ENHANCED EARTHQUAKE PHASE PICKERS IN INDONESIA: EVALUATION AND REFINEMENT OF MACHINE LEARNING MODELS

Brilian Tatag SAMAPTA[1] **Supervisor: Masumi YAMADA**[2]

ABSTRACT

Automatic earthquake detection is vital for reliable seismic monitoring, comprehensive cataloging, and effective disaster mitigation, including tsunami early warning systems. Traditional methods, such as the short-term average to long-term average (STA/LTA) algorithm, are widely used to detect earthquakes but have limitations such as noise susceptibility and imprecise arrival time estimation. Machine learning (ML) is a powerful alternative to such methods, as it offers enhanced detection accuracy and noise resistance through advanced architectures. Herein, the performance of ML models was evaluated for enhanced seismic phase detection using data from the Meteorological, Climatological, and Geophysical Agency of Indonesia (BMKG). Twenty pretrained ML models, including Earthquake Transformer, Generalized Phase Detector, PhaseNet, and Basic Phase Autoencoder, trained on five benchmark datasets (STEAD, GEOFON, NEIC, INSTANCE, and ETHZ), were assessed using the SeisBench program on BMKG dataset containing 65,875 segmented traces and 89,227 manually pick phase arrival times. Results showed that PhaseNet, trained on the NEIC dataset, accurately classified the BMKG earthquake data, demonstrating superior recall and precision while maintaining onset time accuracy. The PhaseNet model was optimized using the BMKG dataset via transfer learning, enhancing its phase identification accuracy. The adapted model was applied to continuous waveform data from March 2023 and generated an earthquake catalog with 120% more events than those in the BMKG manual catalog. The optimized model offers a robust solution for real-time earthquake detection systems, potentially enhancing seismic monitoring capabilities and early warning systems.

Keywords: Machine learning, rapid earthquake detection, phase picker, transfer learning.

1. INTRODUCTION

Automatic earthquake detection is crucial for accurate seismic analysis, disaster response, and mitigation, particularly in regions with high seismic activity like Indonesia. Traditional methods like the STA/LTA (Allen, 1978) algorithm have been widely used for earthquake detection tasks, but they are prone to noise and often struggle with precise phase picking, which is essential for accurate earthquake detection. These limitations are especially pronounced in noisy environments, leading to challenges in creating reliable earthquake catalogs and timely tsunami alerts.

Recent advancements in ML offer promising alternatives to traditional methods. ML models, trained on vast datasets, can effectively identify and analyze seismic events, offering greater precision and resistance to noise. This study determined the most suitable ML algorithm and dataset for BMKG's seismic data. We trained the best model using the BMKG dataset, resulting in an enhanced ML model to improve earthquake detection and response in Indonesia. The ultimate goal is to integrate these findings into the Indonesia Tsunami Early Warning System (InaTEWS) to enhance the accuracy and promptness of earthquake detection, leading to a more reliable earthquake catalog and better disaster preparedness.

[1] Climatological, and Geophysical Agency (BMKG), Indonesia.
[2] Disaster Prevention Research Institute (DPRI), Kyoto University.

2. DATA

ML tools in seismology depend heavily on large, well-curated datasets for accurate earthquake detection and analysis. Herein, earthquake data from January 1st to March 31st, 2023, covering a geographic area from latitudes -12° to 16° and longitudes 93° to 147°, which includes Indonesia and parts of the Philippines, were used. During this period, 3,365 seismic events were recorded (Figure 1), resulting in a rich dataset that includes 65,875 P-phase arrivals and 23,352 S-phase arrivals, manually picked by BMKG officers. The dataset encompasses a wide range of magnitudes up to 7.5, with a magnitude completeness of 3.5, and epicentral distances up to 5,800 km, making it ideal for testing and improving earthquake detection models.

