Infra-Net: A Robust Parallel Decision-Making Network for Discriminating Natural Hazards and Anthropogenic Infrasound Events via Multi-View Feature Learning

Hongru Li, Xihai Li, Jihao Liu, Shengjie Luo, Yun Zhang
Rocket Force University of Engineering, Xi'an, 710025, Shaanxi, China
Correspondence to: Xihai Li (xihai_li@163.com)

Abstract. The accurate classification of infrasound signals is a cornerstone of global geophysical monitoring, essential for both natural hazard early warning systems (e.g., volcanic eruptions, debris flows, and earthquakes) and the verification of the Comprehensive Nuclear-Test-Ban Treaty. However, the development of reliable automated systems is hindered by the inherent scarcity of representative event data, particularly for rare extreme events, as well as the presence of complex, non-stationary background interferences. To maximize the diagnostic value of limited geophysical datasets, this paper proposes Infra-Net, a novel parallel decision-making network driven by multi-view feature learning and a confidence-based fusion mechanism. We introduce a logarithmic wavelet scattering transform to produce robust, mathematically grounded feature representations. Unlike conventional methods that process scattering matrices holistically, our approach treats individual columns as independent feature vectors, providing multiple localized perspectives of the same acoustic event. Architecturally, Infra-Net utilizes a dual-branch structure to simultaneously capture multi-scale spatial features and discriminative temporal patterns. These parallel evaluations are synthesized through a custom confidence-based fusion module, which employs weighted averaging and an inner-product mechanism to ensure a stable and comprehensive final classification. Tested on both public infrasound datasets and real-world CTBTO-measured data, Infra-Net achieved accuracies of 100% and 82.07%, respectively. These results demonstrate that Infra-Net offers a highly robust soft computing solution for enhancing Earth system monitoring and the reliable identification of natural hazards amidst anthropogenic noise.

1 Introduction

Natural hazards such as earthquakes, volcanic eruptions, and debris flows pose significant threats to human society and ecosystems. To improve the timeliness and accuracy of early warning systems, it is essential to conduct multi-dimensional, real-time monitoring of these geodynamic processes. Among various monitoring technologies, infrasound sensing has gained prominence due to its unique physical properties. Infrasound refers to acoustic waves with frequencies below the lower threshold of human hearing, typically defined as being under 20 Hz (Bryan et al. 2018). These waves are characterized by low atmospheric absorption and the ability to propagate over vast distances, making them ideal for detecting remote geophysical events. Such waves are widely generated by natural phenomena, including ground vibrations from earthquakes (Alegria et al. 2015; Li et al. 2016), pressure releases from volcanic eruptions (Liu et al. 2014; Toney et al. 2022), and the geomorphic mass movements associated with debris flows (Kogelnig et al. 2014; Marchetti et al. 2019). Consequently, infrasound signals serve as powerful tools for sensing distant disasters and as vital carriers of information regarding energy exchange within the Earth system. With the collaborative development of sensor networks and signal processing techniques, the use of automated algorithms to classify and recognize infrasound signals has become central to enhancing hazard monitoring effectiveness. However, the accurate identification of these signals in practical applications faces significant obstacles. First, while human activities such as rocket launches (Dai et al. 2021; Wen et al. 2019) and chemical explosions (Koch and Pilger 2018; Park et al. 2018) generate similar acoustic signatures, disaster

mitigation tasks require the precise discrimination of natural hazards from these anthropogenic sources to avoid false alarms. Second, because major natural hazards, particularly extreme events, are sporadic and rare, it is often difficult for researchers to obtain sufficient field observation samples. This "data-scarce" dilemma severely constrains the generalization capabilities of traditional deep learning models, leading to suboptimal performance when encountering complex environmental interferences.

In recent years, the academic community has proposed several schemes for infrasound recognition. For instance, Thüring et al. (2015) utilized Fourier spectra and support vector machines (SVM) to successfully reduce avalanche detection error rates from 65% to 10%. Albert and Linville (2020) achieved an average classification accuracy of only 74% for volcanic activity and seismic events, and a notably lower overall recognition accuracy of 56% for earthquakes, chemical explosions, mine explosions, and volcanic activity using CNNs. Leng et al. (2022) compiled an infrasound dataset including debris flows and lightning, achieving an accuracy of 84.1% using an enhanced LeNet-5 architecture. Pásztor et al. (2023) extracted time-domain, frequency-domain, and other features from infrasound signals generated by quarry blasts and industrial sources. By training SVM and random forest classifiers on these features, they attained F1 scores of 0.88 and 0.93. While these studies have advanced the field, limitations remain. Many models, such as the improved AlexNet proposed by Yuan et al. (2024), rely heavily on data augmentation techniques like slicing and translation to expand samples, which may not accurately reflect the underlying physical distribution of rare natural events. Furthermore, hybrid methods like the LSTM-prototypical network introduced by Zhao et al. (2024) require extensive computational iterations, which may limit their utility in time-sensitive monitoring scenarios where efficiency is paramount.

In summary, the domain of infrasound signal classification is still characterized by four primary bottlenecks. First, deep learning models are often constrained by the scarcity of event data, where traditional augmentation fails to compensate for missing physical information. Second, the opaque nature of many neural networks complicates architecture tuning and necessitates more interpretable feature extraction methods. Third, single-stream architectures often fail to capture the multifaceted spatio-temporal dynamics inherent in non-stationary geophysical signals. Finally, systems lacking redundancy and consensus mechanisms are vulnerable to individual misclassifications, which can compromise the stability of the overall monitoring regime.

To address these challenges, this study introduces Infra-Net, a parallel decision-making framework designed for robust event classification under data-scarce conditions. Instead of relying on conventional data augmentation, this work leverages multi-view feature learning and confidence-based fusion to maximize information gain from limited samples. The specific contributions of this work are as follows:

Multi-view Feature Extraction: A wavelet scattering network (WSN) (Khan et al. 2017) is employed to extract mathematically grounded features. By treating each column of the scattering coefficients as an independent feature vector, the model evaluates the same acoustic event from multiple localized perspectives, thereby enhancing the description of signal non-stationarity without the need for artificial sample expansion.

Feature Contrast Enhancement: A logarithmic transformation is applied to the scattering coefficients to amplify subtle variations between natural hazards and anthropogenic interference sources, forming a highly discriminative feature space.

Parallel Decision Architecture: The Infra-Net architecture comprises two specialized branches, one focused on multi-scale spatial features and the other on critical temporal patterns.

Confidence-Based Fusion: A novel decision-making module is introduced. By performing confidence-weighted averaging followed by a confidence-based inner product, the system dynamically integrates the parallel evaluations from both branches. This consensus-driven approach mitigates the impact of noise and enhances the overall robustness of the classification system.

Extensive evaluations demonstrate that this model effectively addresses the challenges of classifying small-sample infrasound events. This advancement strengthens the capability to characterize critical geophysical phenomena, supporting both natural hazard early warning systems and the verification mission of the CTBT (Lee and Hong 2024).

The remainder of this paper is organized as follows. In the "Materials and Methods" section, we introduce the datasets used and the extraction process of logarithmic wavelet scattering features. The "Model Construction" section presents the proposed method and elaborates on its innovative aspects. The "Results" section details the experimental procedures and offers a comprehensive analysis of the outcomes. In the "Discussion" section, we compare the advantages of our proposed method with those of other existing methods and cast a vision for subsequent endeavors. The "Conclusion" section summarizes the principal findings of this study.

2 Materials and Methods

2.1 Data sources and division

This study utilizes two infrasound datasets containing various interference events: one from an open-source dataset and the other collected through on-site measurements by the CTBTO. The publicly available dataset originates from the open-source library of typical infrasonic signals (LOTIS) (Bryan et al. 2018; Zhao et al. 2024). The data were primarily collected from the infrasound sensor array located in Windless Bight, Antarctica, as shown in Fig. 1.

The empirical foundation of this study consists of two distinct infrasound datasets that capture a wide array of geophysical phenomena and natural hazards. The first dataset is derived from the Library of Typical Infrasonic Signals (LOTIS), an open-source repository designed to facilitate the study of non-stationary atmospheric acoustic waves (Bryan et al. 2018; Zhao et al. 2024). These data were primarily recorded by the infrasound sensor array located at Windless Bight, Antarctica, as illustrated in Fig. 1. This high-latitude monitoring site provides a unique environment for capturing a diverse range of Earth system events, including atmospheric gravity waves (AGW) triggered by auroras, microbaroms (MB), mountain associated waves (MAW), and infrasound generated by volcanic eruptions (VE). By including these natural hazard precursors alongside complex background noise, the dataset provides a rigorous baseline for testing the discriminative power of the proposed model. The dataset division follows the approach described in the reference, as detailed in Table 1.

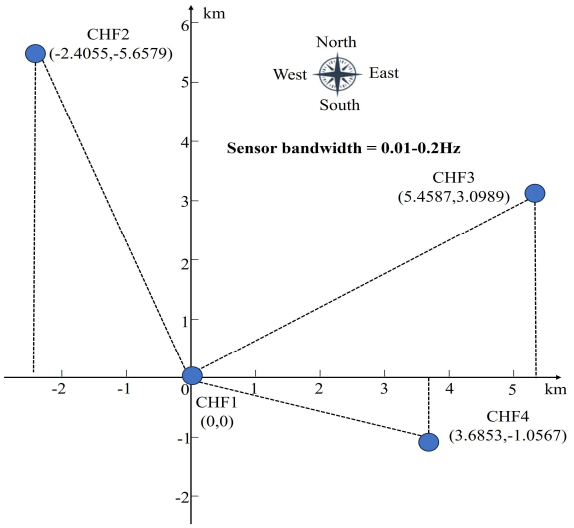


Figure 1: Schematic diagram of the infrasound sensor array located in Windless Bight.

Table 1 The training set and test set division of LOTIS (Li et al. 2026).

	AGW	MAW	MB	VE

training set	81	77	41	46
Test set	34	33	17	19
Total	115	110	58	65

The other infrasound dataset used in this study consists of field-measured data obtained from the CTBTO. Due to high frequency of occurrence and the close resemblance of waveform characteristics and time-frequency distributions to those of nuclear explosions, natural earthquakes and chemical explosions represent significant sources of interference (Wang et al. 2022; Chang and Cui 2013). Therefore, this dataset specifically emphasizes infrasound events generated by natural earthquakes and chemical explosions. Table 2 lists the number of events and waveforms for each category of infrasound in the catalog. Owing to the simultaneous capture of each infrasound event by both multiple sensors within individual station arrays and multiple stations across the monitoring network, a one-to-many relationship is established. Consequently, each event generates multiple waveforms.

Table 2 Infrasound events and waveform numbers (Li et al. 2026).

Event	Number of events	Number of waveforms
Chemical explosions	28	386
Earthquakes	8	403

Partial event information is summarized in Table 3.

Table 3 Partial event information for chemical explosions and natural earthquakes (Li et al. 2026).

