

[Artificial Intelligence in Geosciences 4 (2023) 84–9](https://doi.org/10.1016/j.aiig.2023.05.003)4

|  |  |  |
| --- | --- | --- |
|  | Contents lists available at [ScienceDirect](https://www.keaipublishing.com/en/journals/artificial-intelligence-in-geosciences) |  |
| Artificial Intelligence in Geosciences |
| journal homepage: [www.keaipublishing.com/en/journals/artificial-intelligence-in-geosciences](http://www.keaipublishing.com/en/journals/artificial-intelligence-in-geosciences) |
|  | | |

Original research articles

Blockly earthquake transformer: A deep learning platform for custom phase picking

Hao Maia,∗, Pascal Audeta, H.K. Claire Perryb, S. Mostafa Mousavic, Quan Zhanga

a *Department of Earth and Environmental Sciences, University of Ottawa, Ottawa, Canada, K1N 6N5* b *Canadian Hazards Information Service, Natural Resources Canada, Ottawa, Canada, K1A 0E7* c *Department of Geophysics, Stanford University, Stanford, 94305, USA*

|  |  |  |
| --- | --- | --- |
| A R T I C L E | I N F O | A B S T R A C T |
| *Keywords:*  Earthquake detection  Seismic phase identification Deep learning  Seismology | | Deep-learning (DL) algorithms are increasingly used for routine seismic data processing tasks, including seismic event detection and phase arrival picking. Despite many examples of the remarkable performance of existing (i.e., pre-trained) deep-learning detector/picker models, there are still some cases where the direct applications of such models do not generalize well. In such cases, substantial effort is required to improve the performance by either developing a new model or fine-tuning an existing one. To address this challenge, we present Blockly Earthquake Transformer(BET), a deep-learning platform for efficient customization of deep-learning phase pickers. BET implements Earthquake Transformer as its baseline model, and offers transfer learning and fine-tuning extensions. BET provides an interactive dashboard to customize a model based on a particular dataset. Once the parameters are specified, BET executes the corresponding phase-picking task without direct user interaction with the base code. Within the transfer-learning module, BET extends the application of a deep-learning P and S phase picker to more specific phases (e.g., Pn, Pg, Sn and Sg phases). In the fine-tuning module, the model performance is enhanced by customizing the model architecture. This no-code platform is designed to quickly deploy reusable workflows, build customized models, visualize training processes, and produce publishable figures in a lightweight, interactive, and open-source Python toolbox. |

**1. Introduction**

In the last five years, the global seismological community has wit-nessed the emergence of deep-learning (DL) as a promising candidate for undertaking large-scale seismic processing and modeling tasks in modern seismology (Mousavi and Beroza, 2022a). Among many other seismological applications, DL algorithms for earthquake signal detec-tion and seismic phase picking have had huge success, thanks to the existence of large-scale and publicly-available archived waveforms and phase labels (Mousavi and Beroza, 2022b). DL models are designed to learn intricate patterns concealed in seismic waveforms, identify seis-mic signals and their relations with desired targets such as seismic wave arrival times (i.e., P-arrival, S-arrival, etc.) without human interven-tion. One obvious application of DL phase pickers is in the automation of seismic data processing in near real-time earthquake monitoring workflows that are routinely operated by regional, national and inter-national network operators to improve the accuracy and completeness of earthquake and seismic event catalogues (Yeck et al., 2021). In addition to improving magnitude-frequency distributions, which are critical for seismic hazard assessment and forecasting (Beroza et al.,

∗ Corresponding a[uthor.](mailto:hmai090@uottawa.ca)

*E-mail address:* [hmai090@uottawa.ca](mailto:hmai090@uottawa.ca) (H. Mai).

2021), DL phase pickers may have utility in verification seismology, specifically in the identification of P- and S-phases of very low-yield (e.g., 10−4−10−1kT), low-magnitude (e.g., *𝑚𝑏*1−3) nuclear explosions. Such events are best recorded at local distances (*<*150–200 km) by regional seismic network operators (Koper, 2019). Algorithms based on DL phase pickers like the ones presented here can expedite the creation of reliable earthquake catalogs. Recent reviews of the state-of-the-art applications show that DL-based models outperform classical character-istic function-based models in many scenarios, such as fast detection of P and S arrivals in a global benchmark dataset, and for finding S-waves that are obscured by the coda of earlier phases (Mousavi et al., 2019a; Woollam et al., 2022; Münchmeyer et al., 2022; Ma et al., 2020, 2023). Although DL applications have shown promising performance and resulted in interesting outcomes, the path to developing more effective DL models is not always clear and efforts have been uneven. The model development procedure is often fraught with difficulties with the consequence that any new DL project requires the investment of massive amounts of labor and research resources. DL-based approaches for seismic data processing face similar issues, but most of the steps

<https://doi.org/10.1016/j.aiig.2023.05.003>  
[Received 14 March 2023; Received in revi](https://doi.org/10.1016/j.aiig.2023.05.003)sed form 12 May 2023; Accepted 26 May 2023 Available online 30 May 2023

2666-5441/© 2023 [The Authors. Publishing services by Elsevier B.V.](http://creativecommons.org/licenses/by-nc-nd/4.0/) on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC

BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

*H. Mai et al.*  *Artificial Intelligence in Geosciences 4 (2023) 84–94*

have unique prerequisites that cannot be migrated from previous work. Standard DL workflows can be summarized in four steps:

1. Pre-processing: This step prepares the training and validation data, and includes data collection, cleaning, and labeling. Bench-mark seismic datasets such as STEAD (Mousavi et al., 2019a), INSTANCE (Michelini et al., 2021), and DiTing (Zhao et al., 2022) provide high-quality labeled data for training and model building. However, the number of these benchmark datasets and their applications is limited. Alternatively, applying a cus-tom seismic sample toolbox such as QuakeLabeler (Mai and Audet, 2022) can generate labeled datasets for training based on users’ demands. The quality and distribution of the training data ingested by the DL algorithm have a large impact on the performance of the resulting model. Good training data should have qualities such as relevance, minimal labeling errors, wide distribution representing various real-world classes, few missing or repeated values, etc.

2. Training: In the training step, the pre-processed seismic data are passed to a specific DL architecture to find patterns within the seismic signals and learn to pick seismic phase arrival times. Many DL models have been proposed recently, including U-Net models such as BasicPhaseAE and PhaseNet (Woollam et al., 2019; Zhu and Beroza, 2019). Convolutional and Recurrent Neu-ral Network (CNN and RNN, respectively) earthquake event detectors include CRED and DeepPhasePick (Mousavi et al., 2019; Soto and Schurr, 2021), as well as more complex architec-tures like Earthquake Transformer (EqT) that adds self-attention layers (Mousavi et al., 2020).

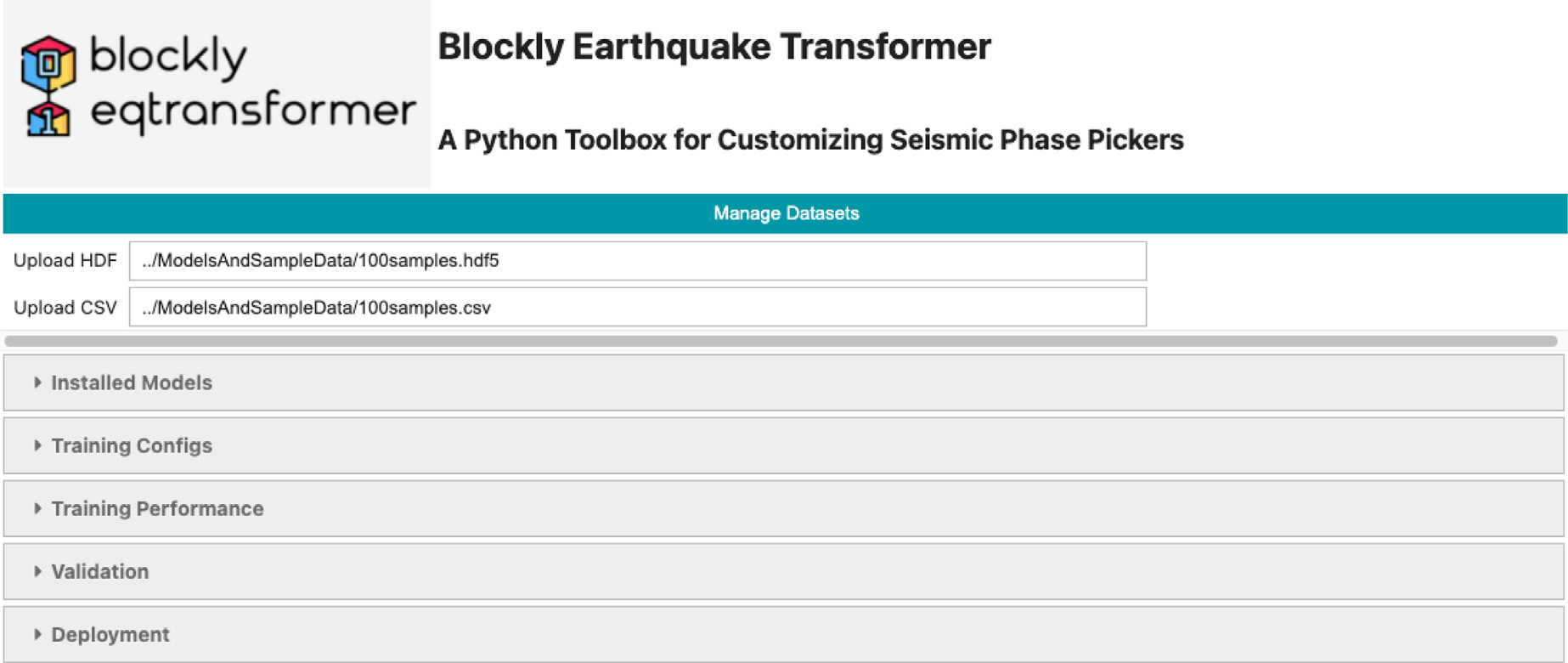
3. Validation: Before building a final phase picker model, the val-idation process allows testing the model against data that have not been used for training. In general, 20% of the entire dataset are retained and excluded from the training dataset. The perfor-mance of the model on this validation set represents how likely the model is to succeed in correctly predicting phases in the face of independent seismic data. A trained model could either be approved (i.e., deemed satisfactory), tuned, or rejected based on the validation results (Livieris et al., 2020). The validation stage should not be time-intensive; however, if the model is rejected, the project returns to the training stage to revise the architecture and parameters. This training and validation loop continues until a model is approved.

