ACCSCIENCE PUBLISHING

Journal of Seismic Exploration

ARTICLE

Microseismic event picking and classification for hot dry rock hydraulic fracturing monitoring using SeisFormer

Mingjun Ouyang[1], Zenan Leng[2], Haotian Hu[3], Zubin Chen[1], Fa Zhao[1], and Feng Sun[1]*

[1]Key Laboratory of Geo-Exploration Instrumentation of the Ministry of Education, College of Instrumentation and Electrical Engineering, Jilin University, Changchun, Jilin, China

[2]Research and Development Center, Changchun UP Optotech Co., Ltd., Changchun, Jilin, China

[3]Department of Research and Development, Jingwei Hirain Co., Ltd., Changchun, Jilin, China

(This article belongs to the *Special Issue: Advanced Artificial Intelligence Theories and Methods for Seismic Exploration*)

Abstract

Accurate seismic monitoring is vital for the safe operation of enhanced geothermal systems in hot dry rock (HDR) reservoirs; however, robust P- and S-wave classification and precise first-arrival picking remain difficult under low signal-to-noise ratios and complex noise conditions. Hence, in this study, we present SeisFormer, a Transformer-based framework that couples adaptive multi-scale windowing with joint time–frequency analysis. It allocates time–frequency resolution on the fly to overcome the limitations of a fixed-window short-time Fourier transform and slowly extracts varying trends and dominant periodicities from waveform sequences. To stabilize the modeling of long-range dependencies, we introduce regularized pseudoinverse attention, which retains the speedups of low-rank approximations while damping amplification in directions associated with small singular values. We evaluated SeisFormer on a unified, multi-site dataset with data from HDR operations in the Qinghai Gonghe Basin and from an unconventional hydraulic-fracturing field in North China. Compared with baselines (EQTransformer, PhaseNet), it exhibited better performance across real-world data, noise-augmented data with non-stationary composite noise, and overlapping multi-event scenarios. On real-world data, it attained 98.30% classification accuracy, with mean arrival-time errors of 1.42 ms (P) and 2.29 ms (S). Ablations show that each component contributes substantially, indicating robustness for near-real-time monitoring and deployment.

Keywords: Microseismic monitoring; Hot dry rock hydraulic fracturing; Picking and classification; Transformer; Adaptive multi-scale windowing; Time–frequency domain

***Corresponding author:**
Feng Sun
(sunfeng@jlu.edu.cn)

Citation: Ouyang M, Leng Z, Hu H, Chen Z, Zhao F, Sun F. Microseismic event picking and classification for hot dry rock hydraulic fracturing monitoring using SeisFormer. *J Seismic Explor*. doi: 10.36922/JSE025290036

Received: July 15, 2025

Revised: September 25, 2025

Accepted: October 24, 2025

Published online: December 2, 2025

1. Introduction

As a clean and renewable energy source, geothermal energy offers low carbon emissions, environmental friendliness, operational stability, high efficiency, and abundant resources. Among geothermal resources, hot dry rock (HDR) has attracted significant attention due to its large heat-storage capacity and development potential, and has become a substantial focus of global exploration.[1] To efficiently extract heat from HDR, hydraulic fracturing is commonly employed to increase reservoir permeability, promote heat flow, and improve energy recovery. Figure 1 illustrates the basic process of heat extraction from HDR:[2] High-pressure fluid is injected to induce rock failure; fractures propagate and release energy, accompanied by microseismic events.

Compared with conventional oil and gas reservoirs, hydraulic fracturing in HDR formations is more complex.[3,4] HDR rocks typically exhibit very low permeability and high strength, and their microseismic signals are broadband with a high-frequency bias. Under high-temperature, high-pressure conditions and elevated injection rates, rock strength is further reduced and microseismicity becomes more complex and heterogeneous in time and space; numerous closely spaced events often occur within short time windows, yielding intricate source distributions and substantially increasing the difficulty of signal processing and interpretation.[5-7] Reliable microseismic monitoring is therefore crucial for assessing fracture-propagation dynamics during HDR stimulation and for providing timely feedback for engineered fracture-extension analysis and field decision-making.[8,9]

In microseismic signal processing, phase identification and first-arrival picking are two core tasks. Traditional methods (e.g., short-time average/long-time average [STA/LTA],[10] and Akaike information criterion[11]) perform well under ideal conditions but are prone to false positives and less effective at low signal-to-noise ratios (SNR) and in complex noise environments, limiting their suitability for HDR field applications.[12-16] Recent advances in deep learning have substantially improved detection and phase picking for seismic and microseismic signals.[17-19] Convolutional and recurrent architectures, such as PhaseNet,[20] PickCapsNet,[21] and a convolutional neural network (CNN) + long short-term memory (LSTM)[22] learn discriminative features but still struggle to model long-range dependencies and cross-scale coupling. Transformer-based models, via self-attention, provide global dependency modeling and have become an important framework for seismic time-series analysis. Representative works include EQTransformer, which jointly models detection and phase picking for regional and teleseismic catalogs; EQCCT, which couples compact CNNs with transformers for efficiency and improves cross-domain robustness via basin-scale transfer learning; SeisT, which uses multitask learning to unify detection, phase classification, and arrival-time estimation; and ICAT-Net, which leverages lightweight attention to balance accuracy and efficiency.[23-28] In mining and engineering scenarios, prior work has also explored handcrafted feature representations and hybrid CNN–transformer classifiers. However, many of these methods target conventional seismic catalogs or relatively stationary noise. In particular, transformer pipelines trained on regional or teleseismic data—characterized by lower event density and more stationary backgrounds—generalize poorly to HDR wavefields featuring overlapping onsets, narrowband harmonics, and low-frequency drift. Moreover, the quadratic cost of full attention and fixed analysis windows can introduce latency and unstable pick times on long streams sampled at 1 kHz with rapid cross-scale variability, motivating a time-frequency-aware architecture with adaptive windowing and a stabilized Nyström attention mechanism.

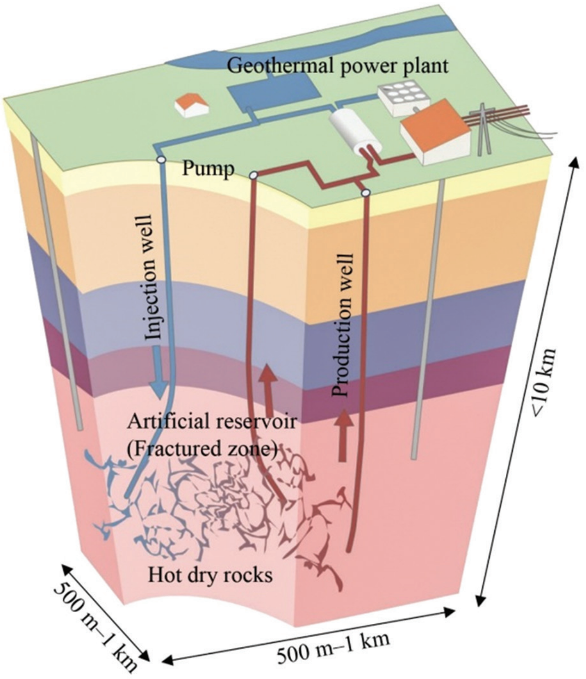

Figure 1. Schematic representation of enhanced geothermal systems using hydraulic fracturing in hot dry rock

In this context, we propose the SeisFormer, a time–frequency transformer framework tailored to HDR hydraulic-fracturing microseismic signals. Our main contributions are as follows:

(i) Time–frequency co-modeling with a transformer. We fuse short-time Fourier transform (STFT)- and two-dimensional (2D)-convolution-derived time–frequency features with self-attention to jointly represent low-frequency trends and high-frequency transients, matching the broadband, high-frequency-biased characteristics of HDR microseismic data.

(ii) Adaptive multi-scale window selection. We adapt window scales based on trend and periodicity cues and employ dynamic grouping and routing for efficient cross-scale modeling in dense-event, strongly non-stationary scenarios.

(iii) Robust and efficient attention. We incorporate a Tikhonov-regularized pseudoinverse into Nyström attention, delivering low-rank speedups while improving numerical stability, thereby supporting million-sample sequences for engineering deployment.

1.1. Time–frequency characteristics

Microseismic data from HDR hydraulic fracturing in enhanced geothermal system reservoirs are complex, non-stationary time series with distinct time–frequency structure.[29] Waveforms comprise background noise, P waves, S waves, and coda, with phase durations and amplitudes varying across operating conditions. In the frequency domain, P waves have higher-frequency and lower-amplitude, whereas S waves have lower-frequency and higher-amplitude.[30,31] Consequently, models must capture both low- and high-frequency content and adapt across multiple time scales to distribution shifts and transient changes.

1.2. Challenges in manual labeling

P- and S-wave arrival times are commonly picked manually, which is labor-intensive and susceptible to inter- and intra-operator variability. There is a clear need for automated classification and picking methods that are efficient, robust, accurate, and less labor-intensive.

1.3. Multi-event scenarios

Fracture propagation and injection fluctuations often trigger closely spaced, overlapping events. Fixed-window approaches are limited in this regime: Short windows truncate long events, whereas long windows include excessive noise. When signal lengths vary widely and event density is high, processing performance degrades and errors propagate downstream to subsequent modeling stages.

2. Proposed method

2.1. Model architecture

We propose the SeisFormer, a time–frequency modeling framework for P/S classification and first-arrival picking. As illustrated in Figure 2, the model (i) performs per-sample adaptive window selection to choose the processing scale, (ii) derives interpretable time–frequency representations via STFT coupled with 2D convolutions, and (iii) models long-range dependencies with a transformer whose self-attention is stabilized by a Tikhonov-regularized pseudoinverse to enhance numerical robustness and computational efficiency.

Section 2.2 introduces the trend- and dominant-frequency-guided complex routing for window selection. Section 2.3 explains how the selected window jointly determines the STFT/2D-convolution hyperparameters and the construction of the time–frequency tensor. Section 2.4 presents Nyström attention with a Tikhonov-regularized pseudoinverse.

