doi:10.26443/seismica.v5i1.1964

SEISMICA

Effectively Distinguishing Blast and Earthquake Sources in Eastern Canada

Justin Chien * [1], **Yajing Liu** [1]

[1]Department of Earth and Planetary Sciences, McGill University, Montréal, QC, Canada

Author contributions: *Conceptualization*: Justin Chien, Yajing Liu. *Methodology*: Justin Chien. *Software*: Justin Chien. *Validation*: Justin Chien. *Formal Analysis*: Justin Chien. *Investigation*: Justin Chien. *Resources*: Yajing Liu. *Writing - Original draft*: Justin Chien. *Writing - Review & Editing*: Yajing Liu. *Visualization*: Justin Chien, Yajing Liu. *Supervision*: Yajing Liu. *Funding acquisition*: Yajing Liu.

Abstract Eastern Canada lies within the stable interior of the North American Plate, yet several regions, including the Western Québec, Charlevoix, Lower St. Lawrence, and Northern Appalachians seismic zones, have experienced damaging intraplate earthquakes. Building reliable earthquake catalogs in these regions is challenging because tectonic earthquakes must be distinguished from industrial blasts, often under low signal-to-noise ratio (SNR) conditions. Here we develop a spectrogram-based source discrimination framework using pretrained convolutional neural network (CNN) image classifiers. Approximately 100,000 three-component waveform records of labeled earthquakes and blasts from the Canadian National Earthquake Database (2000–2024) were converted into standardized three-channel spectrogram inputs, and local event origin time was incorporated as an auxiliary feature. Among the tested models, an EfficientNet-based architecture achieved the best overall performance, yielding the lowest weighted classification cost under the imposed minimum-recall constraint. To account for uneven station coverage, we further developed an event-level framework that combines station-level CNN predictions using reliability-based weights derived from source–station distance and denoised SNR. Applied to an enhanced catalog in the Western Québec Seismic Zone, the framework separated blast-like from earthquake-like events while also identifying noise-dominated detections. This reduces false positives introduced by machine-learning based phase picker and helps produce a more reliable regional earthquake catalog. The proposed framework is efficient, transferable, and well suited for catalog enhancement in Eastern Canada, with clear application potential for other intraplate regions.

Production Editor:
Andrea Llenos
Handling Editor:
Robert Skoumal
Copy & Layout Editor:
Hannah F. Mark

Received:
September 5, 2025
Accepted:
April 27, 2026
Published:
May 9, 2026

1 Introduction

Eastern Canada is located within the intraplate interior of the North American Plate. Seismic activity in this region is primarily driven by regional stress fields and glacial isostatic adjustment, with earthquakes concentrated in areas of crustal weakness (Brooks and Adams, 2020; Ghosh et al., 2019; Reiter et al., 2014; Verdecchia et al., 2022). Each year, the Canadian National Earthquake Database (NEDB) (NRCan, 1985) documents approximately 500 earthquakes across Eastern Canada. This region hosts four major seismic zones: Charlevoix, Lower St. Lawrence (LSL), Northern Appalachians, and the Western Québec Seismic Zone (WQSZ), and includes several major cities, including Québec City, Montréal, Ottawa, Saint John, and Halifax (Fig. 1). Although seismicity in Eastern Canada occurs within a relatively stable intraplate setting compared with plate-boundary regions, historical records include several earthquakes exceeding magnitude 6, including the 1663 Charlevoix earthquake (M_w 7.4–7.9) (Ma and Eaton, 2007; Onwuemeka et al., 2018; Plourde and Nedimović, 2021; Ebel, 2011). Earthquake monitoring in this region is therefore essential for seismic hazard assessment and mitigation.

To improve our understanding of seismicity source characteristics, identify potentially active faults, and produce more accurate seismic hazard assessments, there has been a growing effort to enhance existing earthquake catalogs in many seismogenic regions (Gamboa-Canté et al., 2024; Mancini et al., 2022; Trugman, 2024). A key aspect of catalog enhancement is reliable phase picking, which is necessary to identify P- and S-wave arrivals for smaller-magnitude events that were previously undetected or overlooked. In this context, machine-learning (ML)-based phase pickers have recently gained popularity in seismicity catalog development across numerous study areas (Kubo et al., 2024). After phase picking, accurate phase association is also required to define events based on the consistency between observed and theoretical travel times. However, many seismic events in Eastern Canada occur near areas of intensive mining and quarrying activity (Fig. 1), while detailed information on industrial blasts (e.g., timing and locations) is often proprietary and unavailable on public platforms. In addition, seismic station coverage in Eastern Canada is relatively sparse, with average station spacing of about 50 km, except during the temporary densification associated with the US Transportable Array deployment in 2013–2016. These factors make it challenging to distinguish anthropogenic from tectonic seismic sources accurately.

---

*Corresponding author: shihhan.chien@mail.mcgill.ca

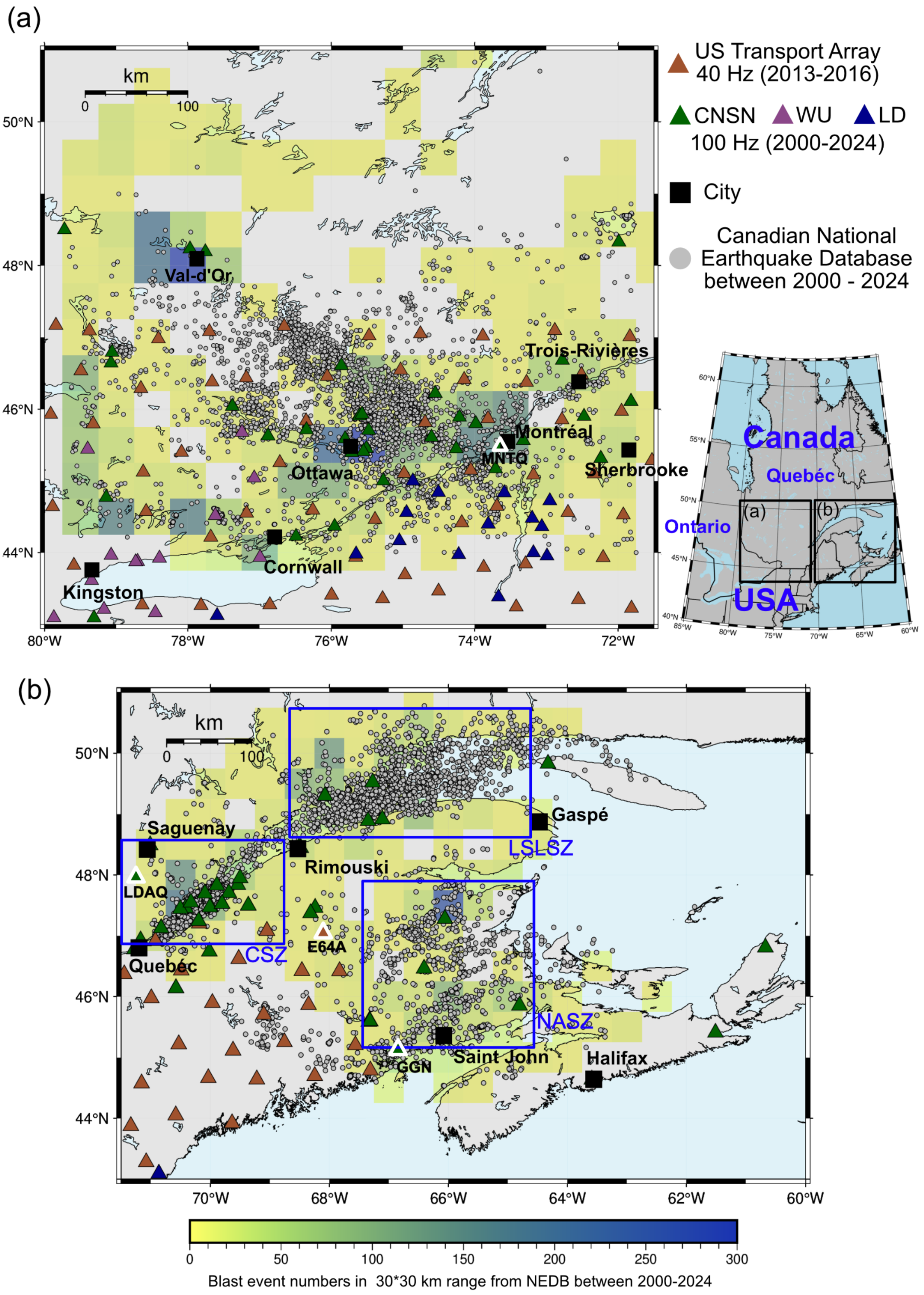


Figure 1 Catalog of earthquakes (gray circles) and blasts (colored grids) recorded in the Canadian National Earthquake Database (NEDB) from 2000 to 2024 in (a) the Western Québec Seismic Zone and (b) the Charlevoix Seismic Zone (CSZ), Northern Appalachians Seismic Zone (NASZ), and Lower St. Lawrence Seismic Zone (LSLSZ). The four highlighted stations—MNTQ, LDAQ, E64A, and GGN—are used for the waveform examples in Fig. 2 and Fig. 3.

As a test in the WQSZ, we applied the widely used ML-based phase picker EqTransformer (Mousavi et al., 2020) to enhance the earthquake catalog (refer to Supplementary Text S1 for the workflow). Our enhanced event catalog revealed an unexpected seasonal variation, with more detected events in the warmer months (Fig. S1). This seasonal pattern initially suggested a possible environmental control on seismicity. However, further investigation indicated that the trend may instead reflect increased industrial blast activity during warmer seasons. Such contamination is plausible because ML phase pickers trained on tectonic earthquakes can also detect quarry blasts when the initial blast onset resembles an earthquake phase arrival. This misidentification may depend on source–station distance and propagation effects, which can modify the apparent waveform characteristics recorded at different stations. For example, we compared the initial ML-enhanced catalog with 578 industrial blast events recorded in the NEDB, between 2020 and 2021, and found that 252 of them matched in origin time and approximate location. Two such examples are shown in Figs. S2 and S3. This indicates that ML pickers trained on earthquake waveforms can sometimes misidentify blast signals. Moreover, most of the newly detected events in our initial enhanced catalog are not listed in the NEDB blast catalog. However, this does not necessarily clear them of industrial blasting sources, because the public NEDB blast catalog is often incomplete due to uncertainties in verifying blast sources. Based on communication with Natural Resources Canada (NRCan) seismologists who maintain the NEDB earthquake and blast catalogs in Eastern Canada, blast events in this region may outnumber earthquakes by approximately 5:1, whereas earthquakes remain the primary focus of routine public cataloging. The public NEDB blast catalog should therefore be considered an analyst-reviewed but incomplete reference for blast events. An effective and efficient classification method is therefore required to distinguish blast sources from tectonic earthquakes, so that the latter can be reliably used in earthquake source-property and seismic hazard studies.

Industrial blast events typically occur within highly localized quarry sites. However, if ML-based phase pickers mistakenly interpret these blasts as earthquakes and fail to identify P- and S-wave arrivals accurately at all stations, the resulting location estimates can be significantly offset because of station-specific uncertainties. In Johnson et al. (2021), mining-induced events located using the original ML phase picker exhibited more diffuse structures and elongated clusters than those obtained using an optimized phase picker retrained on mining-induced seismicity. Unfortunately, no similar retraining procedure currently exists for industrial blasting events because manually labeling a sufficient number of blast phases is extremely time-consuming. As a result, such inaccuracies may produce elongated or spatially diffuse event clusters that could be misidentified as tectonic earthquake sequences. Moreover, if these artificially generated clusters overlap with background seismicity, as is the case in Eastern Canada (Fig. 1), interpretations of seismic hazard and fault structure may be significantly distorted. These considerations further highlight the need for a robust method to distinguish anthropogenic blasts from natural earthquakes, particularly in regions where industrial activity overlaps with natural seismicity.

