Supplementary Model Overview

This supplementary overview summarizes the neural-network models used for seismic event classification. The models compare sequence-based learning from waveform or frequency-domain inputs with image-based learning from spectrogram representations.

Task and Input Data

The classification task was formulated as a binary seismic event classification problem using three-component waveform data:

Item	Description
Input channels	Vertical, north-south, and east-west waveform components (`Z`, `N`, `E`)
Number of classes	2
Train/test split	70% training, 30% testing
Normalization	Per-event z-score normalization
Sequence handling	Waveforms were trimmed to a common length (21 s = 2100 samples or 5,12 s = 512 samples)
Missing values	Samples containing missing waveform values were removed

Signal Representations

Three input representations were considered across the compared models:

Representation	Used by	Description
Raw multichannel waveform	LSTM-FCN	Three-component events were processed directly as multivariate time series.
FFT magnitude spectrum	LSTM	Waveforms were transformed into the

Representation	Used by	Description
		frequency domain and processed as sequences.
STFT spectrogram	Hybrid CNN-ViT	Waveforms were converted into time-frequency images for convolutional and transformer-based models.

Model Summary

Model	Input type	Main idea	Latent / hidden representation
LSTM	FFT sequence	Learns dependencies across the event sequence using recurrent memory cells.	Final LSTM hidden state
LSTM-FCN	Raw waveform sequence	Combines an LSTM branch with a convolutional branch for multivariate time-series classification.	Concatenated recurrent and convolutional features
Hybrid CNN-ViT	STFT spectrogram	Combines a convolutional stem for local spectrogram features with a pretrained vision transformer for global context.	Vision-transformer CLS token

Main Model Parameters

Model	Main parameters
LSTM	3 LSTM layers, hidden size 32 or 128 depending on input representation, dropout 0.3-0.7, binary feedforward output layer
LSTM-FCN	LSTM hidden size 16, convolutional channels 16, parallel LSTM and convolutional branches, binary feedforward classifier
Hybrid CNN-ViT	ViT-Tiny backbone, convolutional stem, 4 transformer blocks, dropout 0.2, binary classification head

Training Setup

Parameter	Value
Optimizer	AdamW
Learning rate	0.001
Epochs	30
Batch size	32 for the FFT-based LSTM model; 64 for raw-waveform and spectrogram-based models
Loss function	Class-weighted cross-entropy
Evaluation metric	F1 score

Preprocessing and Augmentation

The models used a shared basic waveform preprocessing pipeline: removal of missing values, trimming to a shared sequence length, and z-score normalization. During training, waveform-level augmentation was applied to improve robustness, including small time shifts, amplitude scaling, and additive zero-mean Gaussian noise scaled to a random signal-to-noise ratio between 15 and 40 dB. The FFT-based LSTM model additionally used slight time stretching. Spectrogram-based models used random erasing on the spectrogram representation.

GitHub

https://github.com/natural-hazards/earthquakes-nn