acoustics

MDPI

Article

Infrasound Signal Classification Fusion Model Based on Double-Branch and Multi-Scale CNN and LSTM

Hao Yin [1], Yu Lu [1], Yunhui Wu [1], Wei Cheng [2,3], Xinliang Pang [1,*] and Peng Li [1,*]

[1] State Key Laboratory of Chemistry for NBC Hazards Protection, Beijing 102205, China; yinhao1207@163.com (H.Y.); geolyu@163.com (Y.L.); wuyunhui@sklnbcpc.cn (Y.W.)
[2] Institute of Acoustics, Chinese Academy of Sciences, Beijing 100049, China; chengwei@mail.ioa.ac.cn
[3] Key Laboratory of Noise and Vibration Research, Chinese Academy of Sciences, Beijing 100190, China
* Correspondence: pangxinliang@sklnbcpc.cn (X.P.); lipeng@sklnbcpc.cn (P.L.)

Abstract

The accurate classification of infrasound events is significant in natural disaster warning, verification of nuclear test bans and geophysical research. Current deep learning-based classification methods mostly focus on denoised and filtered signals. To simplify the process, avoid information loss, and address the issues of incomplete feature extraction by single-scale convolution kernels and the potential loss of physical information by single models, this paper directly utilizes raw infrasound signals and proposes two fusion classification models based on multi-scale Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM). Experiments were conducted on a typical infrasound signal dataset (comprising four signal types: mountain-associated waves, auroral infrasound waves, volcanic eruptions, and microbaroms). The performances of the two models were compared in terms of accuracy, convergence speed, and stability. The results indicate that both models achieve classification accuracies exceeding 99% with optimal parameter combinations. The dual-branch multi-scale CNN-LSTM model generally outperforms the multi-scale CNN-LSTM model in classification accuracy, while also demonstrating faster convergence speed and better stability. Addressing the class imbalance in the dataset, evaluations using precision, recall, and F1-score further validated the effectiveness of the proposed models. This study demonstrates that the proposed methods can effectively achieve end-to-end classification of raw infrasound signals and are competitive with existing techniques.

Keywords: infrasonic signal; classification recognition; double-branch and multi-scale; convolutional neural network; long short-term memory network

Academic Editors: Michel Darmon and Jian Kang

Received: 22 January 2026
Revised: 28 February 2026
Accepted: 19 March 2026
Published: 24 March 2026

1. Introduction

Infrasound is a low-frequency sound wave that cannot be directly perceived by human ears, with frequencies typically ranging from 0.001 to 20 Hz. Many natural phenomena and human activities are accompanied by the generation of infrasound signals during their occurrence and development, such as nuclear tests [1,2], volcanic eruptions [3,4], earthquakes [5], snow avalanches [6], debris flow [7,8], chemical explosions [9], rocket launches [10], and thunderstorms [11,12]. Infrasound can propagate thousands of kilometers through the atmosphere with low attenuation, making it one of the four technologies (seismic, hydroacoustic, infrasound, and radionuclide) used by the International Monitoring System (IMS) of the Comprehensive Nuclear-Test-Ban Treaty Organization (CTBTO) [13] for monitoring nuclear tests. The key to nuclear explosion infrasound monitoring technology lies in accurately distinguishing the infrasound signals generated by

 https://doi.org/10.3390/acoustics8020021

nuclear explosions from those of other infrasound sources. Accurately identifying these infrasound signals holds significant importance for national security and geophysical research, among other fields.

Early research on infrasound signal classification primarily adopted a "feature extraction + classifier design" paradigm, namely, extracting features in the time domain, frequency domain, or time–frequency domain from the signals and then performing identification using a classifier. The key lies in how to extract salient features to effectively distinguish between different categories of infrasound signals. Initially, researchers employed various single-feature extraction methods such as Mel-frequency cepstral coefficients [14–17], wavelet transform [18,19], short-time Fourier transform [20], Hilbert–Huang transform [21,22], wavelet scattering transform [23,24], information spectrum entropy [25], and empirical mode decomposition [26]. A single feature has limited ability to fully describe the complex characteristics of event signals. To address this, researchers have introduced feature fusion methods, combining multiple features to improve the classification performance of infrasound signals. These studies indicate that feature fusion can effectively enhance the accuracy and robustness of infrasound signal classification [27–30].

Traditional pattern recognition methods require onerous feature construction to obtain features that can uniquely represent infrasonic event categories. However, infrasound signals propagate through time-varying atmosphere and interact with the environment during propagation, making it challenging to extract reliable event features for classification. With the rapid rise of artificial intelligence, infrasonic event classification technology has developed from traditional pattern recognition into the currently popular end-to-end deep learning models. Solomon et al. omitted the feature extraction step and directly used Long Short-Term Memory (LSTM) networks to classify four types of infrasonic signals from the Library of Typical Infrasonic Signals (LOTIS) dataset [31], achieving an accuracy of only 67%. They also employed the VGG model to recognize raw, resampled and band-pass-filtered time-series signals, finding that the recognition accuracy of preprocessed signals was lower than that of raw signals [32]. Leng et al. utilized an improved LeNet-5 network to classify and recognize time–frequency graphs of five types of infrasonic events (debris flows, wind, lightning, rain, and thunder) [33], achieving an accuracy of 84.1%, but the data samples were processed with denoising. Li et al. proposed an infrasound event classification fusion model based on multi-scale SE-CNN and BiLSTM [34], which automatically extracts temporal and spatial features of signals and also realizes the fusion of two types of features, achieving a classification accuracy of over 98% for chemical explosions and earthquakes.

Deep learning-based infrasound signal classification models have simplified the traditional feature extraction process. However, most current research employs denoising and filtering methods for preprocessing to improve the signal-to-noise ratio (SNR) and model performance. Yet, as explicitly noted in the literature [32], denoising and filtering may lead to the loss of inherent physical information within the signals, preventing the acquisition of their complete representation. Furthermore, while CNNs excel at capturing local spatial features in data and perform well when processing time–frequency spectrograms or one-dimensional signals, a standard CNN is unable to learn the temporal dependencies within signals. Conversely, LSTM effectively captures both long-term and short-term dependencies in sequential data, but standard LSTM struggles to extract the local spatial features of signals. Using a single type of network (either CNN or LSTM) alone is insufficient to fully and comprehensively extract the spatial–temporal multi-dimensional features embedded in infrasound signals. In addition to model architecture limitations, infrasound signal classification research faces challenges at the data level—the complexity of raw data and the variability of signal duration. Raw infrasound signals contain complex

https://doi.org/10.3390/acoustics8020021

background noise and interference from propagation effects and may exhibit a low SNR, demanding greater robustness and feature extraction capabilities from models. Moreover, the duration of signals from different infrasound events can vary significantly, requiring models to flexibly handle variable-length input sequences.

Addressing the aforementioned challenges and limitations, particularly the risk information loss due to denoising, the inadequacy of single-model feature extraction, and the need to handle raw data and variable durations, this study undertakes the following primary work:

1. Simplification of the preprocessing pipeline in traditional identification methods. Classification research is conducted directly using infrasound signals without denoising or filtering. This aims to maximally preserve the original physical information within the signals and explore the feasibility of achieving effective classification on complex data closer to real-world scenarios.
2. Significant improvement in the recognition rate. The recognition rate of the intelligent fusion-based processing method has been significantly increased to over 99% in infrasound signal classification.
3. Fusion of CNN and LSTM networks. The combined use of CNN and LSTM networks achieves a more comprehensive characterization of the spatial–temporal features within infrasound signals, thereby enhancing the performance of infrasound signal classification.

2. Materials and Methods

2.1. Dataset

To facilitate comparison with the classification performance of other models, the dataset used in this study is sourced from the open-source Library of Typical Infrasonic Signals [31,35,36]. This dataset includes four types of infrasonic signals, 110 mountain-associated waves (MAWs), 115 Auroral Infrasonic Waves (AIWs), 65 Volcanic Waves (VOLs), and 58 Microbarom Waves (MBs), and the number was 348. This is shown in Table 1. For the purpose of model training in later stages, we performed simple normalization preprocessing on the data. The waveforms of these four categories of signals are illustrated in Figure 1.