4. Deployment: The approved DL picker moves on to the model development stage and is then applied to different target re- gions/datasets, or at various scales ranging from local and re- gional scales to global scales, depending on the prescribed ar- chitecture. It is common that a performing DL model fails to yield satisfactory results using a new benchmark dataset or a new target region (e.g., Woollam et al., 2022), meaning the model has weak generalization. This may be due to different data distributions between the training dataset and the data collected from the regions of interest; DL models cannot guar- antee performance if the new dataset is not statistically similar to the training dataset. Such generalization issues almost always arise, and there is always a limit for the performance of a DL model (Münchmeyer et al., 2022).

To shorten the development cycle of DL phase pickers, transfer learning (TL) has been shown to be an effective solution for efficient re-producibility. TL builds new DL picker models from pre-trained models using partially or fully pre-trained hyper-parameters (weights) and net-work architectures to reduce the training time. This is a rapidly growing field in DL research, which reduces the massive computation require-ments and time resources, and improves upon data-deficient tasks. A few early attempts have proven the success of TL, especially benefiting research projects with too few data to train a full-scale model (Lapins

85

*H. Mai et al.*  *Artificial Intelligence in Geosciences 4 (2023) 84–94*



**Fig. 1.** Screen capture of the blockly earthquake transformer dashboard. BET provides a complete interactive workflow. Users can run BET’s modules by defining their parameters

interactively.

reloaded to continue training or to restore a previous model version) and diagnostic figures. Based on these utilities, users can either launch and apply pre-trained models to new datasets or design new DL or TL models in a short time frame.

*2.1. Data management*

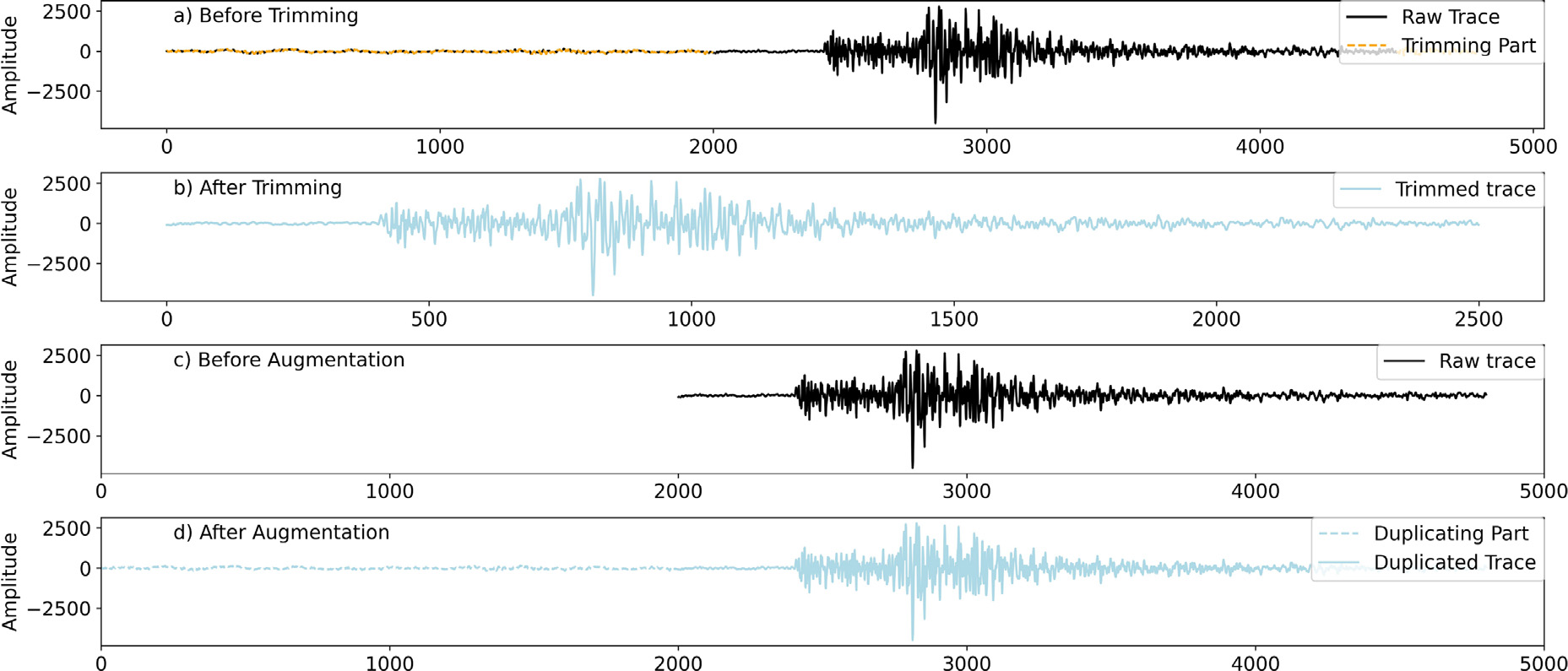
Preparing a good quality dataset for neural network training and model building is a challenging and labor-intensive task. It is often complicated to set up a robust quality control procedure to eliminate erroneous labels and make the distribution of training data as diverse as possible. The BET data management module allows users to upload seismograms, which will be saved in a single Hierarchical Data Format version 5 (HDF5) archive file. The related information (metadata) should be supplied in a comma-separated values (CSV) file. If the file format of the uploaded dataset is compatible with STEAD (Mousavi et al., 2019a), BET can automatically parse it. However, if there are any complications, it is recommended to use QuakeLabeler (Mai and Audet, 2022) first to convert the dataset into the STEAD format. If the configuration of the uploaded data (e.g., sample length, input channel size) conflicts with the model configuration requested, BET’s data management module will initialize the built-in pre-processing functions to attempt to fit the dataset into the model’s prerequisites, if possible. For instance, if the length of raw seismic traces is less than the model’s pre-defined channel length, the data augmentation module will be activated to adaptively duplicate a fragment of the trace away from the seismic signals to pad to the required sample length (see an example in Fig. 2); if the length of raw seismic traces is non-uniform and beyond the required input shape, a trimming method is applied to crop the trace where seismic signals exist (Fig. 2).

*2.2. Model setup*

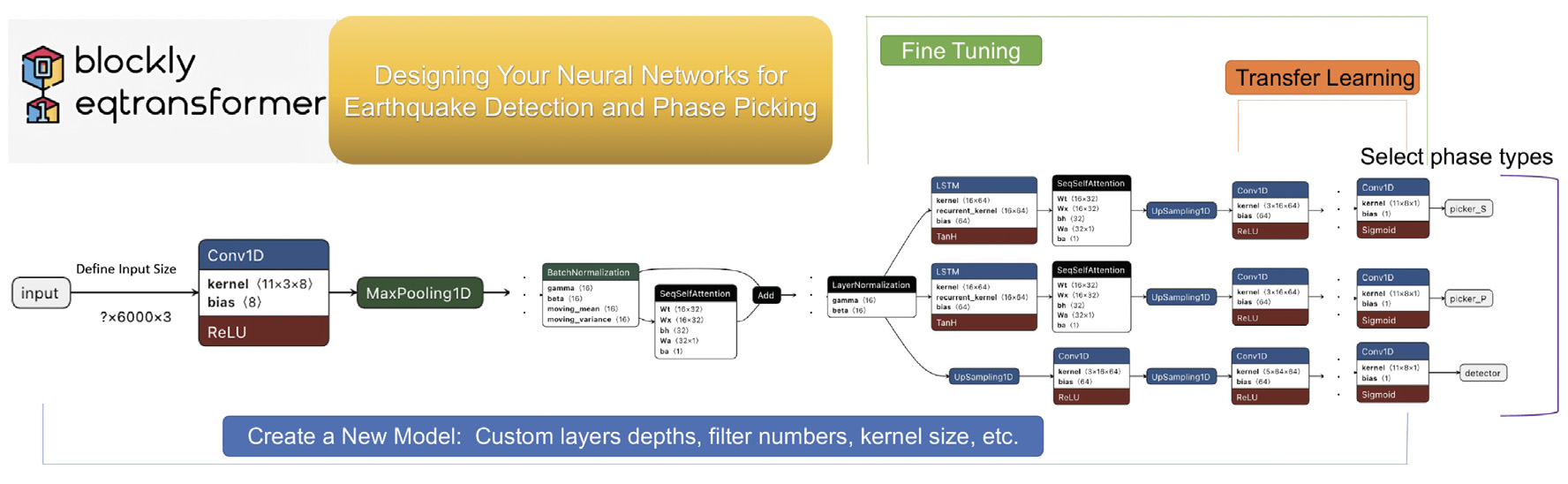
Two types of model installation options are available in BET: (1) load a pre-trained model, and (2) create a new model. The first option will initialize a DL model based on the selected pre-trained network structure and hyper-parameters. The advantage of this option is that training data size can be minimal since the common seismic signal characteristics have been previously understood by the pre-trained models. Later, users can apply transfer learning, fine-tuning, and/or new model methods to customize models using a ‘‘warm start’’ (i.e., an optimal solution to a simpler optimization problem that sets initial values based on parameter information gained from a previous training dataset). The option to create new models is targeted to advanced users

