FASER: Seismic Phase Identifier for Automated Monitoring

Farhan Asif Chowdhury
The University of New Mexico
Albuquerque, USA
fasifchowdhury@unm.edu

M Ashraf Siddiquee
The University of New Mexico
Albuquerque, USA
siddiquee@unm.edu

Glenn Eli Baker
Air Force Research Lab
Albuquerque, USA
glenn.baker.3@us.af.mil

Abdullah Mueen
The University of New Mexico
Albuquerque, USA
mueen@unm.edu

ABSTRACT

Seismic phase identification classifies the type of seismic wave received at a station based on the waveform (i.e., time series) recorded by a seismometer. Automated phase identification is an integrated component of large scale seismic monitoring applications, including earthquake warning systems and underground explosion monitoring. Accurate, fast, and fine-grained phase identification is instrumental for earthquake location estimation, understanding Earth's crustal and mantle structure for predictive modeling, etc. However, existing operational systems utilize multiple nearby stations for precise identification, which delays response time with added complexity and manual interventions. Moreover, single-station systems mostly perform coarse phase identification.

In this paper, we revisit the seismic phase classification as an integrated part of a seismic processing pipeline. We develop a machine-learned model FASER, that takes input from a signal detector and produces phase types as output for a signal associator. The model is a combination of convolutional and long short-term memory (LSTM) networks. Our method identifies finer wave types, including crustal and mantle phases. We conduct comprehensive experiments on real datasets to show that FASER outperforms existing baselines. We evaluate FASER holding out sources and stations across the world to demonstrate consistent performance for novel sources and stations.

CCS CONCEPTS

• **Applied computing** → **Earth and atmospheric sciences**; • **Information systems** → *Data mining*; • **Computing methodologies** → *Neural networks*.

KEYWORDS

earthquakes, seismology, neural networks, geophysical techniques, seismic phase classification

ACM Reference Format:
Farhan Asif Chowdhury, M Ashraf Siddiquee, Glenn Eli Baker, and Abdullah Mueen. 2021. FASER: Seismic Phase Identifier for Automated Monitoring. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '21), August 14–18, 2021, Virtual Event, Singapore.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3447548.3467064

KDD '21, August 14–18, 2021, Virtual Event, Singapore

ACM ISBN 978-1-4503-8332-5/21/08…$15.00
https://doi.org/10.1145/3447548.3467064

1 INTRODUCTION

Real-time seismic signal processing is a key element of the geophysical monitoring required for early warning systems for earthquakes, underground mineral exploration and mining, and nuclear explosion monitoring. Seismic signal processing pipelines involve several sequential steps that start with signal (e.g., from an earthquake) detection from raw seismic signals recorded at a seismic station, and in the end, produce a formalized event bulletin for real-time alarm generation as well as future analysis. Figure 1 shows a typical pipeline. Phase identification is a key step in this pipeline subsequent to the signal detection step, which can be framed as a classification problem that takes a detected seismic signal as input, and outputs the phase label. Phase identification is required for proper utilization of the downstream steps of the pipeline, for example, earthquake location estimation, tomographic studies, and understanding of the Earth's crustal and upper mantle structure [9]. A successful phase classifier must classify a detected seismic waveform into shear or transverse waves (ending with S in Figure 1.right) and compressional or longitudinal waves (ending with P in Figure 1.right), and all their subtypes.

Single Station vs. Array. At present, in operational systems, seismic phase identification is heavily dependent on the use of multiple close-by seismic monitoring stations, forming an *array* of stations [22]. High-quality arrays enable better detection and improved signal-to-noise-ratio, and estimation of phase velocity and direction of arrival; which greatly benefit both phase identification and association. Relative arrival times of seismic phases at different arrays and of different detections at the same array, together with their directions of arrival, are used to accurately classify and associate phases [32]. Unfortunately, most new stations added in dynamic response to changing monitoring needs, such as in oil

fields or novel seismic sources, will be individual stations rather than arrays. This reduces the processing pipeline's performance because phase identification is less accurate for single stations than it is for arrays. In this work, we consider automated phase classification on data collected at a single station to enable rapid deployment addressing dynamic needs.

Fine vs. Coarse Classification. In addition, most existing research work on automated seismic phase classification considers only the two high-level categories, while most monitoring applications require finer classification [7, 30]. For global monitoring, seismic signals are initially classified as teleseismic P (including more complex phases such as PkP and PkikP) or regional P or S (i.e., Pn or Sn). Refinements to the phase identification to add teleseismic S and crustal P and S phases (i.e., S, Pg, and Lg) are made much later in the processing pipeline. In this work, we consider classifying into finer phases at the initial identification step. In addition, phase detection and phase classification are usually sequential steps. Seismic signal detection algorithms are very well developed [25, 26, 33], and signal detection is a key module in the processing pipeline. In this work, we consider phase classification *only after* detection, unlike some recent works that address detection and classification jointly [29, 30].

