual convolutional neural network (ResCNN-18) [45] for local feature extraction; one-dimensional Vision Transformer (1D-ViT) [46], an attention-based model with a global receptive field; lightweight convolutional networks designed for mobile deployment, including MobileNet V3 and ShuffleNet V2; and convolutional neural network-Mamba (CNN-MAMBA), a larger SSM-based benchmark model. The comparison considers both model compactness and classification performance, as reported in Table 5. Because the dataset is severely imbalanced, with noise samples far outnumbering microseismic events, overall accuracy alone is insufficient to characterize the behavior of the classifiers. Therefore, precision, recall, and F1-score are reported as equal-weight macro averages across the three classes, while class-wise results are also analyzed to emphasize the safety-critical microseismic-event category.

Table 5: Classification performance comparison results of different models.

Models	Params (M)	Accuracy	Precision	Recall	F1
CNN-MAMBA	7.17	0.9601	0.8517	0.9452	0.8932
1D-ViT	4.78	0.9431	0.8095	0.9343	0.8626
ResCNN-18	3.85	0.9408	0.7958	0.9419	0.8559
MobileNet V3	3.25	0.9530	0.8336	0.9428	0.8813
ShuffleNet V2	0.34	0.9605	0.8517	0.9358	0.8891
H-NET-mHC	0.35	0.9777	0.9159	0.9661	0.9397

All baseline models were trained and evaluated under a common protocol. Specifically, they used the same 4000-point single-channel waveform-sample input, per-waveform Z-score standardization, 70%/15%/15% sample-level training/validation/test split generated with seed 42, batch size of 128, 20-epoch training budget, AdamW optimizer, initial learning rate of 0.0005, weight decay of 0.01, ReduceLROnPlateau scheduler with a reduction factor of 0.5 and a patience of 3 epochs, maximum gradient norm of 1.0, and random seed of 42. The same dataset ordering and random-split procedure were used for all models. No model-specific search over the data split, optimizer family, learning-rate schedule, or epoch budget was performed. Only architecture-intrinsic settings, such as the patch size and embedding dimension of 1D-ViT, were retained to make each baseline compatible with the common 4000-point input.

As indicated by the quantitative evaluation results in Table 5, the H-NET-mHC model achieves a favorable balance between computational overhead and classification performance. In terms of lightweight design, the total parameter count of H-NET-mHC is 0.35M, substantially fewer than CNN-MAMBA (7.17M) and 1D-ViT (4.78M). With this compact parameter budget, the overall accuracy of H-NET-mHC reaches 0.9777, while the macro-average precision, recall, and F1-score are 0.9159, 0.9661, and 0.9397, respectively. Compared with CNN-MAMBA, which is also based on the SSM architecture, H-NET-mHC improves the macro-average F1 score by 4.65 percentage points despite the reduction in parameters. Furthermore, when compared to the lightweight network ShuffleNet V2, which has a similar parameter count of 0.34M and a macro-average F1-score of 0.8891, H-NET-mHC achieves higher values for all reported classification metrics, supporting the potential use of this architecture for edge-side microseismic signal perception tasks.

When deep networks process microseismic signals with non-stationary characteristics, they face risks of training instability and feature degradation. To assess convergence behavior and discrimination, Fig. 7a,b illustrates the validation-loss and validation-accuracy curves of each model over 20 epochs, while Fig. 7c compares their macro-average ROC curves.

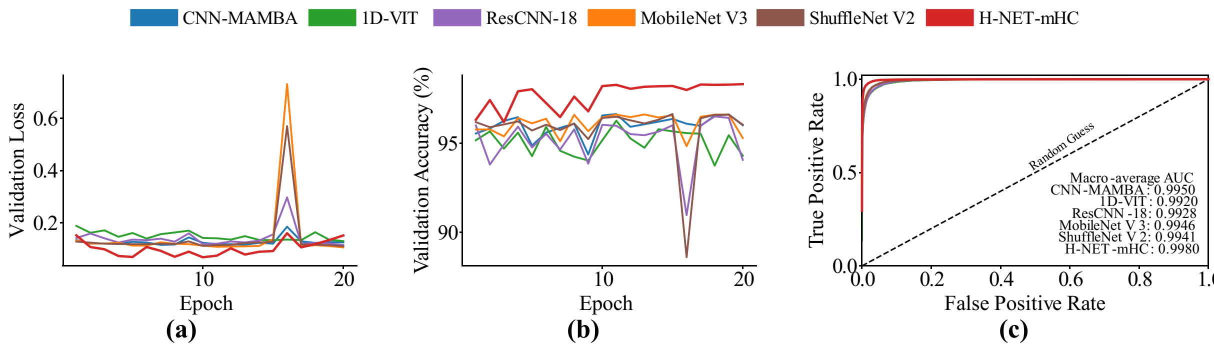

Figure 7: Comparative performance of different models. (**a**) Validation loss; (**b**) Validation accuracy; (**c**) Macro-average ROC curves.

As observed from the loss curves in Fig. 7a, several baseline models, including MobileNet V3, CNN-MAMBA, and 1D-ViT, exhibited pronounced loss spikes during the mid-to-late stages of training, around the 14th epoch. Under the present training protocol, these models appear more susceptible to optimization fluctuations when processing complex noise or abrupt transient signals. Correspondingly, Fig. 7b illustrates that the validation accuracy of these baseline models declined during this phase. In contrast, H-NET-mHC showed smoother convergence over the training cycle, with validation loss decreasing and validation accuracy increasing steadily. This training behavior is consistent with the intended stabilizing role of mHC. Although this pattern does not establish causality, the ablation results support an association between mHC and more stable feature propagation under nonstationary inputs.