86

*H. Mai et al.*  *Artificial Intelligence in Geosciences 4 (2023) 84–94*



**Fig. 2.** Example of trimming and data augmentation. In 2a, the model’s input size requires 2500 sample points, but the raw data has 4500 sample points per trace. The BET trimming module adaptively trims 2000 non-signal sample points from both ends of the raw trace. 2b is the trimmed trace which fits the input size. In 2c, another model’s input size requires 4800 sample points, but the raw trace only has 2500 sample points per trace. The BET data augmentation module automatically adaptively duplicates 2300 non-signal sample points that are pre-appended to the raw trace. 2d shows the trace after data augmentation.



**Fig. 3.** Schematic diagram of BET’s model setup framework. BET provides a visual interface for building and training neural networks. By default, the ’Create a New Model’module allows users to select the number of layers, the number of units in each layer, etc., to fully design a new model. ’Transfer Learning’ and ’Fine Tuning’ modules can load pre-trained model architectures and weights, then select to re-train part of the layers to improve the model’s accuracy in a short time.

long-short-term memory blocks. These arguments control the frame-work of the newly created DL model, which will generate an EqT-like DL picker. This customization can be done via the BET dashboard, which means that no programming is required. Furthermore, BET also provides the complete API hidden in its Jupyter notebook, giving advanced users the possibility to rewrite or extend the other parts of the network architecture, e.g., adding new types of neural network units.

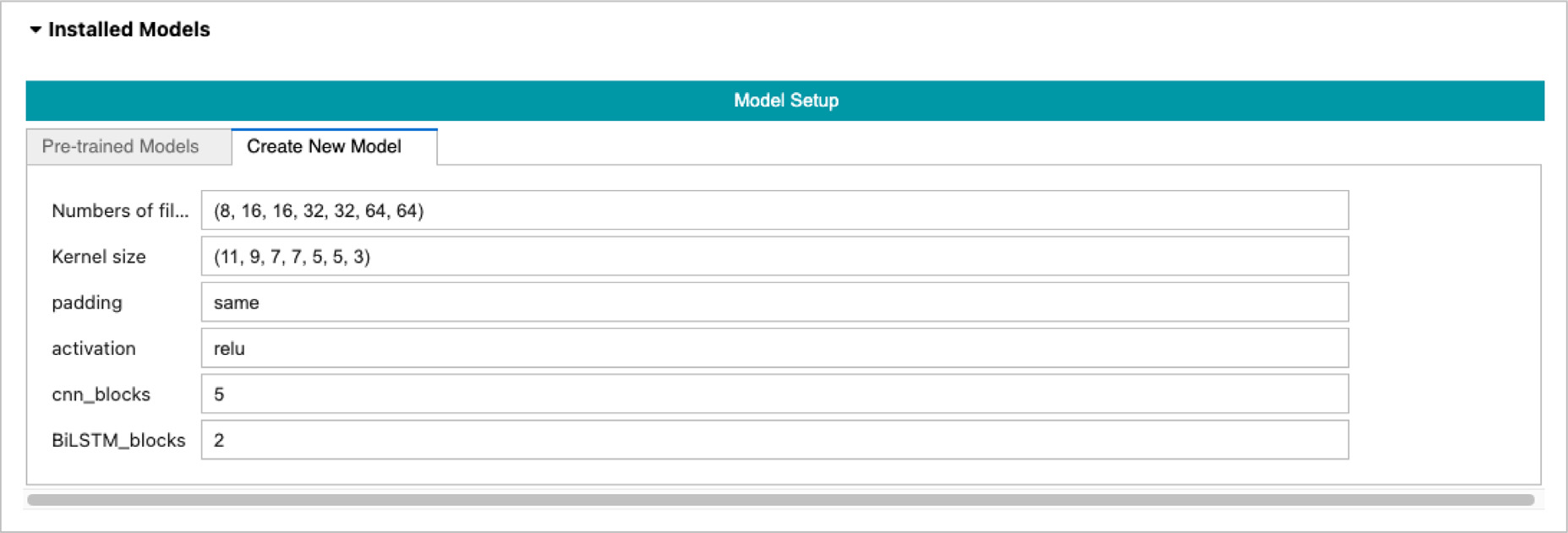
*2.3. Training configuration*

Designing a training configuration is the last step before training a DL model, and can be done interactively through the BET dash-board (‘‘Training Configs’’; see Fig. 5). Most of these arguments are straightforward, and their descriptions are shown in Table 1. The most important part of the training configuration is the designing of input/output (I/O) channels. Here we give a brief overview to help users understand the mechanism of I/O channels in a DL picker.

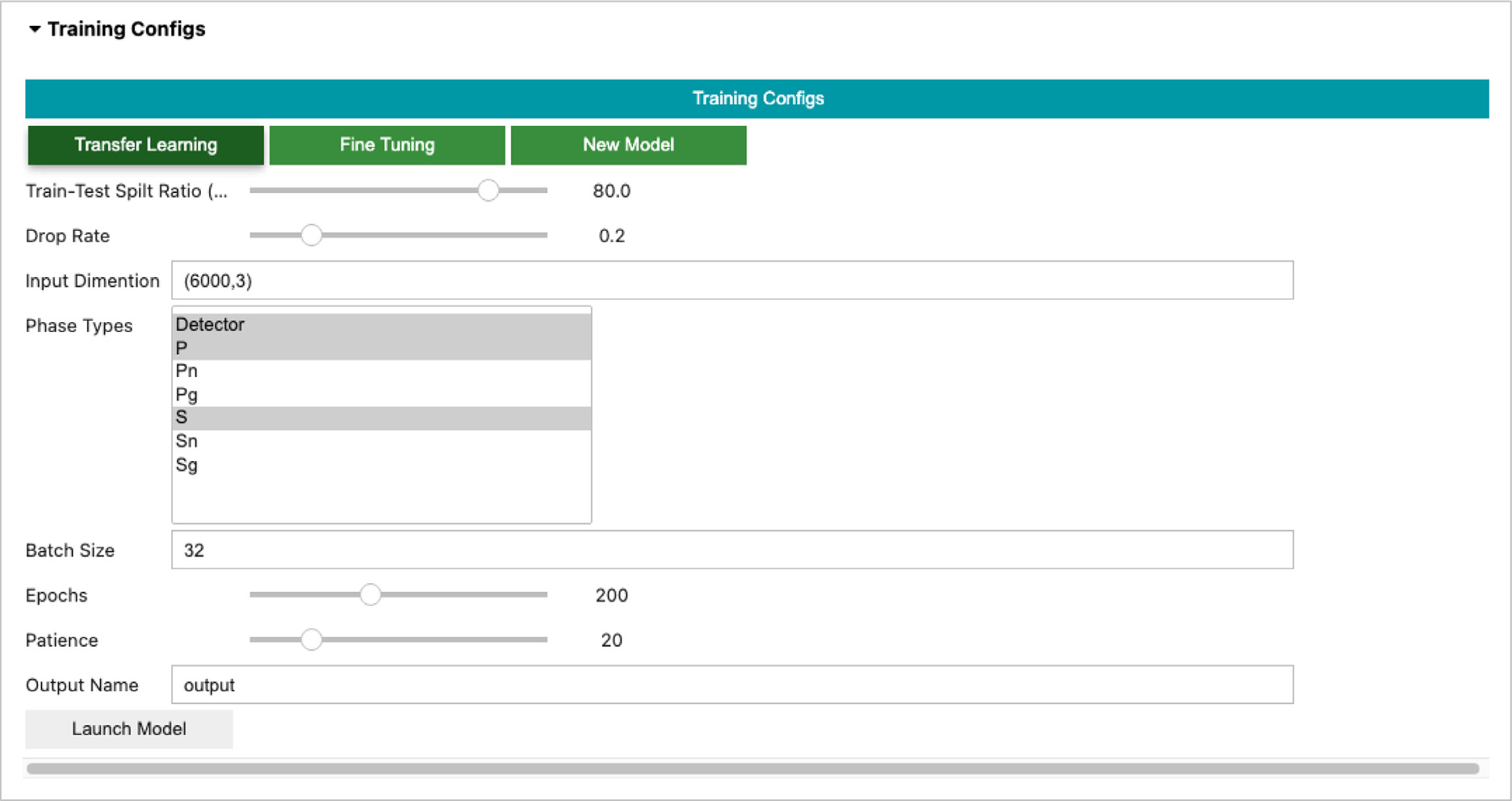
*2.3.1. Input channel*   
 In a DL project, prepared datasets must be compatible with I/O channels of the pre-trained model, e.g., the EqT model receives 3-component seismic traces with 6000 sample points each (6000 ∗ 3). If the new datasets do not satisfy the model’s input shape, errors will be raised during training unless the data have been correctly trimmed

87

*H. Mai et al.*  *Artificial Intelligence in Geosciences 4 (2023) 84–94*



**Fig. 4.** Screen capture of BET’s model setup dashboard, with the ‘‘Create New Model’’ tab activated. Users can customize the model architecture via this dashboard to build new models at different scales.