Global vs. Regional: Lastly, most existing research work focuses on local and near regional methods that only observe two major phases (local P and S, or occasionally regional P and S), and uses these phases to determine source location and origin times, mostly due to constrained focus. However, in a *global* monitoring application, all kinds of teleseismic (>1000km), regional (>200km), and local waves can arrive at various degrees of temporal overlap with arbitrary arrival order. Figure 1 (right) illustrates some of the complexity of global seismic arrivals. Identifying the correct phase from a complex waveform containing multiple arrivals is difficult for global monitoring applications. We use data from the International Monitoring System (IMS) network, a global network operated by the Preparatory Commission of the Comprehensive Nuclear-Test-Ban Treaty Organization (CTBTO). IMS data are processed in near real-time at the International Data Centre (IDC) in Vienna, with initial detection, phase identification, and association performed automatically and then curated by human analysts. Our method can achieve significant classification accuracy even in such a complex application as presented by a global seismic network data.

Challenges to this research. First, there is no publicly available dataset of seismic signals categorized into six-phase classes. The main hindrance of curating such a dataset is the relative rarity of some phases compared to others. For example, in a continuous one-year time frame of the IMS catalog, 128K seismic waveforms are classified as P, and only 1300 are classified as S. Also, the manual labeling of these distinct phase-types requires a depth of knowledge and rigorous training [3]. Second, the phases within broader hierarchical categories share highly similar spectral and amplitude properties. Also, seismic signals originating at different geolocations exhibit different propagation effects, resulting in dissimilar signal properties for the same phase-type. There are also differences among waveforms of a single type at different distances. The current bottleneck in processing is the association of seismic phases with their most likely sources, which could be improved by more accurate initial phase classification.

This work. To tackle these challenges, we focus on a few different aspects. First, we have curated a small-scale yet *balanced* seismic phase dataset collected from IMS network data. From an imbalanced collection of more than 200K seismic events, we have narrowed down to 16K events with finer and balanced phase labeling. We use Continuous Wavelet Transforms (CWT) to obtain a time-frequency representation from raw seismic time-series to utilize both temporal and spectral information. The CWT representation has been shown to be resilient to dynamic noise in waveforms [24]. We design an end-to-end deep neural network to perform phase identification using these CWT representations. We leverage the power of Convolutional Neural Network (CNN) to capture low-level features from the CWT representations that are invariant to frequency, scale, and position [31]. However, due to locally constrained receptive fields, CNNs are inadequate in modeling long-term temporal dependency, where else seismic signals contain distinctive temporal patterns across different phase-types [29]. To mitigate this limitation, we incorporate LSTM on top of the CNN as LSTM can effectively model long term temporal patterns and dependencies.

Our proposed method FASER can perform fine-grained phase identification using single-station data from the global seismic network. Due to minimal preprocessing requirements and instantaneous output generation, it can be readily integrated into the existing real-time seismic signal monitoring pipeline. We show a comprehensive experimental evaluation of FASER using a real dataset in comparison with existing methods to validate improved performance. We justify the generalizability of FASER by demonstrating case studies for applications in novel operating conditions. To the best of our knowledge, this is the *first attempt* to perform *fine-grained phase identification* using single-station seismic signal data.

The rest of the paper is organized as follows. In Section 2, we give a brief overview of related work. In Section 3, we present our proposed phase identification method. We show our experimental results in Section 4. In section 5, we demonstrate two case studies of our method in real-world applications. Finally, we conclude in section 6.

2 RELATED WORK

The existing literature on In general, the methods for seismic phase identification can be broadly categorized into two types, (1) heuristic template matching and statistical analysis based methods, and (2) deep learning based methods.

Statistical and Heuristic Methods. Since the early inception and development of seismic signal monitoring, several rule-based and physics-driven methods have been proposed for seismic phase identification. In [28], data-adaptive polarization filtering method have used for phase detection task. In [2], the difference between the short-term average (STA) and the long-term average (LTA) of the seismic signal has been used for automated detection. Several methods have used higher-order statistics like kurtosis and skewness

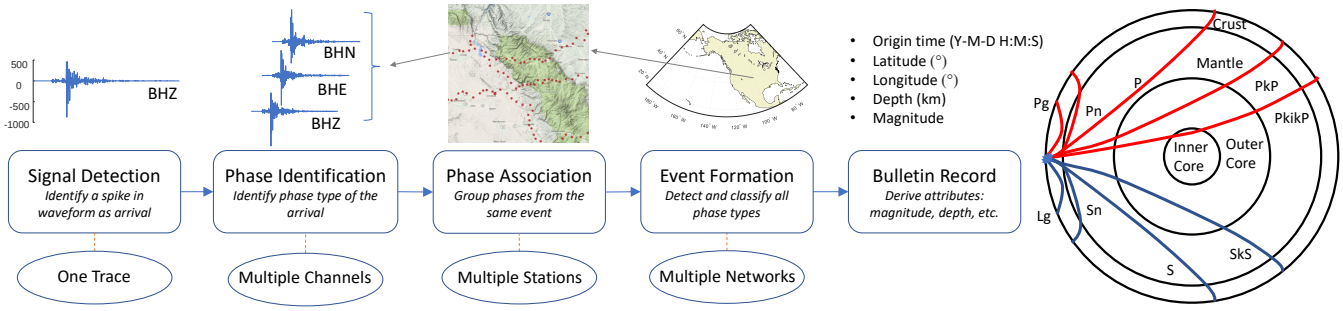


Figure (1) (left) Typical seismic data processing pipeline. Our objective is to develop a Machine Learning model for phase identification. (right) Travel times with respect to distance for various seismic phases. Phases ending with P can commonly be categorized as P, and phases ending with S can commonly be categorized as S.