2.2. Adaptive multi-scale time windows

To capture the multi-scale, time-varying characteristics of microseismic signals, we proposed a dynamic

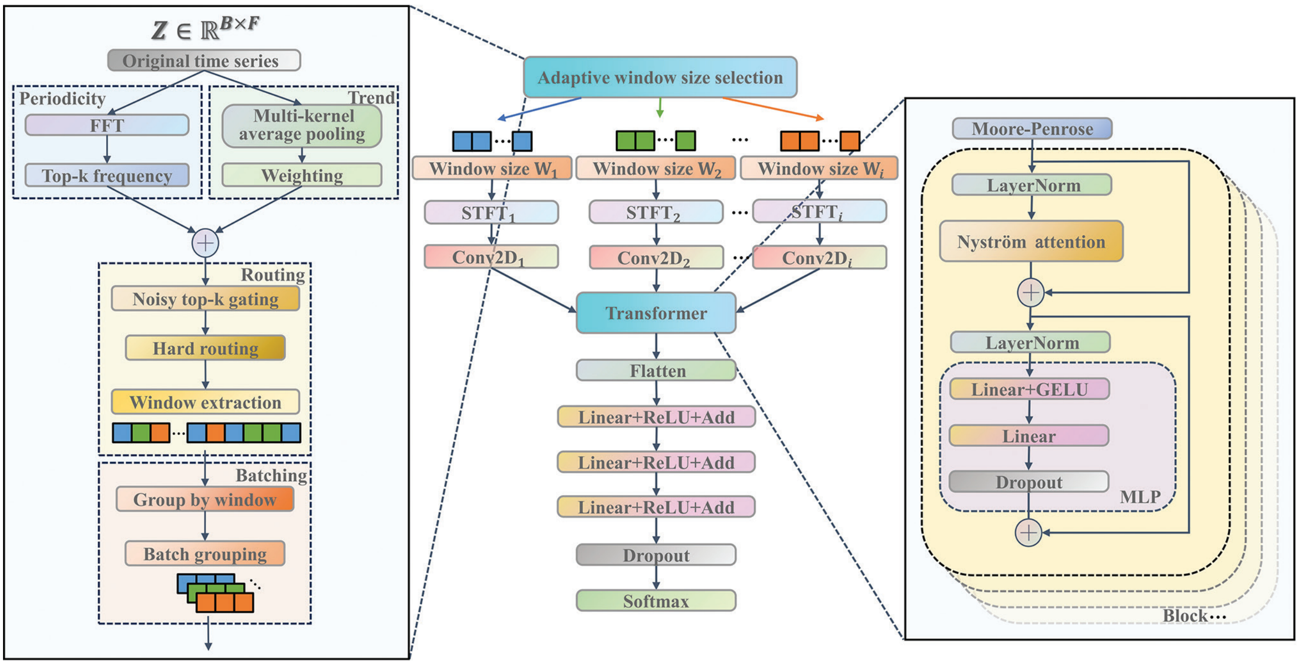

Figure 2. Overview of the SeisFormer model architecture

Abbreviations: Conv2D: Convolutional two-dimensional layer; FFT: Fast Fourier transform; MLP: Multilayer perceptron; STFT: Short-time Fourier transform.

window-selection method that fuses trend and periodicity cues. The technique first extracts multi-scale trend and periodicity representations from the input sequence and compresses them along the temporal axis to obtain a compact indicator vector *f*. We then scored a pre-defined candidate window set $\{W1,\ldots Wn\}$ with *f*; during training, light Gaussian noise was injected into the scores to improve generalization. During inference, a hard-routing strategy (k = 1) selected the optimal window length $W_{i\backslash *}$ to drive the subsequent time–frequency modeling.

Let the input be $z∈R^{B\times W\times D}$, where *B* is the batch size, *W* is the maximum observation window considered by the selector, and *D* = 1 is the channel dimension. Given a scale set $K=\{k_1|,\ldots,k_m|\ \}$, we computed multi-scale moving averages using **Equation (I)**:

$$z_{\mathrm{trend}}^{(k)}=\mathrm{MA}\left(z;k\right)\in\mathbb{R}^{B\times W\times D},\ \forall k\in\mathcal{K} \tag{I}$$

The per-scale trend components were fused with learnable weights, as shown in **Equation (II)**:

$$z_{\mathrm{trend}}=\sum_{k=1}^{m}\alpha_k z_{\mathrm{trend}}^{(k)},\ \alpha_k=\frac{\exp\left(\omega_k\right)}{\sum_{j=1}^{n}\exp\left(\omega_j\right)},$$
$$\omega_k=\mathrm{W}_{\mathrm{trend}}\cdot\mathrm{AvgPool}(z_{\mathrm{trend}}^{(k)})+\mathrm{b}_{\mathrm{trend}} \tag{II}$$

Where AvgPool(·) denotes temporal average pooling and $\{a_k\}$ are softmax-normalized scale weights.

In terms of periodicity features, for each sample, we applied a fast Fourier transform to obtain a complex spectrum Z_{FFT}. The magnitude was calculated using **Equation (III)**:

$$\left|Z_{\mathrm{FFT}}\right|=\sqrt{\mathrm{Re}\left(Z_{\mathrm{FFT}}\right)^2+\mathrm{Im}\left(Z_{\mathrm{FFT}}\right)^2} \tag{III}$$

To emphasize dominant periodic components, we selected the top-*κ* frequency indices by magnitude and retained the corresponding real-valued magnitude features, yielding $Z_{\mathrm{freq}}^{(\kappa)}$. This reduces dimensionality while preserving the principal periodic structure.

For fusion and selection, we performed temporal average pooling on the trend and periodicity features separately and used additive fusion to obtain the indicator vector, as shown in **Equation (IV)**:

$$\mathrm{f}=\mathrm{AvgPool}\left(z_{\mathrm{trend}}\right)+\mathrm{AvgPool}\left(Z_{\mathrm{freq}}^{(\kappa)}\right) \tag{IV}$$

Candidate windows were scored by logits = $W_g f + b_g$. During training, we added zero-mean Gaussian noise $N\ (0,\ \sigma^2)$ to the logits (with *σ* selected on the validation set) to mitigate overfitting near decision boundaries and improve out-of-distribution robustness. During inference, we adopted hard routing by choosing a single window via $i^{\backslash *}$ = arg $\max_i$ logits$_i$, and used $W_{i\backslash *}$ for subsequent time–frequency modeling. The fusion and routing procedure is summarized in Algorithm 1.

For dynamic bucketing and end alignment, since samples within the same batch can select different window lengths, we dynamically grouped (bucketed) samples by their chosen window and formed sub-batches per window. For each group, a sample was fed to the segment obtained

Algorithm 1. Trend-period fusion and hard routing

```
Inputs: $z∈R^{B\times W\times 1}$, kernel set $K = \{k_1|,\ldots, k_m|\}$, top−κ
Outputs: selected index $i^*$ and window length $W_{i^*}$
//Multi-scale trend extraction
For each $k∈K$ do
    $z_{\mathrm{trend}}^{(k)|}\ |\ \leftarrow \mathrm{MovingAverage}(\mathbf{z};\ \mathrm{window}=\mathrm{k})$
end for
//Learnable weighting across scales (softmax on pooled cues)
for each $k∈K$ do
    $u_k \leftarrow \mathrm{AvgPool}(z_{\mathrm{trend}}^{(k)})$ // pool along time
    $\alpha_k \leftarrow W_{\mathrm{trend}}\cdot u_k + b_{\mathrm{trend}}$
end for
$\alpha \leftarrow \mathrm{softmax}([\omega_k]_k)\ //B\times m$, along k
$\bar{Z}_{\mathrm{trend}} \leftarrow \sum_k \alpha_k z_{\mathrm{trend}}^{(k)}\ //B\times W\times 1$
//Periodicity via FFT (keep Top-k complex components)
$Z_{\mathrm{fft}} \leftarrow \mathrm{FFT}(z)$
mag $\leftarrow \sqrt{\Re(Z_{\mathrm{fft}})^2+\Im(Z_{\mathrm{fft}})^2}$
idx ←TopK (mag,κ)//per−sample
$Z_{\mathrm{freq}}$ ←Gather (mag, idx)
//Additive fusion and hard routing
f $\leftarrow \mathrm{AvgPool}(\bar{Z}_{\mathrm{trend}})+\mathrm{AvgPool}(Z_{\mathrm{freq}})$
logits ←$W_g f+b_g$
if training then
    logits ←logits+ε//ε~$N\ (0,\sigma^2)$
    $P$ ←softmax (logits)//optional: for logging/analysis
    $i^*$ ←arg $\max_i p_i$
else
    V//Hard routing at inference: k=1
    $i^*$ ←arg $\max_i$ (logits)
end if
return $i^*$, $W_{i^*}$
```

by slicing from the sequence end leftward with length $W_{i\backslash *}$; the supervision signal (class label and/or arrival-time label) was aligned to the window end. This ensured a one-to-one correspondence among "context length–feature extraction–supervision" while preserving batch efficiency. The selector consisted only of linear transforms and pooling, and dynamic grouping was a tensor reindexing/slicing operation; the overall computational overhead was negligible.

2.3. Time–frequency feature extraction based on STFT and 2D convolution

The STFT preserves both temporal and spectral information and is therefore well suited to low-SNR, compositionally similar, time-varying microseismic signals.[32] In this work, we mapped the input microseismic sequence from the time domain to the time–frequency domain to extract more discriminative spectral features and characterize energy evolution across time. The STFT window length is a key hyperparameter: Increasing the window improves frequency resolution (smaller $\Delta f = f_{s|}/n_{fft}$) but reduces time resolution (larger $\Delta t = H/f_s$)and increases temporal smoothing $\tau_{win} = n_{win}/f_s$; the converse holds for shorter windows.

Let $f_{s|}$ be the sampling rate, n_{fft} the DFT size, n_{win} the window length (we set $n_{fft} = n_{win}$), and $H$the hop size. For a discrete signal $z[n]$, the STFT was calculated using **Equation (V)**:

$$Z(t,f)=\sum_{n=0}^{N-1} z[n]\,\omega[n-tH]\,e^{-j2\pi fn/\mathrm{N}} \tag{V}$$

Where ω[·] is the analysis window (we used a Hann window), t indexes time frames, and f indexes frequency bins. For a batch $z\in \mathrm{R}^{B\times T}$, the STFT produced **Equation (VI)**:

$$\mathrm{Z}_{\mathrm{STFT}} \in \mathbb{C}^{\mathrm{B\times F\times T'}},\ \ \mathrm{F}=\left[\frac{n_{\mathrm{fft}}}{2}\right]+1,\ \ \mathrm{T}' \approx \left[\frac{t-n_{\mathrm{win}}}{H}\right]+1 \tag{VI}$$

with the Nyquist limit $f_{s|}/2$. To retain both magnitude and phase while matching a 2D CNN, we stacked the real and imaginary parts along the channel dimension, forming stack $[\mathrm{Z}_{\mathrm{real}}, \mathrm{Z}_{\mathrm{imag}}]\in \mathrm{R}^{\mathrm{B\times 2\times F\times T'}}$, which was then passed to a 2D convolution followed by ReLU, as shown in **Equation (VII)**:

$$\mathrm{Z}_{\mathrm{conv}} = \mathrm{ReLU}(\mathrm{W} \circledast \mathrm{stack}[\mathrm{Z}_{\mathrm{real}}, \mathrm{Z}_{\mathrm{imag}}]+\mathrm{b}) \tag{VII}$$

where ⊛ denotes 2D convolution on the frequency–time plane. The 2D CNN captures local structures within a single band and cross-band/time dependencies, enabling short-term spectral trend modeling and inter-band coordination.

Following the hard-routing selection in Section 2.2, once the sample-level optimal window $W_i^* \in \{128,256,512\}$ was chosen, we adapted n_{fft} and the hop size H accordingly (50% overlap, $\mathrm{H} \approx \mathrm{n_{fft}}/2$), and proportionally adjusted the number of 2D convolution channels to achieve comparable time–frequency resolution and controlled computation across scales. The mapping used in this paper is shown in **Equation (VIII)**:

$$\left(n_{fft}, H, channels\right)=\begin{cases}(62,31,12), & W_i^*=128\\ (126,63,6), & W_i^*=256\\ (256,127,3), & W_i^*=512\end{cases} \tag{VIII}$$

motivated by setting $n_{\mathrm{fft}} \approx W_i^*/2$ to align $\Delta f = f_{s|}/\mathrm{n_{fft}}$ across scales. Larger windows increase frequency resolution but also the number of frequency bins FFF; hence, we reduced the convolution channels inversely (12→6→3) to offset the growth in feature-map size and stabilize throughput. Modern fast Fourier transform implementations handle non-power-of-two lengths efficiently, so the above choices were numerically and computationally sound. This parameterization was empirically validated as the best-performing configuration, yielding the strongest trade-off among arrival-time accuracy, classification metrics, and efficiency on the validation set, and demonstrating stable behavior in ablation studies.