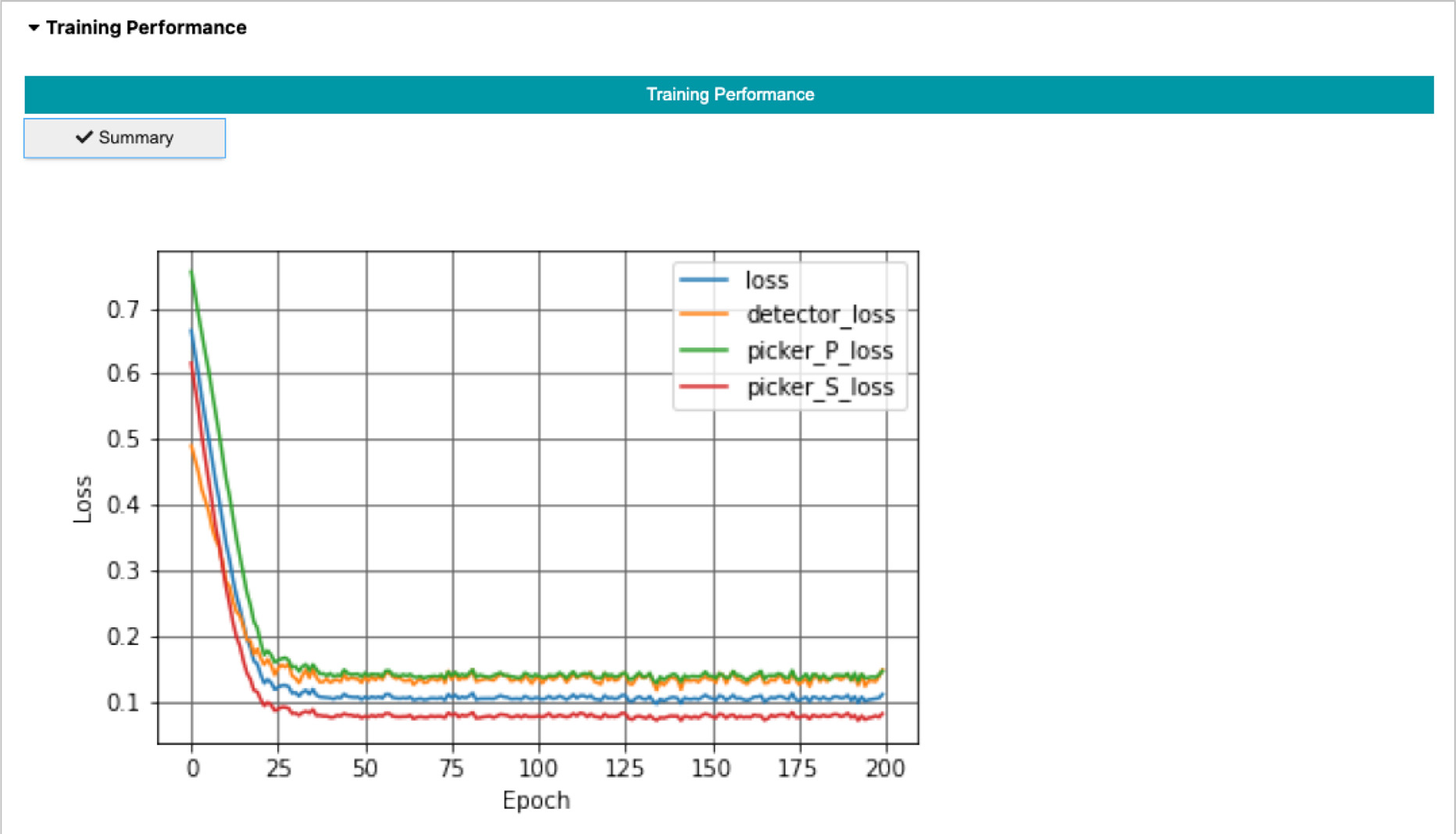
**Fig. 5.** Screen capture of BET’s training configuration dashboard. In this dashboard, users can select different training options, i.e., TL, fine-tuning or training a new model, and other training related arguments.

|  |  |
| --- | --- |
| **Table 1**  Training configuration. |  |
| Argument name | Description |
| *Train–Test Split Ratio Drop Rate*  *Batch Size*  *Epoch*  *Patience*  *Output Name* | Retained data used to test model performance.  Dropping out rates (to avoid over-fitting).  Amount of training examples utilized in one iteration. One training iteration.  The number of epochs to wait before an early stop if there is no progress.  Folder name of the generated DL/TL model. |

Experimental results showed that, in this way, convergence of the model is faster during training. For the phase picker channels, users can select the default P and S pickers, or additional channel identifiers, i.e., P, Pg, Pn, S, Sg, Sn, etc, if those labels are available in the training metadata (i.e., catalog picks). BET will automatically rewrite the output channels based on these selections. Alternatively, users can use the default P and S labels as surrogate for other binary labels, for example, blasts and natural earthquakes, to start training a new model for the discrimination of anthropogenic and natural events.

88

*H. Mai et al.*  *Artificial Intelligence in Geosciences 4 (2023) 84–94*



**Fig. 6.** Screen capture of training performance. The loss function for each output channel and the overall loss decrease linearly with epoch until they reach a stable value, where

loss no longer decreases.

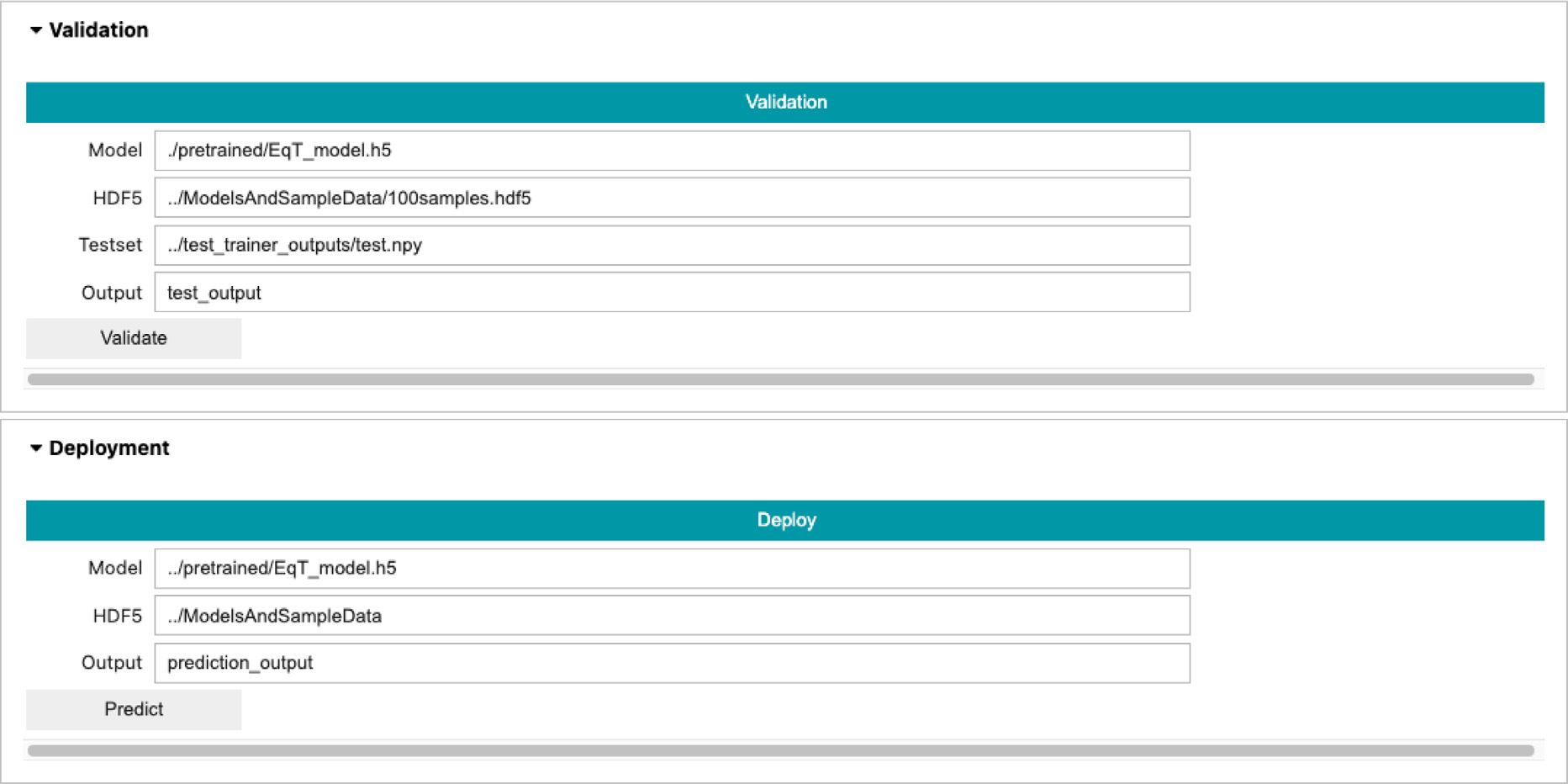
*2.4.2. Validation and deployment*   
 Validating the newly trained DL model is essential to ensure its performance and accuracy (Eelbode et al., 2021). During validation, a sufficient quantity of samples in proportion to the overall size of the dataset, but not considered in the training set, is used to validate the model’s performance. These samples are preserved from the split dataset. If the validation results are as good as in the training dataset, the model has not been over-fitted and the new picker is ready to be deployed. Although there are no standards for evaluating a model, BET provides three types of supporting files to help users evaluate their trained models. In the specified output folder, users can find detailed test results to evaluate the model’s performance, i.e., a summary table (\*.CSV) containing all the picked results, a text file (\*.TXT) reporting the configuration parameters for prediction and model performance, and a figure folder containing the picked arrivals (default: 10 exam-ples of picks). BET also offers a deployment dashboard to quickly apply newly trained pickers to large-scale datasets (see Fig. 7). When the deployment is complete, BET exports several sample graphs and human-readable documents to help users visualize and understand the model predictions (see Fig. 13). These extensions provide ways to visualize and understand model predictions, avoid repetitive labor and let users focus on analyzing performance.