To further evaluate the inter-class discrimination of the models in the multiclass task, this paper conducts a comprehensive analysis combining the normalized confusion matrices in Fig. 8 and the macro-average ROC curves in Fig. 7c. The results show that the class-wise recall of H-NET-mHC for microseismic events, deterministic vibration, and noise is 96.11%, 95.61%, and 98.13%, respectively. Its macro-average AUC is 0.9980. Both the class-wise recall values and the macro-average AUC are the highest among the compared models.

Regarding specific class discrimination, the class-wise recall of 1D-ViT for the three signal types is 93.12%, 92.57%, and 94.59%, respectively, while the recall of MobileNet V3 for the microseismic-event class is 92.63%. In the off-diagonal regions of the confusion matrix, H-NET-mHC exhibits lower off-diagonal error rates; its rate of misclassifying microseismic events as deterministic vibration is 2.22%, which is lower than that of MobileNet V3 (6.19%) and 1D-ViT (4.93%). Regarding decision-boundary discrimination, Fig. 7c shows that the macro-average AUC of H-NET-mHC is higher than that of CNN-MAMBA (0.9950) and 1D-ViT (0.9920), which have substantially more parameters, and is also higher than those of MobileNet V3 (0.9946) and ShuffleNet V2 (0.9941). This indicates that under scenarios of strong ambient background noise and class imbalance, H-NET-mHC provides strong discrimination among the temporal patterns of the three signal classes.

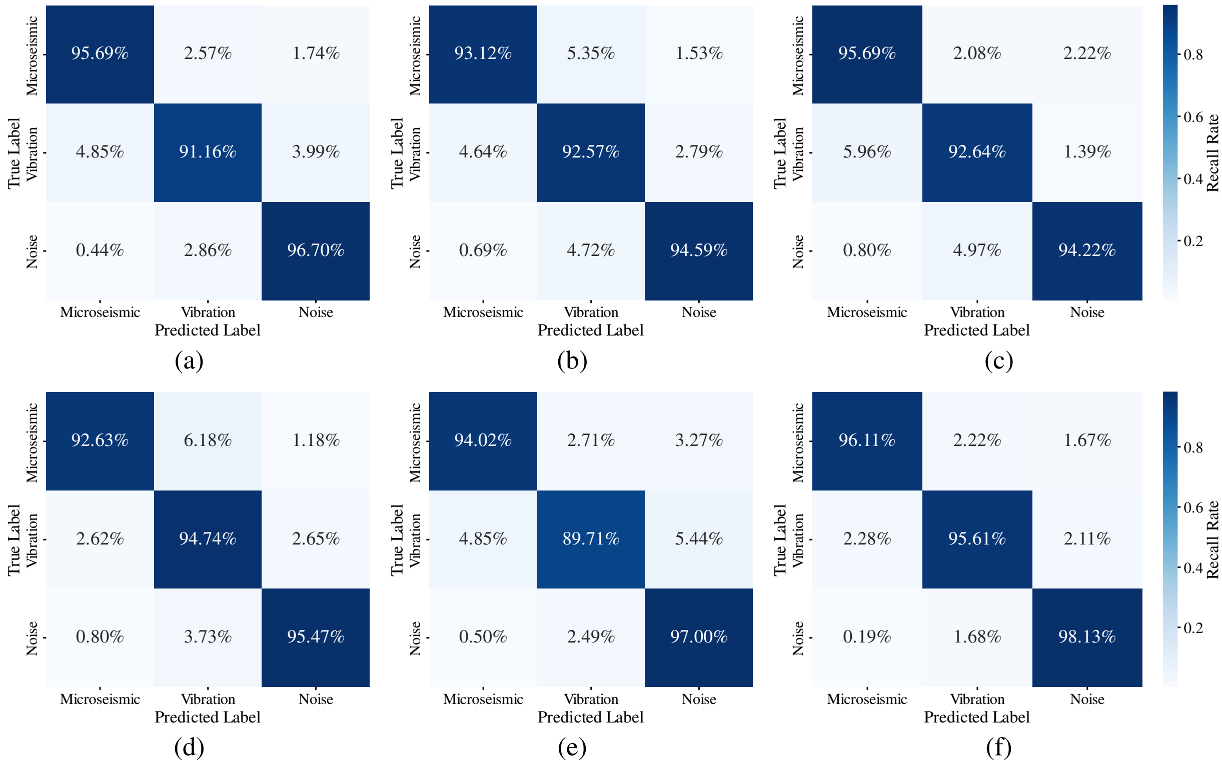

Figure 8: Normalized confusion matrices of different models (**a**) CNN-MAMBA; (**b**) 1D-ViT; (**c**) ResCNN-18; (**d**) MobileNet V3; (**e**) ShuffleNet V2; (**f**) H-NET-mHC.

The bar charts in Figs. 9–11 display the precision, recall, and F1-score of each model across three categories: microseismic events, noise, and deterministic vibration. Due to the dominance of pure noise samples, all evaluated models achieve high scores for this class, as shown in Fig. 10. However, for deterministic vibration, as shown in Fig. 11, and for the microseismic-event category, which accounts for a smaller proportion of the samples, as shown in Fig. 9, the baseline models show lower classification performance. In the recognition task for deterministic vibration, the precision of ResCNN-18 and 1D-ViT is 0.7187 and 0.7241, respectively, reflecting the difficulty traditional models face in effectively separating vibrations from ambient background noise. In the recognition task for microseismic events, the F1-score of ResCNN-18 is 0.7893, and the precision of CNN-MAMBA is 0.7500, indicating the difficulty of extracting sparse event phases under strong noise masking.