Event time	Event trigger time	Latitude	Longitude	Event
20020920	00:38:03.0	-31.0108	136.7756	Woomera Test Explosion
20050321	20:46:22.8	41.131	-112.896	UTTR Missile Detonation
20050323	19:20:00.0	29.374	-94.939	BP Refinery Explosion
20051113	05:41:00.0	43.85	126.57	Jilin Chemical Factory Explosion
20020104	19:38:08.0	32.4	-115.1	Western US Earthquake
20020319	22:14:49.9	30.1	-113.7	Gulf of California Earthquake
20041226	01:00:00.0	5	95	Sumatra M9 Earthquake
20010904	12:45:53.0	37.1	-104.7	Colorado Earthquake
20050410	11:15:00.0	-1.7	99.8	Mentawai M6.5 Earthquake

The amount of infrasound data used in this study is limited. To make more efficient use of the data, the infrasound data are divided into three sub-datasets, and it is ensured that infrasound data from the same event are included in only one subset, thereby preventing the risk of data leakage. Table 4 displays the number of samples and events in the three subsets.

Table 4 Waveform quantity of each subset.

	Number of chemical explosion waveforms	Contains partial chemical explosion events	Number of earthquake waveforms	Contains partial earthquake events
Subset 1	123	Woomera test explosion BP refinery explosion	121	Colorado western US
Subset 2	144	UTTR missile detonation Jilin chemical factory explosion	154	Gulf of California Sumatra
Subset 3	119	Albania arms depot explosions	128	Mentawai

The data division approach is designed to minimize the correlation between the test set and the training set as much as possible, ensuring the reliability of the classification outcomes (Tan et al. 2021). The division methodology is illustrated in Fig. 2, beginning with the division of data into three event-based subsets containing 244, 298, and 247 raw signals respectively. Each subset undergoes log-scattering feature extraction, generating six feature sets per original signal (when the number of scattering paths is six), resulting in a total of 4734 input groups. In the classification experiments, one feature subset was sequentially employed as the test set, while the other two were utilized for training. The final metric was derived by averaging the results from these three independent classification trials.

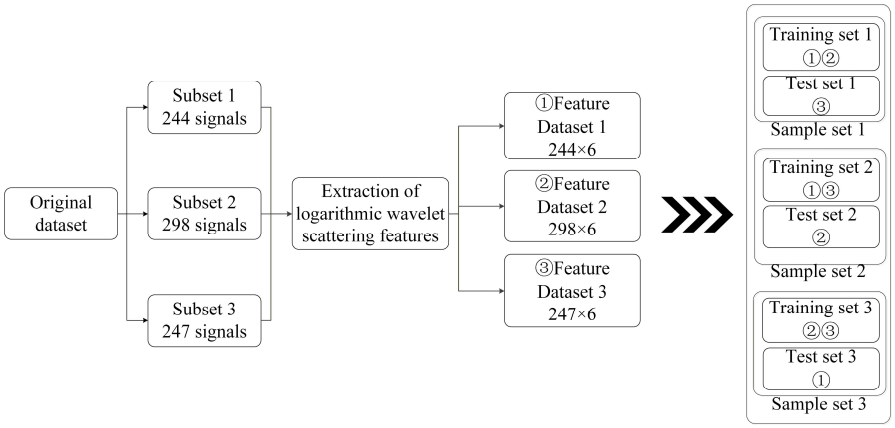


Figure 2: Schematic diagram of dataset division.

2.2 Extraction of logarithmic wavelet scattering features

2.2.1 Wavelet scattering network

Wavelet transform (Lilly and Olhede 2010), as a method for analyzing non-stationary signals, is characterized by multi-resolution and scale variability. However, it lacks translation invariance (Li et al. 2019), which can lead to the loss of certain signal details during analysis. Specifically, when decomposing low-frequency signals, high-frequency components may be overlooked, resulting in the loss of detail in the high-frequency signal [35] (Fan et al. 2022). In contrast, the wavelet scattering transform (WST) computes a semi-discrete wavelet transform of the signal [36] (Souli et al. 2021) and then applies a nonlinear modulus operation. The resulting feature representation possesses desirable properties such as translation invariance and deformation stability, perfectly meeting the basic requirements for feature extractors in machine learning [37–41] (Wiatowski and Bolcskei 2018; Wang et al. 2018a; Wang et al. 2018b; Huang et al. 2018; Liu et al. 2019).

As a geometrically invariant feature extractor [42–44] (Tuan 2023; Buriro et al. 2021; Andén and Mallat 2014), WST maintains robustness against translation, frequency shifts, and scale changes while preserving discriminative structures–crucial for non-stationary infrasound signals affected by atmospheric turbulence. The structure of the WSN is similar to that of a deep CNN and is primarily comprised of three components: signal input, feature extraction, and feature output. The feature extraction component is composed of multiple modules, each of which consists of wavelet convolution, nonlinearity, and averaging. The network's architecture is depicted in Fig. 3. Unlike CNNs, the filters in the WSN are predefined wavelet filters, whose parameters do not require learning through training samples (Ding et al. 2024), eliminating parameter training requirements and preventing overfitting in data-limited scenarios. In scenarios with small samples, WSNs typically achieve lower classification error rates compared to deep CNNs, thus offering certain advantages in the study of the classification for small-sample infrasound events.

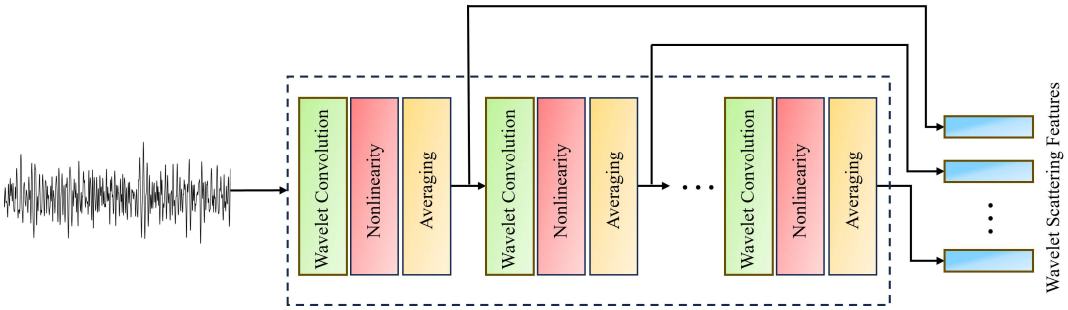


Figure 3: Schematic diagram of the wavelet scattering network.

The WST achieves local translation-invariant descriptors of infrasound signals through time-domain averaging

$$S_0F(t) = F * \phi(t), \tag{1}$$

Where $S_0F(t)$ is the scattering coefficient of $F(t)$. This descriptor eliminates all of the high-frequency information from the infrasound signal, thus extracting its low-frequency content while maintaining translation invariance. The high-frequency information is obtained by utilizing the modulus operation of the wavelet transform.

$$U_kF(t) = \left|F * \psi_k(t)\right|, \tag{2}$$

where $U_kF(t)$ is the scalogram coefficient of $F(t)$, and ψ is the high-frequency wavelet function, and ϕ is the low-pass filter. Equation. (2) represents the high-frequency information at scale k and acquires deformation stability through the modulus operation of the nonlinear wavelet transform (Bruna and Mallat 2011; Bruna and Mallat 2013).

When Eq. (2) is used as the input for the first-order scattering transform, the following equations are obtained (Tan et al. 2024; Liu et al. 2025; Shirodkar et al. 2025):

$$S_1F(t,k_1) = \left|F * \psi_{k_1}(t)\right| * \phi(t), \tag{3}$$

$$U_2F(t) = \left|\left|F * \psi_{k_1}(t)\right| * \psi_{k_2}(t)\right|. \tag{4}$$

By iteratively applying the aforementioned process, scattering coefficients of any order can be obtained. For any $k \geq 1$, the iteration of the wavelet modulus convolution is

$$U_mF(t,k_1,\cdots,k_m) = \left|\left|\left|F * \psi_{k_1}(t)\right| * \cdots\right| * \psi_{k_m}\right|. \tag{5}$$

The m-order scattering coefficient is

$$S_mF(t,k_1,\cdots,k_m) = \left|\left|\left|F * \psi_{k_1}(t)\right| * \cdots\right| * \psi_{k_m}\right| * \phi(t). \tag{6}$$

The final feature vector is composed of the scattering coefficients of each order.

2.2.2 Extraction of logarithmic wavelet scattering features

The spectral distribution of a signal determines the numerical distribution of its wavelet scattering features. If a signal contains abundant high-frequency components, the scattering features generated in the high-frequency region of the wavelet transform may have larger numerical values. The focus of this research is on infrasound signals, which are characterized by relatively low frequencies. Analysis reveals that the frequency distribution of the two types of infrasound event signals under investigation typically ranges from a few tenths to several hertz. Consequently, the wavelet scattering features produced also exhibit relatively small magnitudes. The logarithmic function, being monotonic within its domain, preserves the relative relationships between data points. Moreover, due to the properties of the logarithmic function, it is more sensitive to differences in smaller values compared to differences in larger values, which can further highlight the disparities between features. Therefore, logarithmic wavelet scattering features are utilized as the inputs of the classifier, which enhances the local characteristics of the signal and makes subtle variations more pronounced, thereby improving the expressive power of the features. As shown in Fig. 4, which depicts the wavelet scattering feature extraction framework, the extraction of logarithmic wavelet scattering features can be performed by utilizing the following steps. Following Eq. (1) to (6), third-order wavelet scattering features are extracted. Thereafter, the logarithm of the extracted wavelet scattering features is computed to obtain the final logarithmic wavelet scattering features. If higher-order logarithmic wavelet scattering coefficients are required, Eq. (5) and Eq. (6) can be repeated. However, in practical applications, as the number of layers increases, the energy of the infrasound signal will continuously decrease (Lu et al. 2023), leading to a significant decrease in the discriminability of the features obtained. Therefore, the number of scattering layers is set to three (Zhang et al. 2024).

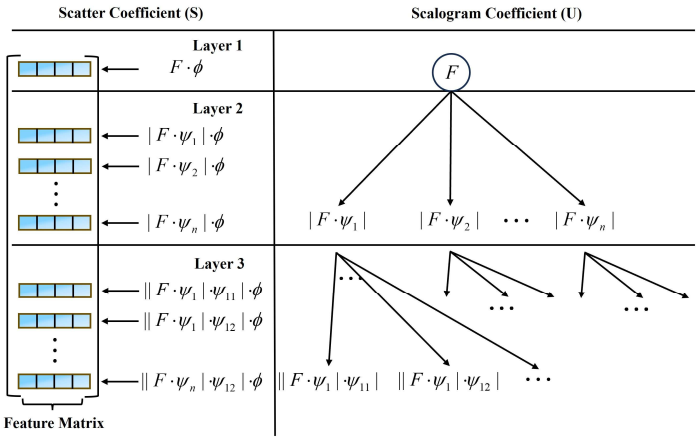


Figure 4: Schematic diagram of the wavelet scattering feature extraction framework.