2.2. Model Construction

2.2.1. Classification Model Based on Multi-Scale CNN and LSTM Fusion

LSTM focuses on the temporal information of signals, but using a single LSTM model may lead to incomplete or ineffective feature extraction, leading to low classification accuracy. Additionally, a CNN emphasizes spatial features of signals. However, due to the varying propagation paths and distances of infrasound signals, along with the influence of atmospheric factors such as wind and temperature during propagation, the corresponding spatial features are not always the same. This is a very key issue for choosing an appropriate convolution kernel size to extract significant features of infrasound signals. If a single-scale convolution kernel is used in the network for feature extraction, it is likely that important information from infrasound signals will be lost. Therefore, without deepening the network structure, stacking convolutional layers with gradually decreasing kernel sizes provides different receptive fields. The network can adaptively select the appropriate receptive field, thus addressing the issue of information loss. First, the convolution module is used to preliminarily extract the local and key information of the infrasonic signal. Then, the LSTM layer is employed to deeply extract the temporal correlation information from the output of the convolution layer. Finally, the output results are obtained through the fully connected layer, as shown in Figure 2.

Table 1. LOTIS dataset.

Type of Infrasonic Signals	Number of Signals	Signal Duration [s]	Number of Samples
MAW	110	7.68	768
AIW	115	7.68	768
VOL	65	7.68	768
MB	58	5.12	512
Totals	348	-	-

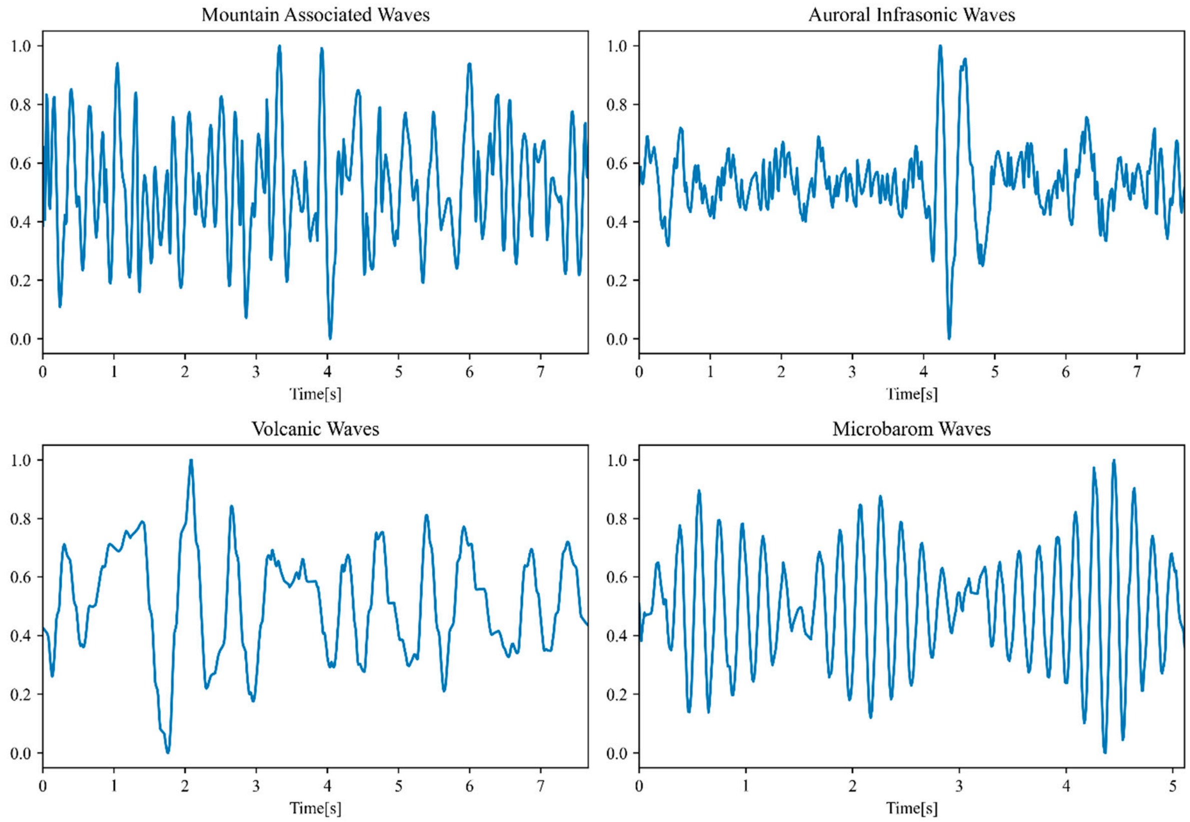

Figure 1. Four types of infrasound patterns in the LOTIS dataset.

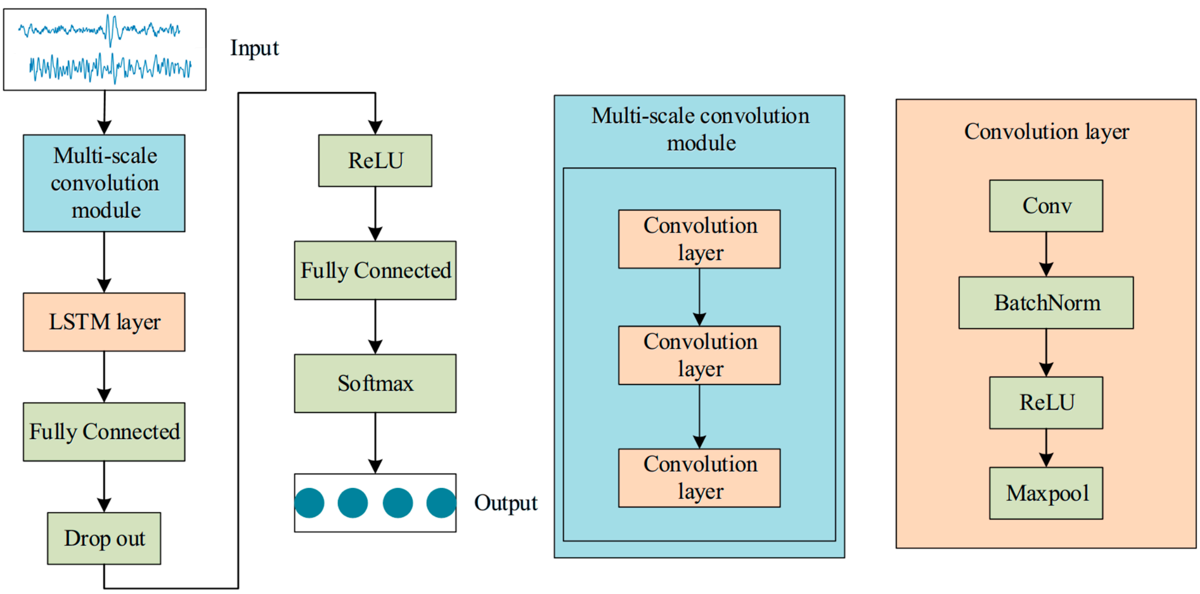

Figure 2. Classification model of multi-scale fusion of CNN and LSTM.

https://doi.org/10.3390/acoustics8020021

The spatial features of the normalized infrasound signals are preliminarily extracted through three convolutional layers. In the shallow layers of the network, a convolutional layer with a kernel size of 7 is used to extract global feature information under a large receptive field. This is followed by successive layers with kernel sizes reduced to 5 and then 3 to perform deep extraction of the subtle features of the infrasound signals. Batch normalization (BN) is added after each convolutional layer to accelerate the convergence speed of the model. The activation function (Rectified Linear Unit, ReLU) is employed to perform nonlinear mapping on the output of the convolution layers, which can not only avoid information loss arisen from feature compression, but also alleviate the vanishing gradient problem. Then, a pooling layer with both stride and a size of 2 is added. In this paper, max pooling is employed to down-sample the features produced by the convolution layers, thus achieving feature dimensionality reduction. This significantly reduces the input dimensions for subsequent network layers, thereby reducing the overall number of the model parameters, and also contributes to improving multiple key performance aspects of the model. After multiple stages of convolution, activation, and pooling, the LSTM layer is employed to further extract the temporal features of the signals. And the dropout operation is introduced to randomly discard a portion of the neurons in the neural network, reducing model complexity and preventing overfitting. Finally, softmax is used as the classifier to complete the classification of the four types of infrasound signals. Table 2 shows the relevant parameters for this network.

Table 2. Relevant parameters of multi-scale CNN and LSTM fusion classification model.