In contrast, H-NET-mHC achieves higher scores in the deterministic-vibration and microseismic-event classes. For deterministic vibration, the F1-score reaches 0.9184. For microseismic events, the precision, recall, and F1-score are 0.8676, 0.9611, and 0.9120, respectively. The relatively low precision for the microseismic-event category is primarily influenced by the inherent class imbalance of the dataset. Because noise and deterministic-vibration samples are much more numerous, the misclassification of even a small fraction of these majority-class samples generates many false positives, thereby lowering the precision of this category. Overall, H-NET-mHC retains sensitivity to weak microseismic features under complex noise.

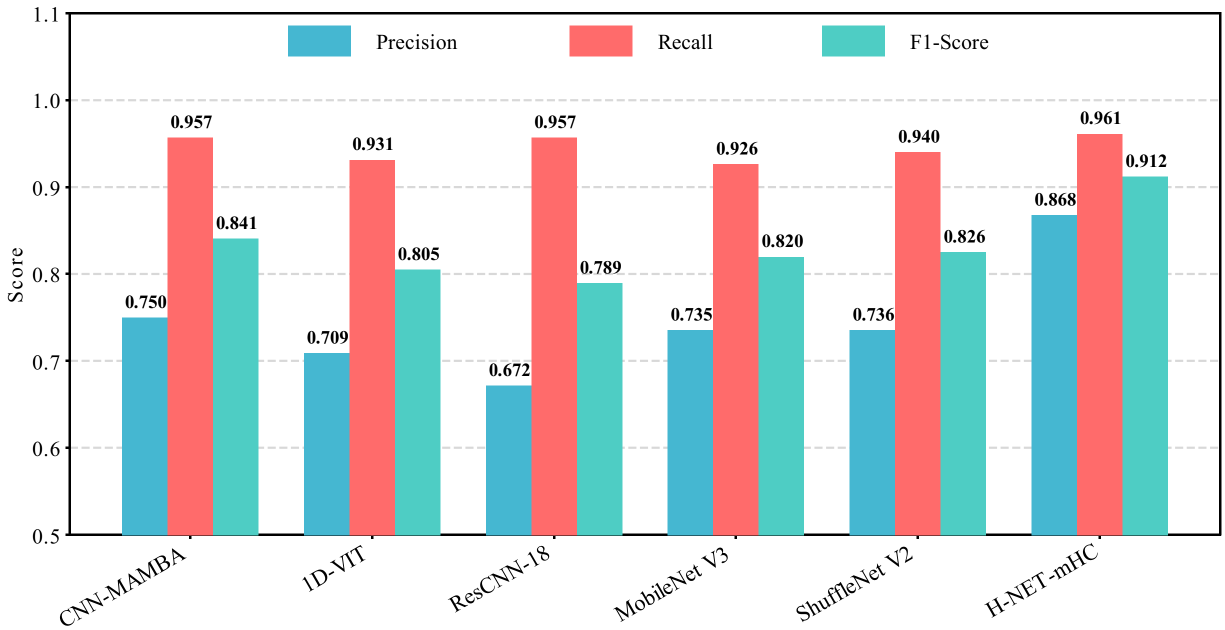

Figure 9: Comparison of different models for microseismic events.

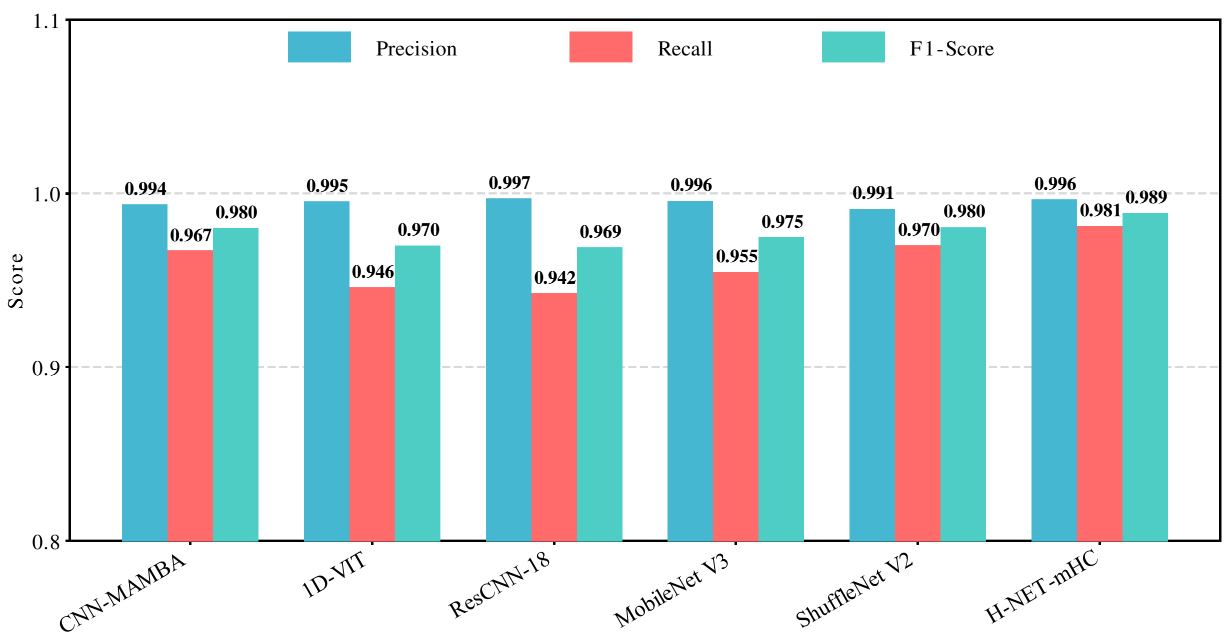

Figure 10: Comparison of different models for noise.