The input infrasound signal has a sampling frequency of 20 Hz and a time-invariant scale size of 150 seconds when an analytic Morlet wavelet is employed (Stéphane 2009). The number of wavelet filters per octave is set to [8 2 1] (Al and Khushaba 2023). Taking a measured infrasound signal as an example, which has a data size of 1 × 3000, is input into the wavelet scattering network to extract the wavelet scattering features, resulting in a data size of 307 × 6, meaning each wavelet scattering feature consists of 6 columns of wavelet scattering coefficients. Each column corresponds to specific wavelet scales and scattering paths.

Unlike conventional approaches that concatenate the six columns of scattering features into a single, high-dimensional input matrix (Lone and Aydin 2023; Priya et al. 2021; Singhal and Kumar 2024), this study processes each column of scattering coefficients as an independent feature vector fed into a parallel evaluation framework. This strategy is grounded in the theoretical foundation of the wavelet scattering transform: scattering coefficients at different levels capture rich, non-stationary information from multi-

scale and multi-resolution perspectives. Each column of scattering features possesses a distinct mathematical expression and physical significance (as shown in Eqs. (1) to (6)), representing specific, complementary attributes of the same acoustic event. By treating these columns as distinct inputs for parallel analysis, the proposed architecture comprehensively preserves the multi-level information of the original signal without the risk of feature redundancy or inter-channel interference that often occurs in flattened inputs. More importantly, this processing strategy enables the generation of multiple independent confidence representations for a single signal. This multi-view characterization lays the essential foundation for the subsequent confidence-based decision-making mechanism, significantly enhancing the system's robustness against complex background interferences. As illustrated in Fig. 5, the visualization of the first three layers of wavelet scattering features from two representative signal samples clearly demonstrates that features at different levels provide complementary and discriminative insights, which are crucial for distinguishing nuclear events from natural phenomena.

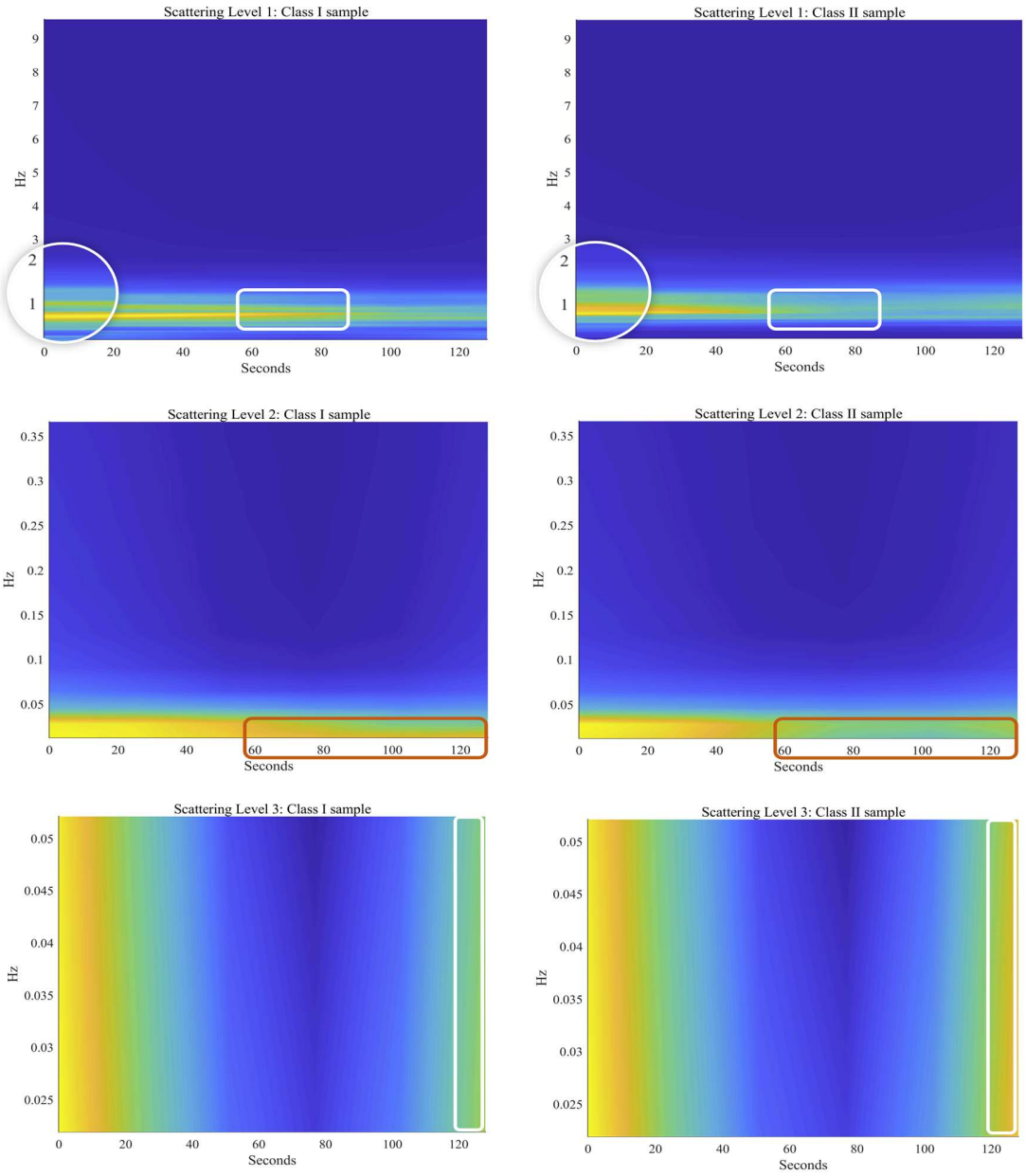


Figure 5: Visualization of the wavelet scattering features in layers 1–3 for a chemical explosion and seismic infrasound signals (The figures on the left represent a chemical explosion, while those on the right depict an earthquake).

As shown in Fig. 5, in the first layer, the characteristic frequency of chemical explosions is generally slightly lower than those of earthquakes, consistent with the more concentrated energy release typical of explosive sources. A clear difference in signal energy appears between 60 and 100 seconds, reflecting the distinct source mechanisms: chemical explosions typically produce impulsive, short-duration signals while earthquakes generate more extended waveforms from complex fault ruptures. Outside this time window, the signal energy distributions show substantial similarity due to common propagation effects.

In the second layer, the range of distinct signal energy distribution expands significantly. From 60 seconds onward, marked differences emerge in the spectral energy distribution, where explosion signals demonstrate energy concentrations tending toward lower frequencies compared to earthquakes. This pattern aligns with the fundamental physical distinction between the relatively simple explosive source mechanism and the more complex double-couple mechanism of earthquakes. However, overall discrimination remains challenging in the first two layers due to limited frequency resolution.

The third layer proves crucial as its enhanced frequency resolution reveals pronounced differences in signal energies after 120 seconds. This improved resolution effectively captures the more rapid amplitude decay characteristic of explosion signals compared to the sustained energy release of earthquake events, thereby facilitating the extraction of subtle discriminative features. The strategic utilization of scattering features from all three layers as parallel inputs enables a comprehensive analysis of these heterogeneous characteristics. Rather than increasing the sample count, this multi-vector approach enhances the granularity of event representation, which is particularly valuable in small-sample-size scenarios where every bit of discriminative information is critical. The layer-specific information content confirms that evaluating scattering features from different layers as distinct feature views effectively captures the full diversity of the signal's physical properties. These results provide a more comprehensive and reliable diagnostic framework for model training, which ultimately strengthens the precision and stability of automated systems dedicated to the real-time detection of natural hazards.

3 Model Construction

Infrasound signals exhibit significant characteristics including long-term dependence, multi-scale features, and non-stationarity. Leveraging these properties, a classification model named Infra-Net is designed, comprising two branch networks and a confidence-based decision-making module to enhance classification accuracy and robustness, as shown in Fig. 6. The Infra-Net consists of five components: input, Branch 1 network, Branch 2 network, confidence-based decision-making module, and output. The design objectives of the principal modules are detailed as follows:

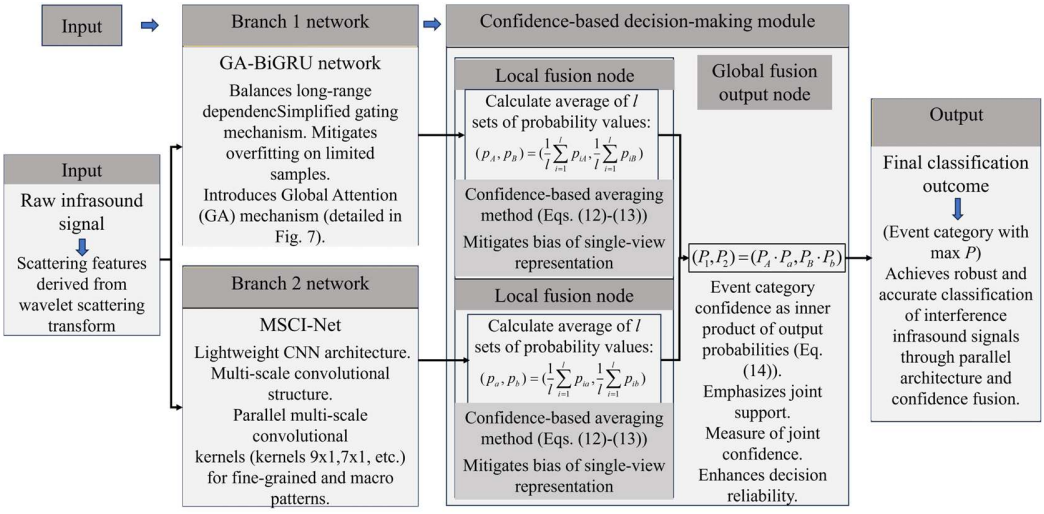


Figure 6: Schematic diagram of Infra-Net.

3.1 Branch 1: GA-BiGRU

For Branch 1, the scattering features derived from the wavelet scattering transform exhibit strict temporal locality preservation, necessitating a sequential modeling network that balances the capture of long-range dependencies with computational efficiency. Unlike the BiLSTM, the BiGRU utilizes a simplified gating mechanism, which significantly reduces parameter complexity while maintaining the ability to model long-term dependencies in sequential data (Cho et al. 2014; Hochreiter and Schmidhuber 1997). This design enhances parameter convergence efficiency on limited-sample infrasound datasets and effectively mitigates the risk of overfitting.

However, the traditional BiGRU architecture utilizes only the final hidden state as the sequential representation, while the intermediate hidden states generated during the process are not directly incorporated into the final output computation. This approach implicitly assumes the Markov property, where is expected to encode the entire historical information. According to the principle of entropy increase in information theory, as the sequence length grows, the information entropy of the hidden state exhibits an exponential decay trend, leading to the gradual loss of early temporal features. Consequently, the final hidden state overly focuses on the final features, affecting the model's ability to model the entire sequence. To address this issue, a Global Attention (GA) mechanism (Luong et al. 2015; Zhang et al. 2024) is introduced, which aims to mitigate information loss and enhance global interactive representation. This mechanism preserves information across both channel and spatial dimensions, thereby improving the performance of the neural network. The schematic diagram of the GA-BiGRU architecture is shown in Fig. 7.