[19, 20]. Also, few methods have used frequency domain information [24, 37]. However, these methods perform poorly in the presence of noise and when the seismic events are of low magnitude. A few other works have proposed a similarity search based template matching method [10, 27]. But, such methods are heavily dependent on a prior collection of sample signal templates and often fail to generalize when used for phase detection at novel stations. Also, the pairwise similarity search with each sample template renders these methods computationally intensive and inefficient for real-time monitoring.

Deep Learning based Methods. Recently, multiple deep learning methods have been proposed to address the aforementioned shortcomings in the context of phase detection and identification [6, 22, 25, 26]. A deep learning based grid-free phase association method for phase identification has been proposed in [30]. In [29], a CNN based architecture has been used for phase identification from one-dimensional seismic signals. More recently in [7], the use of time-frequency representation and CNN has been explored. These methods have shown promising results compared to previously used statistical and heuristic-based methods as CNNs can effectively model low-level structured patterns into high-dimensional embedding. However, due to locally constrained receptive fields, CNN cannot capture the long-term temporal patterns. Thus, these models cannot fully exploit the higher-order temporal structures in seismic signals, distinctive across different phase-types. Moreover, none of these methods perform fine-grained phase identification.

3 METHODS

In practice, the seismic phases are manually labeled by experienced analysts using multi-modal information, i.e., signal amplitude, frequency components in the signal, the distance between event origin and monitoring station, depth of the event origin, etc. However, for generalized and real-time phase classification, there exist a few challenges to producing such information a priori. Depth and source location estimation are intricate regression problems requiring complex analysis. Previously, the effectiveness of time-frequency representations has been demonstrated in a multitude of seismic signal processing tasks [24, 34]. Therefore, we use the CWT to obtain spectral-temporal features, as it produces higher spectral-resolution and more precise temporal-localization than other time-frequency

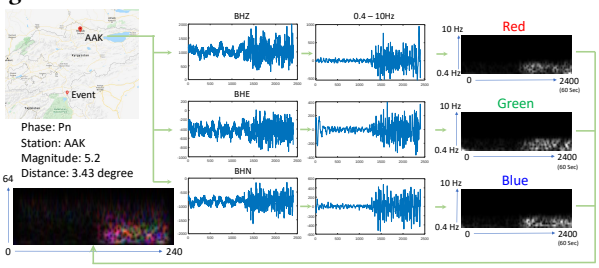


Figure (2) Input images of waveforms are created by taking CWT of individual channels (i.e. BHZ, BHN, BHE).

transformations (e.g., Short-Fourier Transformation) [24]. We use a composite CWT image, where the vertical component CWT coefficients are represented by red brightness, and the two horizontal components are represented by the brightness of green and blue (See Figure 2).

The time-frequency representation of seismic events contains distinct structured features based on their phase-type. However, at a low-level, these features are highly overlapping. CNN have been widely used in the domain of Computer Vision [13, 18], Natural Language Processing [8, 15], Speech recognition [1] and other related domains to learn high-level features from raw structured input for better contrastive representation learning. As CNN's can model the local correlation of spatial and temporal patterns, it is highly suitable for our two-dimensional CWT feature maps, which contain incremental time information on one axis and frequency information on the other.

However, CNN's are inadequate in learning long-range temporal dependencies due to their locally constrained receptive fields [31]. Nevertheless, the long-term temporal patterns in the seismic signals, which are well preserved in CWTs, are vital distinctive features across different phase types as showcased in Figure 3. To circumvent this limitation, Recurrent Neural Networks (RNN) have been instrumental in modeling the temporal dependencies by using the cyclic feedback mechanism from previous time-step inputs. LSTM, an improved variant of vanilla RNNs, are capable of learning and modeling long-term temporal patterns and dependencies [16, 23]. However, It has been shown that higher-level features can be helpful in learning the underlying factors of variations within the input, which should make it easier to learn temporal structures between

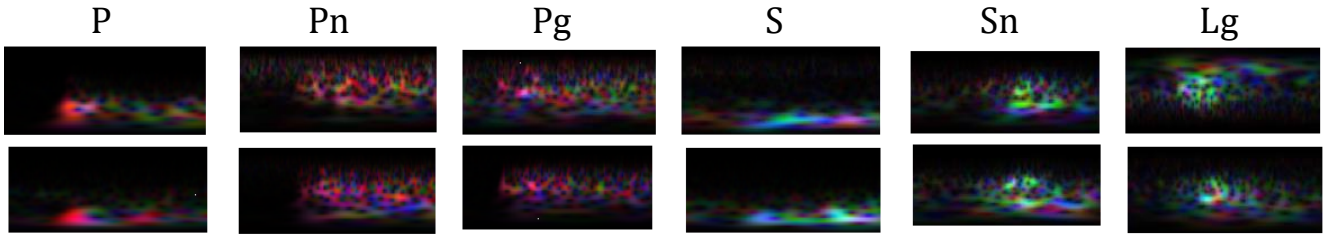


Figure (3) CWT of pairs of examples from all phase types. Compressional or longitudinal waves (P, Pn, Pg) are dominantly red due to high vertical component amplitudes. Transverse waves (S, Sn, Lg) are dominantly green/blue due to high horizontal component amplitudes.

successive time-steps [36]. Thus, oftentimes CNNs have been successfully used as preceding layers before more complex sequential models to reduce the local temporal and frequency variations [31].