Previous studies have applied relative P- and S-phase amplitudes to distinguish different seismic sources (Kim et al., 1993; Yavuz et al., 2019). In theory, explosion sources are not expected to generate significant shear-wave energy. However, as highlighted by a recent synthetic waveform modeling study (Zhang, 2023), isotropic sources can also generate S-wave energy comparable to, or even larger than, P-wave energy because of shallow seismic velocity heterogeneities and/or anisotropy. In many of the NEDB-labeled blast waveforms we examined, larger S-phase amplitudes were clearly observed (Fig. 2(d–f) and Figs. S2 and S3). The relatively large source–station distances involved, averaging $\sim$70 km (see Fig. S4 for additional details), suggest that complex velocity structure along the ray paths may contribute to the S phase amplification.

Event origin time may provide an additional constraint, as industrial blasting is generally expected to occur during local working hours. However, our analysis of NEDB event times shows that this distinction is also imperfect. Earthquakes exhibit a diurnal detectability pattern, with fewer detections during working hours because of elevated daytime cultural noise (Fig. S5a), whereas blast events occur predominantly during standard working hours (local time 8 AM–5 PM; Fig. S5b). Nevertheless, a substantial number of blast events also occur during evening or near-midnight hours, possibly reflecting nighttime or otherwise unregulated blasting activity. These results suggest that source type in this region cannot be robustly distinguished using origin time alone, nor by relying on location estimates or simple P- and S-wave amplitude characteristics in isolation.

Time-frequency information of tectonic earthquakes and NEDB-labeled blast events exhibit distinct patterns. Figure 2 compares spectrograms of earthquakes and blasts with similar magnitudes recorded on different components at different stations with variable sampling rates. For tectonic earthquakes, the S-wave arrival appears as a vertical band spanning the full frequency range up to the Nyquist frequency (Fig. 2(a–c)). In contrast, industrial blast events often show S-wave energy extended over time, forming more horizontal bands (Fig. 2(d–f)).

These spectrogram features may provide a basis for distinguishing between the two event types. For example, Hourcade et al. (2023) developed a custom convolutional neural network (CNN) to differentiate blasts from earthquakes. However, direct application of their pretrained model to eastern Canada, without additional retraining on region-specific data, yielded classification accuracy below 90% on pre-labeled events. One key limitation is that substantially improved performance was only achieved for events with signal-to-noise ratios (SNRs) greater than 3 dB in their study (Hourcade et al., 2023). In Eastern Canada, such conditions are often not met because sparse station coverage and larger source–station distances reduce waveform quality. Consequently, the SNR dependence of the model may limit its applicability in this region. Sampling-rate differences may further reduce performance. The original CNN was trained on 100 Hz data and requires customized $237 \times 50 \times 3$ input spectrograms (time × frequency × three components), whereas many stations in Eastern Canada recorded at 40 Hz. Although resampling is possible, it may introduce artifacts and negatively affect both classification accuracy and model stability. In addition, the spectrogram parameterization, including frequency resolution, may not be optimal for capturing subtle source-dependent patterns in Eastern Canada data. Other CNN-based source discrimination approaches (Kasburg et al., 2024; Maguire et al., 2024; Zhu et al., 2024) have also been proposed, but these models were generally trained on dense seismic arrays in regions where blast activity rarely overlaps with natural seismicity, limiting their transferability to Eastern Canada.

In light of these challenges, this study develops a seismic source discrimination framework tailored to the sparse station coverage and complex intraplate tectonic setting of Eastern Canada. We first compile and preprocess a labeled dataset of tectonic earthquakes and industrial blasts from the NEDB. We then compare waveform and spectrogram characteristics of the two source types and evaluate several CNN-based classification models. Next, we optimize the preferred model and apply it to the initial ML-enhanced catalog in the WQSZ to separate likely blast events from tectonic earthquakes. Finally, we examine how the refined catalog affects the interpretation of seismicity patterns and discuss the broader applicability of the approach to other regions of Eastern Canada.

2 Data and Methods

Many CNN-based image classification models are developed and benchmarked on ImageNet, a large-scale dataset containing more than one million images distributed across 1,000 classes. Well-known architectures such as VGG16 (Qassim et al., 2018), ResNet (Targ et al., 2016), EfficientNet (Koonce and Koonce, 2021), and ConvNeXt (Woo et al., 2023) achieve excellent performance and are widely applied in fields including bioinformatics, medical diagnosis, and space exploration (Ghaffari and Savaria, 2020; Hostallero et al., 2022). These pretrained CNN architectures can be readily implemented in Python using the PyTorch machine-learning library (Paszke et al., 2019) and fine-tuned for new applications.

Given the distinctive spectrogram patterns of industrial blasts and tectonic earthquakes (Fig. 2), we apply a suite of CNN-based classifiers to distinguish between the two source types. The tested models include widely used architectures such as VGG16, ResNet, EfficientNet, and ConvNeXt, as well as several recently developed seismic discrimination models (Ertuncay et al., 2024; Rakotoarisoa and Razafindrakoto, 2025; Hourcade et al., 2023). For consistency, all models use the same three-component spectrogram representation, which allows direct comparison between general-purpose image CNN backbones and seismic-specific architectures. We treat the three-component spectrograms for each event (North, East, and Vertical) as a three-dimensional array with dimensions $time \times frequency \times 3$, analogous to the $height \times width \times 3$ structure of standard RGB images. In the following subsections, we describe how waveform data are preprocessed to match the input format required by CNN image classification models and outline the modifications introduced to improve performance for source discrimination in Eastern Canada.

2.1 Data Acquisition

We select pre-labeled industrial blast and earthquake events from the NEDB public catalog in four seismic zones of Eastern Canada, namely the Northern Appalachians, Charlevoix, LSL, and WQSZ, spanning 2000 to 2024. In addition, we used a non-public internal NEDB blast catalog covering October 2021 to April 2024 that was shared by NRCan seismologists. Only events classified as confirmed blasts were included, whereas entries labeled as suspect, which constitute roughly half of the internal catalog, were not used. For each event, we download all available three-component waveform data from stations within a 1.5° source–station radius. The seismic stations are primarily operated by the U.S. Transportable Array (40 Hz; 2013–2016) and three 100 Hz networks: the Canadian National Seismograph Network (CNSN), the Southern Ontario Seismic Network (WU), and the Lamont–Doherty Cooperative Seismographic Network (LD), which have been active since 2000. Each waveform record includes 30 s before the event origin time and 90 s after.

2.2 Denoising and Spectrogram Enhancement

To prepare the waveform data for spectrogram-based classification, we first standardize all usable records to a common sampling rate of 100 Hz. Continuous traces are merged when gaps are present, and the instrument response is removed to obtain displacement waveforms. The resulting waveforms are then detrended, tapered, and high-pass filtered at 1 Hz to suppress long-period noise and baseline variations. These preprocessing steps stabilize the time–frequency representation and improve the visibility of higher-frequency features relevant to blast and earthquake discrimination.

We next use the DeepDenoiser model of Zhu et al. (2019) to improve SNR. DeepDenoiser, which is based on a deep neural network, effectively attenuates seismic noise even when signal and noise overlap in the same frequency band, while preserving waveform shape under high-noise conditions. As demonstrated by Woollam et al. (2022), this approach is now widely used in catalog enhancement workflows. Although DeepDenoiser is trained primarily on earthquake signals, the results presented below indicate that it also performs well for blast signals.

Following instrument-response removal, filtering, and denoising, we quantify waveform quality using a three-component average SNR. For each station, the signal window extends from 1 s before the estimated P-wave arrival to 10 s after the estimated S-wave arrival.

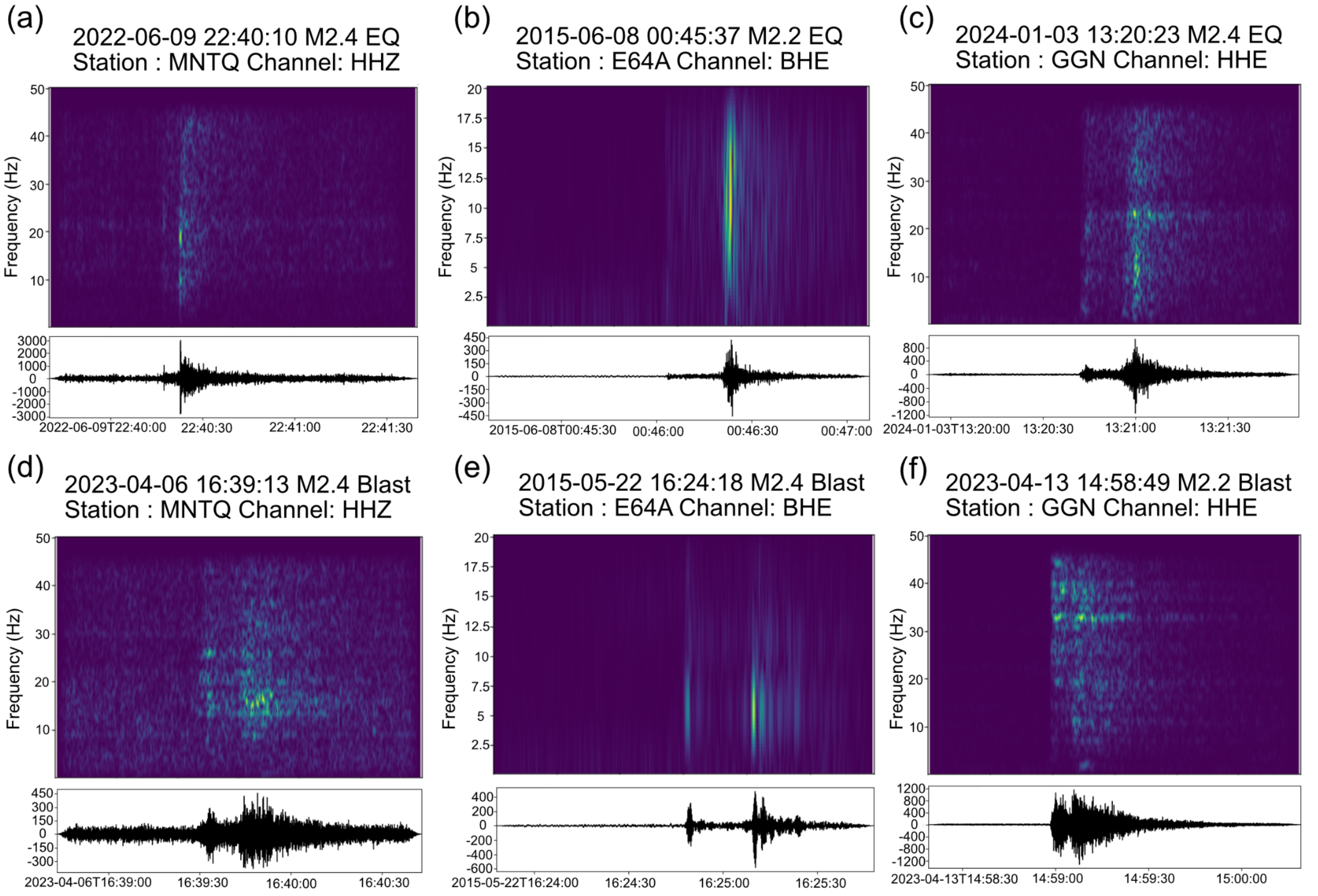


Figure 2 Spectrograms and raw waveforms for earthquakes (a–c) and industrial blasts (d–f). Earthquake and blast event pairs with similar magnitudes of M $\sim$ 2.2–2.4 are show for (a) and (d): vertical component (100 Hz) at station MNTQ in the Western Québec Seismic Zone; (b) and (e) east component (40 Hz) at E64A in the Lower St. Lawrence Seismic Zone; and (c) and (f): east component (100 Hz) at GGN in the Northern Appalachians Seismic Zone.

Arrival times are approximated from source–station distance using average crustal velocities of 6.5 km/s for P waves and 3.7 km/s for S waves, consistent with regional velocity characteristics in Eastern Canada (Kao et al., 2012). Noise is estimated from the remaining parts of the waveform outside the signal window, excluding the first and last 10 s of the trace. The SNR is then calculated as the variance ratio between the signal and noise windows averaged over the three components.

The purpose of these preprocessing steps is to enhance the visibility and separability of diagnostic time–frequency features in the spectrogram representation used for classification, rather than to preserve exact waveform amplitudes or detailed coda characteristics.