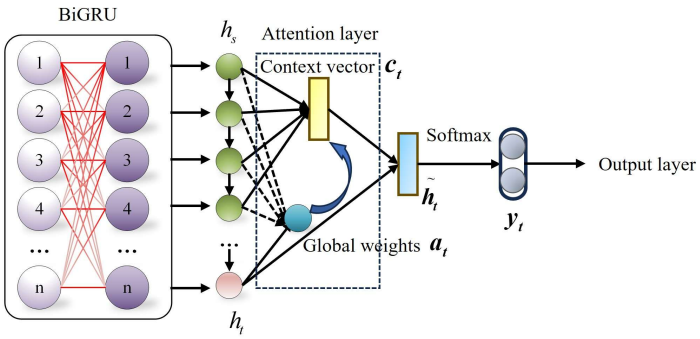


Figure 7: Schematic diagram of the GA-BiGRU.

The scalar $a_t(s)$ represents the weight of the hidden state $\boldsymbol{h}_s$ at time step s relative to all hidden states.

$$a_t(s) = \frac{\exp(\text{score}(h_t, h_s))}{\sum_{s'} \exp(\text{score}(h_t, h_{s'}))} \tag{7}$$

By combining the weights $a_t(s)$ of all t hidden states, the weight vector $\boldsymbol{a}_t$ is obtained. The term $\boldsymbol{c_t(s)}$ denotes the context vector at time step s, which is computed by multiplying the hidden state $\boldsymbol{h}_s$ with its corresponding weight $a_t(s)$. The global context vector $\boldsymbol{c}_t$ for the entire sample is computed by summing across all time steps.

$$\boldsymbol{c}_t = \sum_{s=1}^{T_x} a_t(s) h_s \tag{8}$$

where T_x is the number of sequence length. The target hidden state h_t and context vector $\boldsymbol{c}_t$ are concatenated to form an attention-enhanced vector $\tilde{\boldsymbol{h}}_t$.

$$\tilde{\boldsymbol{h}}_t = \tanh(\boldsymbol{W}_c[c_t; h_t]) \tag{9}$$

Finally, $\tilde{\boldsymbol{h}}_t$ is fed into a Softmax layer to produce the classification probability $\boldsymbol{y}_t$.

$$p(y_t \mid y < t, x) = \text{softmax}(\boldsymbol{W}_s \tilde{\boldsymbol{h}}_t) \tag{10}$$

Through the aforementioned method, GA learns the weights of each time step of the BiGRU, comprehensively considering all hidden states and assigning scores to emphasize important features while suppressing irrelevant ones. This method not only fully leverages the strengths of BiGRU in processing sequential data but also addresses the limitation of traditional methods in insufficiently capturing key features through the attention mechanism, thereby significantly enhancing the model's performance.

3.2 Branch 2: Multi-Scale Convolutional Infrasound Network (MSCI-Net)

The design of Branch 2 aims to leverage the distinct time-frequency feature distributions between the two types of infrasound events (Fig. 5). To this end, this branch incorporates a lightweight CNN architecture, centered around a multi-scale convolutional structure and a hierarchical feature extraction mechanism. Specifically, this branch utilizes parallel multi-scale convolutional kernels to capture different characteristics across varying receptive fields: smaller-scale kernels focus on capturing fine-grained fluctuations and local structures in the signal, while larger-scale kernels are used to extract energy distribution trends and macro patterns over extended time ranges. Through this multi-scale collaborative mechanism, the model can simultaneously respond to both local anomalies and global evolutionary patterns in infrasound signals. To further address the risk of overfitting in limited-data scenarios, this branch abandons traditional deep convolutional networks in favor of a shallower architecture with fewer parameters and a more streamlined structure. This design enhances its sensitivity and robustness to key spatial features (such as energy distribution and structural patterns). Through the aforementioned multi-scale and hierarchical processing, Branch 2 can effectively capture discriminative spatial structures across different time intervals, significantly improving the differentiation capability of infrasound events. The architecture is shown in Fig. 8.

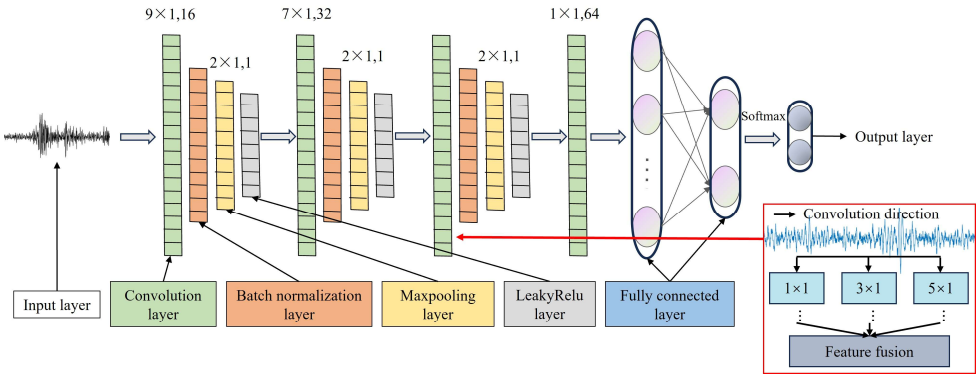


Figure 8: Schematic diagram of the MSCI-Net.

Four convolution layers are utilized to extract features, ensuring the lightweight nature of the MSCI-Net. The first two layers utilize kernels of sizes 9×1 and 7×1 to expand the receptive field, enabling the rapid capture of global features, which are essential for understanding the overall structure and trends of infrasound signals. In the third layer, multiple convolution layers are parallelized with varying kernel sizes at the same network level (Li et al.2024). This allows for the fusion of information filtered at different scales and effectively reduces information loss without increasing the network depth. Before the fully connected layer, a 1×1 convolutional kernel is applied. This is specifically designed for small-sample infrasound signal processing, aiming to reduce parameter counts and enhance nonlinearity, thereby improving the modeling capability for infrasound signals. The LeakyReLU

activation function is introduced after the convolution layers to replace the conventional ReLU. LeakyReLU retains a small gradient in the negative region, mitigating information loss and gradient vanishing issues caused by feature compression, as detailed in Eq. (11).

$$\sigma(x)=\begin{cases} x & \text{if } x \geq 0 \\ \gamma x & \text{if } x < 0 \end{cases} (\gamma = 0.01) \tag{11}$$

Additionally, a small stride is adopted, allowing the convolutional kernels to scan the signal step-by-step, thereby capturing both local details and global features. To enhance translational invariance and improve the MSCI-Net's robustness to minor temporal shifts, MaxPooling layers are incorporated. Furthermore, to address the non-stationarity of infrasound signals and their sensitivity to environmental noise, which complicates the data distribution, a batch normalization layer is employed to accelerate convergence, stabilize the training process, and suppress noise interference. Dropout is also introduced to randomly deactivate a portion of neurons during training, reducing MSCI-Net complexity and preventing overfitting. Ultimately, the MSCI-Net employs two fully connected layers that serve as a critical bridge from feature integration to classification decision-making. While convolutional layers excel at extracting local features, these features remain spatially or temporally fragmented. The first fully connected layer (with 64 neurons) performs global integration of these highly abstract yet scattered "local feature fragments." Through weighted summation and nonlinear activation functions, it learns complex high-order combinatorial relationships among them, thereby forming a compact global representation that incorporates all contextual information. Subsequently, the second fully connected layer maps this 64-dimensional global representation to the final sample label space. Each of neurons directly corresponds to an output category. Subsequently, the Softmax function generates the classification probability for each branch. This design, through multi-scale convolutional kernel configurations, a lightweight architecture, and anti-interference mechanisms, significantly enhances the extraction capability for complex infrasound signal features and the robustness of classification, while maintaining computational efficiency.

Through the strategic use of multi-scale convolutional kernels, a lightweight architecture, and robust anti-interference mechanisms, this branch significantly boosts the Infra-Net's ability to extract complex features from infrasound signals and enhances classification robustness.

3.3 Confidence-based decision-making module

Leveraging information fusion theory (Zhang et al. 2025; Zhang et al. 2024; Dai et al. 2024), Infra-Net employs its novel confidence-based decision-making module to integrate the results of the two branches, achieving enhanced classification robustness. As shown in Fig. 9:

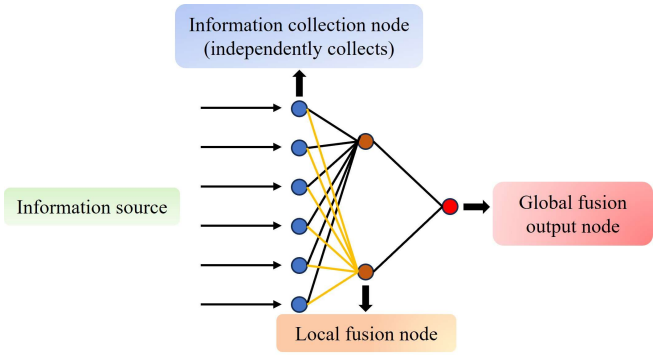


Figure 9: Schematic diagram of information fusion structure.

As shown in Fig. 9, the confidence-based decision-making module is conducted in the following stages. In the information collection stage, after feature extraction through the logarithmic wavelet scattering network, a single $1 \times N$ infrasound signal yields a $q \times l$ set of logarithmic wavelet scattering features. Each $q \times 1$ scattering feature is treated as a distinct feature view for parallel assessment. Subsequently, the logarithmic wavelet scattering features are input into both the MSCI-Net and GA-BiGRU networks, which independently produce l sets of classification probability values corresponding to the same signal, denoted as (p_{iA}, p_{iB}) and (p_{ia}, p_{ib}), respectively, where $(i = 1, 2, \cdots, l)$, q denotes the dimension of the scattering coefficients and l represents the number of scattering paths.

At the local fusion node, for each signal, the average of the l sets of probability values output by each network are calculated, which can be expressed as follows:

$$(p_A, p_B) = (\frac{1}{l}\sum_{i=1}^{l} p_{iA}, \frac{1}{l}\sum_{i=1}^{l} p_{iB}), \tag{12}$$

$$(p_a, p_b) = (\frac{1}{l}\sum_{i=1}^{l} p_{ia}, \frac{1}{l}\sum_{i=1}^{l} p_{ib}). \tag{13}$$

The l collection nodes per branch are merged into one using the confidence-based averaging method (Eqs. (12)-(13)), resulting in local integrated classification probabilities being obtained for each branch. This approach calculates the decision probability for a single signal across two branches by averaging the confidence of each column of scattering coefficients from the same signal, thereby mitigating the potential bias of a single-view representation through ensemble evaluation.