Motivated by these aforementioned successful use-cases, we use a combination of CNN and LSTM in an end-to-end network. First, we utilize CNN to identify low-level spectral-temporal features that are invariant to frequency, scale, and position. Afterward, we organize the output features obtained from CNN into sequential features preserving the temporal ordering. We feed these higher-level sequential representations of low-level structured patterns as input into the LSTM. By utilizing the cyclic feedback mechanism in-between consecutive time steps, LSTM can better model the long-term temporal correlation in the seismic signal. Finally, we feed the output from each time-step into dense layers to make the final output prediction.

3.1 Convolutional Neural Network

CNN [18] perform convolution operations on the input feature map using fixed-size kernels (learned during the training step) to produce higher-order representations. Convolution operations are usually followed by a non-linear activation function and max-pooling layers. The use of an activation function introduces non-linearity, and the max-pooling reduces sensitivity to temporal or spatial variation. CNN's are adept at learning local structural relationships and are invariant to feature scaling, which reduces the dependency on heavy data preprocessing and feature engineering [38].

3.2 Long Short-Term Memory Network.

LSTM networks are an improved variant of traditional RNN [14]. RNNs can model temporal dependencies in the data by utilizing feedback connection by considering both input at the current time step as well as output of the last time step's hidden state. However, vanilla RNNs suffer from the vanishing gradient problem, which prevents the model from learning long range dependencies. LSTM tackles this problem by introducing three gating mechanisms to update the memory cell c_t and hidden state h_t at each step t based on the current time step input x_t and the previous time step's hidden state output h_{t-1}. Each LSTM unit is composed of a memory cell and three main gates: input, output and forget. The input gate i_t, forget gate f_t, output gate o_t, memory cell c_t and hidden state h_t at step t are computed as follows:

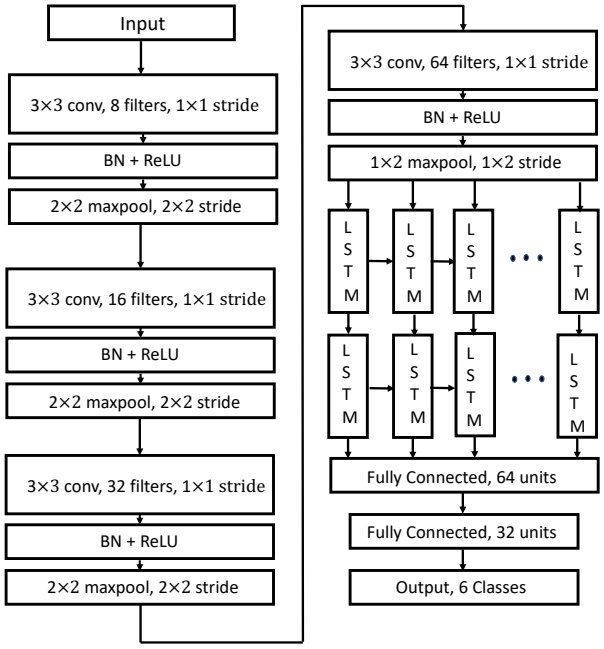


Figure (4) Proposed model architecture.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (1)$$

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f \quad (2)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o \quad (3)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (4)$$

$$h_t = o_t \odot tanh(c_t) \quad (5)$$

Here, W_i, W_f, W_o are the weight matrix and b_i, b_f, b_o are the bias of input, forget and output gate, respectively, and σ is the logistic sigmoid function, tanh is the hyperbolic tangent function, and $\odot$ denotes the element wise multiplication. By this architecture, the LSTM manages to create a controlled information flow by deciding which information it must forget and which information to remember. To understand the mechanism behind the architecture, we can view f_t as the function that controls to what extent the information from the old memory cell is going to be thrown away, i_t controls how much new information is going to be stored in the current memory cell, and o_t controls what to output based on the memory cell c_t.

3.3 Proposed Architecture

Our proposed architecture is comprised of four convolutional layers, followed by two LSTM layers and three fully connected dense layers. In Figure 4, we present our proposed architecture. The input to the networks is the CWT representations obtained from the bandpass filtered seismic signal waveforms. A detailed description of data preprocessing is presented in the following section. In each of the four convolutional layers, we use kernels of size 3×3 with a stride of 1×1 and zero paddings. Each convolution layer is followed by a Batchnormalization layer, a Rectified Linear Unit (ReLU), and a two-dimensional max-pool layer. In the first two max-pool layers, we use a kernel size of 2×2 with a stride of 2×2. However, in the last convolution layer, we perform max-pooling only along the temporal dimension with a kernel of size 1×2 and stride of 1×2. The first convolutional layer has eight filters, and we double the filter number on each subsequent convolutional layer to keep the number of parameters in each convolutional layer the same as we reduce the input image size by half after each convolution layer due to max-pooling.

The output from the final convolution layer is then passed into the LSTM layers preserving the temporal order. The first LSTM layer consists of 32 hidden units and the second LSTM layer consists of 16 hidden units. We use *sigmoid* and *tanh* as the recurrent and output activation function of the LSTM correspondingly. We use a 50% recurrent dropout in the LSTM layers. Both LSTM layers are unrolled for 15 steps as the input feature map to the LSTM has a temporal dimension of fifteen. Both LSTM layers return sequences in each unrolling step. These sequences are flattened before feeding into the dense layers. We stack three dense layers, each with 64, 32, and 6 hidden units consecutively. Each of the dense layers are preceded by Batchnormalization and ReLU activation functions with a 20% dropout rate. We use the softmax activation function in the final dense layer to obtain output probabilities for each phase-type.