Network Layer	Parameter Setting
Conv1d_1	$7 \times 1 \times 32$
Maxpool1d_1	2×1
Conv1d_2	$5 \times 1 \times 64$
Maxpool1d_2	2×1
Conv1d_3	$3 \times 1 \times 128$
Maxpool1d_3	2×1
LSTM layer	Input size: 128
Dropout layer	0.5
FC	Number of neurons: 50
Softmax layer	Number of neurons: 4

2.2.2. Model Based on Double-Branch and Multi-Scale CNN and LSTM

The overall structure of the model designed in this section is largely consistent with the structure presented in the preceding section. The main difference between the two lies in the basic module of the convolutional neural network. As shown in Figure 3, the convolutional neural network basic module designed here is referred to as CBLOCK, which contains two parallel convolution branches and a merge path. The first branch consists of two convolution operations, normalization operations, and activation operations, while the second branch includes one convolution operation, normalization operation, and activation operation. The design idea of the convolution in these two branches is aimed at extracting features of the infrasonic data at different receptive field scales using convolution kernels of different sizes, thereby achieving the purpose of enriching extracted features. The output features after the operation of the two parallel convolution paths will be concatenated through concatenation operation. And the concatenated features will be dimensionally reduced by max pooling, aiming to extract the most relevant features related to the category attributes. The operations following the output of CBLOCK3 are consistent with those described in the preceding section, while the output after two linear operations will be mapped to the category attribute space through softmax. Detailed parameter settings for this network are listed in Table 3.

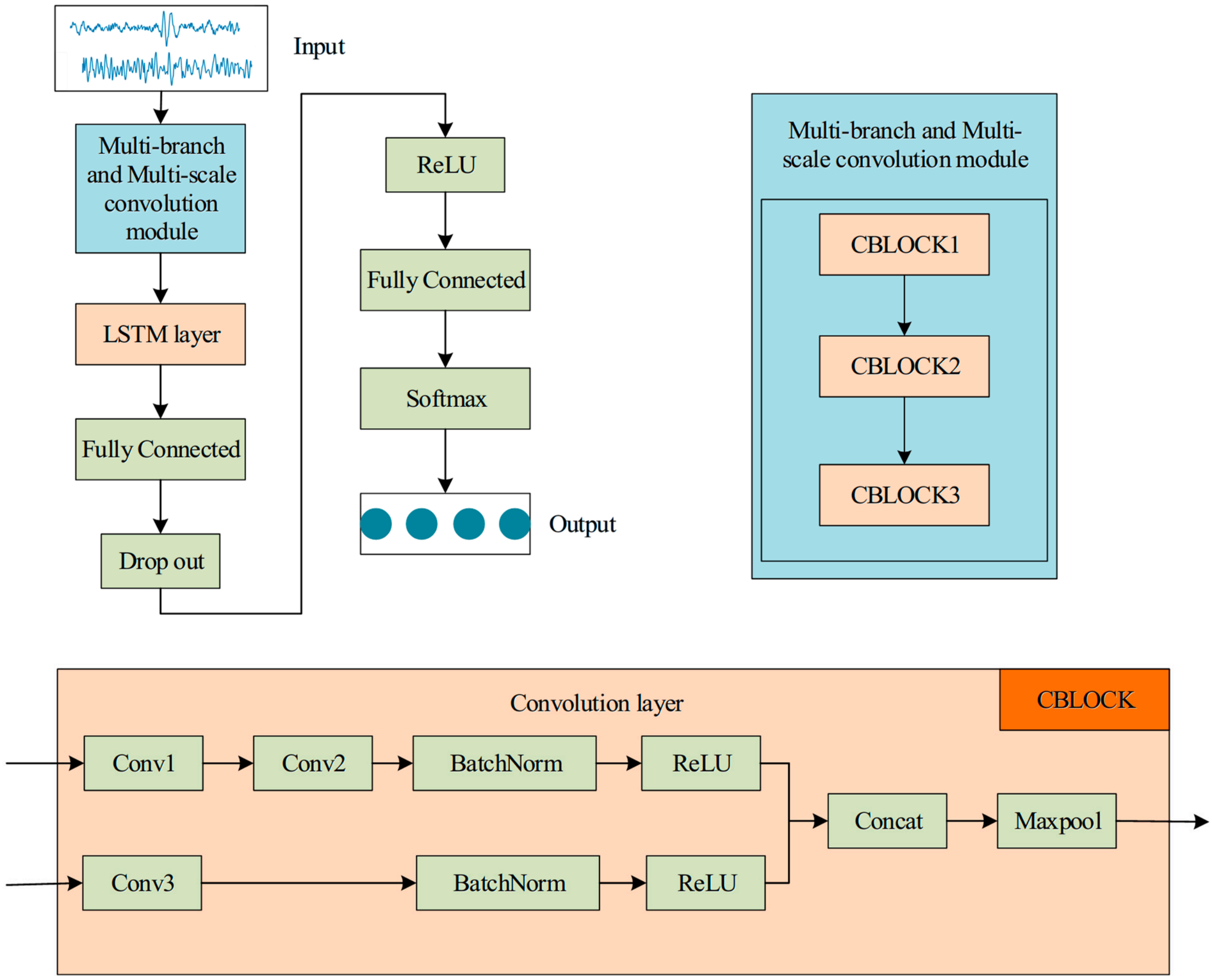

Figure 3. Model structure of double-branch and multi-scale CNN and LSTM.

Table 3. Model parameters of double-branch and multi-scale CNN and LSTM.

Network Layer	Parameter Setting
CBLOCK1	Conv1d_1: 7 × 1 × 32 Conv1d_2: 7 × 32 × 16 Conv1d_3: 1 × 1 × 16 Maxpool1d: 2 × 1
CBLOCK2	Conv1d_1: 5 × 32 × 64 Conv1d_2: 5 × 64 × 32 Conv1d_3: 1 × 32 × 32 Maxpool1d: 2 × 1
CBLOCK3	Conv1d_1: 3 × 64 × 128 Conv1d_2: 3 × 128 × 64 Conv1d_3: 1 × 64 × 64 Maxpool1d: 2 × 1
LSTM layer	Input size: 128
Dropout layer	0.5
FC	Number of neurons: 50
Softmax layer	Number of neurons: 4

2.2.3. Experimental Setup

This paper divides the dataset into three parts, training set, validation set, and test set, at a ratio of 8:1:1. The number of samples in the training set is 278, the validation set contains 34 samples, and the test set has 36 samples, as shown in Table 4. The experimental environment is detailed in Table 5.

Table 4. Experimental dataset partitioning.

Dataset	MAW	AIW	VOL	MB	Total
Training set	88	92	52	46	278
Validation set	11	11	6	6	34
Test set	11	12	7	6	36
Total	110	115	65	58	348

Table 5. Experimental environment.

Test Environment	Type
System	Windows 10 64-bit
Graphics card	NVDIA GeForce RTX 4090 D
CPU	Intel(R) Core(TM) i9-14900K @3.20 GHz
RAM	128 G
Framework	Pytorch 1.12.0
Programming language	Python 3.9.19

The small-batch training strategy was adopted in this experiment. The batch size was set to 16, the learning rate was 0.001, the optimizer was SGD, and the number of iterations was 500. Due to the randomness in dividing the dataset, this study conducted multiple experiments to further verify the generalization ability and robustness of the network model.

3. Results

3.1. Accuracy Comparison of Three Classification Models

To determine the hyperparameter selection of the LSTM layer in the infrasound signal classification model described in Section 2.2, we tested the recognition accuracy of the two models with different numbers of hidden layers and hidden dimensions in the LSTM layer. The LOTIS dataset used in this study is relatively small (348 samples across four classes), with some classes containing fewer than 60 samples. To improve confidence in the results, multiple random splits of the dataset were performed, and the average recognition accuracies obtained are presented in Tables 6 and 7.

Table 6. Accuracy of the model based on multi-scale CNN and LSTM.

Num Layers	Hidden Size			
	32	64	128	256
1	98.90 ± 1.45	99.12 ± 1.37	99.12 ± 1.37	98.68 ± 1.48
2	99.34 ± 1.25	98.68 ± 1.48	99.12 ± 1.37	99.34 ± 1.25
3	98.90 ± 1.45	98.02 ± 2.71	99.12 ± 1.80	99.12 ± 1.80
4	94.28 ± 9.83	97.58 ± 3.66	99.12 ± 2.44	98.24 ± 2.19
5	80.88 ± 20.52	81.32 ± 19.51	74.51 ± 17.40	90.55 ± 6.94

Table 7. Accuracy of the model based on multi-branch and multi-scale CNN and LSTM.

Num Layers	Hidden Size			
	32	**64**	**128**	**256**
1	99.78 ± 0.79	99.78 ± 0.79	99.56 ± 1.07	100.00 ± 0.00
2	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00
3	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00
4	98.46 ± 5.55	98.46 ± 5.55	98.68 ± 3.98	100.00 ± 0.00
5	73.63 ± 18.41	79.12 ± 17.08	81.35 ± 23.94	86.59 ± 12.64

According to the experimental results in Tables 6 and 7, both models achieved relatively good results in classifying infrasound signals when the number of hidden layers in the LSTM layer was four or less. Although good classification results were achieved, when the number of hidden layers in the LSTM layer was four, it consumed a large amount of computational resources, and the overall classification performance was slightly inferior to that with one, two, or three hidden layers.