2.4. Nyström attention with Tikhonov-regularized pseudoinverse

After the 2D convolution, the frequency–time maps were reshaped to form a sequence for the transformer. Let $\mathrm{Z}_{\mathrm{conv}} \in \mathbb{R}^{\mathrm{B\times C\times F\times T''}}$ denote the convolutional output (batch B, channels C, frequency bins F, frames T''). We flatten the (C, F) axes to obtain **Equation (IX)**:

$$\mathrm{X}\in \mathrm{R}^{\mathrm{B\times n\times d}},\ \mathrm{n} \coloneqq T'',\ \mathrm{d} \coloneqq C\times F \tag{IX}$$

and fed X to the transformer (SeisFormer) for further sequence modeling. Multi-head self-attention captured long-range temporal–spectral dependencies, improving microseismic event discrimination.

For notation unification, we set $\mathrm{n} \coloneqq T''$ (sequence length after 2D CNN) and $\mathrm{d} \coloneqq C\times F$ (embedding width before head-splitting). With h heads, the per-head width was $d_h = d/h$. Given $\mathrm{X}\in\mathrm{R}^{\mathrm{B\times n\times d}}$, after linear projections and head splitting, we have $Q,K,V \in \mathbb{R}^{B\times h\times n\times d_h}$. In **Equation (X)**, the scaling $\sqrt{d}$ refers to the per-head width, that is, $d \equiv d_h$.

SeisFormer alternates between self-attention and feed-forward neural network (FFN) blocks and, unlike a standard transformer, employs a Nyström approximation to self-attention for efficiency on long sequences.[33-37] In conventional attention, the row-wise scaled dot-product was calculated using **Equation (X)**:

$$S\left(Q,K\right) = softmax\left(\frac{QK^{\top}}{\sqrt{d}}\right) \in \mathbb{R}^{n\times n} \qquad \text{(X)}$$

We chose $m \ll T$ landmarks with index set M, and defined it as **Equation (XI)**:

$$A \triangleq S(Q_M, K_M) \in \mathrm{R}^{m\times m},\ B \triangleq S(Q, K_M) \in \mathrm{R}^{n\times m},\ C \triangleq S(Q_M, K) \in \mathrm{R}^{m\times n} \qquad \text{(XI)}$$

The classical Nyström approximation is shown in **Equation (XII)**:

$$\hat{S} = BA^{+}C \qquad \text{(XII)}$$

where A^{+} is the Moore–Penrose pseudoinverse. If the true attention S has rank at most m and the landmark submatrices are full rank, we can write $S = UV^{\top}$ with U, $V \in \mathrm{R}^{n\times m}$, which yields $B = UV_M^{\top}$, $C = U_M V^{\top}$ and $A = U_M V_M^{\top}$. Using $(XY)^{+} = Y^{+}\ X^{+}$under the usual full-rank side conditions, we obtain **Equation (XIII)**:

$$BA^{+}C = (UV_M^{\top})((V_M^{+})^{\top}\ U_M^{+})(U_M V^{\top}) = U\underbrace{[V_M^{\top}(V_M^{+})^{\top}]}_{=I_m}\underbrace{[U_M^{+}U_M]}_{=I_m}V^{\top}$$

$$= UV^{\top} = S \qquad \text{(XIII)}$$

Hence, the Nyström reconstruction is exact in this ideal case. In general, $\hat{S}$ still preserves the landmark rows/columns ($\hat{S}_{:,M} = B, \hat{S}_{M,:} = C$)and gives the minimum-norm solution consistent with them.

For numerical stability, we replaced A^{+} with a Tikhonov-regularized pseudoinverse A_{λ}^{+}. Let the SVD of A be the formula shown in **Equation (XIV)**:

$$A = U_A \Sigma V_A^{\top},\quad A^{+} = V_A \Sigma^{+} U_A^{\top},\ \Sigma^{+} = diag\left(\sigma_i^{-1}\ 1_{\{\sigma i > 0\}}\right) \qquad \text{(XIV)}$$

When the landmarks are highly correlated or the subset is skewed, small singular values make the plain inverse amplify noise along those directions. In practice, we used the Tikhonov-regularized pseudoinverse, as shown in **Equation (XV)**:

$$A_{\lambda}^{+} = \left(A^{\top} A + \lambda I\right)^{-1} A^{\top} = V_A diag\left(\frac{\sigma_i}{\sigma_i^{2} + \lambda}\right) U_A^{\top} \qquad \text{(XV)}$$

Whose operator norm satisfies **Equation (XVI)**:

$$\|\ A_{\lambda}^{+}\|_2 = \max_i \frac{\sigma_i}{\sigma_i^{2} + \lambda} \leq \frac{1}{2\sqrt{\lambda}} \qquad \text{(XVI)}$$

thereby avoiding the $1/\sigma_i$ blow-up as $\sigma_i \to 0$. The resulting stable reconstruction is shown in **Equation (XVII)**:

$$\widehat{S_{\lambda}} = BA_{\lambda}^{+}C \qquad \text{(XVII)}$$

Spectrally, the regularization acts as a smooth shrinkage on each singular direction, as seen in **Equation (XVIII)**:

$$AA_{\lambda}^{+}A = U_A diag\left(\frac{\sigma_i^{3}}{\sigma_i^{2} + \lambda}\right) V_A^{\top},\ A - AA_{\lambda}^{+}A$$

$$= U_A diag\left(\frac{\lambda\sigma_i}{\sigma_i^{2} + \lambda}\right) V_A^{\top} \qquad \text{(XVIII)}$$

Hence, the approximation-bias trade-off is monotone in λ. This shrinkage bounds the amplification of perturbations and yields smoother gradients during back-propagation, as the Lipschitz constant along the landmark path is controlled by $\|\ A_{\lambda}^{+}\|_2$.

In terms of complexity, exact attention incurs $O(n^2d)$ time and $O(n^2)$memory, whereas the Nyström scheme requires $O(nmd) + O(m^3)$ to construct B and $C$and to solve a single $m \times m$ system. Under the common regime, $m << n$, the $O(m^3)$term is negligible, and the overall complexity is effectively $O(nmd)$. Replacing A^{+} with A_{λ}^{+} preserved this low-rank acceleration while improving numerical stability for long-range dependency modeling in microseismic signals; this matches our implementation, which computes the (regularized) pseudoinverse on the landmark attention block.

The processed sequence features were flattened and passed to each transformer layer. In each layer, the features were further optimized through the FFN, which consists of linear layers and GELU activations to extract non-linear relationships and enhance feature representation. Residual connections and layer normalization were applied to both self-attention and FFN blocks to accelerate training and prevent gradient vanishing, ensuring stable signal propagation through the network and better adaptation to complex time–frequency structure. The output features then pass through three linear transformations with ReLU and dropout, followed by a final linear layer that maps to the task space; finally, scores were normalized to predict probabilities for P-waves, S-waves, and noise, completing the microseismic event classification.

3. Experimentation

3.1. Parameter configuration

To enhance data representativeness and rigorously evaluate cross-site generalization, we merged data from two independent sites into a joint dataset under a unified organization and labeling protocol: An HDR project in the Gonghe Basin, Qinghai, China, and an unconventional hydraulic-fracturing site in North China. The Qinghai data were acquired in 2020 using an in-house system. Monitoring at well GR-1 (approximately 2 km from GH-02/3) used a 12-level, three-component downhole array (1,100–1,400 m depth; 20 m inter-level spacing) together with a "surface–shallow-well–deep-well" joint layout: 12 surface lines within an 8 × 8 km area centered on GH-02/3 (25 m station spacing; at least 1,512 channels) and 60 three-component shallow-well instruments (10–25 m installation depth), providing coverage to a target depth of ~4,000 m. At the North China site, production wells were arranged in belts, targeting formations at depths of ~3,700–4,300 m. The two sites shared a consistent data organization and labeling protocol: Both used a 1 ms sampling interval, and phases were labeled as noise = 0, P = 1, and S = 2 (Figure 3). Across all annotated frames, class proportions were 70.01% noise, 13.70% P, and 16.29% S.

Modeling was conducted on the joint dataset, comprising a total of 4,000 single-channel time series (Δt = 1 ms)as inputs. Each channel was demeaned and standardized via z-score using statistics computed from the training split of the joint dataset to ensure comparability across sites and channels. Unless otherwise stated, the data were split in a ratio of 8:1:1 into training/validation/test sets. We adopted Adam (initial learning rate 1×10^{-4}) with ReduceLROnPlateau (factor = 0.1, patience = 5) based on validation metrics to promote stable convergence. The training objective combined cross-entropy with L2 regularization (weight decay = 0.003). After each epoch, the model was evaluated on the validation set, and early stopping was applied to curb overfitting and improve generalization.

Given the high noise fraction (70.01%) in the joint dataset, we employed class-weighted cross-entropy during training and assigned a weight of 1.2 to P- and S-phase frames to strengthen discrimination around arrivals, thereby improving picking sensitivity and robustness. All training and evaluation settings were applied uniformly across both sites to ensure fair comparison and reproducibility.

3.2. Comprehensive experimental evaluation

To comprehensively evaluate model performance under different conditions, we designed a series of experiments on a strictly held-out test set from the joint dataset. This test set consisted of 200 data segments from each site (400 in total) and was entirely non-overlapping with the training/validation data. The evaluation scenarios included the real environment, the noise environment, and multi-event cases. To ensure fairness and reproducibility, all methods followed a unified pre-processing pipeline before entering their respective models/algorithms. The evaluation protocol then proceeded in four stages: (i) Multiple methods were compared under the real scenario; (ii) better-performing methods were included in the noise tests; (iii) complex-signal handling was assessed via the multi-event scenario; and (iv) ablation studies were conducted to quantify the contributions of key components.