4 EXPERIMENTS AND RESULTS

In this section, we perform experimental analysis on real seismic data to show the effectiveness of our proposed method in comparison with existing baseline methods.

4.1 Dataset Description

The dataset is curated from 10 years of continuous seismic data collected at the 155 stations of the IMS. These consist of 46 primary stations, 24 of them were arrays, and 105 auxiliary stations, 98 of which are 3-component stations. The dataset includes 80TB uncompressed of seismic waveforms and the comprehensive IMS catalog, with arrival times and phase labels curated by human analysts for over 8 million seismic event detections. This dataset includes the comprehensive IMS catalog, with arrival times and phase labels curated by human analysts for over 8 million seismic event detections. From these 8M seismic events, we filtered out 175K fine-grained seismic phase labeled data. However, among these, the P-phase was predominant, with 128,120 occurrences, while the S-phase had only 1,306 occurrences. To ensure a balanced dataset between crustal, regional, and teleseismic compressional and transverse phase (i.e., Pg, Lg, Pn, Sn, P, and S), we used all labeled S-phases and randomly sampled around 2,500 waveforms from each of the other phases, for a total of 16,304 phases. As the spectro-temporal features of all the phases are highly nuanced, no data augmentation was performed to maintain integrity for practical application scenarios.

4.2 Data Prepossessing.

Our raw input data are 60 seconds long three-channel time-series, sampled at 40Hz. Following conventional seismic signal pre-processing techniques, we filter the waveform from each channel (0.4Hz to 10Hz). As these seismic signals were generated by events of different magnitude and recorded at stations spread across the globe, the amplitude is not relevant to phase identification in this first step. We first detrend each sample and remove the mean. We then max-normalize the data across each-channel, thus retaining the relative amplitudes among components of a station. Afterward, use the CWT to obtain spectral-temporal features representation.

4.2.1 Baseline Methods. In order to validate the performance of our method we have compared our method with the following baseline methods.

- **XGBoost** (XGB) [4]: XGBoost is an optimized ensemble based model that has produced state-of-the-art methods for many classification tasks. We convert the multi-dimensional CWT representations into one-dimensional features as input for XGB. Afterward, we perform standardization across each feature dimension.
- **MLP** [12]: Multi Layer Perception (MLP) is a feed-forward neural network. We use a two-layer MLP with the same input feature as XGB.
- **CNN** [7]: CNN based methods have been previously used in related seismic signal classification tasks. In [7], a CNN based method has been used for two-class phase classification. We use the same CNN architecture used in this paper to compare against our method.
- **LSTM** [14]: LSTM methods are highly suitable for temporal data modeling and have produced state of the art accuracy in many time series classification tasks. We use the most popular stacked LSTM architecture for comparison. The final output is fed into a fully connected layer to generate output label.
- **CRED** [25]: In [25], a ResNet-BiLSTM architecture has been proposed for seismic event detection where it achieved state-of-the-art performance. We use the same architecture for our phase identification task.

4.3 Experimental Settings.

The hyper-parameters of the model were selected empirically by grid-search on the validation set. We use the Adam optimizer [17] with an initial learning rate of 0.01 and with parameters $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-8}$. We apply $L2$ regularization with $\lambda = 0.001$, and we use *categorical cross-entropy* [11] as the loss function

Table (1) Performance metric comparison of FASER against baseline methods. FASER outperforms existing baseline methods in all four metrics: Precision(PR), Recall (RL), F1-Score(F1) and Accuracy (ACC). The best algorithm for each metric is colored in cyan.

Method	PR	RL	F1	ACC
XGB	68.4	67.2	67.2	67.2
MLP	76.2	75.6	75.3	75.6
CNN	75.2	75.2	75.2	75.2
LSTM	75.7	74.3	75.0	75.3
CRED	81.3	80.2	80.7	81.5
FASER	**84.6**	**81.6**	**83.1**	**82.8**

with a training batch size of 256. The XGBoost model was trained until convergence. The neural network models were trained for a maximum of 200 epochs with early stopping on the validation set.

We use 10-fold cross validation to measure the performance of our method, and report the average. In each fold, we use 80% of the data for training, 10% for validation and 10% for testing. We perform random stratification to ensure class balance in the training-validation-test split. All the experiments were performed on a core i5 2.70 GHz desktop computer with 8GB NVIDIA GeForce GTX-1070 GPU.

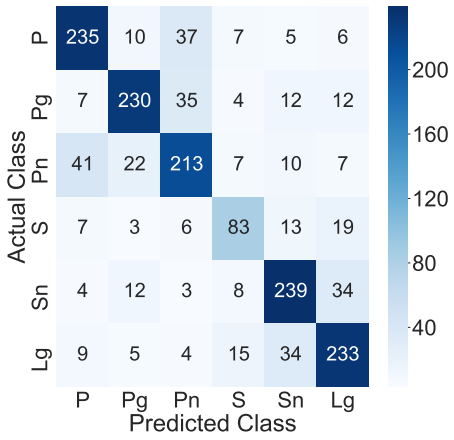
4.3.1 Evaluation Metrics. In our experiments, following conventional practices for classification tasks, we use accuracy as the primary performance metric. However, as there are minor class imbalances in the dataset, we also calculate macro (calculated individually for each class and averaged afterward) precision, recall, and F1-score [5]. Precision, recall, and f1-score are calculated based on true-positive(TP), false-positive(FP), and false-negative(FN) values using the following formulas.