Figure 3 illustrates the effect of these preprocessing steps by comparing the original, high-pass-filtered, and denoised waveforms and their corresponding spectrograms for an M_L 1.1 earthquake and an M_L 1.6 industrial blast in the Charlevoix Seismic Zone, both recorded at a distance of approximately 50 km. Both the SNR and the clarity of time–frequency features are substantially improved after preprocessing. In some low-SNR cases, denoising attenuates late coda amplitudes; however, this does not affect the classification results because the models rely primarily on overall time–frequency image patterns rather than on absolute amplitude information.

We further evaluate denoising performance for original waveforms with SNR > 1 from both earthquakes and blasts. Overall, SNR improves substantially for both event types (Fig. S6). In some cases, however, denoising appears to amplify rather than suppress noise, particularly near the beginning of the waveform, which leads to low or even negative SNR values after denoising. This behavior likely reflects the use of the pretrained DeepDenoiser model without additional retraining for the specific noise conditions and waveform characteristics of earthquakes and industrial blasts in our study region. In other words, the learned noise representation does not fully match our dataset in all cases. Denoised waveforms with negative SNR values are excluded from further analysis, and we adopt a relatively low SNR threshold of 1.5 (approximately 1.8 dB) to retain as many informative signals as possible while maintaining waveform quality. Among records with original SNR > 1, 86.38% of earthquake records and 79.23% of blast records reach denoised SNR > 1.5 after denoising, yielding average denoised SNR values of 13.952 and 11.037, respectively.

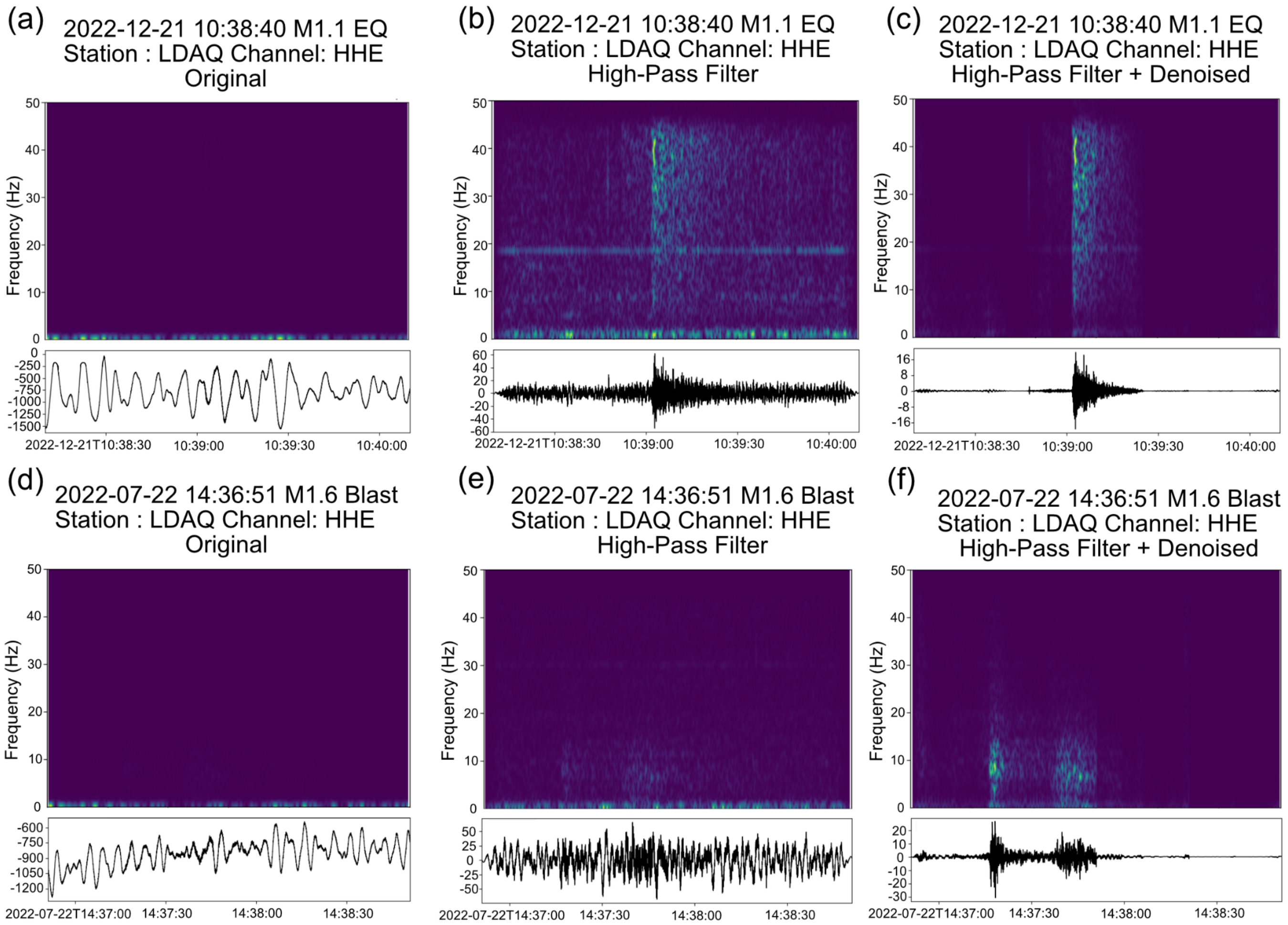


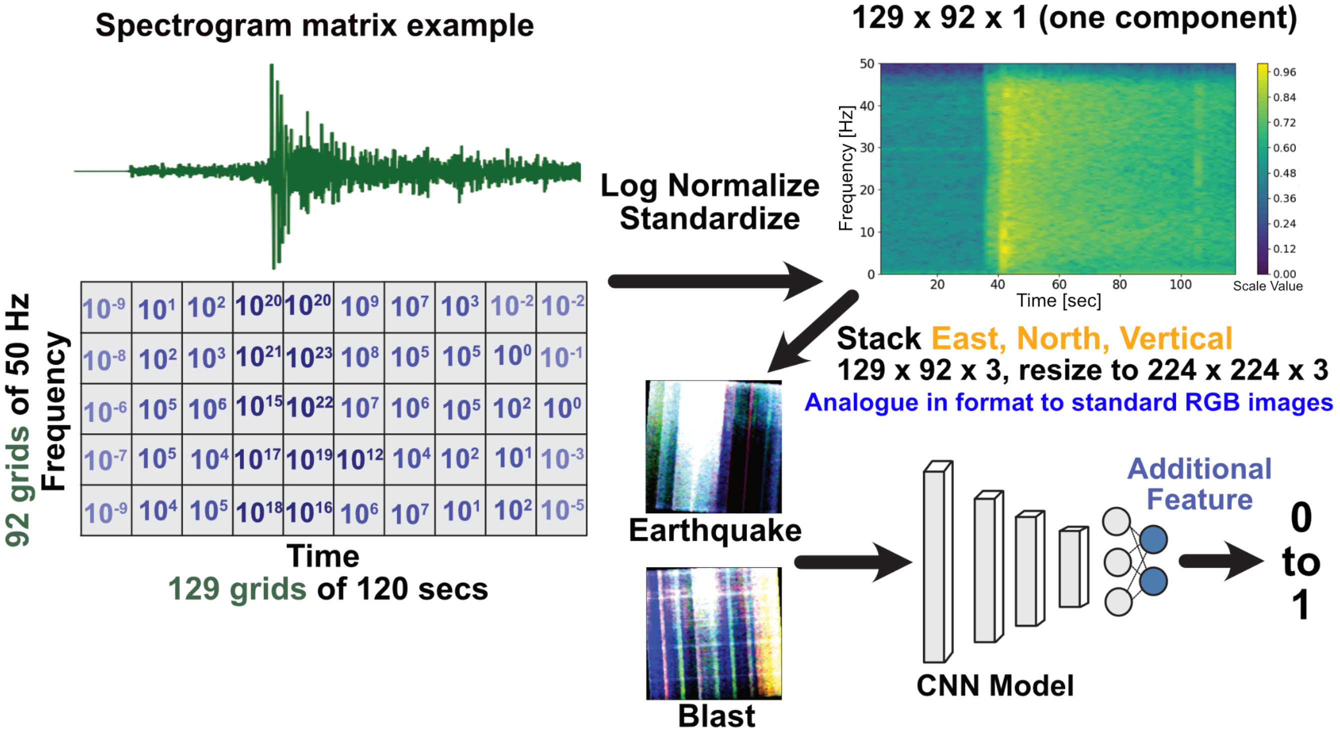
Figure 3 Spectrograms and waveforms in preprocessing steps for a pair of earthquake (a–c; M $\sim$ 1.1) and industrial blast event (d–f; M $\sim$ 1.6) recorded at station LDAQ (east component). (a) and (d): raw waveforms and corresponding spectrograms. (b) and (e) 1 Hz high-pass filtered. (c) and (f) 1 Hz high-pass filtered and denoised.

2.3 From Spectrogram Matrix to CNN Model

Figure 4 illustrates the workflow from denoised waveform to CNN-based source discrimination. We first select three-component waveforms with an average SNR exceeding 1.5. Spectrograms are then generated from 120 s waveform windows using a window length of 256 samples and 128-sample overlap, corresponding to a hop length of 128 samples. This yields a temporal resolution of 3.2 s for 40 Hz data and 1.28 s for 100 Hz data. The corresponding Nyquist frequencies are 20 Hz and 50 Hz, and the associated frequency resolutions are 0.156 Hz and 0.391 Hz, respectively. The resulting spectrograms contain 129 time bins and 92 frequency bins. Near the event origin time, power spectral density values may range from 10^{10} to 10^{20} m^2/Hz, whereas in quieter time windows they may decrease to 10^{-20} m^2/Hz. To accommodate this large dynamic range, we apply logarithmic normalization followed by scaling to the $[0, 1]$ interval. The normalized spectrograms of the North, East, and Vertical components are then combined into a three-channel array of size $129 \times 92 \times 3$, in which the three channels represent the three seismic waveform components rather than RGB color channels derived from a display colormap. Thus, the CNN learns from standardized spectrogram values of the three seismic components, not from artificial color contrasts. The navy-to-yellow color scheme is used only for visualization in the conceptual workflow shown in Fig. 4 and is not part of the model input.

Although the dataset includes two sampling rates (40 Hz and 100 Hz), we do not treat their spectrograms separately. To ensure consistent CNN input dimensions, the 40 Hz waveforms are resampled to 100 Hz before spectrogram generation. This preprocessing step does not recover additional physical information above the original Nyquist frequency of 20 Hz; consequently, the 20–50 Hz portion of the resampled 40 Hz spectrograms contains limited meaningful signal. However, our classification strategy is based primarily on overall spectrogram morphology rather than on artificial high-frequency content. In particular, earthquake records generally retain their vertically concentrated spectral patterns, whereas blast records preserve their more horizontally extended time–frequency characteristics, even after resampling (Fig. S7). Each three-channel spectrogram is then resized to $224 \times 224 \times 3$ to provide a uniform input size for the CNN models. Finally, we apply channel-wise normalization following the standard ImageNet-based initialization protocol to ensure consistent input distributions during training.

To evaluate whether event origin time provides addi-



10^{-9}	10^{1}	10^{2}	10^{20}	10^{20}	10^{9}	10^{7}	10^{3}	10^{-2}	10^{-2}
10^{-8}	10^{2}	10^{3}	10^{21}	10^{23}	10^{8}	10^{5}	10^{5}	10^{0}	10^{-1}
10^{-6}	10^{5}	10^{6}	10^{15}	10^{22}	10^{7}	10^{6}	10^{5}	10^{2}	10^{0}
10^{-7}	10^{5}	10^{4}	10^{17}	10^{19}	10^{12}	10^{4}	10^{2}	10^{1}	10^{-3}
10^{-9}	10^{4}	10^{5}	10^{18}	10^{16}	10^{6}	10^{7}	10^{1}	10^{2}	10^{-5}

Figure 4 Conceptual workflow from spectrogram matrices derived from denoised three-component waveforms to a standardized $224 \times 224 \times 3$ CNN input array. The three channels correspond to the North, East, and Vertical components and are arranged in a format analogous to a conventional RGB image.

tional discriminatory information, we test all CNN models both with and without an added event-time feature. Specifically, we modify the final fully connected layer of each CNN architecture to incorporate the event origin time expressed as local hour of day (see Text S2 for an example using ResNet; Targ et al., 2016). This modification is motivated by the temporal distribution of NEDB cataloged blast events preferentially during local working hours, although a non-negligible number of them also occurred outside this interval.