The class set be {0:Noise, 1:P, 2:S}. Define the confusion matrix be C∈N$^{3\times3}$ with entries C_{ii} = #{samples with true class i predicted as j}, i, j∈{0,1,2}.Then, the overall accuracy is calculated using **Equation (XIX)**:

$$\text{Accuracy} = \frac{\sum_{i=0}^{2} C_{ii}}{\sum_{i=0}^{2}\sum_{j=0}^{2} C_{ij}} \quad \text{(XIX)}$$

For class *i*, **Equation (XX)** was used:

$$TP_i = C_{ii},\ \ FP_i = \sum_{k\neq i} C_{ki},\ \ FN_i = \sum_{k\neq i} C_{ik} \quad \text{(XX)}$$

Class-wise precision, recall, and F1 are shown in **Equation (XXI):**

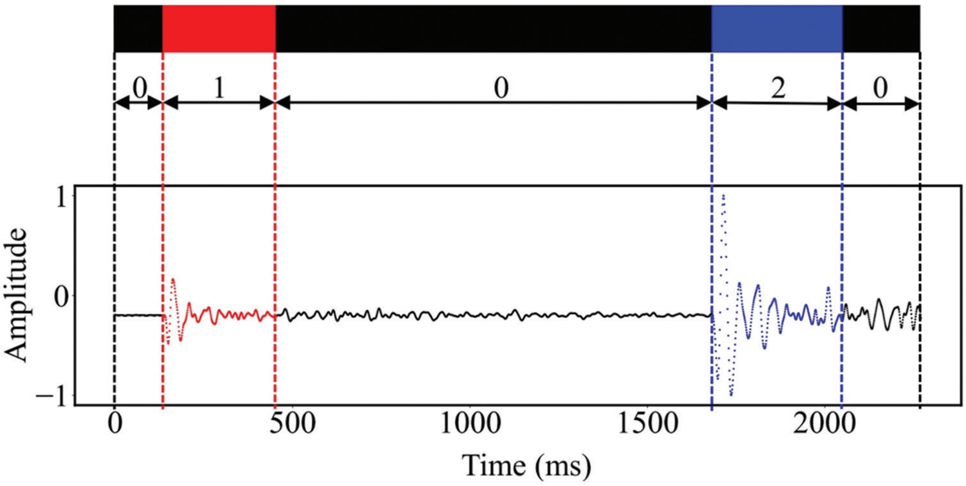

Figure 3. Manual annotation process

$$\mathrm{Precision}_i = \frac{TP_i}{TP_i + FP_i},\ \ \mathrm{Recall}_i = \frac{TP_i}{TP_i + FN_i},$$
$$\mathrm{F1}_i = \frac{2\mathit{Precision}_i\mathrm{Recall}_i}{\mathrm{Precision}_i + \mathrm{Recall}_i} \qquad \text{(XXI)}$$

Weighted aggregates used for class support $n_i = TP_i + FN_i$ (i.e., the number of true samples in class i) with $N = \Sigma_i n_i$, as shown in **Equation (XXII)**:

$$\mathrm{Precision} = \sum_{i=0}^{2}\frac{n_i}{N}\mathrm{Precision}_i,\ \ \mathrm{Recall} = \sum_{i=0}^{2}\frac{n_i}{N}\mathrm{Recall}_i,$$
$$\mathrm{F1} = \sum_{i=0}^{2}\frac{n_i}{N}\mathrm{F1}_i \qquad \text{(XXII)}$$

We also reported mean absolute error (MAE)-P and MAE-S for P/S arrival times, defined as the sample-wise mean absolute difference between the predicted and manually annotated arrivals. Together, these metrics and visualizations quantified both classification and arrival-time picking performance and enabled a consistent comparison of methods across the three scenarios.

3.2.1. Real environment experiment

We evaluated classification performance on real microseismic signals using SeisFormer, EQTransformer,[23] PhaseNet,[20] generalized phase detection (GPD),[38] LSTM, CNN, and STA/LTA.[10] Unless otherwise stated, all models were trained from scratch under a unified pre-processing pipeline, with a sampling rate of 1 kHz, identical train/validation/test splits, and identical label definitions. Model configurations were as follows: SeisFormer—an 8-layer Transformer with eight attention heads, model dimension 64, FFN/multilayer perceptron hidden size 256, dropout 0.5 on attention and FFN, and a multilayer perceptron head with 128 hidden units. EQTransformer was implemented following the public release and original architecture (convolutional encoder, residual CNN stack, 3×BiLSTM, detection decoder branch with multi-stage upsampling). PhaseNet is a one-dimensional U-Net with four down- and upsampled stages (downsampling kernel length 7, stride 4). GPD used four Conv1D layers plus two fully connected layers. The LSTM baseline used a two-layer bidirectional LSTM with 100 hidden units per direction. The CNN baseline is a lightweight one-dimensional CNN with three convolutional blocks and a fully connected head (kernel length 7; channels 32/64/128). STA/LTA is a short/long-window energy-ratio trigger under the same pre-processing/segmentation as the deep models (short/long windows 0.2s/2.0s; threshold tuned on the validation set). For fairness, we used matched optimization, regularization, learning-rate scheduling, early stopping, and random seeds across methods, without modifying baseline architectures.

As shown in Figure 4, SeisFormer, EQTransformer, and PhaseNet clearly outperformed the other baselines. Representative numbers are reported in Table 1: SeisFormer (Accuracy: 98.30%, precision: 97.40%, recall: 97.92%, F1: 97.66%; MAE-P: 1.42 ms, and MAE-S: 2.29 ms), EQTransformer (Accuracy: 96.90%; MAE-P: 1.90 ms, and MAE-S: 3.18 ms), PhaseNet (Accuracy: 94.80%; MAE-P: 4.76 ms, and MAE-S: 6.95 ms), while the remaining baselines lagged substantially behind. Overall, these three models constituted the first tier, with SeisFormer leading in both classification and arrival-time accuracy.

We also benchmarked forward-pass latency on an RTX 4060 (8 GB) + Intel i9-13900HX using single-channel 1 kHz/3 s input (the three-second window was used solely to standardize the latency benchmark), FP32, batch = 1. Results were the median of 100 runs after 20 warm-up iterations, measuring wall-clock time for the model forward only—including in-graph STFT and hard routing, and excluding data loading and disk I/O: SeisFormer ≈ 4.2 ms (GPU)/43 ms (CPU), PhaseNet ≈ 5.1 ms (GPU)/55 ms (CPU), EQTransformer ≈ 8.5 ms (GPU/94 ms (CPU). Under this accuracy–latency trade-off, SeisFormer is the most suitable for near–real-time deployment on the target hardware.

Confusion matrices for each method in Figure 5 further illustrate their strengths and weaknesses. SeisFormer attained an overall true-positive rate of 98.1%, with class-wise rates of 96.1% (P-wave) and 94.3% (S-wave), outperforming all other methods. Notably, most SeisFormer errors arise from small discrepancies between predicted and manually labeled endpoints of P- and S-wave arrivals; such endpoint disagreements have limited impact on microseismic monitoring and are therefore of low significance to the overall evaluation. To further substantiate SeisFormer's advantages, Figure 6A-D shows predictions on representative waveforms from the test sets of both datasets, visually demonstrating efficient classification and accurate arrival-time calibration.

3.2.2. Noise environment experiment

To more faithfully emulate field disturbances and align the evaluation with picking/classification objectives, we calibrated noise intensity using an event-referenced SNR (ER-SNR) and conducted stress tests with non-stationary composite noise that included low-frequency drift, power-line fundamentals and harmonics, impulsive interference, and colored background noise. This design better reflected real HDR noise characteristics than conventional whole-trace SNR and enabled an objective assessment of model robustness under realistic conditions. Concretely, for each record, we constructed an event window E (labels >0) and

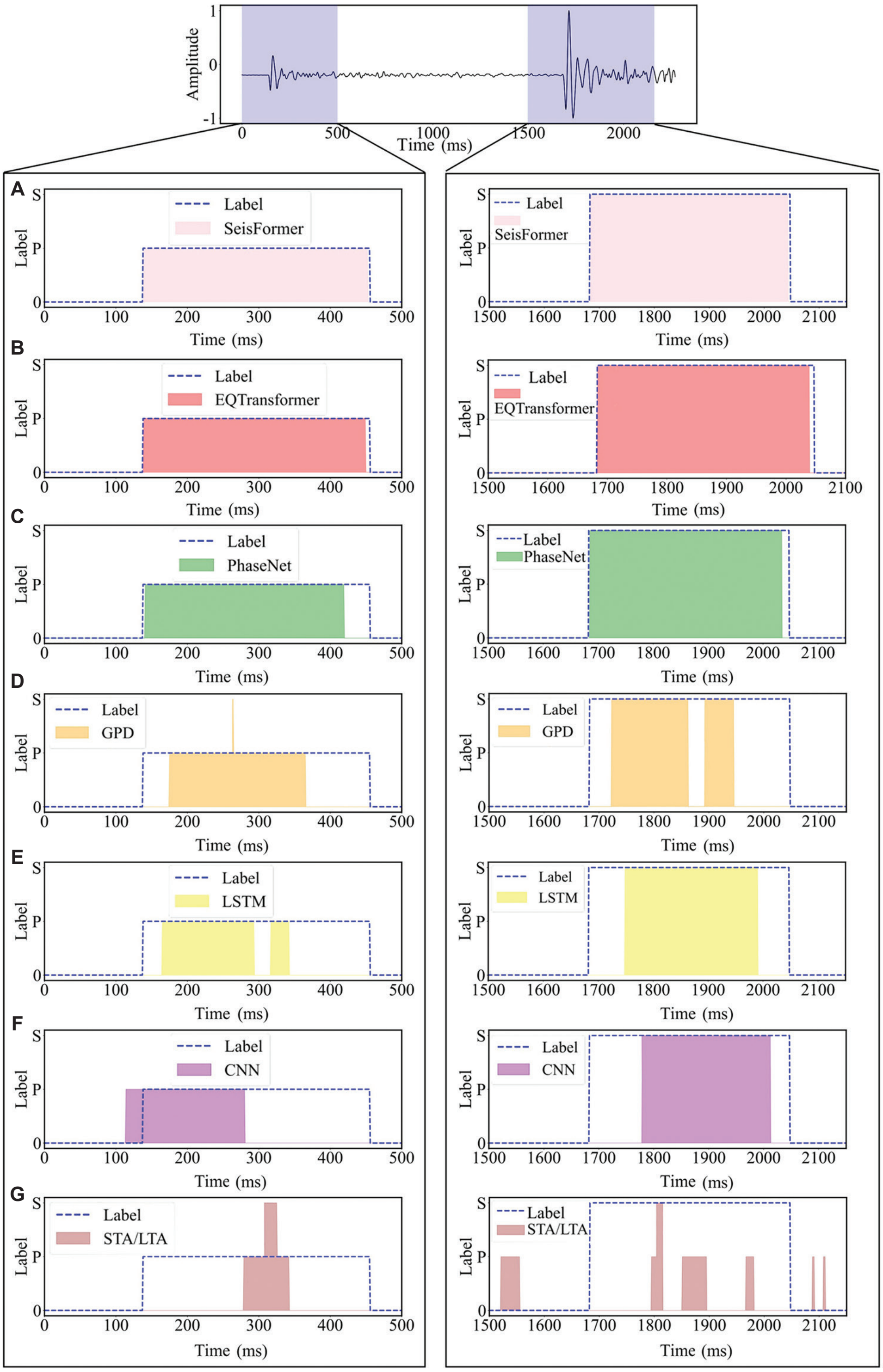

Figure 4. Comparison of classification and picking results from different methods. The same waveform was used to test each model, with the classification results for the 0–500 ms segment extracted to showcase performance in P-wave classification and arrival-time picking, and the 1,500–2,100 ms segment extracted to highlight performance in S-wave classification and arrival-time picking. SeisFormer, EQTransformer, and PhaseNet demonstrated strong performance; further comparisons and evaluations will be conducted in subsequent noise experiments. Classification results of the (A) SeisFormer model, (B) the EQTransformer model, (C) the PhaseNet model, (D) the GPD model, (E) the long short-term memory (LSTM) model, (F) the convolutional neural network (CNN) model, and (G) the short-term average/long-term average (STA/LTA) method.