**3. Prospective applications**

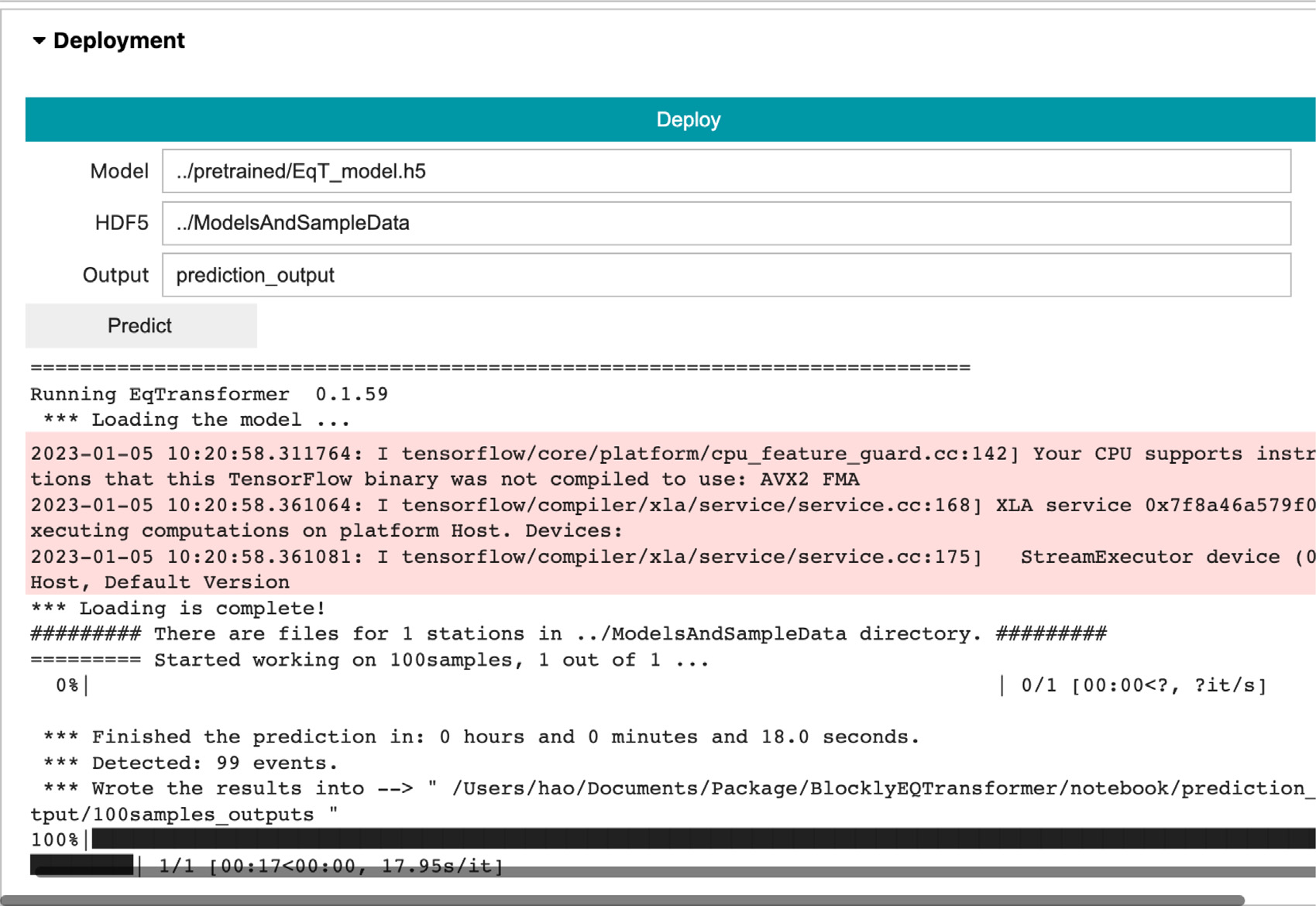
In this section, we present three examples that cater to novice, in-termediate and advanced users, respectively, to illustrate the flexibility and applicability of BET for phase-picking projects. For novice users, BET provides a convenient way to understand how deep learning mod-els can be deployed in earthquake detection and phase-picking projects. For intermediate users, BET can be used to train new deep learning and/or transfer learning models to address specialized phase-picking tasks. For advanced users, they can fully customize deep-learning mod-els to create novel deep-learning pickers. The focus of this section is to give users a quick overview of BET’s capabilities, rather than to showcase the accuracy of any particular trained model. We will focus on various case studies in future work.

89

*H. Mai et al.*  *Artificial Intelligence in Geosciences 4 (2023) 84–94*



**Fig. 7.** Screen capture of validation and deployment dashboard. When the user finishes a training process, the validation and/or deployment dashboard can be activated based on training results, e.g., input training output folder and other user-defined arguments.



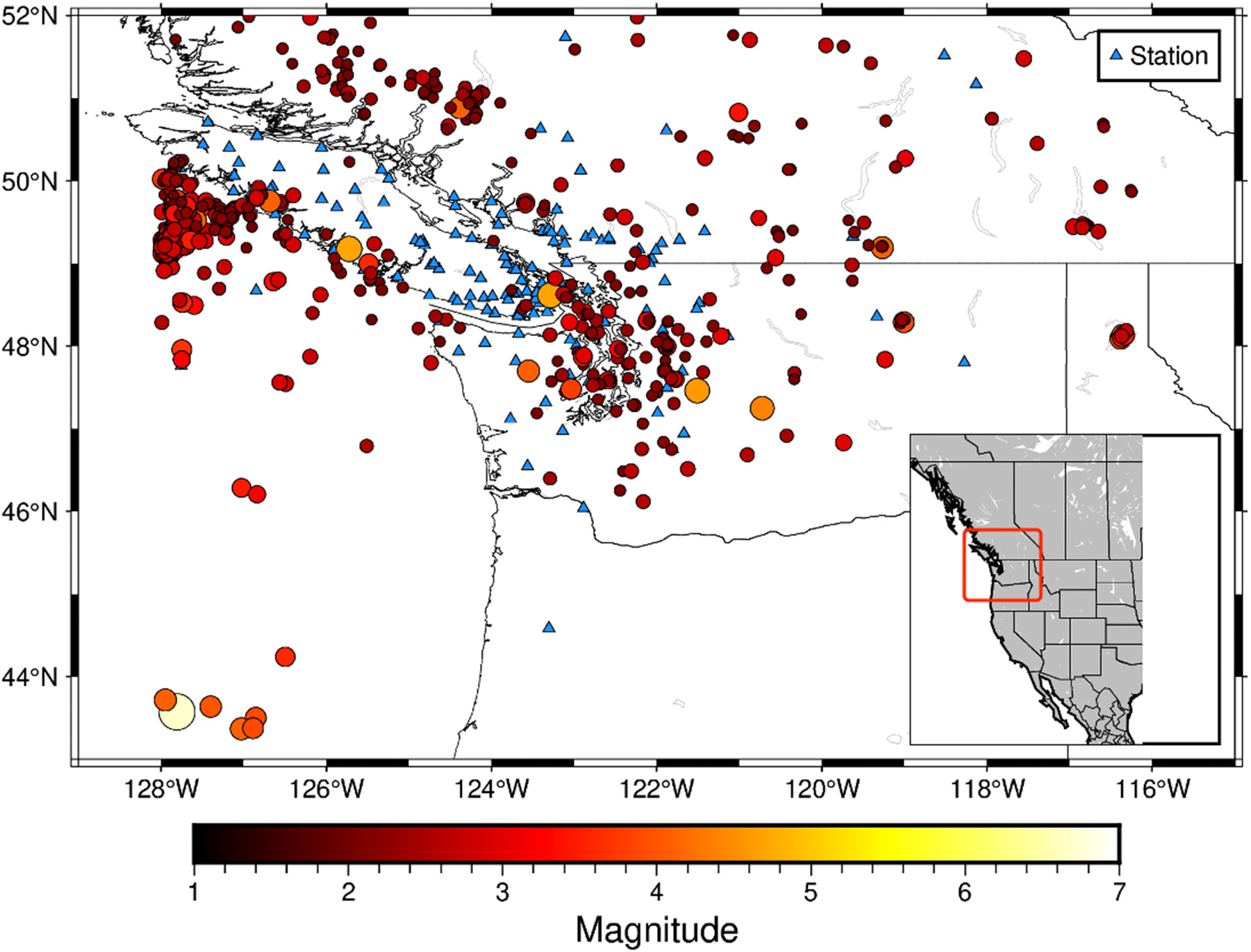
**Fig. 8.** Screenshot of using a pre-trained model to detect and pick earthquake events. Users can use BET’s built-in pre-trained models (and weights) to predict earthquake events. Following the procedure outlined in this paper, the user defines the file path and selects the ‘Predict’ button. BET automatically detects the earthquake events in the dataset and generates results and example visuals in the output folder. The deployment stage generally takes a few minutes, depending on the size of the dataset and computational resources.