Because hour of day is a cyclic variable, we do not use the raw hour value directly. For example, 23:00 and 00:00 are close in time, but their numerical values differ by 23 if treated linearly. To preserve this cyclic relationship, we transform event origin time into two features using sine and cosine functions over a 24-hour period. This representation maps daily time onto the unit circle and allows the model to learn temporal patterns without introducing an artificial discontinuity at midnight.

The final spectrogram dataset contains 105,443 samples and is divided into training, validation, and testing subsets containing 70%, 15%, and 15% of the data, respectively. This corresponds to 73,809 training samples, 15,816 validation samples, and 15,818 testing samples. The training set is used to optimize model parameters, the validation set is used to monitor model performance during training and guide hyperparameter tuning, and the testing set is reserved for final evaluation. Class proportions remain nearly constant across all three subsets, with earthquakes and blasts accounting for approximately 48.7% and 51.3% of the samples, respectively. In the training set, this corresponds to 35,934 earthquake samples and 37,875 blast samples. To improve generalization and reduce overfitting, we apply data augmentation to the training set using random horizontal flips and small rotations between 1° and 10°. Because the CNN models operate on spectrogram images rather than raw waveform amplitudes, these geometric transformations act as regularization for image-based classification rather than as physically meaningful modifications of the seismic signal.

Throughout this study, earthquakes are treated as the positive class and blasts as the negative class. Accordingly, a true positive (TP) denotes an earthquake correctly classified as an earthquake, a true negative (TN) denotes a blast correctly classified as a blast, a false positive (FP) denotes a blast incorrectly classified as an earthquake, and a false negative (FN) denotes an earthquake incorrectly classified as a blast.

Model performance during training was evaluated using the receiver operating characteristic (ROC) curve and its associated area under the curve (AUC). The ROC curve plots the true positive rate (TPR), also referred to as recall ($\mathrm{TPR} = \mathrm{Recall} = TP/(TP + FN)$), against the false positive rate ($\mathrm{FPR} = FP/(FP+TN)$) as the classification threshold varies. The AUC provides a threshold-independent summary of discriminative ability: an AUC of 1 indicates perfect separation between earthquakes and blasts, whereas an AUC of 0.5 indicates performance no better than random guessing. In this study, validation AUC was used to monitor model performance during training and to determine early stopping.

For evaluation, precision is defined as $\mathrm{P} = TP/(TP + FP)$, and the F1 score, $\mathrm{F1} = 2PR/(P + R)$, is used to summarize the balance between precision and recall.

3 Results

In the preliminary model-screening stage, all candidate CNN architectures were trained using the Adam optimizer (Kingma and Ba, 2014) with a batch size of 64. Four learning rates, 10^{-2}, 10^{-3}, 10^{-4}, and 10^{-5}, were tested for both the custom seismic CNN models and the ImageNet-based torchvision models. The candidate models included two recently developed seismic source-discrimination architectures (Ertuncay et al., 2024; Rakotoarisoa and Razafindrakoto, 2025), one of the first CNN architectures specifically developed for blast–earthquake classification using spectrogram inputs (Hourcade et al., 2023), and five widely used image-classification backbones: ResNet18, ResNet50, VGG16, EfficientNet, and ConvNeXt (Qassim et al., 2018; Targ et al., 2016; Koonce and Koonce, 2021; Woo et al., 2023). Each model was trained for up to 15 epochs, with early stopping based on validation AUC and a patience of 3 epochs.

For binary classification, all models used a sigmoid output layer and binary cross-entropy loss (Ruby and Yendapalli, 2020). To assess whether temporal information further improves source discrimination, each model was evaluated under two input configurations: spectrograms only, and spectrograms augmented with the event origin-time feature encoded as sine and cosine components of the local hour. This initial screening was used to identify the model configurations selected for further optimization and robustness analysis.

3.1 Baseline Model Comparison

Figure 5 summarizes the baseline performance of the candidate models for experiments conducted without the encoded origin-time feature (left column) and with the encoded origin-time feature (right column). For each experiment, test-set performance is summarized using the F1 score together with the numbers of FP and FN. We use the F1 score as the primary summary metric in this comparison because it balances precision and recall and is therefore more informative than overall accuracy for this binary classification problem. The FP and FN counts are also reported explicitly because they reflect the classification trade-off most relevant to catalog enhancement. Complete performance metrics for all baseline models and learning-rate settings, including accuracy, AUC, precision, recall, F1 score, FP, and FN, are provided in the Supplementary Material.

Overall, the baseline experiments show that including the origin-time feature improves performance across most model architectures. Excluding results that performed poorly because of unsuitable learning rates, the addition of origin-time information generally increases the F1 score by approximately 1–3% (Fig. 5a–b). The best overall results are obtained for learning rates between 10^{-3} and 10^{-5}, and within this range the ImageNet-based image-classification models generally outperform the seismic-specific CNN architectures.

The improvement associated with the origin-time feature arises mainly from a reduction in FP. As shown in Fig. 5c–d, FP counts decrease substantially, typically by about 50–300 cases, for models trained with learning rates between 10^{-3} and 10^{-5}. FN counts (Fig. 5e–f) also decrease in most cases, although the improvement is generally smaller and less systematic than for FP. Taken together, these results indicate that origin-time information provides useful additional context, primarily by helping the models reject blast-like events that might otherwise be misclassified as earthquakes.

3.2 Model Optimization

Based on the baseline comparison, only the best-performing pretrained image-classification CNNs augmented with the origin-time feature were retained for further optimization. These candidate models were trained for up to 20 epochs using the AdamW optimizer (Loshchilov et al., 2017), with a batch size of 64, a weight decay of 10^{-4}, and architecture-specific dropout for regularization. Early stopping continued to rely on validation AUC, consistent with the baseline experiments, but the patience was extended to five epochs to promote more stable convergence. The validation set was used for early stopping, threshold tuning, and model selection, whereas the test set was reserved exclusively for final performance assessment to avoid test-set leakage.

The final decision threshold for each model was selected on the validation set by minimizing a weighted cost function of the form $1.25 \times FP + 1.0 \times FN$, subject to a minimum recall constraint of 0.97. Only thresholds yielding at least 97% recall for true earthquakes were considered eligible. Among these, the threshold with the lowest weighted cost was selected. The slightly larger penalty assigned to FP reflects their greater practical consequence in this application, because blasts misclassified as earthquakes can contaminate the enhanced catalog and bias subsequent interpretations of seismicity patterns and fault structure. At the same time, FN were retained in the cost function to avoid reducing FP at the expense of an excessive number of missed earthquakes.

Using this validation-based selection criterion, the three best-performing configurations were EfficientNet+hour, ResNet18+hour, and ConvNeXt+hour, all with a learning rate of 5×10^{-4}. Among these, EfficientNet+hour produced the lowest validation cost. To evaluate robustness and reduce the influence of random initialization and stochastic training variability, these three candidate models were retrained using five independent random seeds and compared on the basis of mean test performance. This robustness analysis ensured that final model selection was not determined by a particularly favorable single run, but instead reflected stable behavior across repeated trials. EfficientNet+hour achieved the lowest mean test cost, 580.45 ± 28.52, together with a mean test recall of 0.9719 ± 0.0009 and a mean test AUC of 0.9920 ± 0.0005, and was therefore selected as the final model configuration. Detailed test results for the model optimization stage and the subsequent five-seed analysis are provided in the Sup-

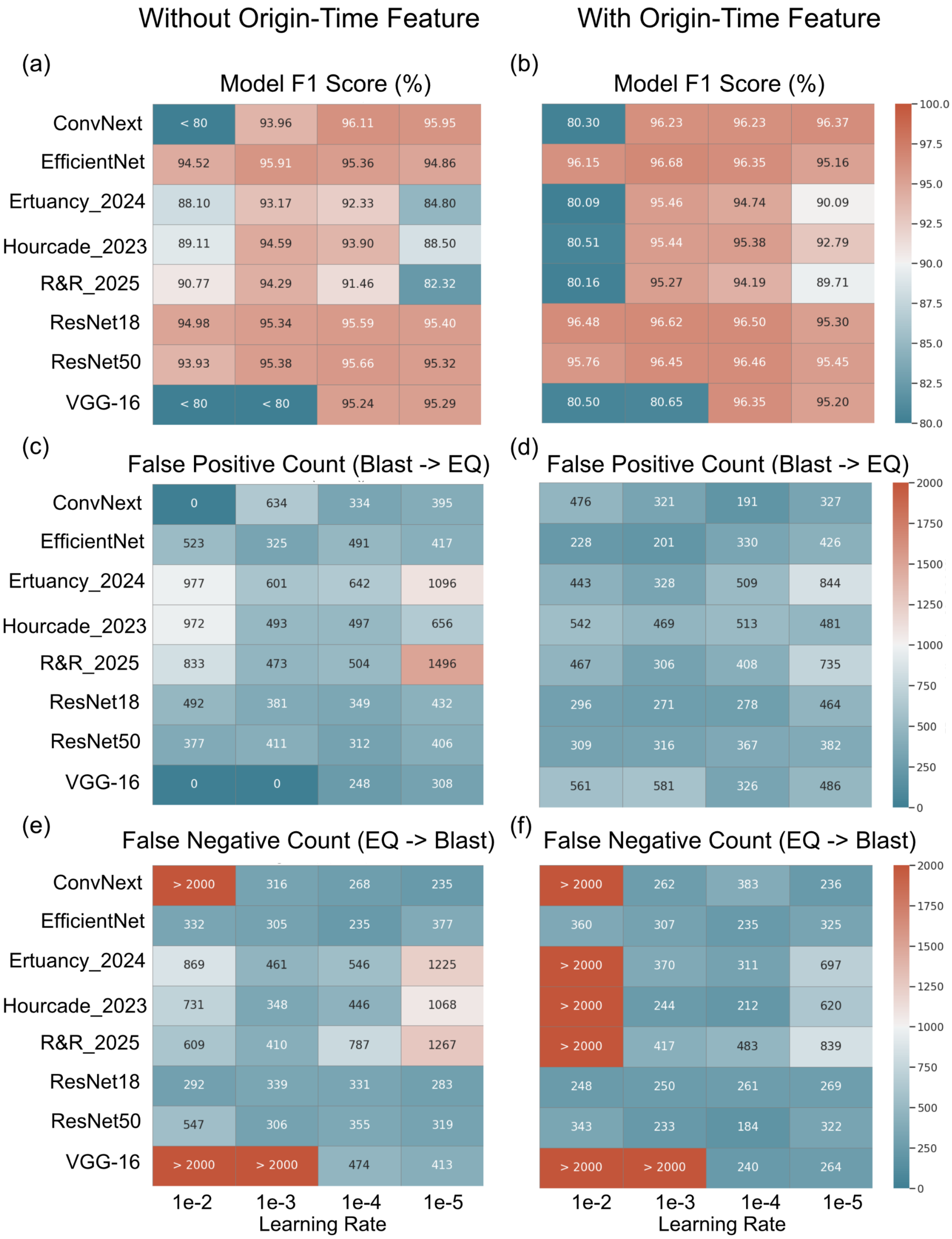


Figure 5 (a–b) F1 scores on the test dataset (∼15,800 spectrogram images), (c–d) false-positive (FP) counts, and (e–f) false-negative (FN) counts, evaluated across learning rates from 10^{-2} to 10^{-5} for baseline models without the origin-time feature (left column) and with the origin-time feature (right column). Ertuncay_2024, Hourcade_2023, and R&R_2025 denote the neural-network structures proposed by Ertuncay et al. (2024), Hourcade et al. (2023), and Rakotoarisoa and Razafindrakoto (2025), respectively.

plementary Material.