$$precision = \frac{TP}{TP + FP}, recall = \frac{TP}{TP + FN}$$
$$F - score = 2 \cdot \frac{precision \cdot recall}{precision + recall}$$

4.4 Results.

In Table 1, we report the performance of our method in comparison with the baseline methods, where the highest performance for each metric is cyan colored. We observe that FASER consistently outperforms all the baseline methods across all performance metrics. To closely probe the performance of FASER across each class, in Figure 5(left), we show the confusion matrix for the test cases of a randomly split 80-10-10 train-validation-test scenario. It is noticeable that the majority of classification error is within the sub-classes of compressional (P, Pg, Pn) and transverse (S, Sn, Sg) waves. This performance is in coherence with the intuitive notion of similar spectral-temporal features within both broader classes.

In Figure 5(right), we plot the t-SNE visualization [21] of the same 10% test cases considering the activation values of the last layer before the prediction layer as deep embedding [35]. The compressional and transverse wave signal samples are well separated in the deep embedding space with high-margin with only a few



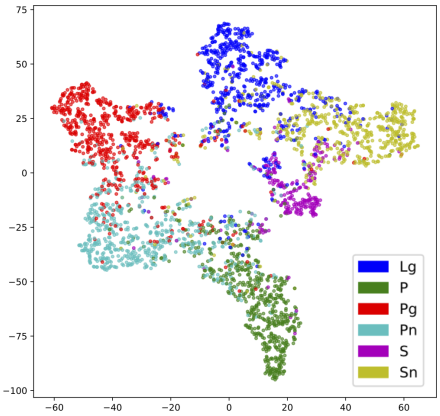


Figure (5) (Top) Confusion matrix for the test cases of a randomly split 80-10-10 train-validation test scenario. (Bottom) t-SNE visualization of the same 10% test cases considering the activation values of the last layer before the prediction layer as deep embedding [35]. The compressional and transverse wave signal samples are well separated.

mispositioned overlaps. However, as the spectral-temporal features within the sub-classes of transverse and compressional waves are often overlapping, we notice soft-boundaries among the intra-sub-classes along with higher overlaps among the samples. Although the separation among finer-phase types is not well-established, it is evident from this embedding projection that our proposed method is adept at learning higher-order separable representations.

5 CASE STUDY: NOVEL OPERATING CONDITIONS

In this section, we demonstrate practical use cases for our developed method in real-world applications when various novel scenarios emerge. In particular, we consider the following two novel scenarios: (1) If a monitoring agency adds a new station at a new location, will our method work without any calibration? (2) If new seismic sources occur in historically aseismic regions, will our method identify novel arrivals from a new source?

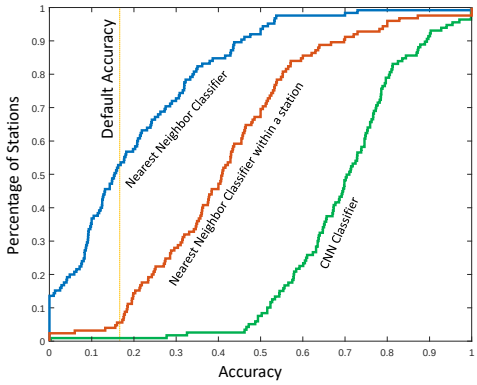


Figure (6) Performance comparison of FASER with naive nearest neighbor classifier for application in novel stations. The plot shows cumulative distribution function (CDF) for percentage of stations having smaller than a given accuracy. FASER achieves an approximately 72% **accuracy for the majority of the stations in contrast to** 17% **for nearest neighbor classifier.**

5.1 Performance at Novel Station

In our dataset, we have phase arrival signals recorded at 125 different stations across the world. To test how our method would perform at a novel station, we hold out signals at one station while training on signals at all other stations. We show empirical Cumulative distribution function (CDF) of stations for various accuracy levels in Figure 6. We compare two classifiers in this experiment. The nearest neighbor classifier that compares a test image with all training images to find the best match under the Euclidean norm, and labels the test image with the phase of the best match. The classifier achieves approximately default classification accuracy of 17% for the majority of the stations. This suggests that the nearest neighbor classifier cannot classify signals at a new station based on historical data at other stations. In contrast, FASER achieves an approximately 72% accuracy for the majority of the stations, suggesting single station analyses of seismic data may be useful in response to dynamically changing monitoring needs. The convolutional layers in our architecture extract local features from the images, unlike relying on a global one-to-one alignment of the images in the Euclidean space, as in the case of the nearest neighbor classifier.

The achieved accuracy of 72% for the majority of the station is significant for the IMS processing pipeline, as IDC analysts relabel 62% of the initial phases detected by the current automated algorithm. Moreover, only 38% of the initial phases remain the same, indicating that the initial phases are correct with 38% accuracy. FASER almost doubles the accuracy for a novel station of the current system's phase identification accuracy for an existing station. We show the held-out performance at each station in Figure 7(left). Most stations achieve higher accuracy (>0.7) when there are several closer stations. In contrast, isolated stations such as the one in the South Pacific suffer from a poorer performance.