*3.2. Training models on new datasets*

Deploying an AI picker in a new scenario (e.g., different input channel requirements, specific regions with no prior training data, etc.) may require the laborious adaptation of previously published DL or TL models. BET is designed to facilitate the training of new deep-learning models on user-defined datasets. The first step for training a new model is dataset preparation in the STEAD format. We illustrate the training

90

*H. Mai et al.*  *Artificial Intelligence in Geosciences 4 (2023) 84–94*



**Fig. 9.** Geographic data distribution of earthquakes of magnitude M ≥ 2, from the training dataset extracted from the NEDB catalog (Natural Resources Canada, 1985) from

2015-01-01 to 2019-12-31. Blue triangles represent the seismic stations. Circles represent earthquakes, with size and color scaled to event magnitudes. The NEDB dataset contains

single trace P- and S-labeled, manually-picked phases and is used in this study as a training dataset to introduce BET’s training module.

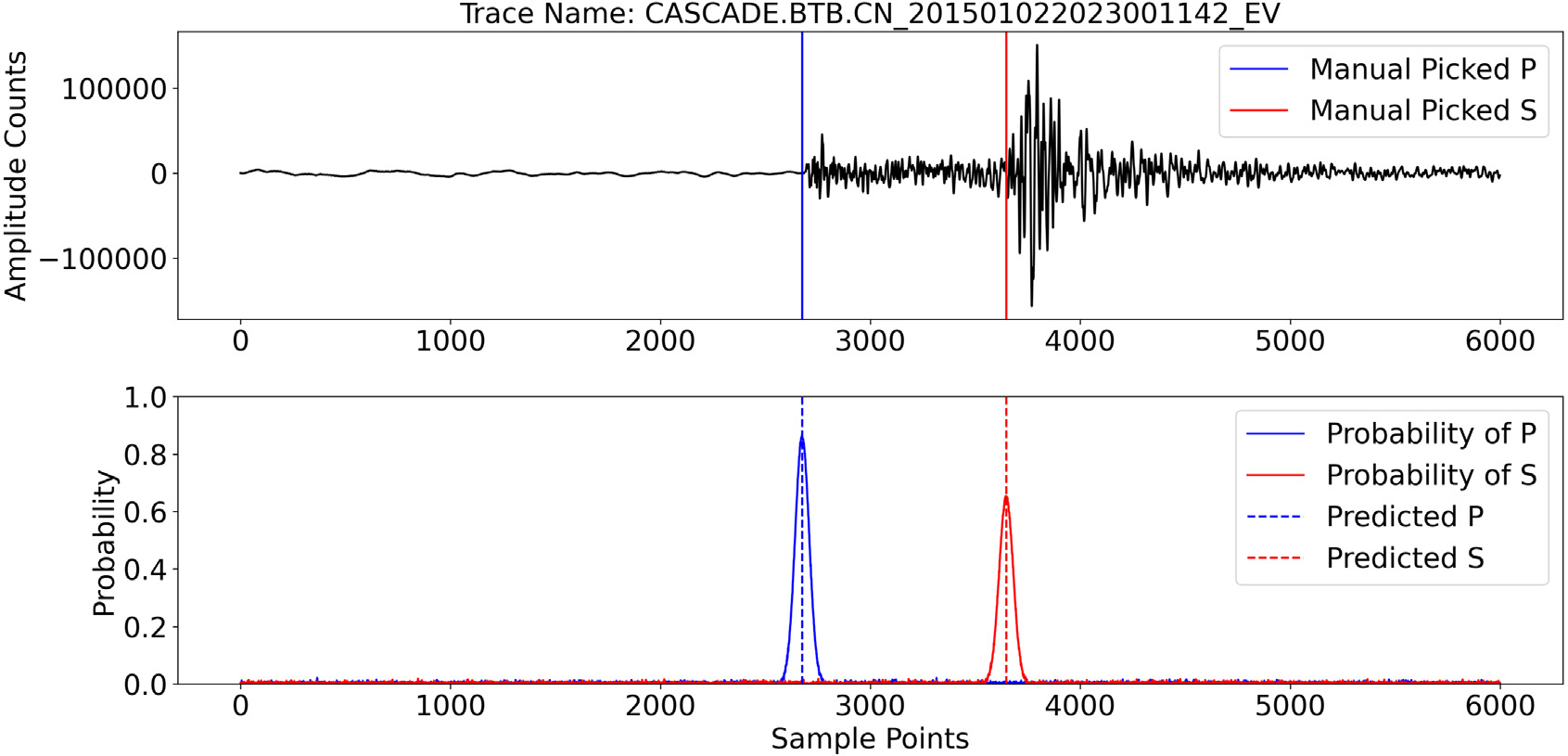
2015-01-01 and 2019-12-31. A magnitude threshold of 2 was chosen based on previous studies (Yeck et al., 2021). We first use this dataset to evaluate the pre-trained EqT model for this new region (Cascadia) and use transfer learning to improve the model.

The second step is to define the architecture of the model. Here we select the ’Transfer Learning’ function for illustration. In the new dataset, each sample uses one channel of input trace; the vertical component for P arrivals and one horizontal component for S arrivals. This input structure differs from most published DL models, making direct application of the pre-trained EqT model challenging. However, for transfer learning on the EqT model, the user needs only to specify that the input shape is that of the original EqT model (i.e., 6000 ∗ 3) and set up other arguments as shown in Table 2. The new dataset will be automatically formatted as the EqT model, and fed into training in BET’s backend. In this case, EqT pre-trained model weights are acceptable as a warm start to accelerate model convergence.

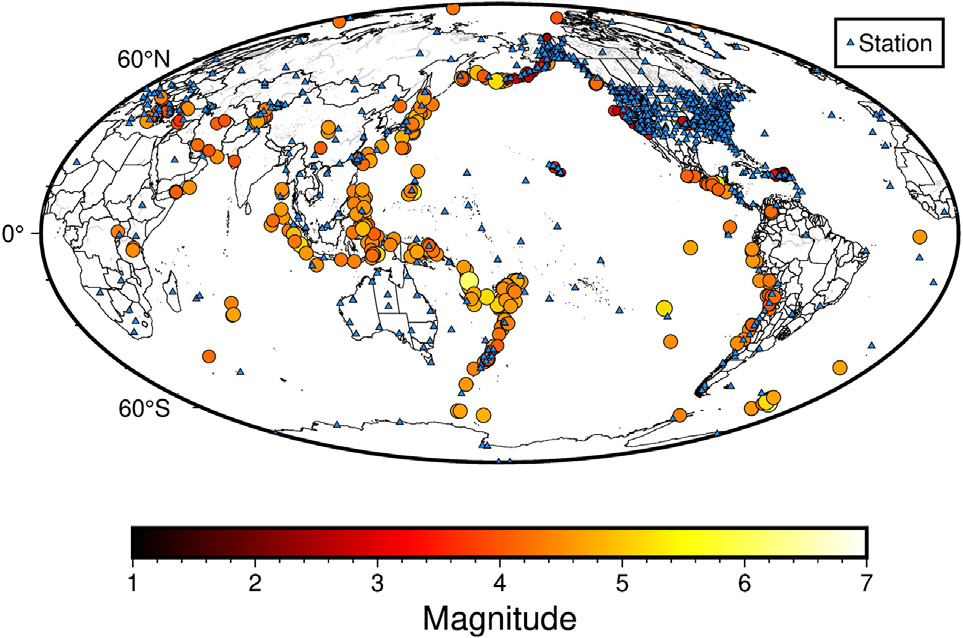
The transfer learning and fine-tuning options are suitable for users who do not have access to arrays of GPUs for computation. A single desktop-level GPU (e.g., GTX 1080I) or workstation with CPU cores (e.g., 8 cores of Intel i7) can complete a training step in a few hours. When training is complete, the next step is to test the accuracy of the trained picker in the validation dashboard. We use the validation dataset (2721 samples), which gets automatically split when loading the dataset, to test BET’s prediction accuracy. Performance comparison of the original EQTransformer (V1.59) model and the BET newly-trained transfer-learning model (Table 3), demonstrates that the new picker has improved detection ability. The original EqT model’s low accuracy suggests that the NEDB data distribution differs from the STEAD dataset. Fig. 10 shows a predicted result from the newly trained

91

*H. Mai et al.*  *Artificial Intelligence in Geosciences 4 (2023) 84–94*



**Fig. 10.** Representative predicted P and S arrival in Cascadia test dataset. The trained picker correctly identifies P and S phase from a N-component seismic trace. The human-reviewed P and S arrival times are taken from the NEDB catalog at station CN.BTB for a magnitude 2.0 event on 2015-01-02 at an epicentral distance of 78 km. The picker predicts a high probability (*>*0.6) at the correct sample position.