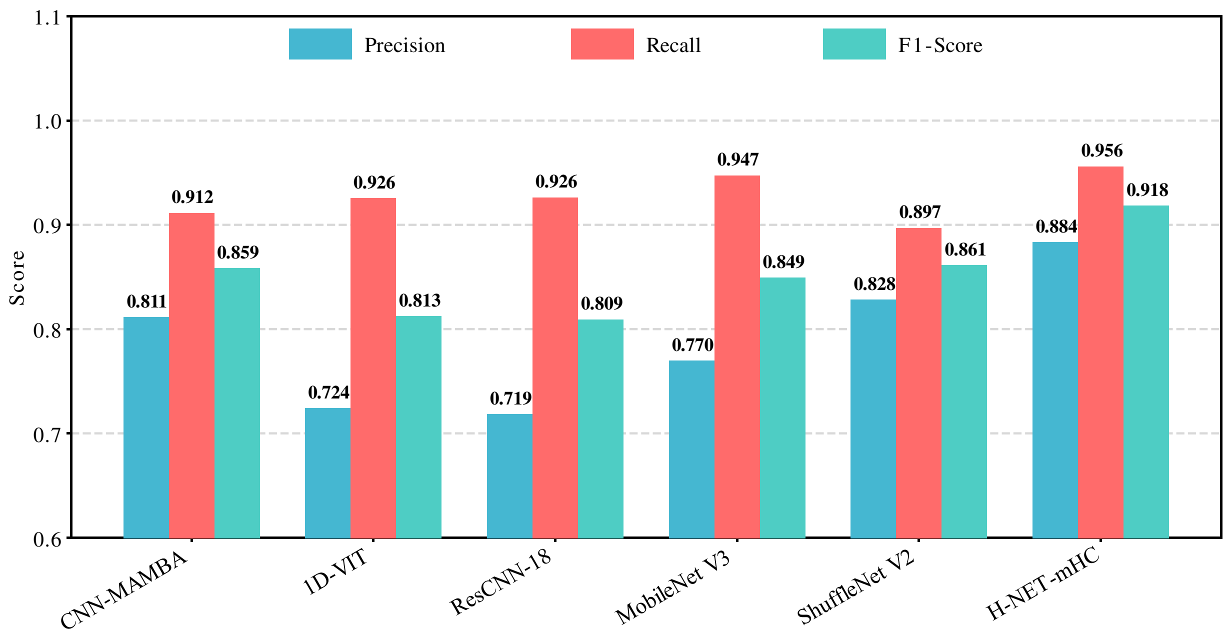

Figure 11: Comparison of different models for deterministic vibration.

3.5 Ablation Studies

To verify the individual effects and complementary recall-oriented behavior of the Content-Aware Dynamic Chunking mechanism and the mHC module within the model architecture, this study designed a set of ablation studies. A standard lightweight SSM without either module served as the baseline. The comparison also included a variant containing only dynamic chunking (H-NET), a variant containing only mHC (Baseline-mHC), and the complete architecture containing both modules (H-NET-mHC). The four network variants were evaluated using unified hyperparameter settings and the same data split. The quantitative results are presented in Table 6, and the normalized confusion matrices of each model on the test set are illustrated in Fig. 12.

Table 6 indicates that the incorporation of both modules changes the balance among the evaluation metrics. Compared with the baseline, H-NET-mHC shows a marginal decrease in accuracy from 0.9780 to 0.9777 (−0.03 percentage points) and macro-precision from 0.9228 to 0.9159 (−0.69 percentage points). This slight reduction is mainly related to the recall-oriented decision tendency introduced by the coupled architecture: after content-aware dynamic chunking enhances the response of potential transient phases and mHC stabilizes the subsequent feature propagation, the model becomes more sensitive to weak microseismic-event patterns. As a result, some borderline or vibration-related samples are more likely to be assigned to the microseismic class, leading to a small increase in false positives and a corresponding decrease in precision. Meanwhile, macro-recall increases from 0.9549 to 0.9661 (+1.12 percentage points), and macro-F1 increases from 0.9383 to 0.9397 (+0.14 percentage points). These results suggest that H-NET-mHC shifts the classification behavior toward fewer missed events and better overall class coverage, which is more consistent with the requirements of safety-critical early warning than a purely precision-oriented operating point.

Table 6: Ablation results for H-NET-mHC.

Models	Accuracy	Precision	Recall	F1
Baseline	0.9780	0.9228	0.9549	0.9383
H-NET	0.9684	0.8943	0.9591	0.9243
Baseline-mHC	0.9786	0.9161	0.9645	0.9391
H-NET-mHC	0.9777	0.9159	0.9661	0.9397