To effectively demonstrate the advantages of the multi-scale and dual-branch design, a test was conducted on a single-scale CNN-LSTM model (with a fixed scale of five), and the recognition accuracies are shown in Table 8. Compared with the two aforementioned models, the single-scale CNN-LSTM model achieved the lowest average accuracy and exhibited poorer stability in test results across random dataset splits, with a standard deviation as high as 25.61.

Table 8. Accuracy of the model based on single-scale CNN and LSTM.

Num Layers	Hidden Size			
	32	**64**	**128**	**256**
1	97.14 ± 0	97.14 ± 0	88.57 ± 14.84	96.19 ± 1.65
2	97.14 ± 0	96.19 ± 1.65	98.09 ± 1.65	96.19 ± 1.65
3	85.71 ± 15.12	89.52 ± 10.82	89.52 ± 5.95	93.34 ± 1.65
4	61.90 ± 25.61	50.47 ± 5.95	74.29 ± 15.91	79.05 ± 15.74
5	48.57 ± 0	48.57 ± 0	48.57 ± 0	48.57 ± 0

3.2. Stability of Classification Models

To verify the stability of the model designed in this paper, the operation of generating the dataset in section A was repeated 10 times to obtain 10 different datasets, and the experimental classification results of the two models are shown in Figures 4 and 5. It can be seen that the classification results of the two models for 10 groups of datasets were less stable when the number of hidden layers was four. Furthermore, the stability of the first model is slightly worse than that of the second model, which is also confirmed by the results shown in Tables 6 and 7.

Based on the above results and analysis, the average recognition accuracy of each infrasonic signal classification model was similar, and the stability of both models was better when the number of hidden layers in the LSTM was set to one, two, or three. However, the recognition performance of the model based on the double-branch multi-scale CNN and LSTM was still better than that of the model based on the multi-scale CNN and LSTM.

3.3. Comparison of Convergence Speeds

Furthermore, we analyzed their convergence speed during the training phase to determine the specific parameters of the number of hidden layers and the dimension for classifiers with similar classification performance. According to the loss curves of each classification model shown in Figures 6 and 7, both classification models can converge, but different numbers of layers and dimensions will affect the convergence speed during training. When the hidden layer dimensions are the same, the convergence speed gradually slows down with an increase in the number of hidden layers. This is related to the increasing complexity of the model with more hidden layers. When the LSTM layer has three hidden layers, the model can converge in the training set but does so at the slowest speed. When the LSTM layer has a single hidden layer, the model converges the fastest in the training set.

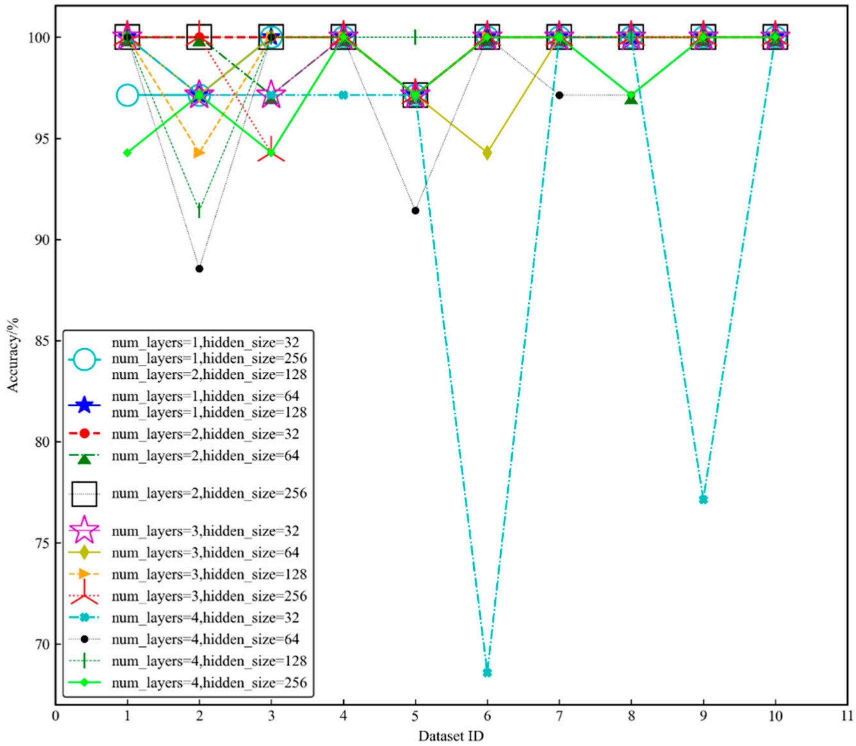

Figure 4. Classification results of multi-scale CNN and LSTM model for 10 datasets.

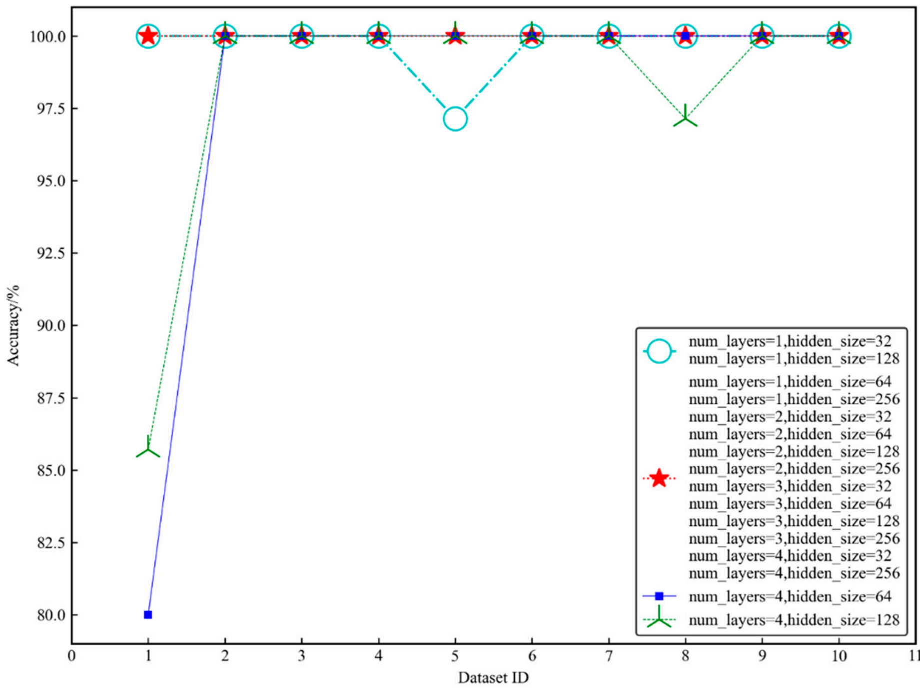

Figure 5. Classification results of double-branch and multi-scale CNN and LSTM model for 10 datasets.

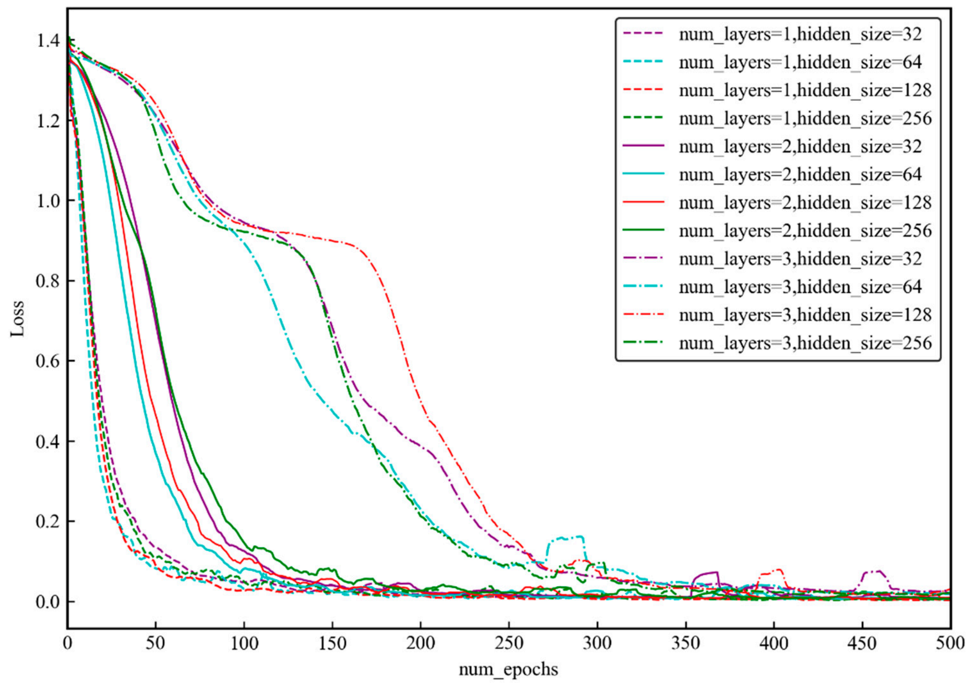

Figure 6. Loss curves based on multi-scale CNN and LSTM model.