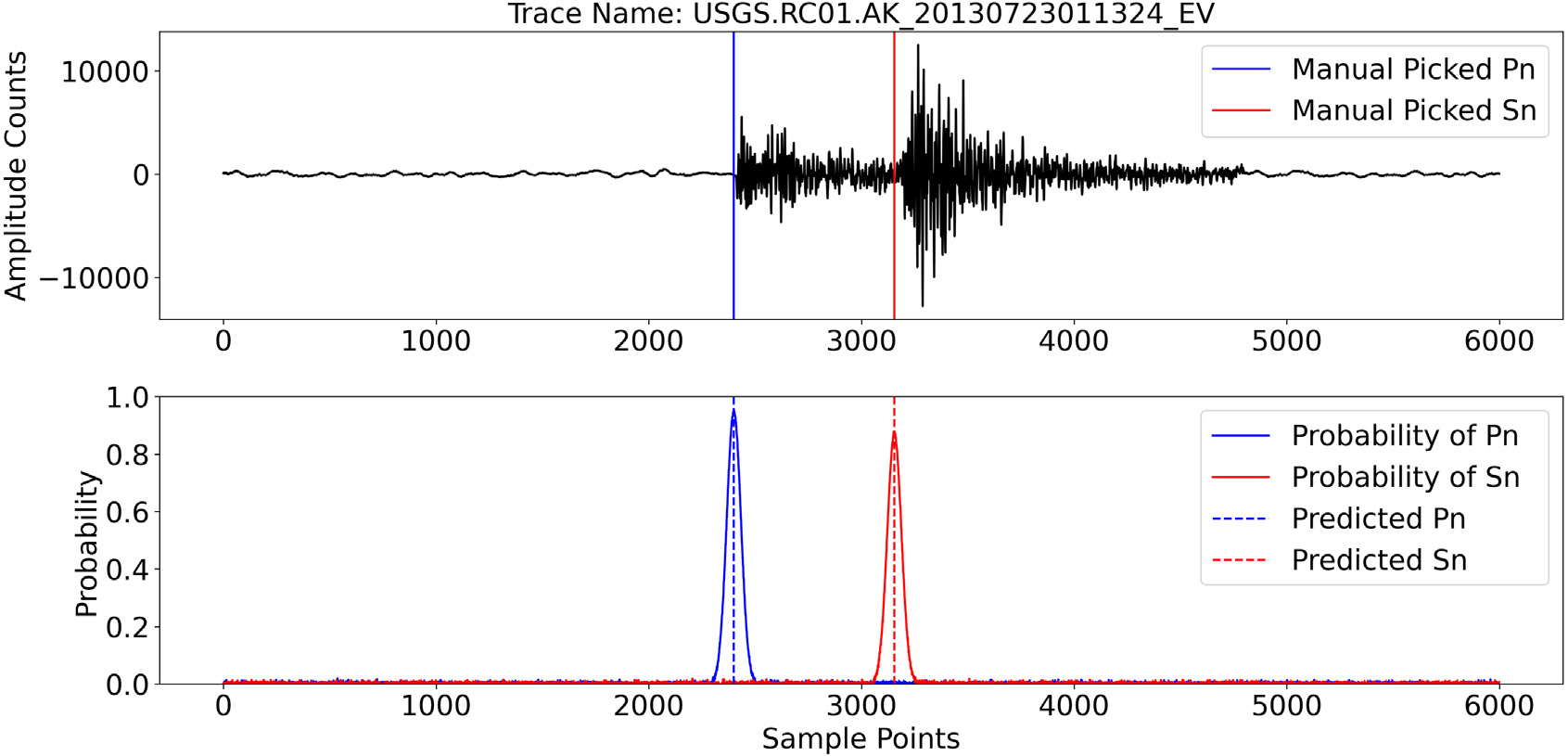
**Fig. 11.** Geographic data distribution of the USGS dataset (Cole and Yeck, 2022) used in this application. The blue triangles represent the available seismic stations. The circles represent earthquakes color-coded by magnitude. The USGS dataset contains human-reviewed P, Pn, Pg, S, Sn and Sg labeled samples, and is used here as a training dataset to test BET’s applicability.

6000 ∗ 3 and use default values for all training settings. We download three months (i.e., from June to September 2013) of data from the U.S. Geological Survey (USGS) machine learning dataset (Cole and Yeck, 2022) and convert it into STEAD format. This USGS dataset is publicly available and contains a wide range of human-reviewed samples of the above 6 phase types. In this project, we set 0.8 as the training-test-split ratio to produce 31,254 training samples and 7814 validation samples (the data distribution map is shown in Fig. 11). In the USGS dataset, each phase is centered in the trace, i.e., the original waveform is sampled at 40 Hz with a start time 60 seconds before the first P-wave arrival, so the arrival time is at the 2400 sample point. Therefore, the original ground truth labels are fixed at a static position, which will cause problems for the model in understanding the spatial features of the picking tasks. However, once the data loading function is activated, BET will automatically pre-process the dataset in its backend, augment sample length to the input shape and randomize arrival positions to avoid biased training.

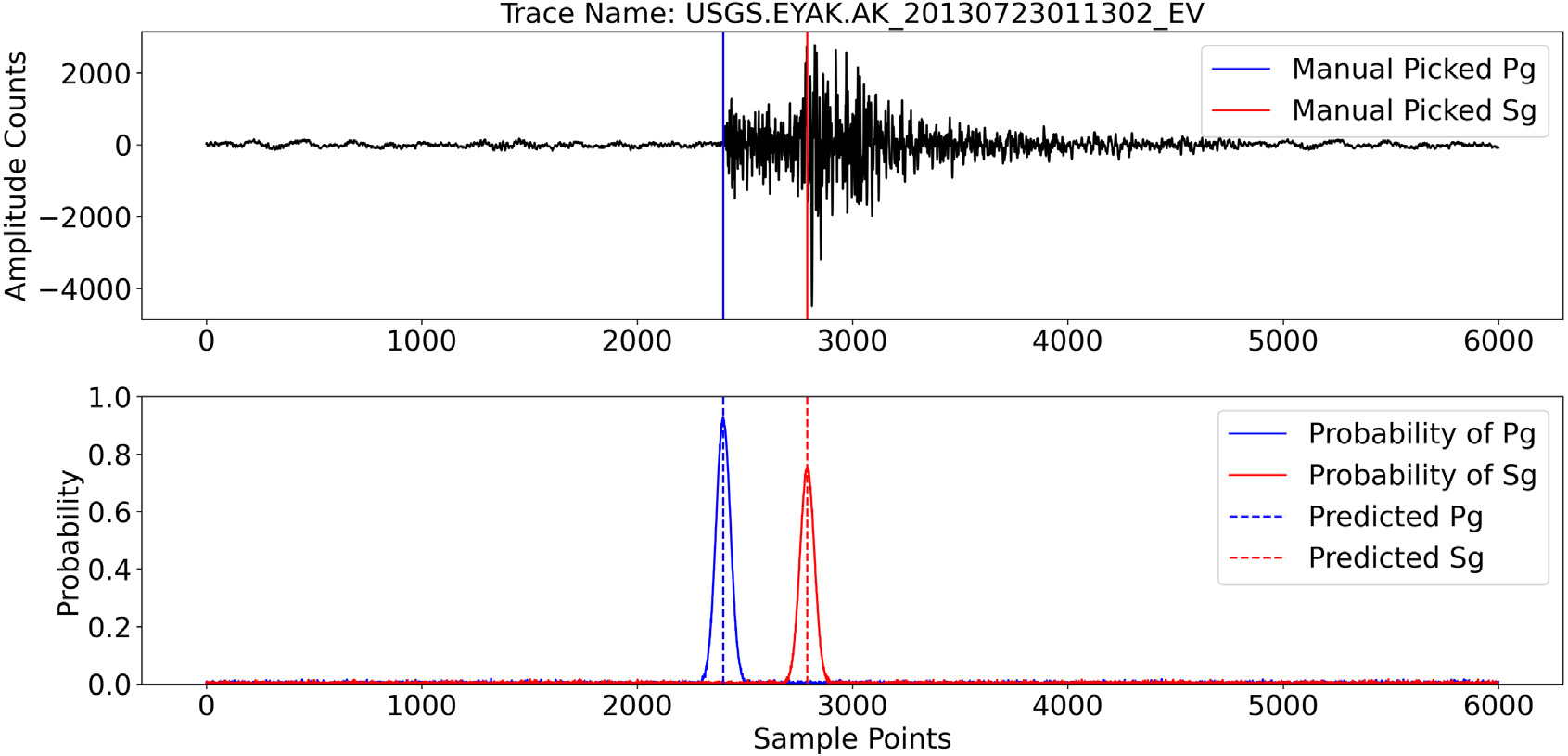
In this example, training reaches its best performance after only 7 epochs, at which point the training is stopped automatically (after 2 h 20 min of training time). The pre-trained model (EqT version 1.59

92

*H. Mai et al.*  *Artificial Intelligence in Geosciences 4 (2023) 84–94*



**Fig. 12.** Representative predicted Pn and Sn arrivals in the USGS dataset (Cole and Yeck, 2022).The fine-tuned picker successfully detects a Pn-arrival and a Sn-arrival at station AK.RC01 from the USGS validation dataset. This event is human-reviewed with a magnitude of 2.9 recorded at a distance of 161 km from the epicenter.