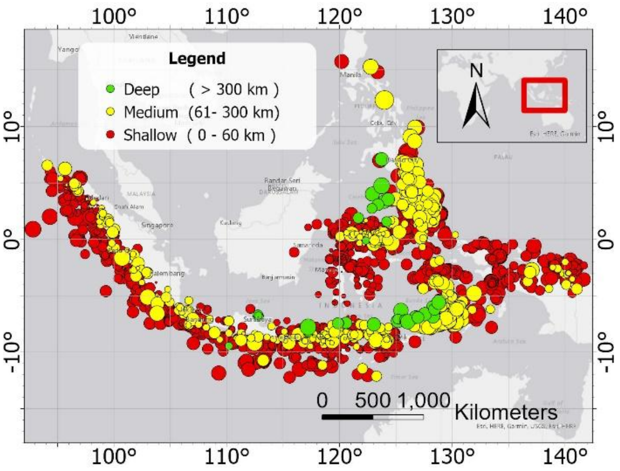

Figure 1. Earthquake distribution map.

We selected 390 seismic stations from the BMKG network that provided continuous and reliable recordings during the research period. The waveform data quality was assessed using a signal-to-noise ratio (SNR) analysis, revealing the dataset's complexity, which includes a significant number of low SNR data.

The SeisBench program (Woollam et al., 2022) provides several pretrained ML models that can be accessed through an Application Programming Interface (API), including PhaseNet (Zhu & Beroza, 2019), Generalized Phase Detector (GPD) (Ross et al., 2018), Earthquake Transformer (Mousavi et al., 2020), and Basic Phase Autoencoder (Woollam et al., 2019). Five benchmark datasets were used to train these models: STEAD (Mousavi et al., 2019), ETHZ (Woollam et al., 2022), GEOFON (Woollam et al., 2022), INSTANCE (Michelini et al., 2021), and NEIC (Yeck et al., 2021). These datasets were selected for their comprehensive seismic data, which is essential for testing the models in Indonesia's unique seismic context. Training each model on each dataset resulted in 20 pretrained ML models, which can be used through the SeisBench program.

3. METHODOLOGY

3.1. Phase detection

The SeisBench program, designed for ML in seismology, was employed to classify seismic phases in the BMKG dataset using ML models. The BMKG dataset was divided into segmented and continuous waveform data. For the segmented data, seismic information from BMKG's network, spanning January 1st to March 31st, 2023, was collected using the Obspy FDSN client (Beyreuther et al., 2010). The manual phase picks were then used to segment continuous waveform recordings into 90-second traces through Python scripting. The Ray module (Moritz et al., 2018) was utilized to improve processing efficiency and manage parallel tasks.

3.2. Performance evaluation

Herein, the accuracy of ML models in detecting seismic phases was evaluated by comparing their predictions to manual picks from the BMKG catalog using five metrics: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), fraction of samples outside acceptable boundaries, detection quantity, and probability patterns. Lower RMSE and MAE values indicate higher accuracy, while the fraction of out-of-bound samples highlights potential errors. Detection quantity is assessed by the number of phases correctly identified or missed, and probability patterns help determine the model's reliability in distinguishing true seismic events from noise, aiding in setting optimal probability thresholds.

3.3. Threshold determination

In phase detection using ML models, the threshold is crucial for determining whether a detected signal corresponds to a seismic phase like P-wave or S-wave. Initially, no threshold was set to capture all possible phases, but after performance evaluation, the optimal threshold was determined using the precision-recall curve. This curve helps balance precision (accuracy of detected phases) and recall (ability to detect all relevant phases). Key metrics include true positives (correctly identified events), false positives (incorrect detections), and false negatives (missed events). Precision measures the accuracy of the model's positive predictions, while recall assesses how well it retrieves actual events. The maximum F score (Eqs. 1) was used to find the optimal threshold that maximizes model performance by choosing the weight for each metric.

$$F_\beta = (1+\beta^2) * \frac{Precision * Recall}{\beta^2 * Precision + Recall} \tag{1}$$

3.4. Model refinement

The model refinement phase focuses on enhancing the best-pretrained model's performance by applying transfer learning, a technique that adapts the model to the BMKG dataset. This involves retraining the model using BMKG's waveform data and catalog and fine-tuning its hyperparameters to better align with the unique characteristics of the local data. The process includes iterative tuning with various configurations of batch sizes, learning rates, and epochs to minimize loss and optimize the model's accuracy. The Adam optimizer (Kingma & Ba, 2014) was employed for efficient gradient descent, reducing the need for extensive hyperparameter adjustments. This refinement ensures the model is more adept at detecting seismic events in the BMKG dataset, improving its real-world application in earthquake detection.