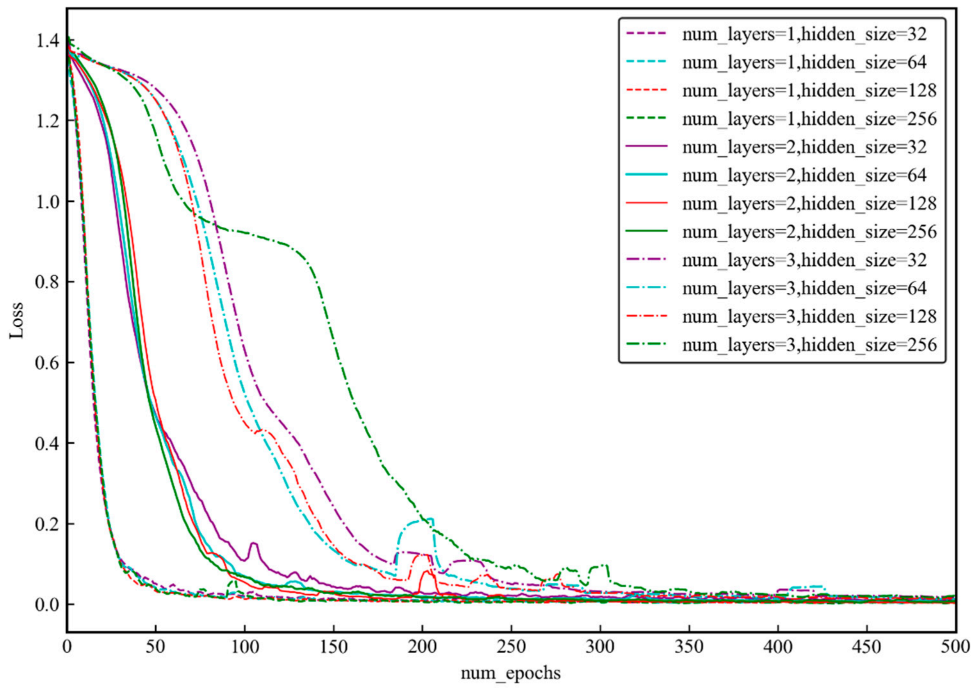

Figure 7. Loss curves based on double-branch and multi-scale CNN and LSTM model.

However, as seen in Tables 6 and 7, the recognition accuracy in the test set is lower at this time. Therefore, the optimal number of hidden layers in the LSTM was determined to be two. At the same time, it can be seen from the figures that models with the same number of hidden layers but different hidden layer dimensions have a small difference in convergence speed. This is partially amplified as shown in Figure 8. Figure 8A shows the partial loss curve of the multi-scale CNN and LSTM model when the number of hidden layers is two, and Figure 8B shows the partial loss curve of the double-branch multi-scale CNN and LSTM model when the number of hidden layers is two. Based on the comparative analysis of Table 6, it can be seen that when the hidden layer dimension is 64, the convergence speed of the model based on multi-scale CNN and LSTM is faster, but the test accuracy is lower. When the hidden layer dimensions are 32 or 256, the test accuracy of model based on multi-scale CNN and LSTM is higher, and both are 99.34%, but the model converges faster when the hidden layer dimension is 32. Table 7 shows that when the number of hidden layers is two, the test accuracy of the model based on double-branch multi-scale CNN and LSTM was consistent, at 100%, yet the model with a hidden layer dimension of 128 converges faster than the others. In conclusion, for the multi-scale CNN and LSTM model, we selected two hidden layers and a hidden layer dimension of 32. For the multi-branch multi-scale CNN and LSTM model, we adopted two hidden layers and a hidden layer dimension of 128. Figure 8C shows the convergence of these two models under the respective parameters; even though the multi-branch multi-scale CNN and LSTM model has a higher hidden layer dimension, it converges faster.

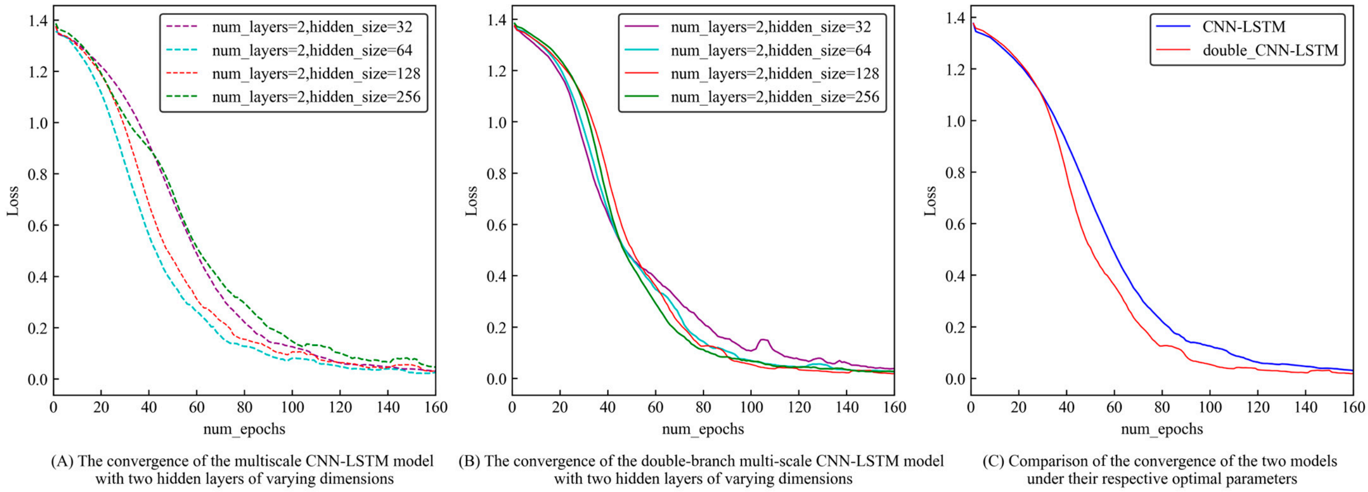

Figure 8. Comparison of convergence of two models when the number of hidden layers is 2.

In addition, we also compared the convergence of the two models with the same number of hidden layers and hidden layer dimensions on the training set. As shown in Figure 9, when the number of hidden layers and the hidden layer dimension are small, the convergence speeds of two models are quite similar. However, as the number of hidden layers and hidden layer dimensions increase, the difference in convergence speeds between the two models gradually increases, and the convergence speed of the double-branch multi-scale CNN and LSTM model is faster than that of the multi-scale CNN and LSTM model.

3.4. Other Performance Metrics of the Two Proposed Classification Models

Due to the class imbalance issue in the dataset, accuracy alone cannot fully reflect model performance. This section employs precision, recall, and F1-score metrics to conduct a more comprehensive evaluation of the proposed models. Based on a comprehensive consideration of accuracy, stability, and convergence speed, this section only presents the precision, recall, and F1-score results for each class under different hidden layer dimensions with a fixed hidden layer count of two, as shown in Tables 9–12. Overall, the performance of the model based on multi-branch multi-scale CNN and LSTM surpasses that of the multi-scale CNN and LSTM model.