**Fig. 13.** Representative predicted Pg and Sg arrivals in the USGS dataset (Cole and Yeck, 2022).The custom BET picker successfully detects a Pg-arrival and a Sg-arrival at station AK.EYAK from the USGS validation dataset. This event is human-reviewed with a magnitude of 2.9 recorded at a distance of 65 km from the epicenter.

new channels can be utilized for training. More complex extensions are also allowable in BET’s scripting level and outlined in the user manual.

**5. Conclusion**

The application and development of DL models are steadily growing in seismology. BET provides a user-friendly platform for engaging a broader range of users, including those without machine learning or coding experience. BET allows researchers to explore DL pickers without having to worry about in-house computational power and other technical concerns. BET will prove useful in improving model performance and lowering the magnitude of completeness of exist-ing seismic catalogues. Similar packages can be developed for other seismological tasks, including those in exploration seismology, where fine-tuning pre-trained models by synthetic data is even more common due to the scarcity of large-scale labeled training datasets. Other appli-cations of BET are in the discrimination between natural earthquakes and anthropogenic events, where separate training datasets may be developed for each event types. This package will be both practical for students to understand deep learning applications and convenient

93

*H. Mai et al.*  *Artificial Intelligence in Geosciences 4 (2023) 84–94*

found at [https://www.sciencebase.gov/catalog/item/6127b30fd34e40 dd9c05094c](https://www.sciencebase.gov/catalog/item/6127b30fd34e40dd9c05094c) [(last accessed](https://www.sciencebase.gov/catalog/item/6127b30fd34e40dd9c05094c) [Dece](https://www.sciencebase.gov/catalog/item/6127b30fd34e40dd9c05094c)mb[er](https://www.sciencebase.gov/catalog/item/6127b30fd34e40dd9c05094c) [2022).](https://www.sciencebase.gov/catalog/item/6127b30fd34e40dd9c05094c) [US](https://www.sciencebase.gov/catalog/item/6127b30fd34e40dd9c05094c)G[S](https://www.sciencebase.gov/catalog/item/6127b30fd34e40dd9c05094c) [dataset provides](https://www.sciencebase.gov/catalog/item/6127b30fd34e40dd9c05094c) [6 types](https://www.sciencebase.gov/catalog/item/6127b30fd34e40dd9c05094c) [of hum](https://www.sciencebase.gov/catalog/item/6127b30fd34e40dd9c05094c)an-reviewed phases, i.e., P, Pg, Pn, S, Sg, Sn.

**Declaration of competing interest**

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

**Acknowledgments**

This research was funded by a Discovery Grant (RGPIN-2018-03752) from the Natural Science and Engineering Research Council of Canada (PA). The authors thank Stephen Crane, Laurel Sinclair, two anonymous reviewers and the editor for their reviews of this manuscript. This is NRCan publication number 20220610.

**References**

[Beroza, G.C., Segou, M., Mostafa Mousavi, S., 2021. Machine learning and earthquake](http://refhub.elsevier.com/S2666-5441(23)00023-0/sb1)  [forecasting—next steps. Nature Commun. 12 (1), 1–3.](http://refhub.elsevier.com/S2666-5441(23)00023-0/sb1)

[Chai](http://refhub.elsevier.com/S2666-5441(23)00023-0/sb2)[, C., Maceira, M., Santos-Villalobos, H.J., Venkatakri](http://refhub.elsevier.com/S2666-5441(23)00023-0/sb1)[shnan, S.V., Schoenball, M., Zhu, W., Beroza, G.C., Thurber, C., Team, E.C., 2020. Using a deep neural network and transfer learning to bridge scales for seismic phase picking. Geophys. Res. Lett. 47 (16), e2020GL088651.](http://refhub.elsevier.com/S2666-5441(23)00023-0/sb2)

[Chri](http://refhub.elsevier.com/S2666-5441(23)00023-0/sb3)[stoffersen, P., Jacobs, K.,](http://refhub.elsevier.com/S2666-5441(23)00023-0/sb2) [2004. The importance of the loss function in option](http://refhub.elsevier.com/S2666-5441(23)00023-0/sb3)  [valuation. J. Financ. Econ. 72 (2), 291–318.](http://refhub.elsevier.com/S2666-5441(23)00023-0/sb3)

Cole[, H., Yeck, W.L., 2022. Global Earthquake](http://refhub.elsevier.com/S2666-5441(23)00023-0/sb3) Machine Learning Dataset: Machine Learning Asset Aggregation of the PDE (MLAAPDE). US Geological Survey, URL <https://www.sciencebase.gov/catalog/item/6127b30fd34e40dd9c05094c>.

[Eelb](http://refhub.elsevier.com/S2666-5441(23)00023-0/sb5)[ode, T., Sinonquel, P., Maes, F., Bisschops, R., 2021. Pitfalls in tra](https://www.sciencebase.gov/catalog/item/6127b30fd34e40dd9c05094c)[ining and validation of deep learning systems. Best Pract. Res. Clin. Gastroenterol. 52, 101712.](http://refhub.elsevier.com/S2666-5441(23)00023-0/sb5)

[Gibson, E., Li, W., Sudre, C., Fidon, L., Shakir, D.I., Wang, G., Eaton-Rosen, Z., Gray, R., Doel, T., Hu, Y., et al., 2018. NiftyNet: a deep-learning platform for medical imaging. Comput. Methods Programs Biomed. 158, 113–122.](http://refhub.elsevier.com/S2666-5441(23)00023-0/sb6)

[Gopalakrishnan, K., Khaitan, S.K., Choudhary, A., Agrawal, A., 2017. Deep convolu-tional neural networks with transfer learning for computer vision-based data-driven pavement distress detection. Constr. Build. Mater. 157, 322–330.](http://refhub.elsevier.com/S2666-5441(23)00023-0/sb7)

[Hua](http://refhub.elsevier.com/S2666-5441(23)00023-0/sb8)[ng, L., Dong, X., Clee, T.E., 2017. A scalable deep learning platfor](http://refhub.elsevier.com/S2666-5441(23)00023-0/sb7)[m for identifying](http://refhub.elsevier.com/S2666-5441(23)00023-0/sb8)  [geologic features from seismic attributes. Lead. Edge 36 (3), 249–256.](http://refhub.elsevier.com/S2666-5441(23)00023-0/sb8)

[Jian](http://refhub.elsevier.com/S2666-5441(23)00023-0/sb9)[g, C., Fang, L., Fan, L., Li, B., 2021. Comparison of the earthquake detect](http://refhub.elsevier.com/S2666-5441(23)00023-0/sb8)[ion abilities of PhaseNet and EQTransformer with the Yangbi and Maduo earthquakes. Earthq. Sci. 34 (5), 425–435.](http://refhub.elsevier.com/S2666-5441(23)00023-0/sb9)

[Jozi](http://refhub.elsevier.com/S2666-5441(23)00023-0/sb10)[nović, D., Lomax, A.,](http://refhub.elsevier.com/S2666-5441(23)00023-0/sb9) [Štajduhar, I., Michelini, A., 2022. Transfer learning: Improving neural network based prediction of earthquake ground shaking for an area with insufficient training data. Geophys. J. Int. 229 (1), 704–718.](http://refhub.elsevier.com/S2666-5441(23)00023-0/sb10)

[Kop](http://refhub.elsevier.com/S2666-5441(23)00023-0/sb11)[er, K.D., 2019. The importance of regional seismic networks](http://refhub.elsevier.com/S2666-5441(23)00023-0/sb10) [in monitoring nuclear](http://refhub.elsevier.com/S2666-5441(23)00023-0/sb11)  [test-ban treaties. In: AGU Fall Meeting Abstracts. Vol. 2019. pp. S14B–05.](http://refhub.elsevier.com/S2666-5441(23)00023-0/sb11)

[Lapi](http://refhub.elsevier.com/S2666-5441(23)00023-0/sb12)[ns, S., Goitom, B., Kendall, J.-M., Werner, M.J., Cashman, K.V., Hammon](http://refhub.elsevier.com/S2666-5441(23)00023-0/sb11)[d, J.O., 2021. A little data goes a long way: Automating seismic phase arrival picking at Nabro volcano with transfer learning. J. Geophys. Res. Solid Earth 126 (7), e2021JB021910.](http://refhub.elsevier.com/S2666-5441(23)00023-0/sb12)

[Livieris, I.E., Stavroyiannis, S., Pintelas, E., Pintelas, P., 2020. A novel validation framework to enhance deep learning models in time-series forecasting. Neural Comput. Appl. 32 (23), 17149–17167.](http://refhub.elsevier.com/S2666-5441(23)00023-0/sb13)

[Ma,](http://refhub.elsevier.com/S2666-5441(23)00023-0/sb14) [Y., Cao, S., Rector, J.W., Zhang, Z., 2](http://refhub.elsevier.com/S2666-5441(23)00023-0/sb13)[020. Automated arrival-time picking using a](http://refhub.elsevier.com/S2666-5441(23)00023-0/sb14)  [pixel-level network. Geophysics 85 (5), V415–V423.](http://refhub.elsevier.com/S2666-5441(23)00023-0/sb14)

94