At the global fusion output node, to achieve a complementary advantage between the two branches, the second step of the confidence-based decision-making module is necessary. The final event category confidence is defined as the inner product of the output probabilities from both networks, which is denoted as

$$(P_1, P_2) = (P_A \cdot P_a, P_B \cdot P_b). \tag{14}$$

The ultimate classification outcome is determined by the event category corresponding to $Max(P_1, P_2)$. The inner product operation in Eq. (14) combines the confidence levels of two networks through multiplication, emphasizing the joint support of both branches for the same category. The magnitude of the inner product value directly reflects the strength of the joint support of the two branches for that category. Therefore, the inner product value can serve as a measure of joint confidence, which helps enhance the reliability of the final decision.

In summary, by integrating temporal, spatial, and logarithmic wavelet scattering features, Infra-Net constructs a more diversified feature space. Subsequently, through a parallel network architecture and a confidence fusion strategy, it effectively enhances the reliability of the results, achieving robust and accurate classification of interference infrasound signals in natural hazards monitoring.

4. Results

4.1 Experimental parameter settings

Before training the Infra-Net, it is necessary to customize the Infra-Net's parameters. The number of hidden layer units in the BiGRU, the learning rate for training, and the maximum number of iterations all impact the recognition results. The optimal combination of these three parameters is determined using the enumeration method and grid search optimization (Abbaszadeh et al. 2022). The BiGRU network includes two layers, which contain 512 and 256 hidden layer units. The Adam optimizer is

360 employed, and cross-entropy is utilized as the loss function. The initial learning rate is 0.001, it is reduced to 0.0001 after 400 training epochs, and the maximum number of epochs is set to 500, with a minimum of 400 training samples per epoch.

4.2 Experimental evaluation index

As shown in Table 5, the confusion matrix allows us to compute four statistical quantities: true positives, false positives, false negatives, and true negatives.

365 **Table 5** Confusion matrix (Li et al. 2026).

Actual/Forecast	Positive	Negative
Positive	True positive (*TP*)	False negative (*FN*)
Negative	False positive (*FP*)	True negative (*TN*)

Three indices are utilized to evaluate the classification performance of the Infra-Net: the accuracy (*ACC*), precision (*P*), and *F1* score. The recall (*R*), the rate of true positive identification, is also considered. The specific definitions of each index are as follows:

$$ACC = \frac{TP + TN}{TP + FP + TN + FN}, \tag{15}$$

370 $$P = \frac{TP}{TP + FP}, \tag{16}$$

$$R = \frac{TP}{TP + FN}, \tag{17}$$

$$F1 = \frac{2 \times R \times P}{R + P}. \tag{18}$$

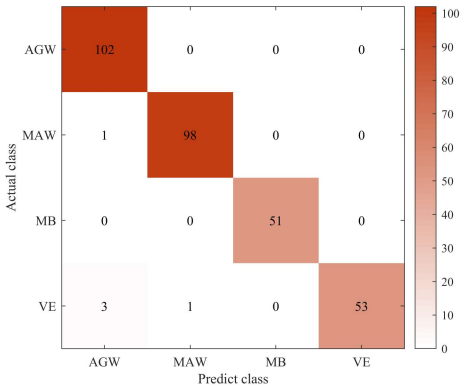
In addition, due to the highly pronounced uneven distribution of samples across various categories in the LOTIS dataset, Cohen's Kappa was further employed as a supplementary metric for evaluating model performance, to prevent the misleadingly optimistic
375 assessment produced by traditional metrics such as accuracy under imbalanced data conditions. Cohen's Kappa is an important statistical measure used to evaluate the consistency and reliability between the predictions of a classification model and the true labels. Unlike accuracy, which only focuses on the proportion of correctly classified samples, this coefficient incorporates a correction for "random agreement", eliminating the influence of random factors in the classification results. Thus, it provides a more robust and accurate measure of model performance. Especially in scenarios with highly skewed class distributions, Cohen's
380 Kappa better reflects the model's true discriminative capability beyond random guessing. Its calculation formula is as follows:

$$Cohen's\ Kappa = \frac{p_o - p_e}{1 - p_e}, \tag{19}$$

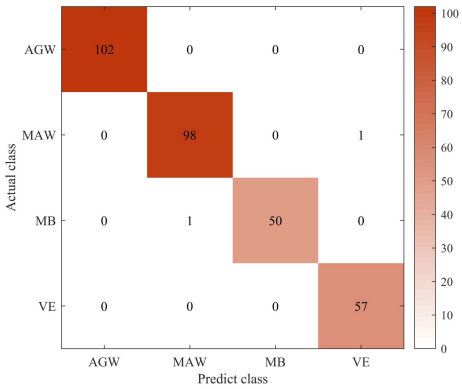
where, p_o is the observed proportional agreement, representing the actual proportion of agreement between evaluators across all samples. p_e is the expected random agreement proportion, calculated assuming evaluators classify independently and randomly.

4.3 Experiment and result analysis

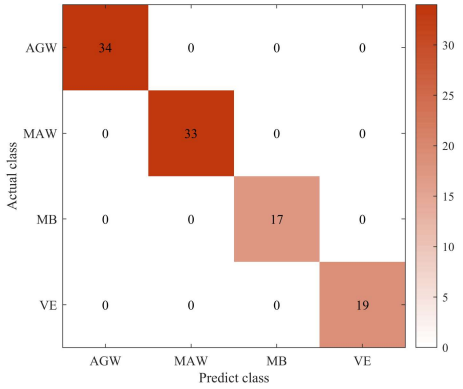
First, experiments were conducted using the LOTIS dataset. Figure 10 shows classification results for Branch 1 (a), Branch 2 (b), and Infra-Net (c) using logarithmic wavelet scattering features. The limited data length only permits a threefold scattering feature path expansion. Experiments (a) and (b) generate three independent classifications per signal without decision fusion, yielding input counts of original signals × 3. Conversely, (c)'s confidence-based decision-making module fuses three results into single outputs, maintaining original signal counts. Thus, (a) and (b) contain threefold more features than (c). It can be seen that the classification accuracy reached above 98% when using only one of the branches designed in this study, which is comparable to or even better than the classification results reported by Bryan et al. (2018) and Zhao et al. (2024) on the same dataset, both in terms of accuracy and time efficiency. Notably, compared with Zhao et al., which requires 10,000 iterations to achieve a maximum accuracy of 99.07%, Branch 2 only needs 300 iterations to reach a result of 99.35%, and the classification accuracy for AGW, MAW and VE events can all reach 100%. Ultimately, when using Infra-Net for classification, the overall classification accuracy stabilizes at 100%, yielding very excellent results. When facing the classification of different categories of infrasound signals, it is only necessary to adjust the number of neurons in the final fully connected layer of Infra-Net to match the corresponding number of categories, without requiring further modifications to the model architecture.



(a) Using Branch 1



(b) Using Branch 2



(c) Using Infra-Net

Figure 10: Classification results for four types of infrasound events obtained using different networks.

For the comparative experiments, state-of-the-art wavelet scattering-based methods along with advanced techniques currently employed in infrasound classification were selected, including WST-AMResNet-18 (Liu et al. 2025), WST-Trans (Shirodkar et al. 2025), WST-BiLSTM (Zhang et al. 2024), and MS-SE-ResNet (Tan et al. 2024). The results are summarized in Table 6.

Table 6 The training set and test set division of LOTIS.

	ACC /%	*Cohen's Kappa /%*
MS-SE-ResNet	96.12	94.65
WST-Trans	98.06	96.00
WST-BiLSTM	98.06	97.32
WST-AMResNet-18	94.17	91.96
Infra-Net	100	100

As can be observed from the table 6, Infra-Net not only achieved the highest classification accuracy but also exhibited a significantly higher Cohen's Kappa compared to the other three models. The elevated Kappa value indicates a high level of agreement between the model's predictions and the ground truth labels, substantially surpassing the level expected by random chance. This demonstrates that the performance advantage of Infra-Net does not stem from biases in the training data distribution but is rather attributable to its stronger feature discrimination capability. Furthermore, it validates the robustness under class-imbalanced conditions.

In summary, experiments with different classification methods on this dataset demonstrate that Infra-Net consistently achieves the best performance, thereby validating its superior capability in multi-class interference infrasound signal recognition. Subsequently, the field-measured CTBTO dataset will be employed for evaluation.

4.3.1 Ablation experiment

To investigate the effect of the parallel structure, specifically the roles of the two branches and the confidence-based decision-making module in the classification of infrasound events, three ablation experiments are conducted:

Testing the classification capability of the GA-BiGRU branch itself: The entire MSCI-Net branch and the confidence-based decision-making module are removed, and only the GA-BiGRU branch is used for classification, denoted as Ma;

Testing the classification capability of the MSCI-Net branch itself: The entire GA-BiGRU branch and the confidence-based decision-making module are removed, and only the MSCI-Net branch is used for classification, denoted as Mb;

Testing the effectiveness of the dual-branch structure: Both the GA-BiGRU and MSCI-Net branches are retained, but the confidence-based decision-making module is replaced with a simple averaging operation, denoted as Mc;

Validating the effectiveness of the confidence-based decision-making module: The complete model proposed in the paper is used for classification, that is Infra-Net.

The radial basis function SVM (RBF-SVM) is established as the benchmark after demonstrating superior accuracy over alternative kernel functions in comparative tests. Corresponding ablation experiment results are presented in Table 7.

Table 7 Results of ablation experiments.

Model	*ACC /%*	*P /%*	*F1*
Baseline	75.73	70.06	77.91
Ma	79.59	75.95	80.44

Mb	81.14	77.78	82.03
Mc	81.89	78.05	82.73
Infra-Net	**82.07**	**78.62**	**82.76**

Note: *ACC*, the accuracy; *P*, precision.

As can be observed from Table 7, all deep learning models demonstrate significantly superior performance compared to the SVM baseline (with the highest accuracy improvement reaching 6.34%), proving the effectiveness of deep learning in the infrasound classification task. Furthermore, ablation experiment results indicate that the dual-branch parallel architecture (Mc) achieves better performance than any single-branch model (Ma or Mb), demonstrating that the feature extraction capabilities of the GA-BiGRU and MSCI-Net branches are complementary and that their fusion is necessary. Moreover, the confidence-based decision-making module proposed in this paper (Md) achieves optimal performance across all evaluation metrics, slightly outperforming the simple averaging fusion strategy (Mc). This indicates that the proposed module can more effectively integrate the predictive information from both branches, enabling more accurate decision-making. These results collectively validate the effectiveness of each module in the proposed architecture.

We also conducted a statistical significance test to more rigorously evaluate the performance difference between Infra-Net and the ablation models (especially Mc). The results are shown in Table 8.

Table 8 The significance test results between Infra-Net and the closest-performing models such as Mc. The results show that the performance improvement of our method is statistically significant across multiple repeated experiments.