Table 1. Comparison of classification performance and arrival time calibration errors for different models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)	Mean P-wave arrival error	Mean S-wave arrival error
SeisFormer	98.30	97.40	97.92	97.66	1.42 ms	2.29 ms
EQTransformer	96.90	96.15	96.48	96.31	1.90 ms	3.18 ms
PhaseNet	95.80	95.02	95.71	95.36	4.76 ms	6.95 ms
GPD	83.70	81.80	82.45	82.12	24.9 ms	30.6 ms
LSTM	85.90	85.10	85.35	86.10	15.1 ms	45.3 ms
CNN	82.10	80.10	81.95	81.30	21.4 ms	54.9 ms
STA/LTA	68.79	61.60	66.42	64.63	152 ms	224 ms

Abbreviations: CNN: Convolutional neural network; GPD: Generalized phase detection; LSTM: long short-term memory; STA/LTA: Short-term average/long-term average.

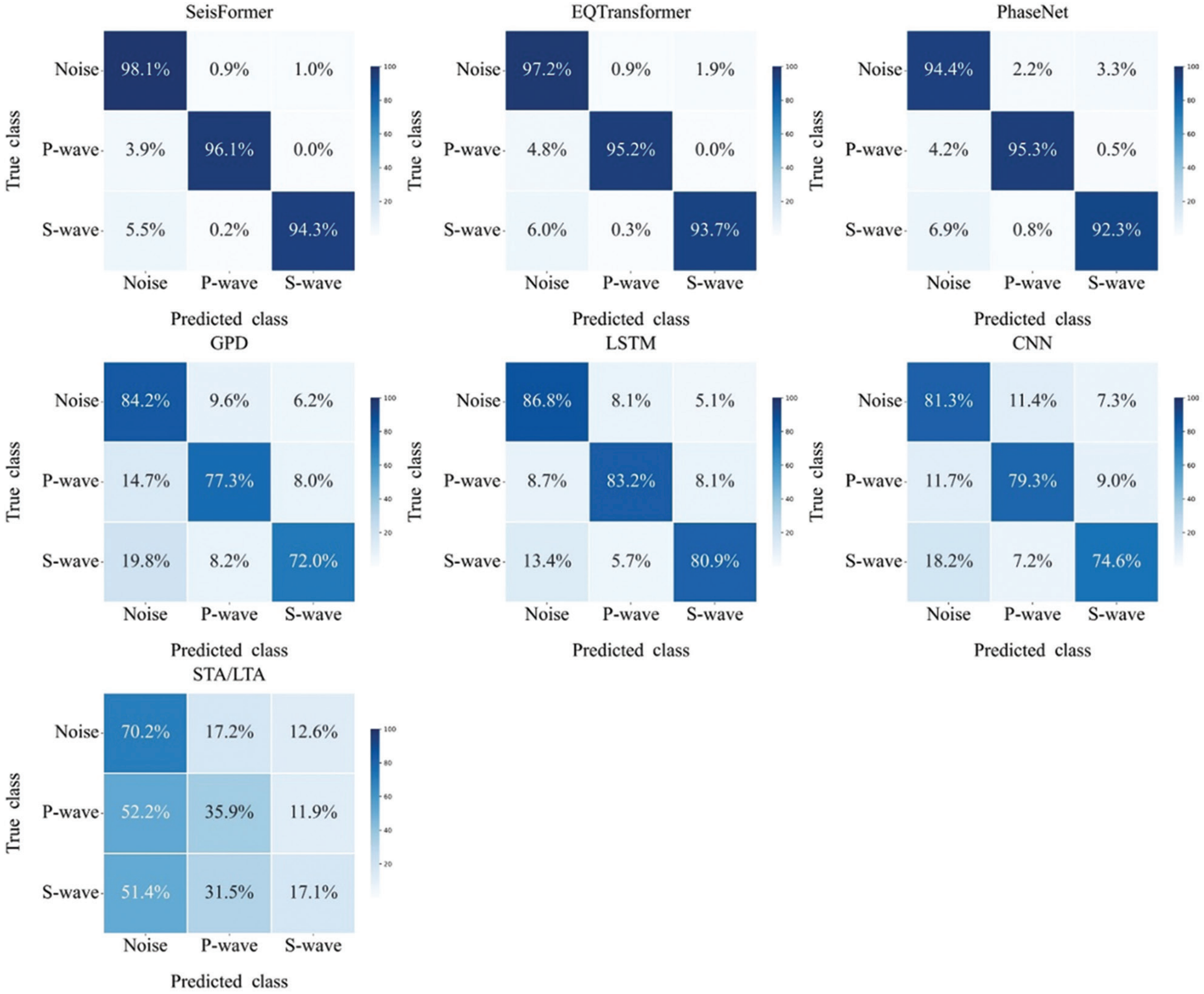

Figure 5. Comparison of confusion matrices for SeisFormer, EQTransformer, PhaseNet, GPD, LSTM, CNN, and STA/LTA on microseismic signal classification
Abbreviations: CNN: Convolutional neural network; GPD: Generalized phase detection; LSTM: long short-term memory; STA/LTA: Short-term average/long-term average.

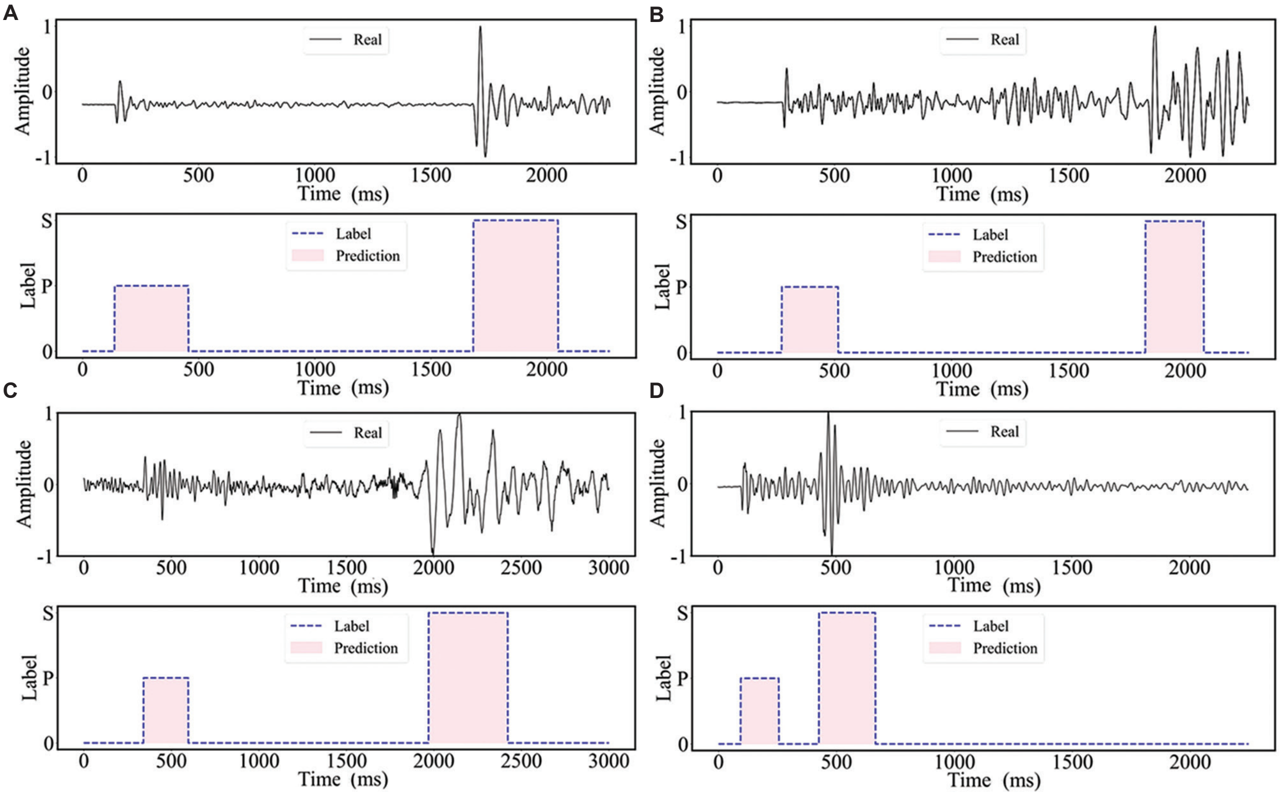

Figure 6. Classification results of the SeisFormer model on representative waveforms from the training set, verifying performance across different data types. (A and B) Qinghai site. (C and D) North China site.

a pre-event baseline window *B* from the second-column labels. The event mask was dilated by approximately ±100 samples to cover onsets and coda. The waveform was then baseline-centered using $\mu_B = \frac{1}{|B|}\sum_{i\in B} s[i]$, yielding $\tilde{s}[i] = s[i] - \mu_B$, which suppresses low-frequency drift in power estimation. In the presence of impulses and non-stationarity, we obtained stable energy estimates by combining trimmed mean of squares with baseline bootstrap length-matching: For any segment we averaged the squared amplitudes after two-sided 10% trimming to reduce the leverage of outliers; as |*B*|≫|*E*| for most records, we repeatedly sampled from *B* subsegments of length |*E*|, computed trimmed power for each replicate, and averaged across *K* = 30 replicates to remove biases due to unequal window lengths. This yielded **Equation (XXIII)**:

$$P_E E = Trim_{0.1}\left(\tilde{s}\left[E\right]^2\right),\ P_B = \mathbb{E}_{boot}\left[Trim_{0.1}\left(\tilde{s}\left[B_{sub}\right]^2\right)\right] \tag{XXIII}$$

and the ER-SNR (in dB) was defined as **Equation (XXIV)**:

$$ER - SNR_{dB} = 10\log_{10}\left(\frac{P_E}{P_B}\right) \tag{XXIV}$$

To match the field noise spectra, we did not add stationary Gaussian white noise. Instead, we synthesized a non-stationary composite of four components—baseline drift (random walk or 1/*f* -like), power-line harmonics (50/60 Hz and overtones with slow AM/PM), sparse impulsive spikes, and colored AR (1) background—and linearly mixed them with fixed relative weights, as shown in **Equation (XXV)**:

$$n_{raw} = w_d n_{drift} + w_h n_{harm} + w_i n_{imp} + w_c n_{col} \tag{XXV}$$

using w_d = 1.0, w_h = 1.0, w_i = 0.6, and w_c = 0.8. The noise powers within *E* and *B*, $P_{nE0}|$ and P_{nB0}, were estimated with the same robust procedure. Given a target ER-SNR level (let γ = $10^{ER-SNRdB/10}$, the composite noise was scaled and added as $x = s + \alpha n_{raw}$ so that the post-augmentation event/baseline power ratio met the target, as shown in **Equation (XXVI)**:

$$\alpha^2 = \frac{P_E - \gamma P_B}{\gamma P_{nB0} - P_{nE0}} \tag{XXVI}$$

If $P_{E|} - \gamma P_B \leq 0$ (the trace is already cleaner than the target) or $P_{nB0} - P_{nE0} \leq 0$ (the noise recipe concentrates relatively more energy in *E* than in *B*, contradicting the