3.3 Classification Framework

The optimization stage described above addresses the first of two related questions: how well can a CNN classify earthquake/blast using an individual spectrogram from a single station? In practice, however, seismic events are commonly recorded by multiple stations, and the final source label should therefore be assigned at the event level rather than from a single station record. The classification framework addresses this second question by combining information from all available stations in a quality-informed manner.

To avoid information leakage between subsets, the training, validation, and test data were separated at the event level rather than at the station level. Thus, all qualified station spectrograms associated with a given event were assigned to the same subset, ensuring that no event contributed records to more than one of the training, validation, or test sets. The full dataset consists of 105,443 station-level records from 17,333 events and remains moderately balanced, comprising 9,268 blast events and 8,065 earthquake events. The final event-based split included 12,133 training events, 2,600 validation events, and 2,600 test events, corresponding to 73,926, 16,101, and 15,416 station-level records, respectively. This corresponds to approximately 70% for training, 15% for validation, and 15% for test. We have verified there is zero event overlap among the training, validation, and test subsets.

In the first stage of the framework, the EfficientNet+hour architecture selected during model optimization was retrained and applied independently to each qualified station spectrogram within the event-based partitions. For each station record, the CNN produced a probability between 0 and 1 representing the likelihood that the waveform was generated by an earthquake rather than a blast. A value close to 1 indicates stronger similarity to earthquake examples in the training set, whereas a value close to 0 indicates stronger similarity to blast examples. This quantity is referred to here as the station-level earthquake probability.

Because station records from the same event are not equally informative, the final event classification was not based on a simple majority vote. Instead, station-level probabilities were combined using a reliability-based weighting scheme, motivated by the expectation that predictions from nearby, high-SNR stations are generally more trustworthy than those from distant or noisy stations. To quantify this effect, the reliability of each station-level CNN prediction was modeled as a function of source–station distance and denoised SNR.

Specifically, a logistic regression model was trained using only the training subset. For each station record, the station-level CNN probability was first converted to a binary class label using a threshold of 0.5, and the predicted label was then compared with the known true label to determine whether the station-level classification was correct. Importantly, this logistic regression model does not predict whether a waveform is an earthquake or a blast. Instead, it estimates the probability that the station-level CNN prediction is correct, conditional on source–station distance D and denoised SNR. The station-specific reliability weight is therefore defined as $w_i = P(\mathrm{correct} \mid \mathrm{D}_i, \mathrm{SNR}_i)$, where larger values of w_i indicate greater expected reliability of the CNN prediction at station i (Fig. S8). For practical application, the resulting reliability surface was discretized on a regular grid spanning D from 0 to 160 km in 5-km increments and denoised SNR values from 1 to 30 dB in 1-dB increments; w_i values were then assigned by lookup on this grid using the nearest distance and SNR values.

The final event-level earthquake probability was computed as the weighted average, $p_j = \sum_i w_i p_i \,/\, \sum_i w_i$, where p_i is the CNN-predicted earthquake probability for station i and w_i is the corresponding reliability weight; j is event index. This formulation gives greater influence to higher-quality station records while down-weighting noisier or less reliable observations. Importantly, this weighting step does not retrain the CNN itself, but instead acts as a post-classification reliability filter applied to the station-level outputs.

To convert the continuous event-level probability into a final class label, an event-level decision threshold was selected separately for each seed using the event-based validation set. This threshold was chosen by minimizing the same weighted cost function used during model optimization, again subject to the minimum recall constraint of $R = 0.97$. Consequently, the optimal event-level threshold was not fixed at the default value of 0.5, but instead varied among seeds and was generally higher, with a mean value of 0.621 and a range of 0.536–0.700 across the 10 experiments. Events with prediction values exceeding this threshold were classified as earthquakes, whereas those below it were classified as blasts. The selected threshold was then applied once to the independent event-based test set for final evaluation. To assess robustness with respect to stochastic model training, the complete classification workflow, including CNN training, station-level prediction, reliability weighting, and event-level threshold selection, was repeated over 10 independent random seeds using the same fixed event-based training, validation, and test partitions. Application of the final EfficientNet+hour framework yielded event-level classifications for 2600 test events, with mean FP and FN counts of 36.1 and 34.1, respectively, across the 10-seed experiments. Full results for all random seeds are provided in the Supplementary Material.

4 Application

Using a catalog-enhancement workflow that combines ML-based phase picking with phase association, we first constructed an initial ML-enhanced catalog containing ~3,600 events in the WQSZ between January 2020 and December 2022. We begin the enhancement in 2020 because CNSN station coverage in the WQSZ increased from 14 to 22 stations at that time. Each event in the initial enhanced catalog was required to contain at least eight total phase picks, including both P- and S-phase picks from at least three stations, thereby reducing the likelihood of false associations. The resulting catalog contains approximately eight times more events than the NEDB earthquake catalog over the same interval. However, it still includes a mixture of earthquakes, blasts, and false events (noise) associated with erroneous ML phase picks and phase associations (see Figs. S2 and S3 for examples of enhanced-catalog events corresponding to NEDB-documented blasts).

Following the same procedure used to construct the training dataset, station-level classification was applied only to waveforms with denoised SNR $>$ 1.5 (equivalent to $\sim$1.8 dB) and epicentral distance within $1.5°$. Event-level probabilities were then calculated as reliability-weighted averages of the qualified station-level probabilities. Application of the final optimized EfficientNet model to the initial enhanced WQSZ catalog yielded a strong bimodal distribution of event-level prediction values (Fig. S9). Specifically, 2806 events fall between 0.0 and 0.1 and 468 events fall between 0.9 and 1.0, so that these two end-member clusters together account for 91.3% of all event predictions. These clusters likely correspond to highly convincing blast-like and earthquake-like signals, respectively. The blast/earthquake ratio is also close to the NRCan internal catalog estimate of 5 to 1.

Because the enhanced catalog includes false associations and noise-dominated waveforms, intermediate event-level prediction values should not be interpreted solely as evidence of ambiguous source type. In many cases, they instead reflect low-confidence classification caused by weak or inconsistent station-level spectrogram features, erroneous phase picking, incorrect phase association, or noisy waveforms.

To better understand the full prediction range, we randomly selected 20 events from each 0.1-wide prediction bin between 0 and 1 and visually examined their waveforms, phase picks, travel-time ray paths, and the number of stations whose spectrograms exceeded the denoised SNR threshold within a $1.5°$ radius. This inspection showed that events with fewer than three qualified station records are most commonly associated with noise waveforms (Figs. S10–S11). By contrast, events with at least three qualified stations generally show clearer and more internally consistent travel-time ray-path patterns (Figs. S12–S13), supporting more confident event-level classification. Accordingly, events with only one or two qualified stations were assigned to the noise category by default. Exceptions were made only when either the original event contained at least 10 ML-picked phases, counting P and S separately, or the event-level prediction was extreme, with values > 0.975 or < 0.025 (Figs. S14–S17). Such exceptions likely correspond to small-magnitude events recorded by only a few stations, events occurring in areas of limited station coverage where location uncertainty affects SNR estimation, or cases in which multiple nearby events occur within the same 120-s time window.

Out of the $\sim$3,600 detected events, 2,862 were classified as blasts, 512 as tectonic earthquakes, and the remainder as noise-dominated or likely false detections. By comparison, the NEDB catalog for the WQSZ contains 1070 blasts and 504 earthquakes over the same time interval. Figure 6 shows the month-by-month distribution of these event classes from 2020 to 2022, demonstrating that the ML-enhanced catalog contains substantially more blast detections than the NEDB, while also identifying additional tectonic earthquakes and a distinct set of noise. The figure also indicates that temporal variations in the enhanced catalog are largely controlled by blast activity, underscoring the importance of source discrimination prior to tectonic interpretation.

In particular, Figs. 7a and 8a show the spatial distribution of the CNN-classified events together with the corresponding NEDB entries in the Montréal region and Ottawa Valley, respectively, overlaid on mapped geological faults (Lamontagne et al., 2020). Without removing blast events, the enhanced catalog could lead to misleading tectonic interpretations. For example, apparent clusters north of Montréal Island near the November 1983 M3.4 event (Fig. 7a) might be misinterpreted as newly activated fault structures. Similarly, alignments of newly detected events east and south of Ottawa (Fig. 8a) could be mistaken for evidence of active faulting. The CNN discriminator therefore plays a critical role in filtering blast-like and noise-related detections and in preventing erroneous tectonic interpretations of the ML-enhanced catalog.

We also applied the same enhancement and classification workflow to the Charlevoix and Northern Appalachian seismic zones during three-month summer intervals in two different years, when industrial activity was highest based on NEDB-labeled blast occurrence. In Charlevoix (Fig. 9), the workflow identified 130 high-confidence earthquakes, including all of the 32 listed in the NEDB catalog, and about 120 additional blasts. However, the locations of some newly classified blast events should be interpreted cautiously, as several appear within the river and likely reflect location uncertainty rather than true blast occurrence there. In contrast, the Northern Appalachian region (Fig. 10) yielded only a small number of newly classified events together with a comparable number of missed detections, likely owing to sparse and uneven station coverage combined with lower levels of both seismic and industrial activity. These results indicate that station density and azimuthal coverage strongly influence both detection sensitivity and classification confidence.

Panels b of Figures 7–10 also show strongly bimodal distributions of event-level prediction values, indicating stable model performance across the tested regions. Moreover, panels c of Figures 7–10 present Venn diagrams comparing the ML-enhanced and NEDB catalogs across the four tested regions. Although many new events were identified in the Ottawa and Montréal sectors of the WQSZ, several NEDB-documented earthquakes were not recovered, largely because of sparse station coverage and our event-detection criteria requiring at least eight total phase picks, including both P- and S-phase picks from at least three stations. Most of these missed detections are concentrated in the northeastern part of Ottawa, where station coverage is limited. A complete breakdown of ML-enhanced and NEDB event counts for the WQSZ regions is provided in Table S1. Classification performance was poorer in the Northern Appalachians, where larger average inter-station distances and uneven station distribution further limited detection and classification capability. By contrast, in Charlevoix none of the NEDB earthquakes were missed and only three blast events went undetected, reflecting the benefits of denser and more uniformly distributed stations, including smaller azimuthal gaps.

Figure 6 Monthly histograms comparing machine-learning (ML)-classified events and events listed in the Canadian National Earthquake Database (NEDB) in the Western Québec Seismic Zone (area shown in Fig. 1a). From top to bottom, panels (a), (b), and (c) present the monthly event counts for 2020, 2021, and 2022, respectively. The bars show the numbers of ML-classified blasts, ML-classified noise events, ML-classified earthquakes, NEDB blasts, and NEDB earthquakes for each month.

5 Discussion

In many ML classification studies, the primary objective is to maximize overall performance through extensive hyperparameter tuning. In contrast, the aim of this study is to develop a method that is both highly effective and straightforward to implement within a standard Google Colab environment, without requiring specialized high-performance GPU resources. Our practical goal is to distinguish industrial blasts from tectonic earthquakes under realistic constraints imposed by sparse station coverage and variable source–station distances. Because final source identification is performed at the event level using information from multiple stations, modest station-level classification errors do not necessarily compromise the final event classification. In this context, a station-level accuracy of about 97% is sufficient to support reliable event-level discrimination.

Several additional approaches may further improve classification performance. First, more sophisticated but computationally more demanding optimization schemes, such as stochastic gradient descent (SGD) or SGD with momentum (Bottou, 2012), could be explored. Second, additional physically motivated features, such as the corner-frequency ratio between P and S waves, may provide useful complementary information, as demonstrated by Korrat et al. (2022) for source discrimination in Egypt. Third, because industrial blasts tend to occur preferentially on weekdays and during warmer months (Fig. S18), temporal information beyond the hour of day used here may also improve classification in some regions, although such temporal patterns may not be transferable across all study areas. Finally, further tuning of hyperparameters such as dropout rate, learning rate, and weight decay may yield incremental improvements in model performance.