Model	Source	Performance Metric	p-value	Significant?
Infra-Net vs Mc	Table 7	*ACC/P/F1*	$p < 0.05$	Yes
Infra-Net vs Mb	Table 7	*ACC/P/F1*	$p < 0.05$	Yes
Infra-Net vs Ma	Table 7	*ACC/P/F1*	$p < 0.05$	Yes

To intuitively demonstrate the effectiveness of this fusion mechanism, we have supplemented the visualization of the probability change process as you requested. It should be noted that when both branches make correct predictions or both make incorrect predictions, the fusion result is consistent with the branch results and cannot reflect the decision correcting capability of the fusion module. Therefore, we focus on representative samples where the two branches produce inconsistent classification conclusions. Figure 11 shows a sample where Branch 1 is correct and Branch 2 is incorrect; after inner product fusion, the joint confidence favors the true class, and the final output is a correct recognition result. Figure 12 shows the opposite case (Branch 1 incorrect, Branch 2 correct); the fusion module similarly corrects the influence of the erroneous branch and outputs a correct recognition result. These two cases fully demonstrate that the multi view learning strategy combined with the confidence-based fusion module proposed in this paper can effectively leverage the complementarity of the two branches in feature perspectives, suppress the adverse effects of single branch misjudgment, and thus significantly improve the accuracy and reliability of recognition results.

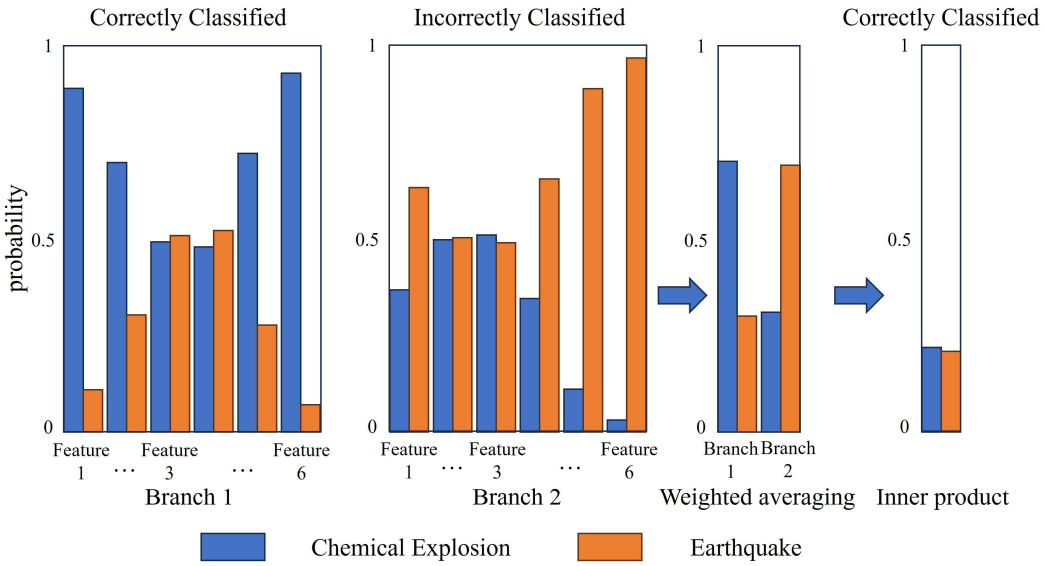


Figure 11: Confidence fusion process when Branch 1 is correct and Branch 2 is incorrect.

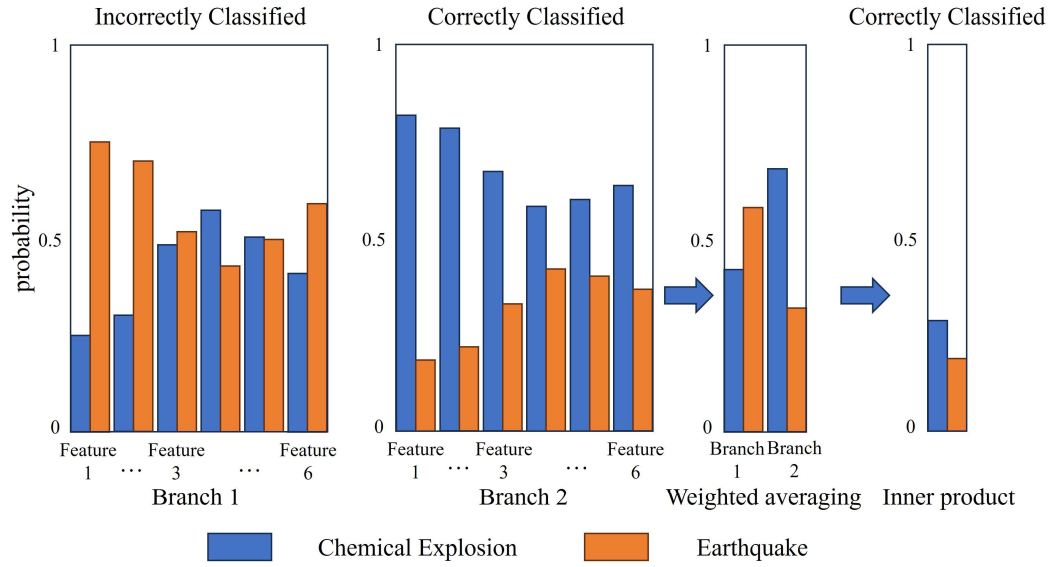


Figure 12: Confidence fusion process when Branch 1 is incorrect and Branch 2 is correct.

4.3.2 Comparative experiment

Since the layer depth of the wavelet scattering network and the time-invariant scale can affect the dimensionality of the resulting wavelet scattering coefficients, which in turn influence the classification outcomes, comparative experiments were conducted to determine the optimal selection of network layers and time-invariant scales. These experiments aimed to derive general conclusions about the parameter settings. According to Wang et al. (2017), the optimal dimensionality of the feature matrix is achieved when the product of the scattering decomposition sampling frequency and the time-invariant scale is approximately equal to the total number of sampling points N. Therefore, the time-invariant scale was primarily analyzed by considering cases in which it is equal to the signal duration and half the signal duration. Additionally, one-way analysis of variance (ANOVA-1) (Ai et al. 2008) was performed to verify that taking the logarithm of the wavelet scattering features can effectively enhance the separability. Finally, experiments were conducted by replacing the GA-BiGRU network in the Infra-Net with a GA-GRU network. This was done to validate the superior feature extraction capability of the GA-BiGRU compared to the GA-GRU.

(1) Two-layer wavelet scattering networks were utilized, with the time-invariant scales set to 150 seconds and 75 seconds. Infrasound signals with a data size of 1 × 3000 were input into the wavelet scattering networks to extract the logarithmic wavelet scattering features with data sizes of 6 × 168 and 12 × 123 as the inputs. These cases are referred to as methods Md1 and Md2.

(2) A three-layer wavelet scattering network was employed, with the time-invariant scale set to 75 seconds (this study sets it to 150 seconds). Infrasound signals with a data size of 1 × 3000 were input into the wavelet scattering network to extract the logarithmic wavelet scattering features with a data size of 191 × 12 (this study yields a data size of 307 × 6) as inputs. This case is referred to as method Me.

(3) The GA-BiGRU network was replaced with a GA-GRU network. This case is referred to as method Mf.

Figure 13 presents a comparison of the classification results for experiments (1) – (3).

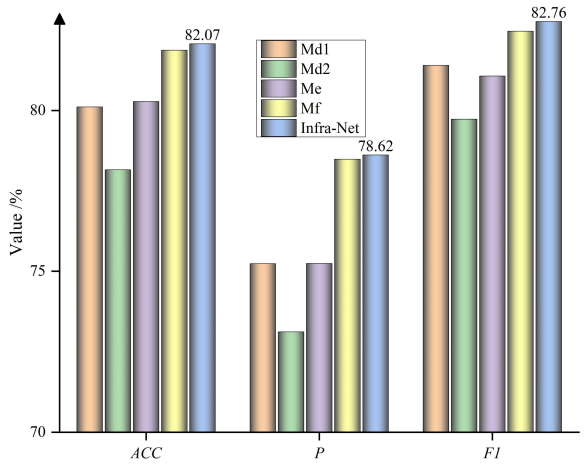


Figure 13: Comparison of the experimental results.

By comparing Md1 with Md2 and Me with the Infra-Net, it was found that when the number of network layers is fixed, the time-invariant scale directly affects the sample expansion multiple and the length of the data. Moreover, the data length of the scattering features has a direct impact on the classification results. Shorter data lengths are associated with poorer classification results, indicating that even though the number of extracted scattering feature paths grows exponentially, the amount of effective information they carry remains limited. Therefore, a balance must be struck between the dilation factor and the feature length. The experimental results lead to the final conclusion that setting the product of the scale parameter and the sampling frequency close to the total number of sampling points ensures sufficient information retention in the feature matrix while avoiding redundancy caused by excessive dimensionality, thus achieving better classification results. Additionally, it was found that when a three-layer network is used, the classification accuracy is superior to the case when a two-layer network is used, suggesting that a three-layer wavelet scattering network is generally more effective in extracting signal features. Finally, comparing Mf with the Infra-Net revealed that the GA-BiGRU network generally has an advantage over the GA-GRU network in terms of processing time series data.

In this study, after extracting the wavelet scattering features, their logarithm was taken to ascertain whether this transformation positively influenced the experimental outcomes. ANOVA-1 was conducted on the wavelet scattering features before and after taking the logarithm. The threshold for statistical significance was set to $p < 0.05$. This conventional level indicates that observed differences are unlikely to occur by chance.

Features were extracted via a 150-second time-invariant scale scattering network. Each input raw signal (1×3000) generates six feature vectors (307×1), denoted as Features 1 to 6. One-way ANOVA was performed on these six features both before and after logarithmic transformation, with results presented in Table 9. The comparisons reveal that with the exception of Feature 3, the

significance of the differences was greater after the logarithm of the remaining Features was taken. Furthermore, there are three Features, which initially lacked significant differences, became significantly different after the logarithmic transformation. Additionally, random sampling analysis across all of the data confirmed the universality of this conclusion, thereby demonstrating that taking the logarithm of the wavelet scattering features, as in the proposed method, can enhance their separability and consequently improve the classification accuracy.

Table 9 Results of the ANOVA-1 analysis before and after logarithmic transformation of the wavelet scattering features.

	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6
Before logarithm	**0.0016**	**0.0573**	0.0034	**0.0182**	**0.0807**	**0.0518**
After logarithm	**0.0009**	**0.0016**	0.0950	**0.0033**	**6.42E−06**	**3.31E−09**

Furthermore, to substantiate the reliability of the classification performance of Infra-Net, comparative experiments have been conducted with more current advanced methods. These include advanced methods used in the field of infrasound signal classification, namely the CNN network (Bryan et al. 2018), the improved CNN-5 network (Tan et al. 2021), the improved LeNet-5 (Leng et al. 2022), the improved Alex-Net (Yuan et al. 2024), Prototype network (Zhao et al. 2024), and MS-SE-ResNet (Tan et al. 2024). The comparison also includes methods such as WST-AMResNet-18 (Liu et al. 2025), WST-Trans (Shirodkar et al. 2025), and WST-BiLSTM (Zhang et al. 2024). For baseline comparison, SVM was employed as a traditional method.