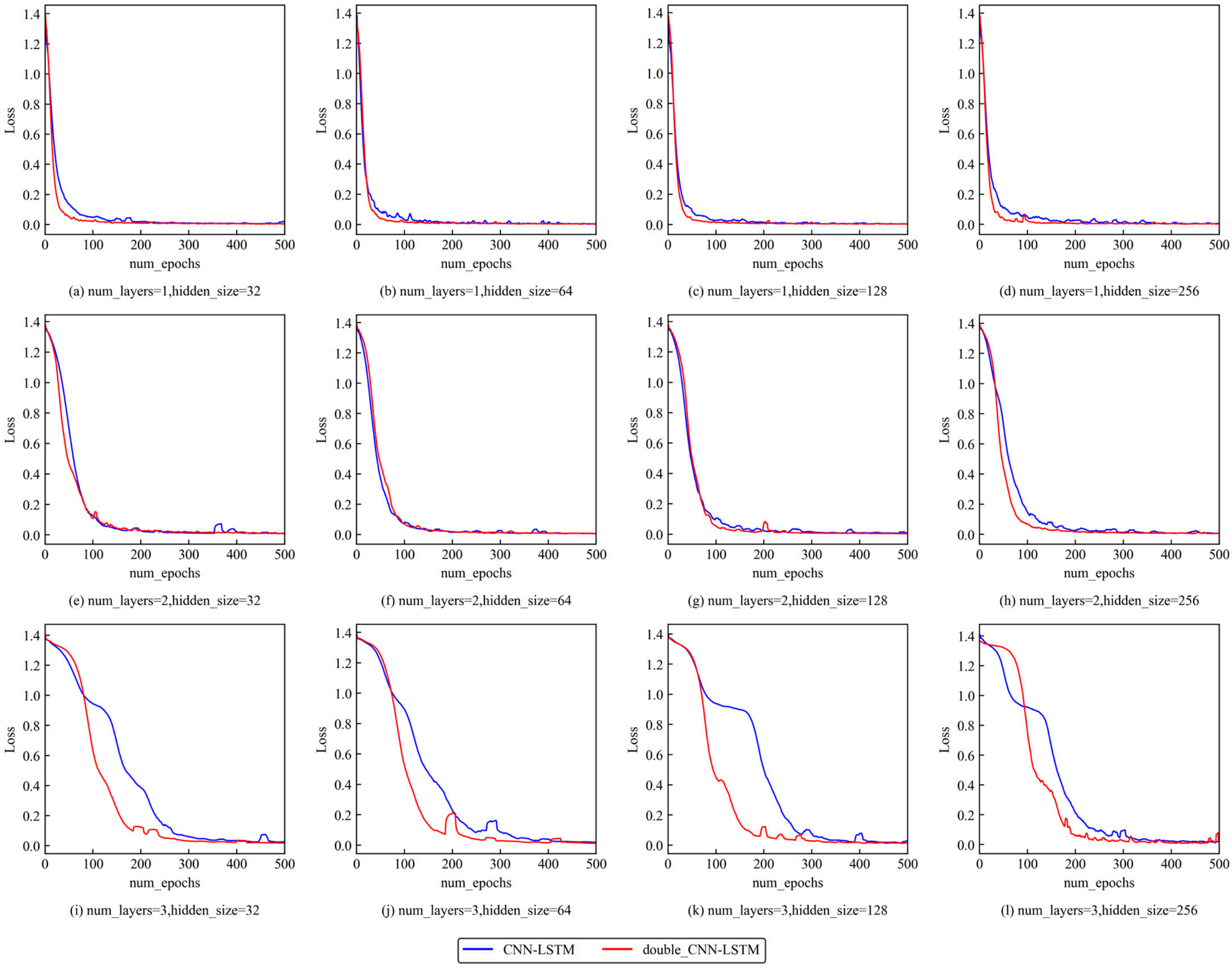

Figure 9. Comparison of convergence of two models.

Table 9. Precision of the model based on multi-scale CNN and LSTM.

Classes	Hidden Size			
	32	64	128	256
AIW	0.99 ± 0.02	0.98 ± 0.03	0.98 ± 0.03	0.99 ± 0.02
MAW	1 ± 0	0.99 ± 0.02	1 ± 0	1 ± 0
MB	1 ± 0	1 ± 0	1 ± 0	1 ± 0
VE	0.98 ± 0.04	0.98 ± 0.04	0.99 ± 0.03	0.98 ± 0.04

Table 10. Recall of the model based on multi-scale CNN and LSTM.

Classes	Hidden Size			
	32	64	128	256
AIW	0.98 ± 0.03	0.97 ± 0.03	0.99 ± 0.02	0.98 ± 0.03
MAW	1 ± 0	0.99 ± 0.02	0.99 ± 0.02	1 ± 0
MB	1 ± 0	1 ± 0	1 ± 0	1 ± 0
VE	0.98 ± 0.03	0.97 ± 0.05	0.97 ± 0.05	0.98 ± 0.03

Table 11. F1-scores of the model based on multi-scale CNN and LSTM.

Classes	Hidden Size			
	32	**64**	**128**	**256**
AIW	0.98 ± 0.02	0.97 ± 0.02	0.98 ± 0.02	0.98 ± 0.02
MAW	1 ± 0	0.99 ± 0.01	0.99 ± 0.01	1 ± 0
MB	1 ± 0	1 ± 0	1 ± 0	1 ± 0
VE	0.98 ± 0.03	0.97 ± 0.03	0.98 ± 0.03	0.98 ± 0.03

Table 12. Precision, recall, and F1-score of the model based on multi-scale multi-branch CNN and LSTM.

Classes	Hidden Size			
	32	**64**	**128**	**256**
AIW	1 ± 0	1 ± 0	1 ± 0	1 ± 0
MAW	1 ± 0	1 ± 0	1 ± 0	1 ± 0
MB	1 ± 0	1 ± 0	1 ± 0	1 ± 0
VE	1 ± 0	1 ± 0	1 ± 0	1 ± 0

3.5. Comparison with Existing Findings

In order to validate the advantages of the proposed model, comparisons were made between the CNN-LSTM fusion model and existing infrasound signal classification methods based on the LOTIS dataset. The results are shown in Table 13, and it can be seen that the proposed model shows better performance on the same dataset.

Table 13. Comparison with other methods.

Study	Year	Method	Accuracy (%)
Bryan et al. [23]	2018	WST-SVM	98.7
Solomon et al. [31]	2018	LSTM	66.88
Zhao et al. [37]	2023	Prototypical network-LSTM	97.96
Proposed work	2025	CNN-LSTM	99.34

4. Discussion

Infrasound signal is non-stationary. Its frequency components of different times and the distribution of a particular frequency are different. Moreover, the signal characteristics of different types of infrasound sources are also distinct. In our work, there was no need to filter the infrasound signals, as we directly used the raw waveform data as the input for the model for classifying the LOTIS dataset.

Compared with the existing research using deep learning to classify infrasound signals, the method proposed in this paper achieves an average correct recognition rate of 99.34% on raw data. Complex propagation processes inevitably cause distortion and noise contamination in infrasound signals, yet the signals that are collected by monitoring stations still retain certain critical physical information. Noise typically manifests as irregular short-term fluctuations, while effective event signals exhibit inherent correlations. The convolutional operation in a CNN performs spatial filtering, and the fusion of multi-scale convolutional kernels facilitates the model's extraction of deep features that are sensitive to signal class differentiation and insensitive to noise. LSTM selectively memorizes key temporal correlation features and discards irrelevant information through the gating mechanisms. The fusion of spatially invariant features extracted by a CNN and the temporal variation features captured by LSTM resolves the problem of insufficient and incomplete features extracted by a single model. Without relying on "clean" waveforms, the proposed spatio-temporal fusion model learns physically meaningful discriminative features.

Results from Tables 6 and 7 indicate that when the LSTM network has fewer layers (one layer), the model can only capture local fluctuation patterns of the signals, potentially leading to insufficient feature extraction and consequently slightly lower classification accuracy. In contrast, models with more layers (2–3 layers) can construct hierarchical feature representations, where the lower-level network extracts local temporal patterns, while the higher-level network integrates these lower-level features to capture long-term temporal dependencies. However, increasing the number of layers (4–5 layers) may trigger gradient explosion, causing difficulties in training deep models, manifested as reduced accuracy and slower convergence. Therefore, selecting an optimal number of layers balances feature extraction capability with training stability, enabling effective fusion with the CNN's spatial features. The dimensionality of the LSTM network determines its cell state vector capacity, which directly affects information-bearing capability. Low dimensionality may excessively compress the feature space, losing critical details and causing severe confusion between similar events. High dimensionality enhances nonlinear fitting capacity but risks introducing redundancy. Moreover, model parameters increase exponentially with dimensionality, making overfitting likely. Moderate-to-high dimensionality allows for sufficient encoding of multi-scale temporal features, complementing a CNN's spatial features to enhance joint representations while balancing model performance and computational cost, achieving the fastest convergence.

A comparison between Tables 6 and 8 further illustrates that multi-scale convolutional kernels can capture more detailed spatial information. Comparative analysis of Tables 6 and 7 reveals that the dual-branch multi-scale CNN-LSTM classification model generally outperforms the multi-scale CNN-LSTM model. Multi-scale convolutional kernels extract finer spatial details. The dual-branch architecture forms parallel sub-networks, which can not only cross-suppress noise through mutual constraint but also enable interactive feature learning to preserve effective physical patterns. The parallel structure gives the model a multi-path gradient flow during backpropagation, alleviating vanishing gradients in single paths and accelerating convergence of the networks. Notably, when LSTM parameters increase substantially (five layers), the multi-scale CNN-LSTM model achieves higher average classification accuracy than the dual-branch counterpart, which is potentially attributable to overfitting in the latter architecture.