We also evaluated the sensitivity of model performance to class imbalance during training by varying the retained blast fraction from 0.05 to 1.0 (0.05, 0.2, 0.5, and 1.0) and by comparing runs with and without positive-class weighting in the loss function. Across all tested configurations, validation and test recall remained highly stable at approximately 0.975–0.977, in-

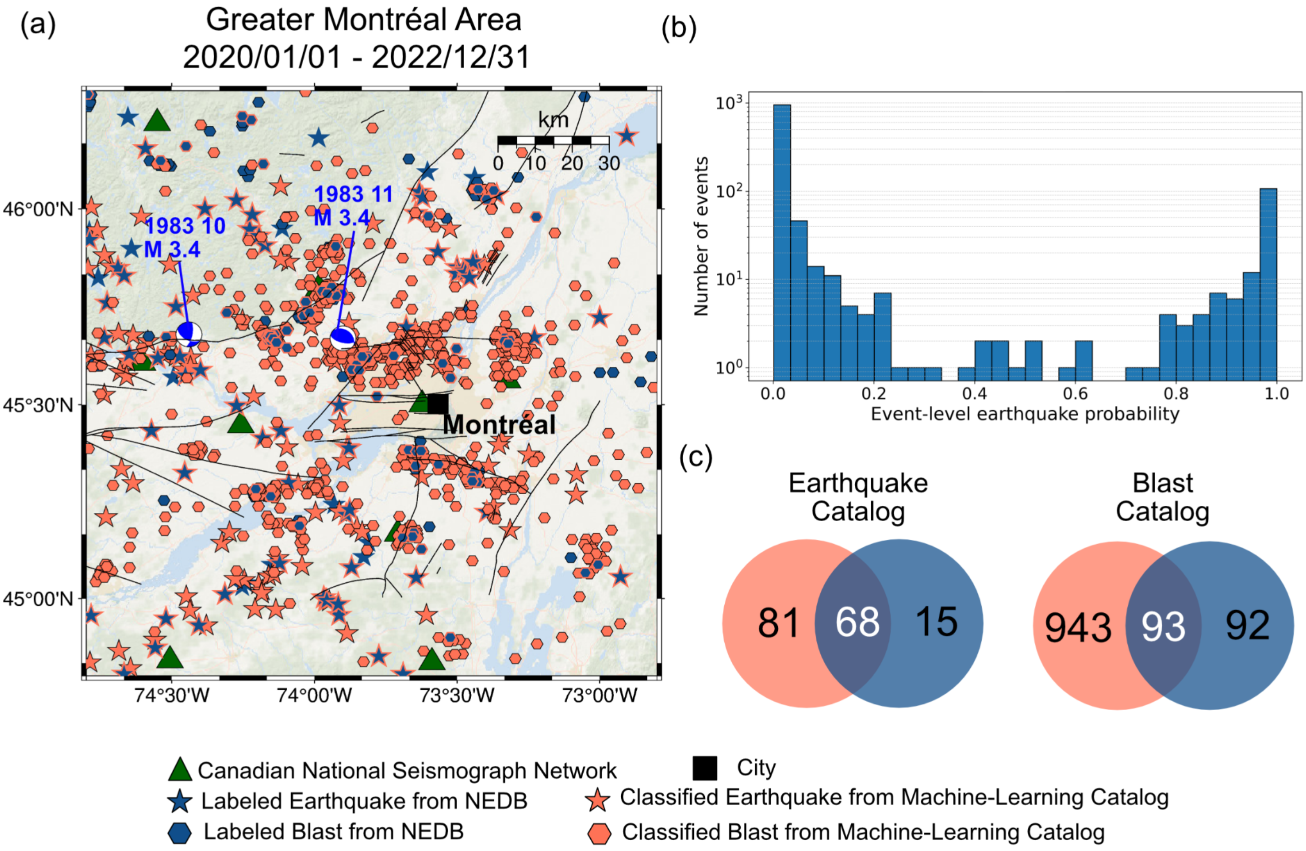


Figure 7 (a) Spatial distribution of blast and earthquake events identified from the machine-learning (ML) enhanced catalog using the optimized EfficientNet model (orange), and from the NEDB pre-labeled catalog (dark blue) in Western Québec Seismic Zone (WQSZ) near Montréal, overlaid with focal mechanism solutions for earthquakes of magnitude ≥ 3 from Mazzotti and Townend (2010). The mapped faults in WQSZ are from Geological Survey of Canada (Lamontagne et al., 2020). (b) Histogram of ML predicted earthquake probability values for events shown in panel (a), 0 is for blast events, 1 for tectonic earthquakes. (c) Venn diagrams comparing ML-enhanced catalog events (orange) against NEDB catalog events (dark blue); events are matched if their epicenters lie within 0.1° and their origin times differ by at most 5 s.

dicating that earthquake detection performance is robust to substantial changes in class balance. By contrast, FP rate, precision, and the weighted cost function $(1.25 \times \mathrm{FP} + 1.0 \times \mathrm{FN})$ were more sensitive to the retained blast fraction (see Fig. S19): performance generally deteriorated as fewer blast examples were retained, and this degradation became pronounced when the retained blast fraction fell below 0.2. For retained blast fractions of 0.2 and above, performance differences were comparatively small. Positive-class weighting likewise did not yield a significant performance gain. Full results are provided in the Supplementary Material.

Designing a new CNN architecture from scratch to achieve even lower FP and FN rates in Eastern Canada is certainly possible, but it would introduce substantial additional complexity. In particular, the heterogeneous station distribution and variable waveform sampling rates would likely require architecture-specific preprocessing choices and more extensive hyperparameter tuning. Such development would also become especially time-consuming given the volume of waveform data involved across a large geographic region. By contrast, CNNs pretrained on large-scale image datasets provide a much more efficient strategy for spectrogram-based classification. These models already contain robust low- and mid-level feature representations learned from diverse image data, and fine-tuning them on seismic spectrograms converges rapidly with relatively little effort (Figs. S20 and S21). This transfer-learning strategy therefore reduces both computational cost and methodological complexity relative to designing and optimizing an entirely new architecture.

Future work could extend the framework by including random-noise waveforms as an explicit third class and by training directly on raw spectrograms without denoising, since noise spectrograms differ markedly from both earthquake and blast signals. In the present study, an SNR threshold of 1.5 was applied during model training to ensure that the earthquake and blast examples used for supervision were of sufficient quality and were not dominated by pure noise. The same station-level SNR criterion was also applied during catalog classification: only spectrograms with SNR > 1.5 were retained for station-level prediction and subsequent event-level aggregation. Rather than imposing a separate event-level SNR cutoff, however, we evaluated the number of qualified stations contributing to each event-level prediction and used this in-

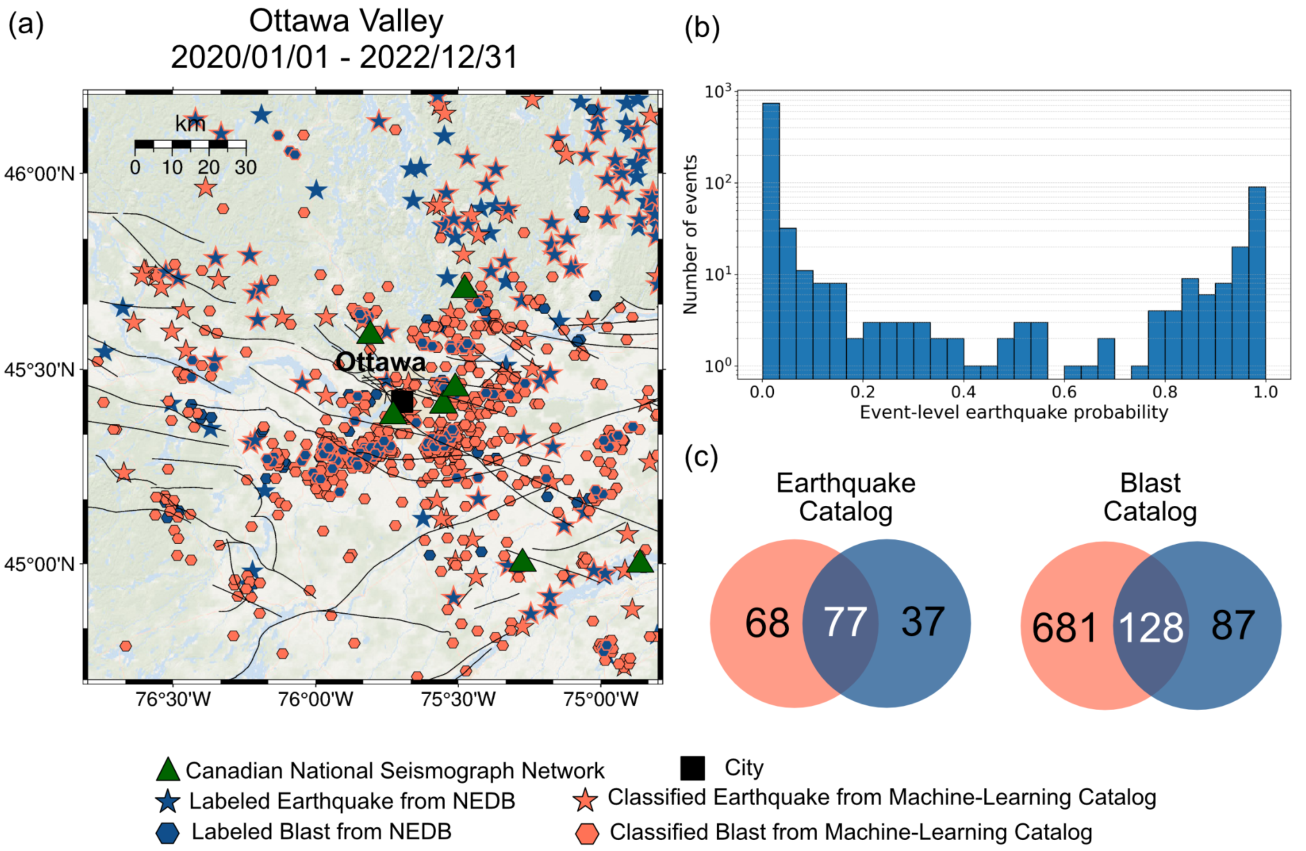


Figure 8 Same as Figure 7, for the Ottawa area in Western Québec Seismic Zone.

formation to identify likely false detections. This approach was adopted because stricter event-level screening could unintentionally exclude legitimate small-magnitude events or closely spaced events occurring within the same two-minute window. At the same time, the phase-association criteria used here are relatively strict, requiring at least eight total phase picks, including both P- and S-phase picks from at least three stations. It is therefore unlikely that a genuine nearby event would systematically fail the station-level SNR criterion at all available stations. Nevertheless, incorporating a dedicated noise class in future work would likely further improve rejection of spurious detections generated by ML-based phase pickers. More broadly, this classification workflow could also serve as a useful benchmark for evaluating popular ML phase-picking algorithms such as PhaseNet (Zhu and Beroza, 2019), EqTransformer (Mousavi et al., 2020), and GPD (Ross et al., 2018).

The large number of newly classified blast events demonstrates that, although ML-based phase pickers are trained primarily on earthquake waveforms, they still generate picks for many blast signals. However, the number of missed blast events relative to the NEDB catalog remains nontrivial (Panels c in Figures 7–10). This comparison should be interpreted with caution, because our training dataset is mainly based on the public NEDB blast catalog, which likely provides an incomplete record of industrial blasting activity in eastern Canada. Some missed blast events are therefore expected, especially because the phase pickers were not trained specifically to recognize blast signatures and may fail to detect a considerable fraction of industrial explosions. Furthermore, of the 768 NEDB-only blast events, 46% are concentrated near Val-d'Or (Fig. 1(a)), one of the largest mining districts in Quebec. The relatively sparse station coverage in this area, with only two nearby stations, VDQ, operating from 1980 to 2002, and VLDQ, operating since 2002, likely contributes to the large number of missed blast detections by limiting detection sensitivity and reducing the reliability of phase association. In regions with especially intense blasting activity, one possible improvement would be to retrain phase pickers using both earthquake and blast waveforms. In practice, however, assembling a sufficiently large and well-labeled blast training set would require substantial manual effort and might not generalize well to other regions. The present discrimination framework therefore offers a practical alternative by separating blasts from earthquakes after detection, without requiring extensive retraining of the phase pickers themselves.