Figure 14 presents the classification results of the different models for two types of infrasound events. Compared with traditional methods, the classification accuracy of Infra-Net is improved by about 9%. Even compared with the MS-SE-ResNet, which has the highest accuracy among them, Infra-Net still achieves an accuracy improvement of 3.55%. When compared with other advanced methods related to the combination of wavelet scattering features and deep learning in this study, the accuracy is also significantly increased, which fully demonstrates that the confidence-based decision-making module has also played an important role in improving classification performance. It can also be observed that, compared with traditional infrasound classification methods, wavelet scattering based methods mostly achieve better classification results, indicating that it has great application potential in feature representation and small sample classification tasks, and will provide a reliable methodological and technical support for following research.

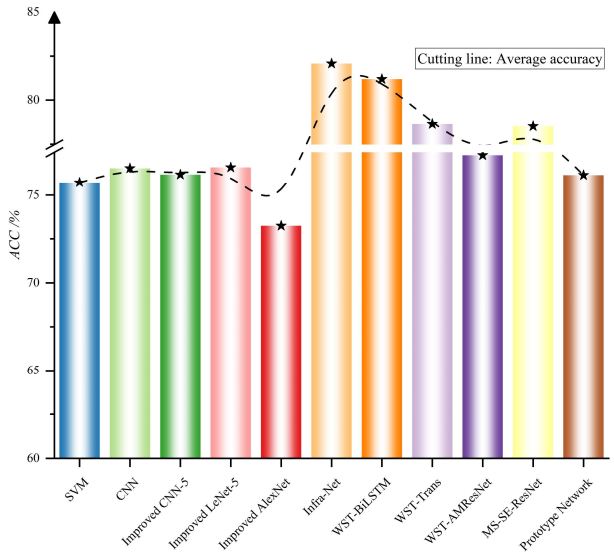


Figure 14: Classification results for two types of infrasound events obtained using different models.

5 Discussion

This section discusses the performance differential between traditional wavelet scattering-based methods and the proposed approach, benchmarking it against several widely-adopted techniques.

The first comparative experiments were conducted: with fixed network architecture and parameters, event classification was performed using scattering features both before and after logarithmic processing as inputs, while simultaneously comparing the performance differences between column-wise decomposition of the scattering feature matrix into independent inputs versus traditional entire matrix input. Experimental results are detailed in Table 10.

Table 10 Comparative Experimental Results

Input method	*ACC* (%)
Infra-Net	**82.07**
Without logarithmic processing	78.91
Entire matrix input	77.89

As shown in Table 10, the Infra-Net achieves superior accuracy compared to two conventional approaches. This result strongly validates that employing logarithmic transformation to amplify feature distinctions is an effective strategy for enhancing classification performance of low-frequency signals. Furthermore, the column-wise decomposition method for scattering matrices significantly outperforms traditional entire matrix input in accuracy. This advantage stems from a dual mechanism: the decomposition of scattering feature paths enhances training sufficiency, while information fusion theory enables comprehensive analysis of the results. Their synergistic effect significantly optimizes both classification accuracy and robustness.

To highlight the superiority of Infra-Net, in addition to the comparative experiments shown in Table 10, the classification accuracies of the LSTM, BiLSTM, WST-LSTM, WST-BiLSTM and WST-BiGRU models were also compared. Furthermore, to validate the efficacy of the purpose-built lightweight design in MSCI-Net for small-sample scenarios, the proposed structure was substituted with deeper network architectures, including Alex-Net (Krizhevsky et al. 2017), Visual Geometry Group (VGG)-16, and VGG-19, to compare the classification performances of the shallow and deep networks under small-sample conditions. The comparison results are presented in Table 11, and all results are the average values of each subset. The accuracy and other indices of the direct classification obtained using preprocessed data are significantly lower than those obtained using logarithmic wavelet scattering features, with improvements of nearly 20% across all of the indices when these features are used, thus confirming their effectiveness. The results also indicate that replacing MSCI-Net with deeper networks did not enhance the accuracy. Although the classification efficiency values are not listed, the deep networks required significantly longer training time compared to MSCI-Net. This suggests that shallow networks can better and more rapidly learn generalized features from limited data in small-sample scenarios.

Table 11 Classification results of the different methods.

Model	*ACC* (%)	*P* (%)	*F1* (%)
LSTM	52.90	52.19	50.16
BiLSTM	55.11	54.97	50.89
WST-LSTM	77.38	75.42	77.49

WST-BiLSTM	77.33	76.24	77.41
WST-BiGRU	77.87	72.84	80.29
Replacing MSCI-Net with Alex-Net	79.60	75.43	80.75
Replacing MSCI-Net with VGG-16	81.19	76.96	82.16
Replacing MSCI-Net with VGG-19	80.30	76.62	82.19
Infra-Net	**82.07**	**78.62**	**82.76**

As evidenced by the experimental results, the proposed method achieves a maximum classification accuracy of 82.07% on the CTBTO measured dataset. Although this performance is lower than that obtained on the public LOTIS dataset, it maintains a clear advantage when compared horizontally with other approaches evaluated on the same CTBTO dataset. The performance discrepancy primarily stems from the CTBTO dataset's inherent characteristics: it contains infrasound events with a wide range of source distances and demonstrates greater susceptibility to noise interference in real monitoring environments, collectively increasing classification difficulty. Crucially, despite these challenges, the proposed method consistently delivers superior and stable classification performance across both distinct datasets, thereby validating its effectiveness and robustness in complex practical scenarios.

6 Conclusion

This study addresses the challenge of classifying infrasound signals with limited sample sizes in natural hazard monitoring through the Infra-Net model. Through systematic experimentation and analysis, key conclusions were drawn regarding data processing, model construction, and application efficacy, demonstrating its effectiveness in modeling complex systems for robust geophysical monitoring and disaster decision support.

In data processing, compared to conventional methods that use the entire matrix as input, this study improves classification accuracy by 8.3%. Furthermore, using the extracted logarithmic wavelet scattering features as inputs, rather than raw signals, directly enhances classification precision by over 20%. These results indicate that the data processing method significantly improves classification performance in small-sample scenarios.

In model construction, compared to the baseline model, Infra-Net improves all classification metrics by an average of 6.58%, fully validating the effectiveness of its multi-branch collaborative modeling strategy in enhancing the discriminative capability of the model.

In application efficacy, based on the conclusion that the optimal feature dimension satisfies the parameter setting of scattering sampling frequency × time-invariant scale ≈ signal sampling points, this method is expected to achieve predictable and robust performance in analogous application scenarios. Despite the above results, this study still has several limitations that urgently need to be addressed in future work.

First, the performance gap between the public LOTIS dataset and the CTBTO field-measured data indicates that the current model's recognition capability is significantly constrained when dealing with signal distortions caused by extremely long-range propagation, and its cross-scene generalization ability still needs improvement. Furthermore, the ablation experiment results show that the confidence-based fusion strategy proposed in this paper achieves only limited improvement over simple averaging fusion. To address these limitations, we will focus on the following research directions. First of all, we will introduce domain adaptation strategies, for example, reducing the model's dependence on specific monitoring environments through transfer learning to narrow the generalization gap. Second, we will collect and annotate more diverse infrasound event data (including various natural and

anthropogenic events at different propagation distances and signal-to-noise ratios) to increase the coverage of training samples. Third, we will explore uncertainty-aware fusion methods to better handle conflicts among predictions from different views, thereby improving decision robustness and supporting more reliable natural hazard early warning.

CRediT authorship contribution statement

Hongru Li: Conceptualization, Methodology, Software, Writing-Original draft preparation; **Xihai Li**: Data curation; **Jihao Liu**: Supervision; **Shengjie Luo**: Software ; **Yun Zhang**: Software, Validation

Declaration of Competing Interest

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

Acknowledgments

I would like to express my sincere gratitude to Professor Yang Jun and his team for their significant assistance with the data aspect of this study.

Data Availability

The datasets supporting the findings of this study are available from the corresponding author upon reasonable request. The LOTIS dataset can be accessed through the associated publication at https://doi.org/10.11684/j.issn.1000-310X.2024.06.002. The CTBTO infrasound data are available via the official website of the Preparatory Commission for the CTBTO subject to their standard application and authorization procedures.

References

Abbaszadeh M, Soltani SM, Ahmed AN. 2022. Optimization of support vector machine parameters in modeling of Iju deposit mineralization and alteration zones using particle swarm optimization algorithm and grid search method. Comput Geosci. 165:105140. https://doi.org/10.1016/j.cageo.2022.105140.

Ai L, Wang J, Wang X. 2008. Multi-features fusion diagnosis of tremor based on artificial neural network and D–S evidence theory. Signal Process. 88(12):2927–2935. https://doi.org/10.1016/j.sigpro.2008.06.018.

Albert S, Linville L. 2020. Benchmarking current and emerging approaches to infrasound signal classification. Seismol Res Lett. 91(2A):921–929. https://doi.org/10.1785/0220190116.

Alegria OC, et al. 2015. Empirical wavelet transform-based detection of anomalies in ULF geomagnetic signals associated to seismic events with a fuzzy logic-based system for automatic diagnosis. In: Wavelet Transform and Some of Its Real-World Applications, InTech. p 111–124. https://doi.org/10.5772/61163.

Al JA, Khushaba RN. 2023. Deep hand gesture recognition: a wavelet scattering alternative to convolutional networks. In: 2023 IEEE Statistical Signal Processing Workshop (SSP). p 438–442. https://doi.org/10.1109/SSP53291.2023.10208011.

Andén J, Mallat S. 2014. Deep scattering spectrum. IEEE Trans Signal Process. 62(16):4114–2418. https://doi.org/10.1109/TSP.2014.2326991.

Bruna J, Mallat S. 2011. Classification with scattering operators. In: CVPR 2011. p 1561–1566. https://doi.org/10.1109/CVPR.2011.5995635.

Bruna J, Mallat S. 2013. Invariant scattering convolution networks. IEEE Trans Pattern Anal Mach Intell. 35(8):1872–1886. https://doi.org/10.1109/tpami.2012.230.

Bryan KJ, et al. 2018. Deep wavelet scattering features for infrasonic threat identification. In: Chemical, Biological, Radiological, Nuclear, and Explosives Sensing XIX. p 1062901–1062918. https://doi.org/10.1117/12.2304544.

Buriro A, et al. 2021. Classification of alcoholic eeg signals using wavelet scattering transform-based features. Comput. Biol Med. 139:104969. https://doi.org/10.1016/j.compbiomed.2021.104969.

Chang TY, Cui HL. 2013. Determination of direction of arrival of seismic wave by a single Tri-axial fiber optic geophone. Def Technol. 9(1):1–9. https://doi.org/10.1016/j.dt.2013.02.001.

Cho K, et al. 2014. Learning phrase representations using RNN encoder–decoder for statistical machine translation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). p 1724–1734. https://doi.org/10.3115/v1/D14-1179.

Dai R, et al. 2024. VTNet: A multi-domain information fusion model for long-term multi-variate time series forecasting with application in irrigation water level. Appl Soft Comput. 167(A):112251. https://doi.org/10.1016/j.asoc.2024.112251.

Dai Y, Teng P, Lyu J, Ji P, Cheng W. 2021. Analysis of the infrasound signals from rocket launch. J Appl Acoust. 40(5):676–683. https://doi.org/10.11684/j.issn.1000-310X.2021.05.004.