The results in Table 13 show that, compared to methods such as WST-SVM and LSTM, both of the proposed models exhibit higher classification performance and robustness. For WST-LSTM, the classification performance is influenced by its own parameters and manually set features. The classification process is also affected by the quality and length of the data, and preprocessing of the data is inevitably required. Additionally, careful design of feature vectors and fine-tuning of hyperparameters are necessary to ensure optimal recognition performance [23]. For LSTM, while it can capture temporal features over long periods, if the chosen hyperparameters are not suitable, the classification performance will naturally be worse [31]. For prototypical network-LSTM, there is no need to consider the duration of the signal, but the relevant parameters of the prototype network's encoding unit LSTM need to be determined, and the choice of the distance metric function for the metric computation unit must also be considered [36,37].

The physical characteristics of infrasound signals are determined by their source mechanisms. Signals generated by different sources (such as volcanic eruptions, nuclear explosions, mudslides, and auroras) exhibit unique properties in the time–frequency domain. Deep learning models capture mathematical mappings of these physical characteristics through convolutional kernels of different scales. Small-scale feature extractors (such as small one-dimensional convolutional kernels) focus on capturing the local fluctuation patterns of the signal. These features directly correspond to "first arrival" or "high-frequency details" in the waveform, reflecting the physical essence of instantaneous energy release from the source. The intermediate features of the model, formed by combining local fluctuations, characterize specific frequency bands. The medium-scale receptive field enables the network to identify the "timbre" or "texture" of the signal, which precisely corresponds to the spectrum or dominant modal distribution in physical analysis, reflecting the physical characteristics of sustained oscillation processes within the source. A major characteristic of infrasound waves is their ability to propagate thousands of kilometers through the atmosphere. The propagation process is strongly influenced by factors such as atmospheric temperature, wind speed, and wind direction, leading to signal dispersion, multipath interference, and attenuation. This constitutes the core physical reason for the "multi-scale" spatiotemporal variations observed in infrasound signals. Due to variations in atmospheric sound speed with altitude, different frequency components of infrasound waves travel along distinct paths (such as stratospheric ducts and thermospheric ducts), arriving at the receiver at different times, thus forming the so-called "dispersion curve." To capture such long-range dependencies, the model must rely on large-scale (global) feature extractors. These large-scale features are capable of integrating energy segments of different frequencies arriving at different times to form a hypothesis of the complete propagation path. The global contextual information learned by the model actually encodes the physical propagation order of "which frequency component arrives first and which arrives later," which directly corresponds to the specific atmospheric duct structure. Infrasound signals often arrive simultaneously via multiple atmospheric paths, resulting in observed waveforms that are superpositions of multiple wavelets. Multi-scale networks simulate this physical process through the interaction of features across different scales. This fusion of features enables the model to understand that seemingly chaotic waveforms are, in fact, the result of superimposing multiple physical replicas in both time and amplitude, thereby separating different propagation modes within the feature space.

Although this paper has demonstrated the excellent performance of the dual-branch multi-scale CNN and LSTM fusion model in terms of classification accuracy and convergence speed and has presented the test results of the experimental dataset under different random splits, there are still certain limitations in demonstrating the model's generalization ability. Therefore, in future work, we can attempt to compare multi-branch models by relying on different datasets or increasing the amount of data.

5. Conclusions

This paper proposes two multi-scale deep learning models based on the fusion of CNN and LSTM for infrasound signal classification, performing end-to-end feature extraction and classification directly on raw signals after simple normalization. The main conclusions are as follows:

1. In the classification task of four types of infrasound events (MAW, AIW, VOL, MB), the proposed multi-branch multi-scale CNN-LSTM fusion model achieves significantly higher classification accuracy than the standard multi-scale CNN-LSTM model under the same hyperparameters, reaching up to 100%. This indicates that the complex fusion structure combined with multi-scale feature learning can more effectively represent the essential characteristics of raw infrasound signals.

2. Stability tests on different datasets show that the classification results of the multi-branch multi-scale CNN-LSTM model exhibit smaller fluctuations and demonstrate stronger robustness, making it more suitable for raw signal processing in real-world monitoring environments.

https://doi.org/10.3390/acoustics8020021

3. Convergence analysis indicates that this model achieves a faster convergence speed when reaching comparable optimal accuracy. Moreover, as the LSTM hidden layer dimension or number of layers increases, its advantage in learning efficiency further expands, demonstrating higher feature learning efficiency.

This study confirms the feasibility of achieving high-precision classification directly on raw infrasound data, avoiding the potential information loss associated with traditional denoising preprocessing and thereby enhancing the practicality and reliability of infrasound monitoring systems in real-world noisy environments. Limited by current experimental conditions, the sample size and number of categories are relatively small. Future work will validate the model's generalization capability on larger-scale and more diverse datasets and further optimize the network structure to reduce computational complexity, promoting the application of this model in real-time infrasound monitoring systems.

Author Contributions: Conceptualization, H.Y.; methodology, H.Y.; validation, H.Y.; formal analysis, H.Y., Y.L. and X.P.; investigation, H.Y.; data curation, W.C.; writing—original draft preparation, H.Y.; writing—review and editing, H.Y.; supervision, Y.W. and P.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data that support the findings of this study are available from publicly available data (LOTIS), and it can be downloaded via https://github.com/mitchell-solomon/Infrasound-Classification-with-LSTM-RNN/tree/master/infra_tflow/src/data (accessed on 27 August 2024).

Conflicts of 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.

Abbreviations

The following abbreviations are used in this manuscript:

CNN	Convolutional neural network
LSTM	Long- and short-time memory network
IMS	International Monitoring System
CTBTO	Comprehensive Nuclear-Test-Ban Treaty Organization
SVM	Support Vector Machines
LOTIS	Library of Typical Infrasonic Signals
SE-CNN	squeeze excitation–convolution neural network
BiLSTM	bidirectional long short-term memory network
SNR	signal-to-noise ratio
MAWs	Mountain-Associated Waves
AIWs	Auroral Infrasonic Waves
VOLs	Volcanic Waves
MBs	Microbarom Waves
BN	Batch normalization
ReLU	Rectified Linear Unit

References

1. Che, I.-Y.; Kim, K.; Le Pichon, A.; Park, J.; Arrowsmith, S.; Stump, B. Illuminating the North Korean Nuclear Explosion Test in 2017 Using Remote Infrasound Observations. *Geophys. J. Int.* **2021**, *228*, 308–315. [CrossRef]
2. Park, J.; Assink, J.; Stump, B.; Hayward, C.; Arrowsmith, S.; Che, I.-Y. Atmospheric Model Inversion Using Infrasound Signals from the North Korean Underground Nuclear Explosion and the Subsequent Collapse Event in 2017. *Geophys. J. Int.* **2022**, *232*, 902–922. [CrossRef]
3. Iozzia, A. The Influence of Volcano Topographic Changes on Infrasound Amplitude: Lava Fountains at Mt. Etna in 2021. *Bull. Volcanol.* **2023**, *85*, 54. [CrossRef]