5.2 Performance on Novel Sources

Most earthquakes originate along fault lines, while the rest of the earth is quieter. Novel seismicity in previously undocumented areas is intriguing. Hence, we evaluate our model by holding out regions of the earth for testing. For each held-out region, we train our model with data from the rest of the world and test the performance of our model on seismic events in that region. For this experiment, we divide the earth into 12× 12-degree grids. If a grid cell is not seismically active (i.e., not enough data), we exclude the region from testing. In Figure 7(right), we show the world map where the shaded grid cells are held out, one at a time. The average hold-out accuracy is 77.27%, with a standard deviation of 4.13%. More importantly, this suggests that our model is well suited for novel seismicity with little or no prior recorded events. We test on 46 cells of the 12×12 degree grids, which cover most of the known seismic events recorded at the NEIC (National Earthquake Information Center) for a three year period.

6 CONCLUSION

In this paper, we present a method to perform fine-grained seismic phase identification, which can be readily integrated into existing seismic signal processing pipelines. As seismology evolves into a big-data-driven science, deep learning methods are becoming an indispensable part of next-generation seismic monitoring systems. This work shows a practical example of integrating deep-learning methods in an existing semi-autonomous system to achieve complete autonomy. We demonstrate empirical evaluation of our method with a real-world dataset where it outperforms existing methods. Our method reduces the dependency on using array-based methods, which inhibits precise monitoring for regions with limited monitoring stations. It also reduces the dependency on large-collections of manually curated template sets and presents the opportunity of using transfer-learning for stations with limited labeled data. Due to the minimal preprocessing requirements and faster prediction generation, it is highly suitable for a real-time monitoring pipeline. In the future, with the use of larger datasets, more complex models would produce higher accuracy as well as better generalizability. Moreover, the introduction of interpretable models would be highly suitable for downstream analysis.

REFERENCES

[1] Ossama Abdel-Hamid, Abdel-rahman Mohamed, Hui Jiang, Li Deng, Gerald Penn, and Dong Yu. 2014. Convolutional neural networks for speech recognition. *IEEE/ACM Transactions on audio, speech, and language processing* 22, 10 (2014), 1533–1545.

[2] Rex V Allen. 1978. Automatic earthquake recognition and timing from single traces. *Bulletin of the Seismological Society of America* 68, 5 (1978), 1521–1532.

[3] Peter Bormann, Klaus Klinge, and Siegfried Wendt. 2009. Data analysis and seismogram interpretation. In *New manual of seismological observatory practice (NMSOP)*. Deutsches GeoForschungsZentrum GFZ, 1–102.

[4] Tianqi Chen and Carlos Guestrin. 2016. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. 785–794.

[5] Farhan Asif Chowdhury, Satomi Suzuki, and Abdullah Mueen. 2019. Structured Noise Detection: Application on Well Test Pressure Derivative Data. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2952–2960.

[6] Joshua Dickey, Brett Borghetti, William Junek, and Richard Martin. 2019. Beyond Correlation: A Path-Invariant Measure for Seismogram Similarity. *Seismological*

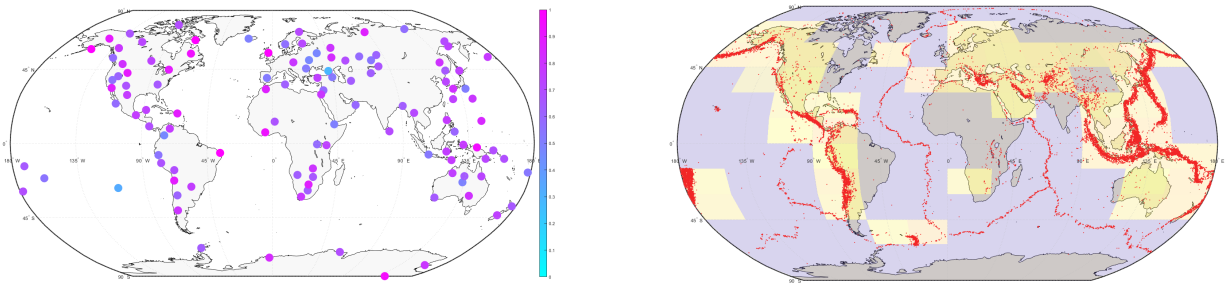


Figure (7) (Left) Testing with held-out stations. Each dot is an IMS station. The color bar shows held-out accuracy for a station. (Right) Testing with held-out regions. Each red dot is an event recorded at the NEIC (National Earthquake Information Center). We hold out shaded regions. Average hold-out accuracy across shaded regions is 77.27%, with a standard deviation of 4.13%, suggesting model efficacy at novel source regions.

Research Letters (11 2019). https://doi.org/10.1785/0220190090

[7] Ramin MH Dokht, Honn Kao, Ryan Visser, and Brindley Smith. 2019. Seismic event and phase detection using time–frequency representation and convolutional neural networks. *Seismological Research Letters* 90, 2A (2019), 481–490.

[8] Cicero Dos Santos and Maira Gatti. 2014. Deep convolutional neural networks for sentiment analysis of short texts. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*. 69–78.

[9] Luz García, Isaac Álvarez, Carmen Benítez, Manuel Titos, Ángel Bueno, Sonia Mota, Angel De la Torre, José C Segura, Gerardo Alguacil, Alejandro Díaz-Moreno, et al. 2016. Advances on the automatic estimation of the P-wave onset time. *Annals of Geophysics* 59, 4 (2016), 0434.