Ding ZW, et al. 2024. Recognition method of coal–rock reflection spectrum using wavelet scattering transform and bidirectional Long–Short-Term Memory. Rock Mech Rock Eng. 57:1353–1374. https://doi.org/10.1007/s00603-023-03600-z.

Fan X, et al. 2022. Intelligent recognition of coal mine microseismic signal based on wavelet scattering decomposition transform. J China Coal Soc. 47(7):2722–2731. https://doi.org/0253-9993(2022)07-2722-10.

Hochreiter S, Schmidhuber J. 1997. Long short-term memory. Neural Comput. 9(8):1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735.

Huang W, Zou XY, Qiao R. 2018. Scattering wavelet based deep network for image classification. In: Communications in Computer and Information Science. Springer Singapore. p 459–469. https://doi.org/10.1007/978-981-10-8530-7_45.

Khan AA, Dhawan A, Akhlaghi N, Majdi JA, Sikdar S. 2017. Application of wavelet scattering networks in classification of ultrasound image sequences. In: 2017 IEEE International Ultrasonics Symposium (IUS). p 1–4. https://doi.org/10.1109/ULTSYM.2017.8091649.

Koch K, Pilger C. 2018. Infrasound observations from the site of past underground nuclear explosions in North Korea. Geophys J Int. 216(1):182–200. https://doi.org/10.1093/gji/ggy381.

Kogelnig A, Hübl J, Suriñach E, Vilajosana I, McArdell BW. 2014. Infrasound produced by debris flow: propagation and frequency content evolution. Nat. Hazard. 70:1713–1733. https://doi.org/10.1007/s11069-011-9741-8.

Krizhevsky A, Sutskever I, Hinton GE. 2017. ImageNet classification with deep convolutional neural networks. J Commun ACM. 60(6):84–90. https://doi.org/10.1145/3065386.

Lee S, Hong SG. 2024. Classification of nuclear activity types for neighboring countries of South Korea using machine learning techniques with xenon isotopic activity ratios. Nucl Eng Technol. 56(4):1372–1384. https://doi.org/10.1016/j.net.2023.11.042.

Leng X, et al. 2022. Debris flow infrasound recognition method based on improved LeNet-5 network. Sustainability-Basel. 14(23):15925. https://doi.org/10.3390/su142315925.

Li HR, et al. 2024. Infrasound event classification fusion model based on multiscale SE-CNN and BiLSTM. Appl Geophys. 21:579–592. https://doi.org/10.1007/s11770-024-1089-4.

Li HR, et al. 2026. Machine learning-driven classification of natural disasters via parallel confidence fusion. Mach. Learn.: Sci. Technol. 7: 015028. https://doi.org/10.1088/2632-2153/ae3c58.

Li J, et al. 2019. Heart sound signal classification algorithm: a combination of wavelet scattering transform and twin Support Vector Machine. IEEE Access. 7:179339–179348. https://doi.org/10.1109/access.2019.2959081.

Lilly JM, Olhede SC. 2010. On the analytic wavelet transform. IEEE Trans Inf Theory. 56(8):4135–4156. https://doi.org/10.1109/tit.2010.2050935.

Li M, Liu X, Liu X. 2016. Infrasound signal classification based on spectral entropy and support vector machine. Appl Acoust. 113:116–120. https://doi.org/10.1016/j.apacoust.2016.06.019.

Liu L, Wu J, Li D, Senhadji L, Shu H. 2019. Fractional wavelet scattering network and applications. IEEE Trans Biomed Eng. 66(2):553–563. https://doi.org/10.1109/tbme.2018.2850356.

Liu X, Li M, Tang W, Wang S, Wu X. 2014. A new classification method of infrasound events using Hilbert-Huang Transform and Support Vector Machine. Math Probl Eng. 2014(3):1–6. https://doi.org/10.1155/2014/456818.

Liu Y, et al. 2025. Underwater acoustic classification using wavelet scattering transform and convolutional neural network with limited dataset. Appl Acoust. 232:110564. https://doi.org/10.1016/j.apacoust.2025.110564.

Lone AW, Aydin N. 2023. Wavelet scattering transform based doppler signal classification. Comput Biol Med. 167:107611. https://doi.org/10.1016/j.compbiomed.2023.107611.

Luong T, Pham H, Manning CD. 2015. Effective approaches to attention-based neural machine translation. In: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP). p 1412–1421. https://doi.org/10.18653/v1/D15-1166.

Lu X, Liu S, Chen X, Zhang K. 2023. Identification of the lubrication state of journal bearings based on acoustic emission and WST-CNN collaboration. J Vib Shock. 42(2):71–77. https://doi.org/10.13465/j.cnki.jvs.2023.22.008.

Marchetti E, et al. 2019. Infrasound array analysis of debris flow activity and implication for early warning. J Geophys Res-Earth. 124(2):567–587. https://doi.org/10.1029/2018jf004785.

Park J, et al. 2018. Characteristics of infrasound signals from North Korean underground nuclear explosions on 2016 January 6 and September 9. Geophys J Int. 214(3):1865–1885. https://doi.org/10.1093/gji/ggy252.

Pasyanos ME and Kim K. 2019. Seismoacoustic analysis of chemical explosions at the nevada national security site. J Geophys Res-Sol Ea. 124(1):908–924. https://doi.org/10.1029/2018JB016705.

Pásztor M, Czanik C, Bondár I. 2023. A single array approach for infrasound signal discrimination from quarry blasts via machine learning. Remote Sens.15(6):1657. https://doi.org/10.3390/rs15061657.

Priya BL, Jayalakshmy S, Pragatheeswaran JK, Saraswathi D, Poonguzhali N. 2021. Scattering convolutional network based predictive model for cognitive activity of brain using empirical wavelet decomposition. Biomed Signal Process Control. 66:102501. https://doi.org/10.1016/j.bspc.2021.102501.

Shirodkar VR, Edla DR, Kumari A. 2025. Advancing motor imagery EEG classification through wavelet scattering transforms and 1D transformers. Procedia Comput Sci. 258:2860–2869. https://doi.org/10.1016/j.procs.2025.04.546.

Singhal S, Kumar M. 2024. SPTDMD-WST: Arrhythmia classification from spatiotemporal modes of dynamic mode decomposition using wavelet scattering transform. Biomed Signal Process Control. 92:105983. https://doi.org/10.1016/j.bspc.2024.105983.

Stéphane M. 2009. CHAPTER 7-Wavelet Bases. In: A Wavelet Tour of Signal Processing. Academic Press. p 263–376. https://doi.org/10.1016/B978-0-12-374370-1.00011-2.

Souli S, Amami R, Yahia SB. 2021. A robust pathological voices recognition system based on DCNN and scattering transform. Appl Acoust. 177:107854. https://doi.org/10.1016/j.apacoust.2020.107854.

Tan XF, et al. 2024. Classification method of infrasound events based on the MVIDA algorithm and MS-SE-ResNet. Appl Geophys. 21:667–679. https://doi.org/10.1007/s11770-024-1112-9.

Tan XF, Li XH, Liu JH, Li G, Yu XT. 2021. Classification of chemical explosion and earthquake infrasound based on 1-D convolutional neural network. J Appl Acoust. 40(3):457–467. https://doi.org/10.11684/j.issn.1000-310X.2021.03.018.

Thüring T, Schoch M, Herwijnen A, Schweizer J. 2015. Robust snow avalanche detection using supervised machine learning with infrasonic sensor arrays. Cold Reg Sci Technol. 111:60–66. https://doi.org/10.1016/j.coldregions.2014.12.014.

Toney L, Fee D, Witsil A, Matoza RS. 2022. Waveform features strongly control subcrater classification performance for a large labeled volcano infrasound dataset. Seism Rec. 2(3):167–175. https://doi.org/10.1785/0320220019.

Tuan DP. 2023. Classification of motor-imagery tasks using a large eeg dataset by fusing classifiers learning on wavelet-scattering features. IEEE Trans Neural Syst Rehabil Eng. 31:1097–1107. https://doi.org/10.1109/TNSRE.2023.3241241.

Wang H, Li S, Zhou Y, Chen S. 2018a. SAR automatic target recognition using a roto-translational invariant wavelet-scattering convolution network. Remote Sens. 10(4):501. https://doi.org/10.3390/rs10040501.

Wang L, Liu B, Wang H, Zhong G, Dong J. 2018b. Deep gabor scattering network for image classification. In: Pattern Recognition and Computer Vision. Springer International Publishing. p 332–343. https://doi.org/10.1007/978-3-030-03335-4_29.

Wang W, Gao X, Le L. 2017. Study of the similarities in scale models of a single-layer spherical lattice shell structure under the effect of internal explosion. Shock Vib. 4:1–13. https://doi.org/10.1155/2017/9181729.

Wang ZX, et al. 2022. Monitoring and prediction of the vibration intensity of seismic waves induced in underwater rock by underwater drilling and blasting. Def Technol. 18(1):109–118. https://doi.org/10.1016/j.dt.2020.10.007.

Wen Y, et al. 2019. Analysis of infrasound signals generated by CZ-7 teleport 2 launch vehicle. J Ordnance Equip Eng. 40(9):70–73. https://doi.org/10.11809/bqzbgcxb2019.09.015.

Wiatowski T, Bolcskei H. 2018. A mathematical theory of deep convolutional neural networks for feature extraction. IEEE Trans Inf Theory. 64(3):1845–1866. https://doi.org/10.1109/TIT.2017.2776228.

Yuan L, Liu DL, Sang XJ, Zhang SJ, Chen Q. 2024. Debris flow infrasound signal recognition approach based on improved AlexNet. Computer and Modernization. (3):1–6. https://doi.org/10.3969/j.issn.1006-2475.2024.03.001.

Zhang A, Chun S, Cheng Z, Zhao P. 2024. Predicting the core thermal hydraulic parameters with a gated recurrent unit model based on the soft attention mechanism. Nucl Eng Technol. 56(6):2343–2351. https://doi.org/10.1016/j.net.2024.01.045.

Zhang H, et al. 2024. Classification of motor imagery EEG signals using wavelet scattering transform and Bi-directional long short-term memory networks. Biocybern Biomed Eng. 44(4):874–884. https://doi.org/10.1016/j.bbe.2024.11.003.

Zhang Q, Wang P, Pedrycz W, Li Z. 2024. Neighborhood entropy guided by a decision attribute and its applications in multi-source information fusion and attribute selection. Appl Soft Comput. 167(B):112380. https://doi.org/10.1016/j.asoc.2024.112380.

Zhang X, Shen J, Li J, Liu X. 2025. An instance-oriented multi-source information fusion technique based on neighborhood granules. Appl Soft Comput. 181:113483. https://doi.org/10.1016/j.asoc.2025.113483.

Zhao ZJ, et al. 2024. A method for classification of few-shot infrasound signals applying prototype network. J Appl Acoust. 43(6):1193–1202. https://doi.org/10.11684/j.issn.1000-310X.2024.06.002.