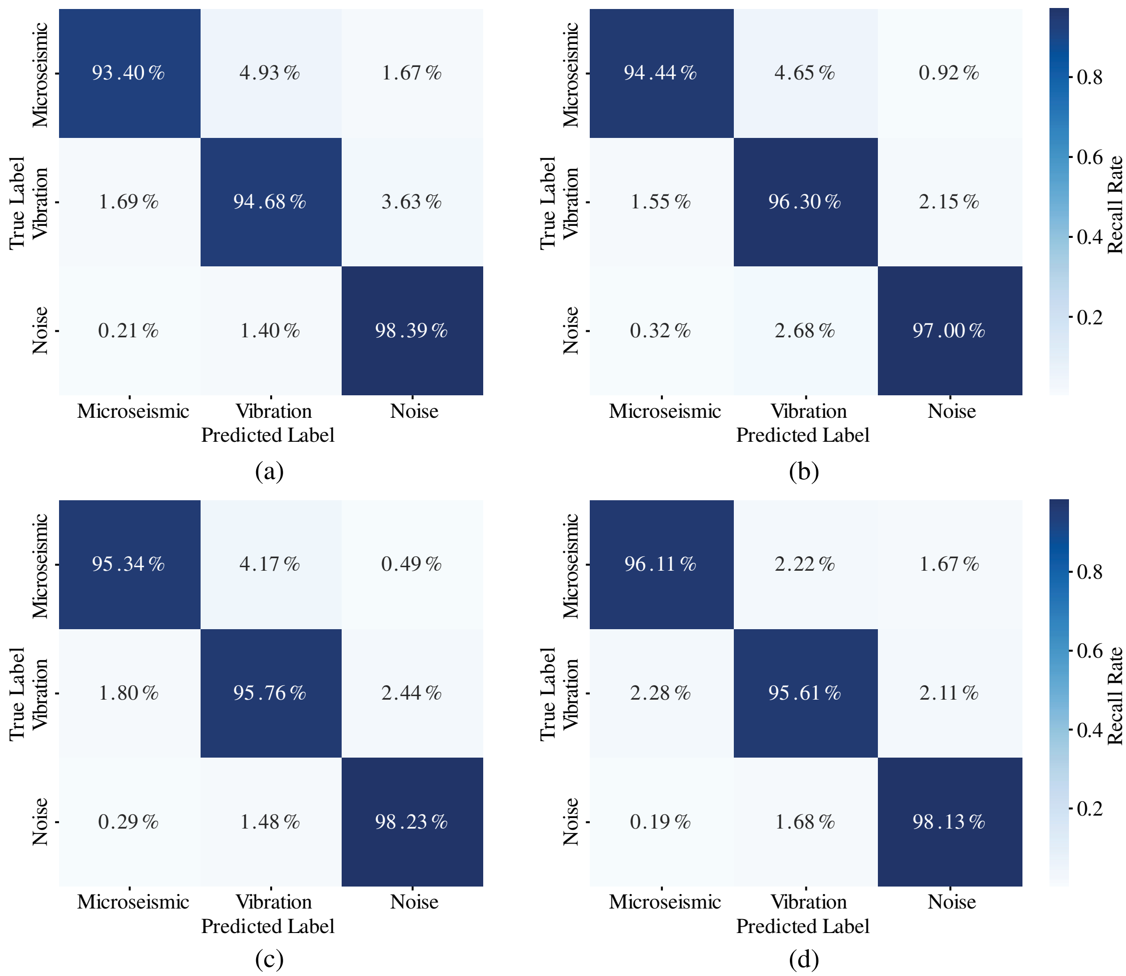

Figure 12: Normalized confusion matrices of the ablation study. (**a**) Baseline; (**b**) H-NET; (**c**) Baseline-mHC; (**d**) H-NET-mHC.

Further analysis of the normalized confusion matrices in Fig. 12 provides a class-level explanation of the ablation results. For the safety-critical microseismic-event category, the baseline model achieves a recall of 93.40%, while H-NET and Baseline-mHC increase this value to 94.44% and 95.34%, respectively. The complete H-NET-mHC architecture further improves the recall of this category to 96.11%. This trend indicates that both modules contribute to the recognition of critical microseismic patterns, but they play different roles in the feature evolution process. Dynamic chunking enhances the response of informative transient phases and suppresses low-importance temporal regions, but this selective compression may also change the variance and temporal density of features entering deeper SSM stages. mHC, in contrast, constrains feature propagation and stabilizes Mamba updates, but when used alone, it may also regularize unfiltered background components. In the complete architecture, dynamic chunking first improves the signal-to-noise ratio and reduces the sequence length, while mHC subsequently stabilizes the fusion of the compressed representation with each Mamba update. This sequence may explain why the complete model is more sensitive to weak microseismic-event features and achieves higher recall for the critical class. In the monitoring of unfilled goafs affecting urban public safety, missed microseismic events may lead to more serious safety consequences than additional false alarms and manual review. Therefore, the recall improvement achieved by H-NET-mHC reflects a more suitable operating behavior for early-warning applications.

For the safety-critical microseismic class, recall increases from 93.40% in the baseline to 96.11% in H-NET-mHC. This improvement is accompanied by an increase in deterministic-vibration samples classified as microseismic events from 1.69% to 2.28%. The model therefore accepts a small increase in false alarms and manual-review burden in exchange for fewer missed microseismic events. This precision-recall trade-off is appropriate for an early-warning setting in which missed events can have more serious safety consequences than additional expert review.

4 Discussion

H-NET-mHC achieved an overall accuracy of 97.77% on the real-world dataset from the Changde Gypsum Mine, and the recall of microseismic events reached 96.11%. Under conditions of class imbalance, the model maintained a low false-positive rate, misclassifying only 80 pure-noise samples as microseismic signals. The experimental results show strong inter-class discrimination for temporal features exhibiting non-stationary characteristics. The relatively low precision (0.8676) for the microseismic-event category is mainly because the large base of pure-noise and deterministic-vibration samples amplifies the proportion of false positives. In engineering applications involving underground monitoring of an unfilled goaf, missing microseismic events increases safety risks such as surface subsidence. In contrast, a high-recall operating profile better aligns with the safety requirements of early-warning systems for rock-mass instability.