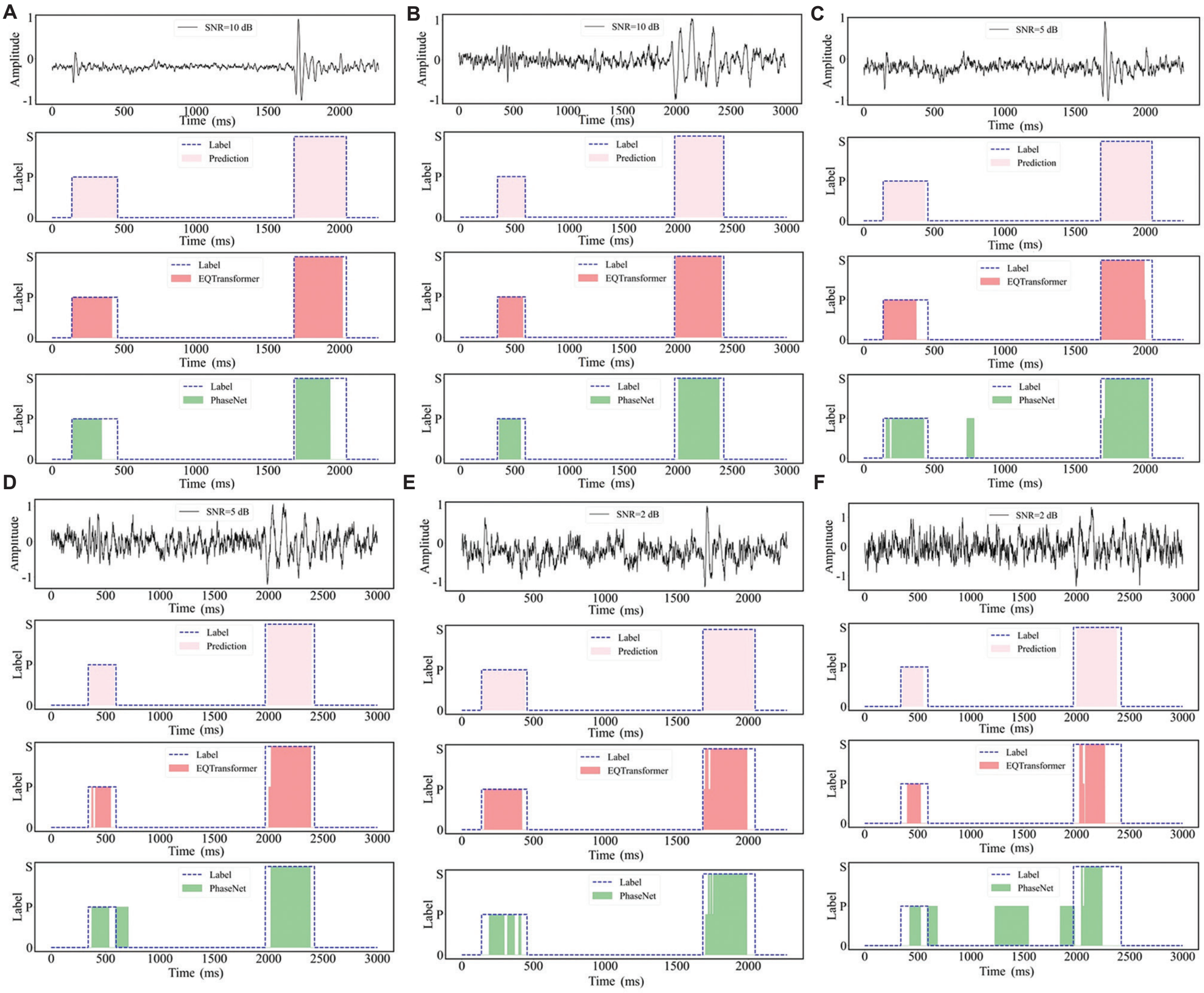

Figure 7. Phase classification and arrival-time picking across event-referenced signal-to-noise ratio (ER-SNR) levels. (A, C, and E) Qinghai. (B, D, and F) North China. Each column shows the same representative record under three noise settings (10/5/2 dB); rows are, from top to bottom, raw/noisy waveform (with ER-SNR), SeisFormer, EQTransformer, and PhaseNet. Performance degrades as ER-SNR decreases; SeisFormer consistently exhibits more precise and temporally coherent P/S predictions, smaller picking bias, and slower growth in false/missed detections across both datasets.

target balance), we kept the original waveform to avoid unrealistic distortion. All augmentations used fixed random seeds for reproducibility and fairness, and for each record and level, we computed and logged the achieved ER-SNR to verify calibration error against the target.

We adopted three ER-SNR levels (10 dB, 5 dB, and 2 dB), corresponding to moderate, strong, and extreme degradation, with a no-noise condition as the baseline. For each level, waveforms, spectrograms, and augmented samples were generated under fixed random seeds for inference and visualization. To provide representative comparisons, we selected two records from Figure 6 (panel A: Qaidam; panel D: North China) and evaluated SeisFormer, EQTransformer, and PhaseNet across the four noise conditions (no noise, 10 dB, 5 dB, 2 dB). All results were reported as raw values (accuracy, precision, recall, F1, and P/S arrival MAE), accompanied by corresponding waveforms and STFT spectrograms to visually demonstrate the degradation trend with increasing noise. Spectrograms used frame-wise adaptive STFT: At each time position, the window length was chosen by the selector in Section 2.2, and the STFT for that frame was computed as in Section 2.3; frames from different windows are interpolated onto a unified time–frequency grid and concatenated to form a continuous spectrogram. We also overlaid a window-identifier ribbon aligned to the time axis to indicate the time–frequency resolution used per segment. Related visualizations and noise-robustness curves are shown in Figures 7 and 8.

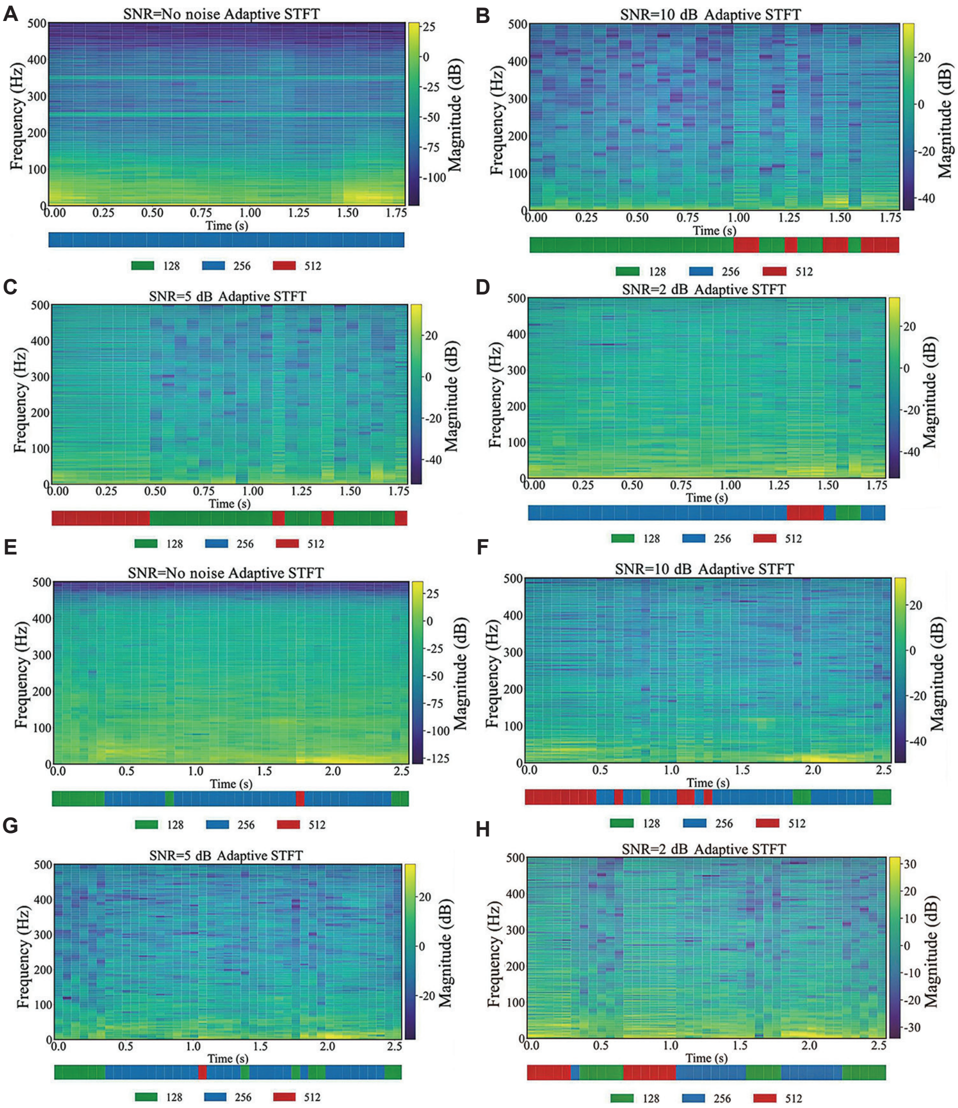

Figure 8. Adaptive short-time Fourier transform (STFT) time–frequency spectrograms with window-selection indicator bars. (A-D) Qinghai. (E-H) North China. Shown are the adaptive STFT magnitude spectra of the same two records as in Figure 7 under different noise/scenario settings (vertical axis: Frequency/Hz; horizontal axis: time/s; color scale: magnitude/dB). Below each spectrogram, the colored bar indicates the frame-wise window selection (green = 128, blue = 256, red = 512) with pixel-wise alignment to the spectrogram's time axis.
Abbreviation: SNR: Signal-to-noise ratio.

As shown in Figure 7A-F, when ER-SNR decreased from 10 dB to 5 dB and 2 dB, the classification and arrival-time accuracy of all three methods degraded: Baseline elevation and impulsive interference raised false alarms on non-event segments and introduced systematic delays in the picks. In contrast, SeisFormer consistently produced more temporally coherent and better-aligned P/S predictions on both datasets while maintaining a tighter temporal window—at 10 dB it nearly coincided with the annotations; at 5 dB it still stably covered the main energy of the events with markedly fewer false positives than EQTransformer and PhaseNet; and under the extreme 2 dB condition, although slightly contracted, its onset/offset remained broadly consistent with the labels, whereas the baselines exhibited fragmented or drifting predictions, leakage of energy into the baseline, or P/S confusion. Detailed metrics are presented in Table 2.

The STFT spectrograms in Figure 8 make the non-stationarity and narrowband harmonics, as well as their evolution with SNR, visually explicit, and empirically demonstrate that frame-wise adaptive windowing dynamically allocated time–frequency resolution: Under high noise it favored longer windows to enhance frequency resolution and suppressed harmonics and low-frequency drift, whereas at higher SNR it adopted shorter windows to preserve onset transients—thereby highlighting the band-limited event energy even at low SNR. These visualizations provide direct and interpretable evidence for the robustness of the proposed time–frequency strategy. Overall, as noise intensifies, SeisFormer exhibited slower growth in false/missed detections and smaller arrival-time drift, indicating stronger noise resilience.

3.2.3. Multi-event scenario experiment

To assess the model's adaptability in complex signal conditions, we evaluated it on a dense-event window containing multiple consecutive P/S arrivals. As shown in Figures 9 and 10, the proposed model preserved clear P/S boundaries between adjacent events, with onsets and offsets closely matching the annotations. When inter-event intervals shortened and energy overlaps arose, it still robustly localized phase breakpoints and effectively suppressed cross-segment leakage. The spectrogram reveals that the model adaptively switches to shorter windows at rapid energy transitions to retain transient details, while favoring longer windows in regions with background undulations or strong narrowband interference to stabilize spectral structure. Consequently, in dense multi-event scenarios, the model achieved a favorable balance between arrival-time accuracy and noise robustness.