In addition to under-detection, many newly classified blast events also exhibit substantial location uncertainty. Although most cluster near NEDB-documented blast locations, some remain more sparsely distributed (Figs. 7–9a), and a small subset likely corresponds to noise-related false detections that were nevertheless assigned blast-like classifications. Further external validation, for example using satellite imagery or industrial

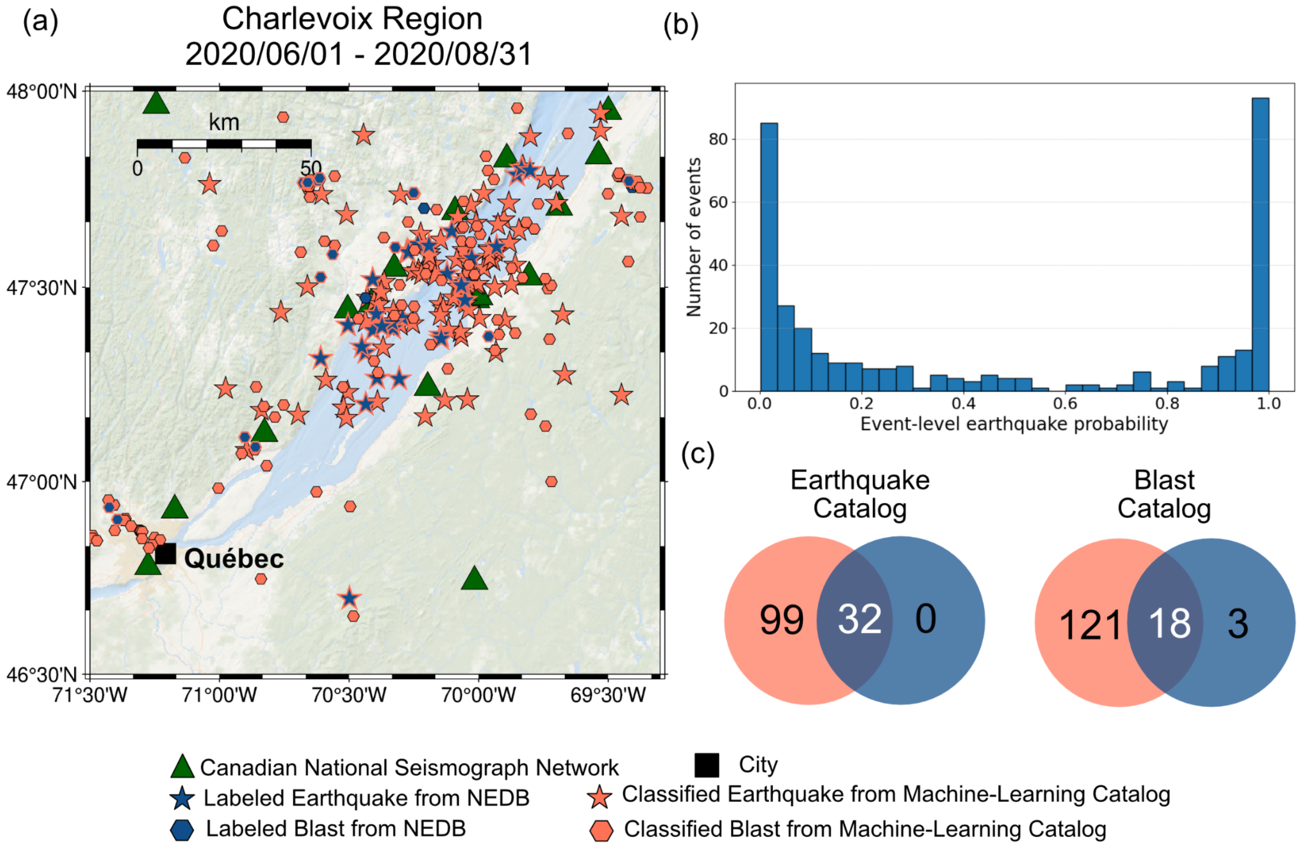


Figure 9 Same as Figure 7, for the Charlevoix Seismic Zone.

land-use information, would be useful in future work to better assess the origin of these spatially scattered blast-like events. In the Charlevoix region, some classified blast events are even located under the river. This likely occurs because ML phase pickers trained primarily on earthquake waveforms do not always identify blast P- and S-phase arrivals consistently across all available stations, leading to large phase-association and location errors. In Charlevoix, the horizontal location uncertainty of blast events (Fig. S22) reaches up to ∼30 km, and many events exceed 10 km. Such uncertainties are sufficient to displace events that likely originated near the coast into the river, as suggested by their proximity to NEDB-documented blast locations (Fig. 9a).

Over the three-year study period, our ML-enhanced catalog and CNN discriminator identified 273 new tectonic earthquakes and 2,560 new blast events in the WQSZ, but failed to recover about 261 NEDB-documented earthquakes (Table S1; see the Supplementary Material for the comprehensive catalog). Most of these missed events are concentrated in the remote northwestern sector of the WQSZ, adjacent to the Ottawa–Bonnechere Graben fault system (Fig. S23), where only two seismic stations lie within a 100-km radius and detection sensitivity is therefore substantially reduced. In addition, the NEDB catalog recorded more small-magnitude earthquakes in 2022, especially events with $M_L < 1.5$ (Fig. S24). Many of these events appear to have been recorded by only one or two stations, making them difficult to pass our phase association criteria. Consequently, the number of earthquakes listed in the NEDB is higher than the number of ML-classified earthquakes in 2022 (Fig. 6(c)), and more than 50% of the NEDB-only earthquakes occurred during that year. Likewise, in the Northern Appalachian region (Fig. 10), sparse and uneven station coverage, together with the phase-association limitations, resulted in numerous undetected NEDB events. Relaxing phase-association parameters, such as reducing the required phase picks and/or common stations, would likely reduce missed detections in these sparse coverage areas, but also increases false detection rates in other regions where station coverage is adequate. For example, in the Charlevoix Seismic Zone (Fig. 9) and around major WQSZ population centers such as Montréal (Fig. 7), the CNN discriminator indeed reliably identifies both new earthquakes and blast events with high confidence.

6 Conclusion

Traditional seismic event discrimination approaches often perform poorly in Eastern Canada with sparse and/or uneven station coverage. In contrast, our results show that pretrained CNN image classifiers applied to denoised spectrogram representations provide an effective and practical alternative. We processed approximately 100,000 three-component waveforms from NEDB earthquakes and blasts recorded between 2000 and 2024, transformed each waveform into a standardized RGB-like CNN input, and incorporated event origin

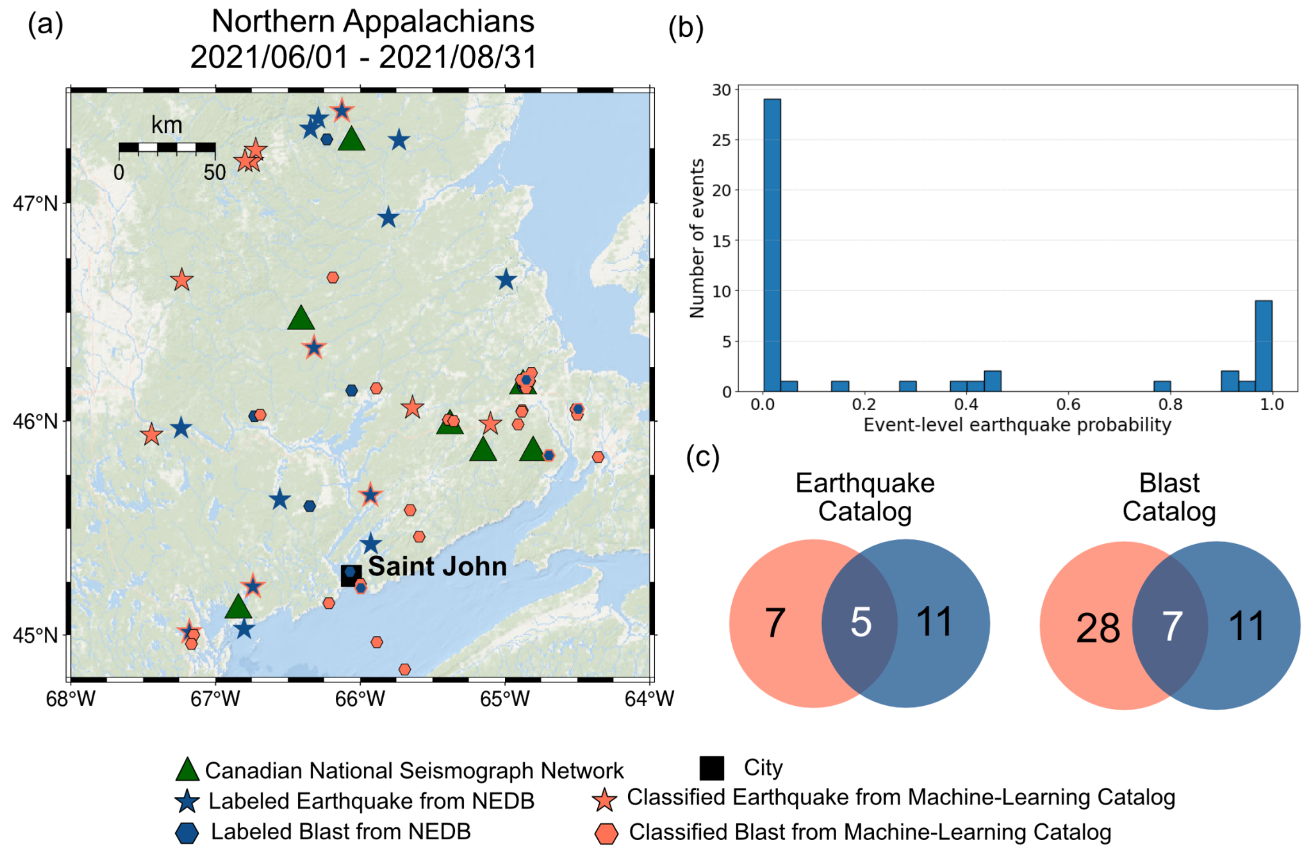


Figure 10 Same as Figure 7, for the Northern Appalachians Seismic Zone.

time as an auxiliary feature. Among the tested models, an EfficientNet-based architecture achieved the best overall performance under a weighted cost function of $1.25 \times \mathrm{FP} + 1.0 \times \mathrm{FN}$ for distinguishing tectonic earthquakes from industrial blasts.

To address uneven station coverage, we developed an event-level classification framework that combines station-level CNN predictions using reliability-based weights derived from source–station distance and denoised SNR. This approach assigns greater influence to station records expected to be more reliable while reducing the impact of lower-quality or more distant observations. The resulting event-level earthquake probabilities are strongly bimodal, with most events assigned values close to 0 or 1, corresponding to highly convincing blast-like and earthquake-like signals, respectively, and only a relatively small fraction occupying the intermediate range associated with greater classification uncertainty.

When applied to the WQSZ enhanced catalog, where ML phase pickers generate numerous additional detections, this event-level discrimination framework effectively filters out blast-dominated and noise-related detections while retaining earthquake-like events for further analysis. In this way, the method helps convert a noisy ML-enhanced detection catalog into a more interpretable and seismologically useful regional earthquake catalog that can support future studies of source properties and seismic hazard in Eastern Canada. Given its limited task-specific customization and the advantages of image-based pretraining, this framework provides a practical benchmark for future ML-driven catalog-enhancement workflows and shows clear potential for application in other regions characterized by sparse station coverage and mixed occurrences of tectonic earthquakes and industrial blasts.

Acknowledgements

This study was supported by NSERC Discovery Grant RGPIN-2024-05737 awarded to Liu. The authors thank Christopher Boucher of Natural Resources Canada for providing the NEDB internal blast catalog from October 2021 to April 2024. We also gratefully acknowledge discussions with Dr. Rebecca M. Harrington and members of the Earthquake Processes Group at McGill University and Ruhr University Bochum.