The ablation results suggest that content-aware dynamic chunking and mHC have complementary effects under severe class imbalance, mainly by shifting the model toward higher recall for safety-critical microseismic events rather than by improving every aggregate metric. As shown in Fig. 13, the content-aware dynamic chunking layer evaluates local sequence importance through a lightweight routing function. The blue curve represents the input microseismic signal, and the red curve represents the output dynamic-weight score. In the arrival regions of transient phases, the weighting network adaptively generates local response peaks. Through element-wise multiplication, this data-driven masking mechanism assigns higher weights to informative phases during feature downsampling, thereby reducing the propagation of ambient noise into deeper layers.

Deep SSM architectures face risks of hidden-state divergence and feature degradation when modeling sequences with non-stationary characteristics. The feature-norm evolution analysis in Fig. 14 shows that when traditional residual connections are employed, the Euclidean (L2) norm of the feature tensor increases markedly as network depth increases. After the introduction of the mHC module, the feature flow is mixed through a doubly stochastic matrix generated by the Sinkhorn-Knopp algorithm. This architecture helps limit excessive growth of feature norms, ensuring the stable evolution of high-dimensional features during deep nonlinear stacking.

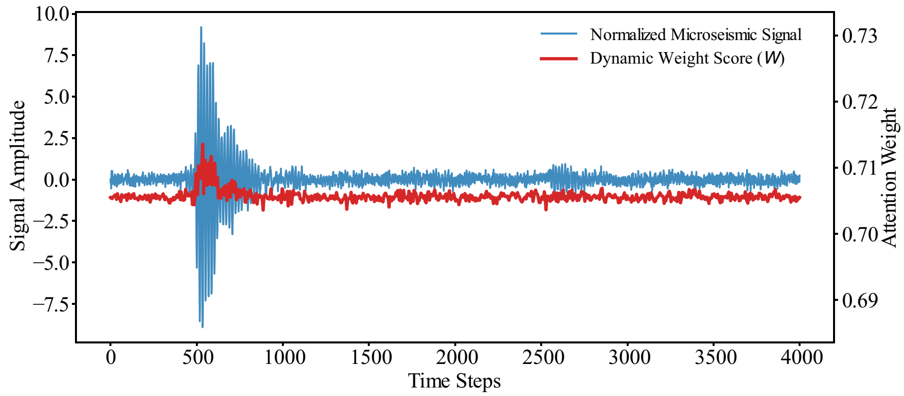

Figure 13: Content-aware dynamic chunking response.

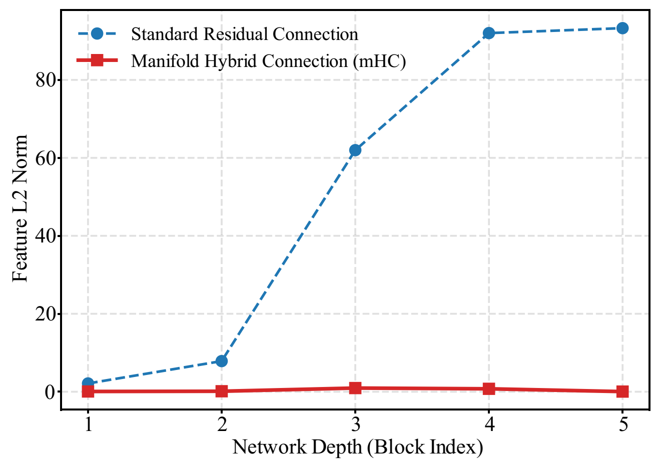

Figure 14: Feature norm evolution analysis.

The complete model contains 353,221 trainable parameters and requires only about 1.35 MiB for FP32 parameters. In the single-thread CPU test described in Section 3.3, H-NET-mHC required 41.11 ms on average for one 4000-point sample and processed 24.32 samples/s. These measurements complement the parameter count with an implementation-level runtime reference and show that the trained model can run on a general-purpose CPU with a compact storage footprint. However, latency and energy consumption depend on the processor, runtime, quantization, and operator implementation. Future work will extend the evaluation to low-power Arm processors and dedicated mine-monitoring terminals and will investigate quantization and operator optimization.

This study primarily models three classes of single-channel (1D) signals: microseismic events, deterministic vibration, and ambient noise. Although the model extracts features consistently under strong-noise conditions, there is still room for improvement in the boundary-localization accuracy of the content-aware dynamic chunking layer for weak phases under extremely low signal-to-noise ratios. The current architecture does not yet capture the spatial correlation implied by the physical topology of the sensors. In addition, the experiments were conducted using one fixed random split and one seed. Therefore, the comparative results should be interpreted as performance under a unified sample-level split rather than as split-independent superiority. Future work will integrate multisensor arrays, repeat experiments with multiple random seeds, and validate the model under independent acquisition conditions. It will also examine whether the doubly stochastic mixing constraints of mHC extend to high-dimensional spatial feature flows.

5 Conclusion

This study presented H-NET-mHC, a compact hierarchical SSM for classifying non-stationary microseismic signals under severe class imbalance. Content-aware dynamic chunking performs phase-sensitive temporal compression, while mHC constrains the subsequent feature-flow evolution. The 353,221-parameter model achieved 97.77% overall accuracy, 96.11% microseismic-event recall, and 0.9397 macro-F1 under a unified fixed-split protocol. In the single-thread CPU benchmark, the model required only 1.35 MiB of FP32 parameter memory and achieved a mean latency of 41.11 ms per 4000-point sample. The compact footprint supports the edge-oriented design objective and provides a CPU runtime reference. Future work should evaluate split robustness using multiple seeds and independent acquisition conditions.

Acknowledgement: Not applicable.