3.2.4. Ablation experiment

To quantify the contribution of each component to overall performance, we conducted ablation studies under realistic settings with a unified training/evaluation protocol (Table 3). Replacing Nyström attention with exact dot-product attention (without Nyström) reduced accuracy/F1 to 91.72%/91.93% and increased P/S arrival MAE to 6.36/7.71 ms, indicating that the Tikhonov-regularized pseudoinverse within the Nyström block

Table 2. Classification performance and arrival time calibration errors of the SeisFormer under different signal-to-noise ratios

Event-referenced signal-to-noise ratio	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)	Mean P-wave arrival error	Mean S-wave arrival error
None	98.30	97.40	97.92	97.66	1.42 ms	2.29 ms
10dB	95.73	94.75	94.69	95.38	3.02 ms	5.37 ms
5dB	92.88	93.13	92.90	93.96	6.33 ms	8.41 ms
2dB	87.02	86.76	87.06	87.17	12.48 ms	18.49 ms

Table 3. Comparison of classification performance and arrival time calibration errors under different model configurations

Method	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)	Mean P-wave arrival error (%)	Mean S-wave arrival error (%)
SeisFormer	98.30	97.40	97.92	97.66	1.42 ms	2.29 ms
Without Nyström	91.72	92.05	91.68	91.93	6.36 ms	7.71 ms
Frequency domain only	90.21	92.11	92.41	92.16	4.79 ms	5.34 ms
Time domain only	86.02	87.77	88.32	89.18	8.81 ms	9.05 ms
Without an adaptive window	93.82	92.19	90.71	90.30	10.21 ms	14.16 ms

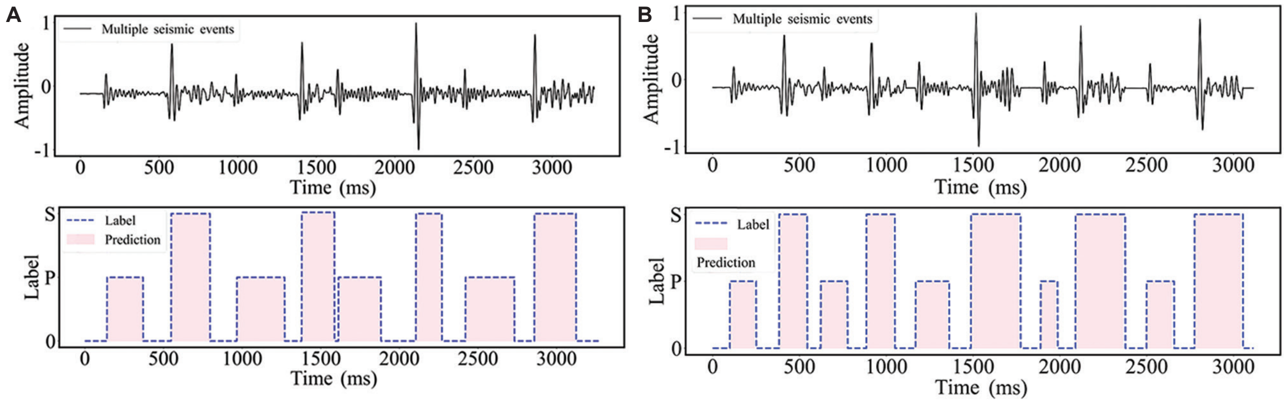

Figure 9. Classification results of the SeisFormer model in scenarios with multiple events occurring within a short time window. Each waveform segment has a duration of 3,000–4,000 ms.

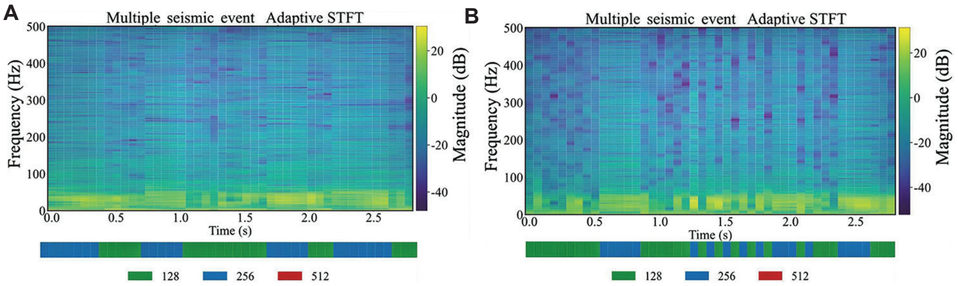

Figure 10. Adaptive short-time Fourier transform time–frequency spectrogram with frame-level window-selection indicator bars (corresponding to Figure 9)

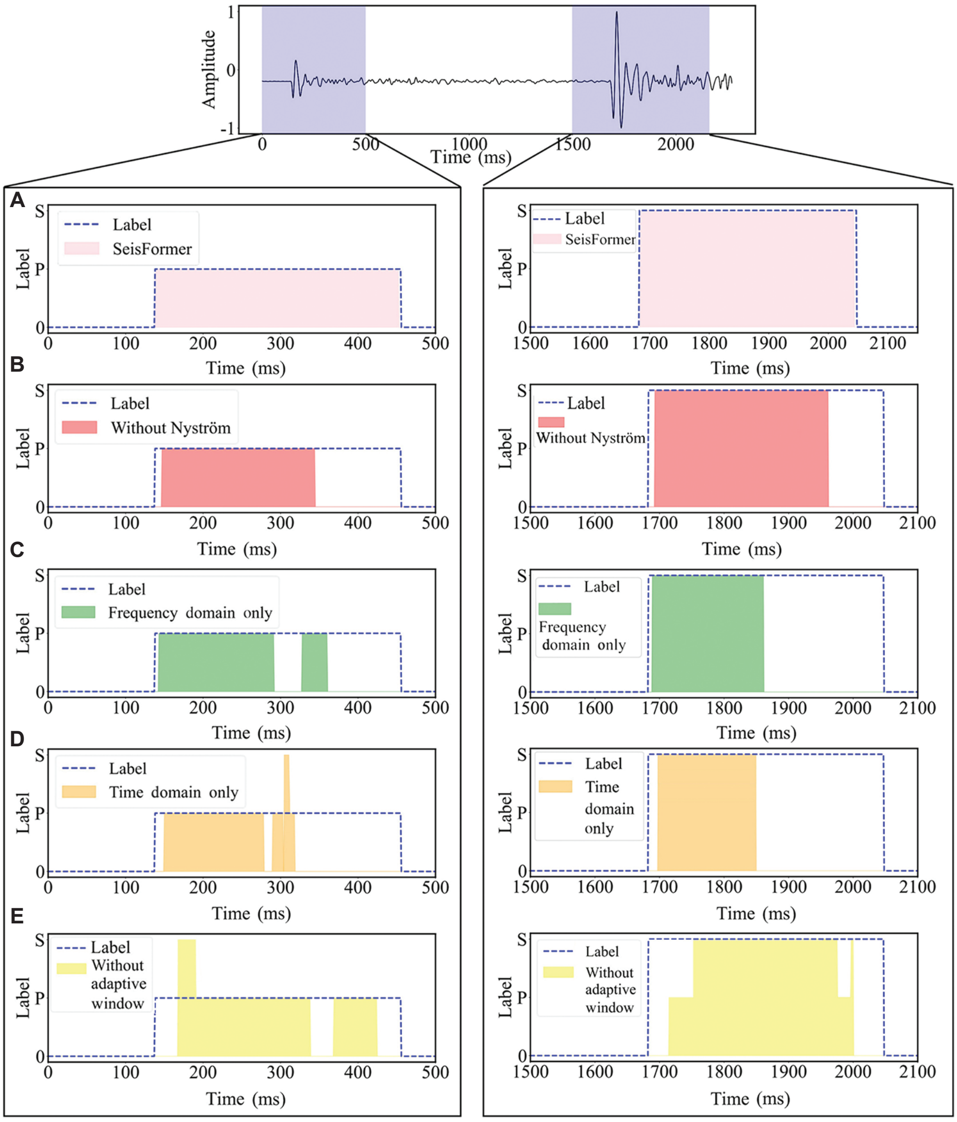

Figure 11. Comparison of classification and picking results across ablation variants. (A) SeisFormer; (B) without Nyström; (C) frequency domain only; (D) time domain only; and (E) without adaptive window.

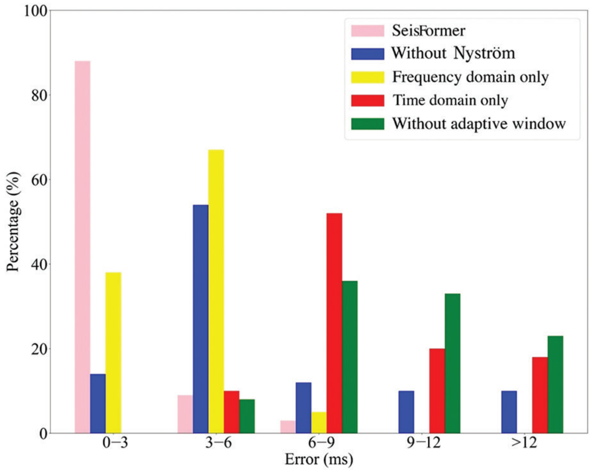

Figure 12. Distribution of average P-wave arrival time errors across different model configurations

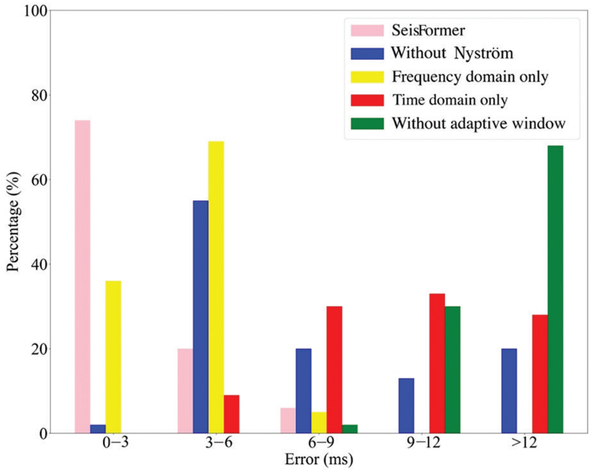

Figure 13. Distribution of average S-wave arrival time errors across different model configurations

(in the full model) constrained small-singular-value directions, suppressed noise amplification, and stabilized the weight distribution. Restricting the representation to a single domain markedly weakened modeling capacity: The frequency domain only variant, while closer to the full model, attained only 90.21%/92.16% (accuracy/F1) with MAE 4.79/5.34 ms; the time domain only variant further degraded to 86.02%/89.18% with MAE 8.81/9.05 ms, underscoring that a single domain cannot simultaneously capture transient onsets and band-limited structure—joint time–frequency modeling is critical for robust picking and classification. Removing the adaptive windowing mechanism (without adaptive windowing) still yielded 93.82% accuracy, but F1 dropped to 90.30% and arrival errors increased to 10.21/14.16 ms, demonstrating that frame-wise window selection, in combination with STFT/convolutional parameters, is essential for mitigating low-frequency drift and harmonic interference while preserving onset alignment.