Data and code availability

The supplemental materials include waveform examples and a schematic of our catalog-enhancement workflow, all models result, and initial ML-classified catalog compared with Canadian National Earthquake Database (NEDB). Additional catalogs and code for discrimination model training, preprocessing, and implementation details are publicly available at (Chien, 2025). The NEDB is available at https://www.earthquakescanada.nrcan.gc.ca/stndon/NEDB-BNDS/bulletin-en.php. The analysis packages used in this study are open source and freely available, including PyGMT, ObsPy, NonLinLoc (Lomax et al., 2000), and PyOcto (Münchmeyer, 2023).

Competing interests

The authors have no competing interests to declare.

References

Bottou, L. Stochastic gradient descent tricks. In *Neural Networks: Tricks of the Trade: Second Edition*, pages 421–436. Springer, 2012. doi: 10.1007/978-3-642-35289-8_25.

Brooks, G. R. and Adams, J. A review of evidence of glacially-induced faulting and seismic shaking in eastern Canada. *Quaternary Science Reviews*, 228:106070, 2020. doi: 10.1016/j.quascirev.2019.106070.

Chien, J. SABER – Spectrogram-based Algorithm for Blast and Earthquake Recognition. [Software], 2025. doi: 10.5281/zenodo.16905883.

Ebel, J. E. A new analysis of the magnitude of the February 1663 earthquake at Charlevoix, Quebec. *Bulletin of the Seismological Society of America*, 101(3):1024–1038, 2011. doi: 10.1785/0120100190.

Ertuncay, D., Lorenzo, A. D., and Costa, G. Seismic signal discrimination of earthquakes and quarry blasts in North-East Italy using deep neural networks. *Pure and Applied Geophysics*, 181(4): 1139–1151, 2024. doi: 10.1007/s00024-024-03440-0.

Gamboa-Canté, C., Arroyo-Solórzano, M., Benito, B., Aguilar, J., Arroyo, I. G., Camacho-Astigarrabia, E., Castro, D., Flores, O., Linkimer, L., Marroquin, M. G., et al. Seismicity in Central America (1520–2020) and Earthquake catalog compilation for seismic hazard assessments. *Bulletin of Earthquake Engineering*, 22(15): 7201–7234, 2024. doi: 10.1007/978-94-015-9536-0_5.

Ghaffari, A. and Savaria, Y. CNN2Gate: An implementation of convolutional neural networks inference on FPGAs with automated design space exploration. *Electronics*, 9(12):2200, 2020. doi: 10.3390/electronics9122200.

Ghosh, A., Holt, W. E., and Bahadori, A. Role of large-scale tectonic forces in intraplate earthquakes of central and eastern North America. *Geochemistry, Geophysics, Geosystems*, 20(4): 2134–2156, 2019. doi: 10.1029/2018GC008060.

Hostallero, D. E., Li, Y., and Emad, A. Looking at the BiG picture: incorporating bipartite graphs in drug response prediction. *Bioinformatics*, 38(14):3609–3620, 2022. doi: 10.1093/bioinformatics/btac463.

Hourcade, C., Bonnin, M., and Beucler, É. New CNN-based tool to discriminate anthropogenic from natural low magnitude seismic events. *Geophysical Journal International*, 232(3): 2119–2132, 2023. doi: 10.1093/gji/ggac441.

Johnson, S. W., Chambers, D. J., Boltz, M. S., and Koper, K. D. Application of a convolutional neural network for seismic phase picking of mining-induced seismicity. *Geophysical Journal International*, 224(1):230–240, 2021. doi: 10.1093/gji/ggaa449.

Kao, H., Shan, S.-J., Bent, A., Woodgold, C., Rogers, G., Cassidy, J. F., and Ristau, J. Regional centroid-moment-tensor analysis for earthquakes in Canada and adjacent regions: An update. *Seismological Research Letters*, 83(3):505–515, 2012. doi: 10.1785/gssrl.83.3.505.

Kasburg, V., Müller, J., Eulenfeld, T., Breuer, A., and Kukowski, N. Cross-Regional Seismic Event Discrimination via Convolutional Neural Networks: Exploring Fine-Tuning and Ensemble Averaging. *Bulletin of the Seismological Society of America*, 114(2): 842–856, 2024. doi: 10.1785/0120230198.

Kim, W.-Y., Simpson, D., and Richards, P. G. Discrimination of earthquakes and explosions in the eastern United States using regional high-frequency data. *Geophysical Research Letters*, 20 (14):1507–1510, 1993. doi: 10.1029/93GL01267.

Kingma, D. P. and Ba, J. Adam: A Method for Stochastic Optimization. arXiv preprint 1412.6980, 2014. doi: 10.48550/arXiv.1412.6980.

Koonce, B. and Koonce, B. EfficientNet. *Convolutional neural networks with swift for Tensorflow: image recognition and dataset categorization*, pages 109–123, 2021.

Korrat, I., Lethy, A., ElGabry, M., Hussein, H., and Othman, A. S. Discrimination between small earthquakes and quarry blasts in Egypt using spectral source characteristics. *Pure and Applied Geophysics*, 179(2):599–618, 2022. doi: 10.1007/s00024-022-02953-w.

Kubo, H., Naoi, M., and Kano, M. Recent advances in earthquake seismology using machine learning. *Earth, Planets and Space*, 76(1):36, 2024. doi: 10.1186/s40623-024-01982-0.

Lamontagne, M., Brouillette, P., Grégoire, S., Bédard, M., and Bleeker, W. Faults and lineaments of the western Quebec seismic zone, Quebec and Ontario, 2020. doi: 10.4095/321900.

Lomax, A., Virieux, J., Volant, P., and Berge-Thierry, C. Probabilistic earthquake location in 3D and layered models: Introduction of a Metropolis-Gibbs method and comparison with linear locations. In *Advances in Seismic Event Location*, pages 101–134. Springer, 2000. doi: 10.1007/978-94-015-9536-0_5.

Loshchilov, I., Hutter, F., et al. Fixing weight decay regularization in adam. *arXiv preprint arXiv:1711.05101*, 5, 2017.

Ma, S. and Eaton, D. W. Western Quebec seismic zone (Canada): Clustered, midcrustal seismicity along a Mesozoic hot spot track. *Journal of Geophysical Research: Solid Earth*, 112(B6), 2007. doi: 10.1029/2006JB004827.

Maguire, R., Schmandt, B., Wang, R., Kong, Q., and Sanchez, P. Generalization of Deep-Learning Models for Classification of Local Distance Earthquakes and Explosions across Various Geologic Settings. *Seismological Research Letters*, 95(4):2229–2238, 2024. doi: 10.1785/0220230267.

Mancini, S., Segou, M., Werner, M. J., Parsons, T., Beroza, G., and Chiaraluce, L. On the use of high-resolution and deep-learning seismic catalogs for short-term earthquake forecasts: Potential benefits and current limitations. *Journal of Geophysical Research: Solid Earth*, 127(11):e2022JB025202, 2022. doi: 10.1029/2022JB025202.

Mazzotti, S. and Townend, J. State of stress in central and eastern North American seismic zones. *Lithosphere*, 2(2):76–83, 2010. doi: 10.1130/L65.1.

Mousavi, S. M., Ellsworth, W. L., Zhu, W., Chuang, L. Y., and Beroza, G. C. Earthquake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking. *Nature Communications*, 11(1):3952, 2020. doi: 10.1038/s41467-020-17591-w.

Münchmeyer, J. PyOcto: A high-throughput seismic phase associator. *arXiv preprint arXiv:2310.11157*, 2023. doi: 10.26443/seismica.v3i1.1130.

NRCan. Natural Resources Canada. Canadian National Earthquake Database [Dataset]. Canadian Hazards Information Service, 1985. doi: 10.17616/R3TD24.

Onwuemeka, J., Liu, Y., and Harrington, R. M. Earthquake stress drop in the Charlevoix Seismic Zone, eastern Canada. *Geophysical Research Letters*, 45(22):12226–12235, 2018. doi: 10.1029/2018GL079382.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019.

Plourde, A. P. and Nedimović, M. R. Earthquake depths, focal mechanisms, and stress in the Lower St. Lawrence Seismic Zone. *Seismological Research Letters*, 92(4):2562–2572, 2021. doi: 10.1785/0220200429.

Qassim, H., Verma, A., and Feinzimer, D. Compressed residual-VGG16 CNN model for big data places image recognition. In *2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC)*, pages 169–175. IEEE, 2018. doi: 10.1109/CCWC.2018.8301729.

Rakotoarisoa, A. T. and Razafindrakoto, H. N. Application of Convolutional Neural Networks for Discriminating Anthropogenic and Natural Earthquakes in Madagascar. *Bulletin of the Seismological Society of America*, 115(4):1642–1657, 2025. doi: 10.1785/0120240244.

Reiter, K., Heidbach, O., Schmitt, D., Haug, K., Ziegler, M., and Moeck, I. A revised crustal stress orientation database for Canada. *Tectonophysics*, 636:111–124, 2014. doi: 10.1016/j.tecto.2014.08.006.

Ross, Z. E., Meier, M.-A., Hauksson, E., and Heaton, T. H. Generalized seismic phase detection with deep learning. *Bulletin of the Seismological Society of America*, 108(5A):2894–2901, 2018. doi: 10.1785/0120180080.

Ruby, U. and Yendapalli, V. Binary cross entropy with deep learning technique for image classification. *Int. J. Adv. Trends Comput. Sci. Eng*, 9(10), 2020. doi: 10.30534/ijatcse/2020/175942020.

Targ, S., Almeida, D., and Lyman, K. Resnet in resnet: Generalizing residual architectures. *arXiv preprint arXiv:1603.08029*, 2016.

Trugman, D. T. A High-Precision Earthquake Catalog for Nevada. *Seismological Research Letters*, 95(6):3737–3745, 2024. doi: 10.1785/0220240106.

Verdecchia, A., Onwuemeka, J., Liu, Y., and Harrington, R. M. Depth-dependent crustal stress rotation and strength variation in the Charlevoix seismic zone (CSZ), Québec, Canada. *Geophysical Research Letters*, 49(22):e2022GL100276, 2022. doi: 10.1029/2022GL100276.

Woo, S., Debnath, S., Hu, R., Chen, X., Liu, Z., Kweon, I. S., and Xie, S. ConvNeXt V2: Co-designing and scaling ConvNets with masked autoencoders. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16133–16142, 2023. doi: 10.1109/CVPR52729.2023.01548.

Woollam, J., Münchmeyer, J., Tilmann, F., Rietbrock, A., Lange, D., Bornstein, T., Diehl, T., Giunchi, C., Haslinger, F., Jozinović, D., et al. SeisBench—A toolbox for machine learning in seismology. *Seismological Research Letters*, 93(3):1695–1709, 2022. doi: 10.1785/0220210324.

Yavuz, E., Sertçelik, F., Livaoğlu, H., Woith, H., and Lühr, B.-G. Discrimination of quarry blasts from tectonic events in the Armutlu Peninsula, Turkey. *Journal of Seismology*, 23(1):59–76, 2019. doi: 10.1007/s10950-018-9793-2.

Zhang, M. P/SV Amplitude Ratios of Shallow Isotropic Explosions and Earthquakes Could Be Indistinguishable at Local Distances: Insights from Single-Station Waveform Simulations. *The Seismic Record*, 3(1):48–56, 2023. doi: 10.1785/0320220044.

Zhu, J., Fang, L., Miao, F., Fan, L., Zhang, J., and Li, Z. Deep learning and transfer learning of earthquake and quarry-blast discrimination: applications to southern California and eastern Kentucky. *Geophysical Journal International*, 236(2):979–993, 2024. doi: 10.1093/gji/ggad463.

Zhu, W. and Beroza, G. C. PhaseNet: a deep-neural-network-based seismic arrival-time picking method. *Geophysical Journal International*, 216(1):261–273, 2019. doi: 10.1093/gji/ggy423.

Zhu, W., Mousavi, S. M., and Beroza, G. C. Seismic signal denoising and decomposition using deep neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 57(11):9476–9488, 2019. doi: 10.1109/TGRS.2019.2926772.