[10] Steven J Gibbons and Frode Ringdal. 2006. The detection of low magnitude seismic events using array-based waveform correlation. *Geophysical Journal International* 165, 1 (2006), 149–166.

[11] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. 2016. *Deep learning*. MIT press.

[12] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. 2009. *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media.

[13] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.

[14] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation* 9 (1997), 1735–1780.

[15] Baotian Hu, Zhengdong Lu, Hang Li, and Qingcai Chen. 2014. Convolutional neural network architectures for matching natural language sentences. In *Advances in neural information processing systems*. 2042–2050.

[16] Fazle Karim, Somshubra Majumdar, Houshang Darabi, and Shun Chen. 2017. LSTM fully convolutional networks for time series classification. *IEEE access* 6 (2017), 1662–1669.

[17] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

[18] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*. 1097–1105.

[19] L Küperkoch, Tobias Meier, J Lee, W Friederich, and EGELADOS Working Group. 2010. Automated determination of P-phase arrival times at regional and local distances using higher order statistics. *Geophysical Journal International* 181, 2 (2010), 1159–1170.

[20] Nadège Langet, Alessia Maggi, Alberto Michelini, and Florent Brenguier. 2014. Continuous Kurtosis-Based Migration for Seismic Event Detection and Location, with Application to Piton de la Fournaise Volcano, La RéunionContinuous Kurtosis-Based Migration for Seismic Event Detection and Location. *Bulletin of the Seismological Society of America* 104, 1 (2014), 229–246.

[21] Laurens van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-SNE. *Journal of machine learning research* 9, Nov (2008), 2579–2605.

[22] Ian W. McBrearty, Joan Gomberg, Andrew A. Delorey, and Paul A. Johnson. 2019. Earthquake Arrival Association with Backprojection and Graph Theory. (4 2019). http://arxiv.org/abs/1904.00980

[23] Abdel-rahman Mohamed, Geoffrey Hinton, and Gerald Penn. 2012. Understanding how deep belief networks perform acoustic modelling. In *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 4273–4276.

[24] S. Mostafa Mousavi, Charles A. Langston, and Stephen P. Horton. 2016. Automatic microseismic denoising and onset detection using the synchrosqueezed continuous wavelet transform. *GEOPHYSICS* 81, 4 (7 2016), V341–V355. https://doi.org/10.1190/geo2015-0598.1

[25] S. Mostafa Mousavi, Weiqiang Zhu, Yixiao Sheng, and Gregory C. Beroza. 2019. CRED: A Deep Residual Network of Convolutional and Recurrent Units for Earthquake Signal Detection. *Scientific Reports* 9, 1 (12 2019), 10267. https://doi.org/10.1038/s41598-019-45748-1

[26] Thibaut Perol, Michaël Gharbi, and Marine Denolle. 2018. Convolutional neural network for earthquake detection and location. *Science Advances* 4, 2 (2 2018), e1700578. https://doi.org/10.1126/sciadv.1700578

[27] Christopher Rawles and Clifford Thurber. 2015. A non-parametric method for automatic determination of P-wave and S-wave arrival times: application to local micro earthquakes. *Geophysical Journal International* 202, 2 (2015), 1164–1179.

[28] Anya M Reading, Weijian Mao, and David Gubbins. 2001. Polarization filtering for automatic picking of seismic data and improved converted phase detection. *Geophysical Journal International* 147, 1 (2001), 227–234.

[29] Zachary E Ross, Men-Andrin Meier, Egill Hauksson, and Thomas H Heaton. 2018. Generalized seismic phase detection with deep learning. *Bulletin of the Seismological Society of America* 108, 5A (2018), 2894–2901.

[30] Zachary E Ross, Yisong Yue, Men-Andrin Meier, Egill Hauksson, and Thomas H Heaton. 2018. PhaseLink: A Deep Learning Approach to Seismic Phase Association.

[31] Tara N Sainath, Oriol Vinyals, Andrew Senior, and Haşim Sak. 2015. Convolutional, long short-term memory, fully connected deep neural networks. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 4580–4584.

[32] Peter M Shearer. 2019. *Introduction to seismology*. Cambridge university press.

[33] M Ashraf Siddiquee, Zeinab Akhavan, and Abdullah Mueen. 2019. SeiSMo: Semi-supervised Time Series Motif Discovery for Seismic Signal Detection. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*. 99–108.

[34] Satish Sinha, Partha S Routh, Phil D Anno, and John P Castagna. 2005. Spectral decomposition of seismic data with continuous-wavelet transform. *Geophysics* 70, 6 (2005), P19–P25.

[35] Junyuan Xie, Ross Girshick, and Ali Farhadi. 2016. Unsupervised deep embedding for clustering analysis. In *International conference on machine learning*. 478–487.

[36] SHI Xingjian, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-Kin Wong, and Wang-chun Woo. 2015. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In *Advances in neural information processing systems*. 802–810.

[37] Haijiang Zhang, Clifford Thurber, and Charlotte Rowe. 2003. Automatic P-wave arrival detection and picking with multiscale wavelet analysis for single-component recordings. *Bulletin of the Seismological Society of America* 93, 5 (2003), 1904–1912.

[38] Chunting Zhou, Chonglin Sun, Zhiyuan Liu, and Francis Lau. 2015. A C-LSTM neural network for text classification. *arXiv preprint arXiv:1511.08630* (2015).