As shown in Figure 11, the ablated variants exhibited more diffuse probability responses, window spillover, and larger onset drifts relative to the baseline. Figures 12 and 13 further corroborate this trend quantitatively: Per-trace error curves showed systematic increases in P/S arrival errors whenever a component was removed, with the largest growth observed without an adaptive window. Taken together, Nyström with a Tikhonov-regularized pseudoinverse + joint time–frequency representation + adaptive windowing acted synergistically: Adaptive windowing yielded the most significant gains in arrival-time precision, the regularized pseudoinverse secured numerical/training stability, and time–frequency complementarity set the upper bound and robustness of both classification and picking.

4. Conclusion

We proposed SeisFormer, a P/S-wave classification and first-arrival picking network for HDR hydraulic fracturing. SeisFormer combines adaptive multi-scale windowing with joint time–frequency modeling and introduces a stabilized Nyström attention module to enhance long-range dependency modeling and feature discriminability. Evaluated on a joint multi-site dataset constructed from HDR stimulation in the Qinghai Gonghe Basin and unconventional hydraulic fracturing in North China, SeisFormer achieved state-of-the-art performance on real data, noise-augmented data with non-stationary interference, and dense multi-event windows, demonstrating robustness across operating conditions and strong generalization. In field settings, the classification accuracy reached 98.30%, with mean arrival-time errors of 1.42 ms (P) and 2.29 ms (S). Under low SNR and complex signal environments, the model maintained stable classification and picking accuracy. Ablation studies further confirmed the significant contributions of the key components to overall performance.

Based on measured results on a NVIDIA GeForce RTX 4060 (8 GB) + Intel Core i9-13900HX platform—where the method attained P/S arrival-time errors ”.2.5 ms—future work can refine the unified pre-processing and end-to-end inference pipeline and conduct systematic robustness and fault-tolerance evaluations under complex, non-stationary noise, and dynamic operating conditions. In parallel, SeisFormer can be migrated to edge-computing modules and portable platforms to support near-real-time field monitoring and facilitate engineering deployment.

Acknowledgments

None.

Funding

This work was supported by the National Key R&D Program of China (Grant Nos. 2018YFB1501803 and 2019YFC1804805-4) and the China Geological Survey Project (Grant No. DD2019135).

Conflict of interest

The authors declare they have no competing interests.

Author contributions

Conceptualization: Mingjun Ouyang, Feng Sun
Formal analysis: Mingjun Ouyang
Investigation: Zenan Leng, Haotian Hu, Zubin Chen, Fa Zhao
Methodology: Mingjun Ouyang
Validation: Mingjun Ouyang
Visualization: Mingjun Ouyang
Writing–original draft: Mingjun Ouyang
Writing–review & editing: Mingjun Ouyang, Feng Sun

Availability of data

The microseismic dataset used in this study, obtained from hydraulic fracturing operations in HDR formations in the Qaidam Basin, Qinghai, China, is subject to a non-disclosure agreement and cannot be made publicly available.

References

1. Dong S, Jiao J, Zhou S, Lu P, Zeng Z. 3-D gravity data inversion based on enhanced dual U-Net framework. *IEEE Trans Geosci Remote Sens*. 2023;61:1-11.

   doi: 10.1109/TGRS.2023.3306980

2. Cloetingh S, Sternai P, Koptev A, *et al*. Coupled surface to deep Earth processes: Perspectives from TOPO-EUROPE with an emphasis on climate- and energy-related societal challenges. *Global Planet Change*. 2023;226:104140.

   doi: 10.1016/j.gloplacha.2023.104140

3. Ma W, Wang Y, Wu X, Liu G. Hot dry rock (HDR) hydraulic fracturing propagation and impact factors assessment via sensitivity indicator. *Renew Energy*. 2020;146:2716-2723.

   doi: 10.1016/j.renene.2019.08.097

4. Xie J, Cao H, Wang D, Peng S, Fu G, Zhu Z. A comparative study on the hydraulic fracture propagation behaviors in hot dry rock and shale formation with different structural discontinuities. *J Energy Eng*. 2022;148(6):04022040.

   doi: 10.1061/(asce)ey.1943-7897.0000856

5. Cheng Y, Zhang Y, Yu Z, Hu Z, Yang Y. An investigation on hydraulic fracturing characteristics in granite geothermal reservoir. *Eng Fract Mech*. 2020;237:107252.

   doi: 10.1016/j.engfracmech.2020.107252

6. Cheng Y, Zhang Y, Yu Z, Hu Z. Investigation on reservoir stimulation characteristics in hot dry rock geothermal formations of China during hydraulic fracturing. *Rock Mech Rock Eng*. 2021;54(8):3817-3845.

   doi: 10.1007/s00603-021-02506-y

7. Cheng Y, Zhang Y. Experimental study of fracture propagation: The application in energy mining. *Energies*. 2020;13(6):1411.

   doi: 10.3390/en13061411

8. Wang H, Alkhalifah T, Waheed UB, Birnie C. Data-driven microseismic event localization: An application to the Oklahoma Arkoma Basin hydraulic fracturing data. *IEEE Trans Geosci Remote Sens*. 2022;60:1-12.

   doi: 10.1109/TGRS.2021.3120546

9. Warpinski N. Microseismic monitoring: Inside and out. *J Petrol Technol*. 2009;61(11):80-85.

   doi: 10.2118/118537-JPT

10. Allen RV. Automatic earthquake recognition and timing from single traces. *Bull Seismol Soc Am*. 1978;68(5):1521-1532.

    doi: 10.1785/bssa0680051521

11. Akaike H. A new look at the statistical model identification. *IEEE Trans Autom Control*. 1974;19(6):716-723.

    doi: 10.1109/tac.1974.1100705

12. Allen R. Automatic phase pickers: Their present use and future prospects. *Bull Seismol Soc Am*. 1982;72(6B):S225-S242.

    doi: 10.1785/BSSA07206B0225

13. Álvarez I, Garcia L, Mota S, *et al*. An automatic P-phase picking algorithm based on adaptive multiband processing. *IEEE Geosci Remote Sens Lett*. 2013;10(6):1488-1492.

    doi: 10.1109/lgrs.2013.2260720

14. Earle PS, Shearer PM. Characterization of global seismograms using an automatic-picking algorithm. *Bull Seismol Soc Am*. 1994;84(2):366-376.

    doi: 10.1785/bssa0840020366

15. Akazawa TA. Technique for Automatic Detection of Onset Time of P- and S-Phases in Strong Motion Records. In: *Proceedings 13th Conference on Earthquake Engineering*. Vancouver, Canada. Paper No. 786; 2004.

16. Kurz JH, Grosse CU, Reinhardt HW. Strategies for reliable automatic onset time picking of acoustic emissions and of ultrasound signals in concrete. *Ultrasonics*. 2005;43(7):538-546.

    doi: 10.1016/j.ultras.2004.12.005

17. Anikiev D, Birnie C, Waheed UB, *et al*. Machine learning in microseismic monitoring. *Earth Sci Rev*. 2023;239:104371.

doi: 10.1016/j.earscirev.2023.104371

18. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521:436-444.

doi: 10.1038/nature14539

19. Mousavi SM, Beroza GC. Deep-learning seismology. *Science*. 2022;377:725-729.

doi: 10.1126/science.abm4470

20. Zhu W, Beroza GC. PhaseNet: A deep-neural-network-based seismic arrival-time picking method. *Geophys J Int*. 2019;216(1):261-273.

doi: 10.1093/gji/ggy423

21. He Z, Peng P, Wang L, Jiang Y. PickCapsNet: Capsule network for automatic P-wave arrival picking. *IEEE Geosci Remote Sens Lett*. 2021;18(4):617-621.

doi: 10.1109/lgrs.2020.2983196

22. Zhao, Y, Xu, H, Yang, T, Wang D, Sun D. A hybrid recognition model of microseismic signals for underground mining based on CNN and LSTM networks. *Geomat Nat Hazard Risk*. 2021;12(1):2803-2834.

doi: 10.1080/19475705.2021.1968043

23. 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;11:3952.

doi: 10.1038/s41467-020-17591-w

24. Saad OM, Chen Y, Siervo D, *et al*. EQCCT: A production-ready earthquake detection and phase-picking method using the compact convolutional transformer. *IEEE Trans Geosci Remote Sens*. 2023;61:1-15.

doi: 10.1109/TGRS.2023.3319440

25. Li S, Yang X, Cao A, *et al*. SeisT: A foundational deep-learning model for earthquake monitoring tasks. *IEEE Trans Geosci Remote Sens*. 2024;62:1-15.

doi: 10.1109/tgrs.2024.3371503

26. Li XN, Chen FJ, Lai YP, Tang P, Liang X. ICAT-net: A lightweight neural network with optimized coordinate attention and transformer mechanisms for earthquake detection and phase picking. *J Supercomput*. 2025;81:191.

doi: 10.1007/s11227-024-06664-y

27. Peng P, Lei R, Wang JM. Enhancing microseismic signal classification in metal mines using transformer-based deep learning. *Sustainability*. 2023;15(20):14959.

doi: 10.3390/su152014959

28. Zhang X, Wang X, Zhang Z, Wang Z. CNN-transformer for microseismic signal classification. *Electronics*. 2023;12(11):2468.

doi: 10.3390/electronics12112468

29. Zhu L, Chuang L, McClellan JH, Liu E, Peng Z. *A Multi-Channel Approach for Automatic Microseismic Event Association using RANSAC-Based Arrival Time Event Clustering (RATEC)*. [arXiv Preprint]; 2017.

doi: 10.48550/arXiv.1702.01856

30. Li Z, Gou X, Jin W, Qin N. Frequency features of microseismic signals. *Chin J Geotech Eng*. 2008;30(6):830-834.

31. Aki K, Richards PG. *Quantitative Seismology*. 2nd ed. Sausalito, CA: University Science Books; 2002.

32. Ma C, Ran X, Xu W, *et al*. Fine classification method for massive microseismic signals based on short-time Fourier transform and deep learning. *Remote Sens*. 2023;15(2):502.

doi: 10.3390/rs15020502

33. Xiong Y, Zeng Z, Chakraborty R, *et al*. Nyströmformer: A Nyström-based algorithm for approximating self-attention. *Proc AAAI Conf Artif Intell*. 2021;35(16):14138-14148.

doi: 10.48550/arXiv.2102.03902

34. Piao X, Chen Z, Murayama T, *et al*. Fredformer: Frequency Debiased Transformer for Time Series Forecasting. In: *Proceedings 30th ACM SIGKDD Conference Knowledge Discovery and Data Mining (KDD 2024)*. Long Beach, CA, USA; 2024. p. 1234-1245.

doi: 10.1145/3637528.3671928

35. Soto-Quiros P. A fast method to estimate the Moore-Penrose inverse for well-determined numerical rank matrices based on Tikhonov regularization. *J Math Comput Sci*. 2024;37(1):59-81.

doi: 10.22436/jmcs.037.01.05

36. Longhas PRA, Abdul AM. Application of Tikhonov regularization in generalized inverse of adjacency matrix of undirected graph. *Int J Math Trends Technol*. 2022;68(2):1-6.

doi: 10.14445/22315373/ijmtt-v68i2p501

37. Gulliksson ME, Wedin PÅ, Wei Y. Perturbation identities for regularized Tikhonov inverses and weighted pseudoinverses. *BIT Numer Math*. 2000;40:513-523.

doi: 10.1023/A:1022319830134

38. Ross ZE, Meier MA, Hauksson E, Heaton TH. Generalized seismic phase detection with deep learning. *Bull Seismol Soc Am*. 2018;108(5A):2894-2901.

doi: 10.1785/0120180080