Funding Statement: This research was funded by the National Natural Science Foundation of China (Grant No. 52374168), the Science and Technology Innovation Program of Hunan Province (Grant No. 2023RC3069), the National Key Research and Development Program of China (Grant No. 2022YFC2904105), and the Research Fund of the State Key Laboratory for Fine Exploration and Intelligent Development of Coal Resources, China University of Mining and Technology (Grant No. SKLCRSM20KFO03).

Author Contributions: Pingan Peng: Writing—review & editing, Funding acquisition, Conceptualization. Qi Zhang: Validation, Software, Methodology, Investigation. Ya Liu: Writing—review & editing, Data curation. Yue Han: Validation, Software. Zhida Jiang: Validation, Formal analysis. Linli Chen: Resources, Investigation. Xuefeng Huo: Software, Methodology. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The data that support the findings of this study are available from the corresponding author, Qi Zhang, upon reasonable request.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Zhang X, Li W, Li T, Li Z, Cai G, Shen Z, et al. Stability analysis and numerical simulation of foundation in old goaf under building load. Front Earth Sci. 2023;11:1063684. doi:10.3389/feart.2023.1063684.
2. Chen J, Shan Y, Li S, Huang Q, Dan Z, Yu H. Research and engineering practice of high stage and long span complex goaf filling treatment technology. Sci Rep. 2025;15(1):43033. doi:10.1038/s41598-025-24224-z.
3. Du Z, Liu Z, Liu R, Wang S, Li F. Influencing factors for the instability and collapse mode of the goaf structure in a gypsum mine. Adv Civ Eng. 2021;2021(1):5577287. doi:10.1155/2021/5577287.
4. Shi Y, Wang H, Tan X, Jin Y, Wang J, Tang B. A stability analysis of an abandoned gypsum mine based on numerical simulation using the itasca model for advanced strain softening constitutive model. Appl Sci. 2023;13(23):12570. doi:10.3390/app132312570.
5. Ma K, Sun XY, Tang CA, Yuan FZ, Wang SJ, Chen T. Floor water inrush analysis based on mechanical failure characters and microseismic monitoring. Tunn Undergr Space Technol. 2021;108(4):103698. doi:10.1016/j.tust.2020.103698.
6. Ma K, Tang CA, Liang ZZ, Zhuang DY, Zhang QB. Stability analysis and reinforcement evaluation of high-steep rock slope by microseismic monitoring. Eng Geol. 2017;218(4):22–38. doi:10.1016/j.enggeo.2016.12.020.

7. Franek P, Plaza-Faverola A, Mienert J, Buenz S, Ferré B, Hubbard A. Microseismicity linked to gas migration and leakage on the western svalbard shelf. Geochem Geophys Geosyst. 2017;18(12):4623–45. doi:10.1002/2017gc007107.
8. Saragiotis CD, Hadjileontiadis LJ, Panas SM. PAI-S/K: a robust automatic seismic P phase arrival identification scheme. IEEE Trans Geosci Remote Sens. 2002;40(6):1395–404. doi:10.1109/TGRS.2002.800438.
9. Kuyuk HS, Susumu O. Real-time classification of earthquake using deep learning. Procedia Comput Sci. 2018;140(8):298–305. doi:10.1016/j.procs.2018.10.316.
10. Zhu W, Tai KS, Mousavi SM, Bailis P, Beroza GC. An end-to-end earthquake detection method for joint phase picking and association using deep learning. JGR Solid Earth. 2022;127(3):e2021JB023283. doi:10.1029/2021jb023283.
11. Saad OM, Hafez AG, Soliman MS. Deep learning approach for earthquake parameters classification in earthquake early warning system. IEEE Geosci Remote Sens Lett. 2021;18(7):1293–7. doi:10.1109/LGRS.2020.2998580.
12. Mousavi SM, Ellsworth WL, Zhu W, Chuang LY, Beroza GC. Earthquake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking. Nat Commun. 2020;11(1):3952. doi:10.1038/s41467-020-17591-w.
13. Zhang S, Ku B, Ko H. Learnable maximum amplitude structure for earthquake event classification. IEEE Geosci Remote Sens Lett. 2022;19(1):3005805. doi:10.1109/LGRS.2022.3145387.
14. Cui Y, Bai M, Wu J, Chen Y. Earthquake signal detection using a multiscale feature fusion network with hybrid attention mechanism. Geophys J Int. 2024;240(2):988–1008. doi:10.1093/gji/ggae423.
15. Ku B, Kim G, Ahn JK, Lee J, Ko H. Attention-based convolutional neural network for earthquake event classification. IEEE Geosci Remote Sens Lett. 2021;18(12):2057–61. doi:10.1109/LGRS.2020.3014418.
16. Wei C, Huang H, Wang T, Zheng B, Feng Z. Seismic signal classification research based on multi-attention mechanism residual network. In: 2024 6th International Conference on Communications, Information System and Computer Engineering (CISCE); 2024 May 10–12; Guangzhou, China. p. 344–8. doi:10.1109/CISCE62493.2024.10653071.
17. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. Red Hook, NY, USA: Curran Associates Inc.; 2017. p. 6000–10.
18. Huang Y, Xie Y, Liu W, Ma Y, Miao F, Zhang G. Multi-scale convolution networks for seismic event classification with windowed self-attention. J Seismol. 2025;29(1):257–68. doi:10.1007/s10950-024-10262-z.
19. Meng F, Ren T, Zhang H, Wang X, Chen H. SeisParaNet: a novel multitask network for seismic source characterization in earthquake early warning. IEEE Trans Geosci Remote Sens. 2024;62:5913711. doi:10.1109/TGRS.2024.3376776.
20. Meng F, Ren T, Wang P, Liu X, Xiang W, He X. A two-stage earthquake event classification model based on diffusion probability model. IEEE Trans Geosci Remote Sens. 2024;62(2):5933309. doi:10.1109/TGRS.2024.3479327.
21. Wang P, Ren T, Liu J, Meng F, Liu X, Liu W. Neural network light weighting approach using multi-metric evaluation of convolution kernels. IEEE Access. 2023;11:89067–76. doi:10.1109/ACCESS.2023.3305984.
22. Pascanu R, Mikolov T, Bengio Y. On the difficulty of training recurrent neural networks. arXiv:1211.5063. 2012.
23. Náprstek J, Fischer C. Non-stationary response of structures excited by random seismic processes with time variable frequency content. Soil Dyn Earthq Eng. 2002;22(9–12):1143–50. doi:10.1016/S0267-7261(02)00141-0.
24. Garcia SR, Romo MP, Alcántara L. Analysis of non-linear and non-stationary seismic recordings of Mexico city. Soil Dyn Earthq Eng. 2019;127(5):105859. doi:10.1016/j.soildyn.2019.105859.
25. Kumar R, Zhao W, Singh V. Joint time-frequency analysis of seismic signals: a critical review. Struct Durab Heal Monit. 2018;12(2):65–83. doi:10.1002/9781119892373.
26. Zhou HL, Wang CC, Marfurt KJ, Jiang YW, Bi JX. Enhancing the resolution of non-stationary seismic data using improved time-frequency spectral modelling. Geophys J Int. 2016;205(1):203–19. doi:10.1093/gji/ggv553.
27. Ditommaso R, Mucciarelli M, Ponzo FC. Analysis of non-stationary structural systems by using a band-variable filter. Bull Earthq Eng. 2012;10(3):895–911. doi:10.1007/s10518-012-9338-y.
28. Turco F, Azevedo L, Herold D. Geostatistical interpolation of non-stationary seismic data. Comput Geosci. 2019;23(4):665–82. doi:10.1007/s10596-019-9812-6.