4. Morelli, R.S.; Gheri, D.; Campus, P.; Coppola, D.; Marchetti, E. Long Range Infrasound Monitoring of Yasur Volcano. *J. Volcanol. Geotherm. Res.* **2022**, *432*, 107707. [CrossRef]
5. Scamfer, L.T.; Anderson, J.F. Exploring Background Noise with a large-N Infrasound Array: Waterfalls, Thunderstorms, and Earthquakes. *Geophys. Res. Lett.* **2023**, *50*, e2023GL104635. [CrossRef]
6. Watson, L.M.; Carpenter, B.; Thompson, K.; Johnson, J.B. Using Local Infrasound Arrays to Detect Plunging Snow Avalanches along the Milford Road, New Zealand (Aotearoa). *Nat. Hazards* **2022**, *111*, 949–972. [CrossRef]
7. Belli, G.; Walter, F.; McArdell, B.; Gheri, D.; Marchetti, E. Infrasonic and Seismic Analysis of Debris-Flow Events at Illgraben (Switzerland): Relating Signal Features to Flow Parameters and to the Seismo-Acoustic Source Mechanism. *JGR Earth Surface* **2022**, *127*, e2021JF006576. [CrossRef]
8. Liu, D.; Wu, Q.; Dong, H.; Leng, X.; He, L.; Gao, Y. Construction and Application of Debris Flow Infrasound Real-Time Monitoring and Warning Visualization Platform. *Nat. Hazards* **2022**, *112*, 521–543. [CrossRef]
9. Bowman, D.C.; Krishnamoorthy, S. Infrasound from a Buried Chemical Explosion Recorded on a Balloon in the Lower Stratosphere. *Geophys. Res. Lett.* **2021**, *48*, e2021GL094861. [CrossRef]
10. Pilger, C.; Hupe, P.; Gaebler, P.; Ceranna, L. 1001 Rocket Launches for Space Missions and Their Infrasonic Signature. *Geophys. Res. Lett.* **2021**, *48*, e2020GL092262. [CrossRef]
11. Rusz, J. Locating Thunder Source Using a Large-Aperture Micro-Barometer Array. *Front. Earth Sci.* **2021**, *9*, 614820. [CrossRef]
12. Sorokin, A.; Dobrynin, V. Method of Studying Infrasound Waves from Thunderstorms. *Sol.-Terr. Phys.* **2022**, *8*, 62–68. [CrossRef]
13. Coyne, J.; Bobrov, D.; Bormann, P.; Duran, E.; Grenard, P.; Haralabus, G.; Kitov, I.; Starovoit, Y. CTBTO: Goals, Networks, Data Analysis and Data Availability. In *New Manual of Seismological Observatory Practice 2 (NMSOP2)*; GeoForschungsZentrum GFZ: Potsdam, Germany, 2012; p. 877.
14. Ham, F.M.; Leeney, T.A.; Canady, H.M.; Wheeler, J.C. An Infrasonic Event Neural Network Classifier. In *Proceedings of the IJCNN'99. International Joint Conference on Neural Networks. Proceedings (Cat. No.99CH36339), Washington, DC, USA, 10–16 July 1999*; IEEE: Piscataway, NJ, USA, 1999; Volume 6, pp. 3768–3773.
15. Ham, F.M.; Rekab, K.; Acharyya, R.; Lee, Y.C. Infrasound Signal Classification Using Parallel RBF Neural Networks. *Int. J. Signal Imaging Syst. Eng.* **2008**, *1*, 155. [CrossRef]
16. Ham, F.M.; Iyengar, I.; Hambebo, B.M.; Garces, M.; Deaton, J.; Perttu, A.; Williams, B. A Neurocomputing Approach for Monitoring Plinian Volcanic Eruptions Using Infrasound. *Procedia Comput. Sci.* **2012**, *13*, 7–17. [CrossRef]
17. Park, S.; Ham, F.M.; Lowrie, C.G. Discrimination of Infrasound Events Using Parallel Neural Network Classification Banks. *Nonlinear Anal. Theory Methods Appl.* **2005**, *63*, e859–e865. [CrossRef]
18. Chilo, J.; Jabor, A.; Liszka, L.; Eide, A.J.; Lindblad, T.; Bergkvist, L.P.N.-O.; Stahlsten, T.; Andersson, B.L.; Karasalo, I.; Cederholm, A. Filtering and Extracting Features from Infrasound Data. In *Proceedings of the 14th IEEE-NPSS Real Time Conference, Stockholm, Sweden, 4–10 June 2005*; IEEE: Piscataway, NJ, USA, 2006; pp. 451–455.
19. Eckel, F.; Langer, H.; Sciotto, M. Monitoring Sources of Volcanic Activity at Mount Etna Using Pattern Recognition Techniques on Infrasound Signals. *Geophys. J. Int.* **2022**, *232*, 1–16. [CrossRef]
20. Fields, M.P.; Bennett, H.; Scoggins, R. Machine Learning for Source Classification Utilizing Infrasound Data. In *Proceedings of the Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications III*; Pham, T., Solomon, L., Hohil, M.E., Eds.; SPIE: Bellingham, WA, USA, 2021; p. 87.
21. Liu, X.; Li, M.; Tang, W.; Wang, S.; Wu, X. A New Classification Method of Infrasound Events Using Hilbert-Huang Transform and Support Vector Machine. *Math. Probl. Eng.* **2014**, *2014*, 456818. [CrossRef]
22. Luo, X.; Ming, S.; Li, M.; Tang, W. Infrasonic Signals Clustering Based on Hilbert-Huang Transform and K-Means. *Am. J. Inf. Sci. Comput. Eng.* **2017**, *3*, 1–6.
23. Bryan, K.J.; Smith, K.E.; Solomon, M.; Clauter, D.A.; Smith, A.O.; Peter, A.M. Deep Wavelet Scattering Features for Infrasonic Threat Identification. In *Proceedings of the Chemical, Biological, Radiological, Nuclear, and Explosives (CBRNE) Sensing XIX*; Fountain, A.W., Guicheteau, J.A., Howle, C.R., Eds.; SPIE: Orlando, FL, USA, 2018; Volume 10629, pp. 1–11.
24. Smith, K.; Solomon, M.; Bryan, K.J.; Smith, A.O.; Peter, A.M. Near-Field Infrasound Classification of Rocket Launch Signatures. In *Proceedings of the Chemical, Biological, Radiological, Nuclear, and Explosives (CBRNE) Sensing XIX*; Fountain, A.W., Guicheteau, J.A., Howle, C.R., Eds.; SPIE: Orlando, FL, USA, 2018; p. 51.
25. Li, M.; Liu, X.; Liu, X. Infrasound Signal Classification Based on Spectral Entropy and Support Vector Machine. *Appl. Acoust.* **2016**, *113*, 116–120. [CrossRef]
26. Tan, X.; Li, X.; Liu, J.; Li, G.; Yu, X. Classification of chemical explosion and earthquake infrasound based on 1-D convolutional neural network. *J. Appl. Acoust.* **2021**, *40*, 457–467. [CrossRef]
27. Madankan, Z. Infrasound Source Identification Based on Spectral Moment Features. *Int. J. Intell. Inf. Syst.* **2016**, *5*, 37. [CrossRef]
28. Wu, Y.; Zhao, Z.; Chen, X.; Zou, S. Research on Deep Learning Method of Atmospheric Low Frequency Acoustic Signal Recognition. *J. Univ. Electron. Sci. Technol. China* **2020**, *49*, 758–765.

29. Zhang, Y.; Chen, Q.; Su, P.; Liu, D.; Cui, J.; Chen, J.; Ma, J.; Xing, Q.; Xu, F.; Fan, Y.; et al. Research on the Response Characteristics and Identification of Infrasound Signals in the Jialongcuo Ice Avalanche, Tibet. *Remote Sens.* **2023**, *15*, 4482. [CrossRef]
30. Albert, S.; Linville, L. Benchmarking Current and Emerging Approaches to Infrasound Signal Classification. *Seismol. Res. Lett.* **2020**, *91*, 921–929. [CrossRef]
31. Solomon, M.; Smith, K.; Bryan, K.; Smith, A.O.; Clauter, D.A.; Peter, A.M. Infrasound Threat Classification: A Statistical Comparison of Deep Learning Architectures. In *Proceedings of the Chemical, Biological, Radiological, Nuclear, and Explosives (CBRNE) Sensing XIX*; Fountain, A.W., Guicheteau, J.A., Howle, C.R., Eds.; SPIE: Orlando, FL, USA, 2018; p. 42.
32. Solomon, M.; Billouin, M.; Smith, A.O.; Zeineddine, J.; Chow, K.; Rangarajan, A.; Peter, A.M. Regional Infrasonic and Seismic Event Classification with Machine Learning. In *Proceedings of the Chemical, Biological, Radiological, Nuclear, and Explosives (CBRNE) Sensing XXIII*; Guicheteau, J.A., Howle, C.R., Eds.; SPIE: Orlando, FL, USA, 2022; Volume 12116, p. 28.
33. Leng, X.; Feng, L.; Ou, O.; Du, X.; Liu, D.; Tang, X. Debris Flow Infrasound Recognition Method Based on Improved LeNet-5 Network. *Sustainability* **2022**, *14*, 15925. [CrossRef]
34. Li, H.; Li, X.; Tan, X.; Liu, T.; Zhang, Y.; Niu, C.; Liu, J. Infrasound Event Classification Fusion Model Based on Multiscale SE-CNN and BiLSTM. *Appl. Geophys.* **2024**, *21*, 579–592. [CrossRef]
35. Bishop, J.W.; Blom, P.S.; Webster, J.; Reichard-Flynn, W.; Lin, Y. Deep Learning Categorization of Infrasound Array Data. *J. Acoust. Soc. Am.* **2022**, *152*, 2434–2445. [CrossRef]
36. Zhao, Z. *Classification and Recognition of Variable Duration Infrasound Signals with Small Samples*; University of Chinese Academy of Sciences: Beijing, China, 2023.
37. Zhao, Z.; Cheng, W.; Ji, P.; Teng, P.; Lv, J.; Yang, J. A method for classification of few-shot infrasound signals applying prototype network. *J. Appl. Acoust.* **2024**, *43*, 1193–1202. [CrossRef]