29. Peng K, Tang Z, Dong L, Sun D. Machine learning based identification of microseismic signals using characteristic parameters. Sensors. 2021;21(21):6967. doi:10.3390/s21216967.
30. Capilla C. Application of the Haar wavelet transform to detect microseismic signal arrivals. J Appl Geophys. 2006;59(1):36–46. doi:10.1016/j.jappgeo.2005.07.005.
31. Rosca M, Wu Y, Qin C, Dherin B. On a continuous time model of gradient descent dynamics and instability in deep learning. arXiv:2302.01952. 2023.
32. Gilmer J, Ghorbani B, Garg A, Kudugunta S, Neyshabur B, Cardoze D, et al. A loss curvature perspective on training instability in deep learning. arXiv:2110.04369. 2021.
33. Kloberdanz E, Kloberdanz KG, Le W. DeepStability: a study of unstable numerical methods and their solutions in deep learning. In: Proceedings of the 44th International Conference on Software Engineering; 2022 May 22–27; Pittsburgh, PA, USA. p. 586–97. doi:10.1145/3510003.3510095.
34. Roy P, Ghosh S, Bhattacharya S, Pal U. Effects of degradations on deep neural network architectures. arXiv:1807.10108. 2018.
35. Gu A, Goel K, Ré C. Efficiently modeling long sequences with structured state spaces. arXiv:2111.00396. 2021.
36. Gu A, Dao T. Mamba: linear-time sequence modeling with selective state spaces. In: Proceedings of the First Conference on Language Modeling (COLM); 2024 Oct-7-9; Pittsburgh, PA, USA.
37. Zhu L, Liao B, Zhang Q, Wang X, Liu W, Wang X. Vision mamba: efficient visual representation learning with bidirectional state space model. arXiv:2401.09417. 2024.
38. Howard A, Sandler M, Chen B, Wang W, Chen LC, Tan M, et al. Searching for MobileNetV3. In: 2019 IEEE/CVF International Conference on Computer Vision (ICCV); 2019 Oct 27–Nov 2; Seoul, Republic of Korea. p. 1314–24. doi:10.1109/ICCV.2019.00140.
39. Ma N, Zhang X, Zheng HT, Sun J. ShuffleNet V2: practical guidelines for efficient CNN architecture design. In: Computer Vision—ECCV 2018. Cham, Switzerland: Springer International Publishing; 2018. p. 122–38. doi:10.1007/978-3-030-01264-9_8.
40. Gu A, Dao T, Ermon S, Rudra A, Re C. HiPPO: recurrent memory with optimal polynomial projections. In: Proceedings of the 34th International Conference on Neural Information Processing Systems; 2020 Dec 6–12; Vancouver, BC, Canada.
41. Hwang S, Wang B, Gu A. Dynamic chunking for end-to-end hierarchical sequence modeling. arXiv:2507.07955. 2025.
42. Xie Z, Wei Y, Cao H, Zhao C, Deng C, Li J, et al. mHC: manifold-constrained hyper-connections. arXiv:2512.24880. 2025.
43. Sokolova M, Lapalme G. A systematic analysis of performance measures for classification tasks. Inf Process Manag. 2009;45(4):427–37. doi:10.1016/j.ipm.2009.03.002.
44. Fawcett T. An introduction to ROC analysis. Pattern Recognit Lett. 2006;27(8):861–74. doi:10.1016/j.patrec.2005.10.010.
45. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016 Jun 27–30; Las Vegas, NV, USA. p. 770–8. doi:10.1109/CVPR.2016.90.
46. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. An image is worth 16×16 words: transformers for image recognition at scale. arXiv:2010.11929. 2020.