3.5. Phase association

The phase association process used the Gaussian Mixture Model Association (GaMMA) (Zhu et al., 2022) method to organize detected seismic phases into associated events, generating a new catalog. GaMMA, designed to handle the complexity of seismic data in regions with intense activity, combines Gaussian mixture modeling with Bayesian principles for enhanced accuracy. The process also identifies unassociated phases, which are treated as false picks, and compares the new catalog with the BMKG catalog to validate improvements. The phase association process used the IASP91 velocity model (Kennett & Engdahl, 1991), which was the same velocity model used to create the BMKG earthquake catalog.

4. RESULTS AND DISCUSSION

4.1. Performance evaluation

The evaluation of 20 pretrained ML models on BMKG's dataset highlighted the GPD-INSTANCE, EQTransformer-INSTANCE, and PhaseNet-NEIC models as top performers in detecting seismic events. These models showed high accuracy in onset time detection, with the GPD model excelling in RMSE and MAE metrics, despite some precision issues. In contrast, the BasicPhaseAE model performed poorly, generating a high number of false positives and displaying significant delays in phase detection. The PhaseNet and EQTransformer models generally performed well, particularly with the INSTANCE and

NEIC datasets, where they achieved high true positive rates and low false positive rates. The probability pattern analysis confirmed the GPD model's reliability, with a clear distinction between true detections and noise, making these models suitable for further application in seismic detection tasks.

4.2. Optimal performance

Through performance evaluation, EQTransformer with the INSTANCE dataset, GPD with the INSTANCE dataset, and PhaseNet with the NEIC dataset were identified as the top-performing pretrained models. Initial evaluations without thresholds highlighted the models' full potential, but applying thresholds improved detection accuracy by focusing on catalog detections. The highest F1 scores for each model guided these thresholds, balancing precision and recall. Applying these thresholds reduced the number of detections outside the catalog and improved onset time accuracy. The EQTransformer, GPD, and PhaseNet models each saw varying reductions in out-of-catalog detections and improvements in onset time accuracy, as detailed in Figures 2 and 3. Overall, the PhaseNet model with the NEIC dataset demonstrated the best average performance across all evaluated factors.

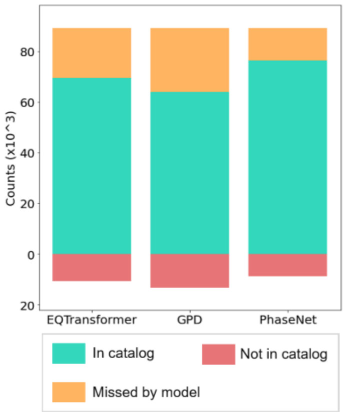

Figure 2. Detection quantities after applying thresholds.

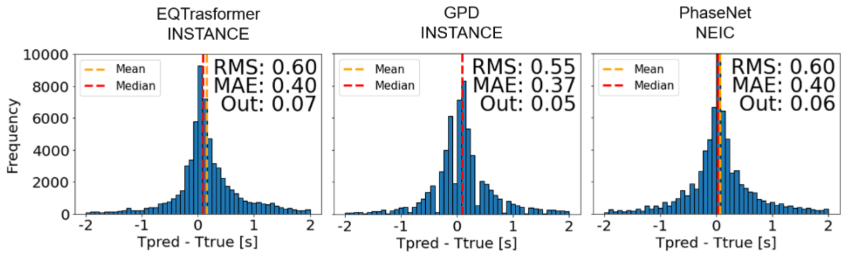

Figure 3. Histogram of residual times after applying thresholds.

4.2. Transfer learning

The PhaseNet-NEIC model, identified as the best-pretrained model, was further enhanced through transfer learning using a BMKG dataset of 31,839 traces and 42,312 phase arrival times from January to February 2022. This dataset was split into training and development sets, and multiple hyperparameters were tested to minimize loss on the development set. The model achieved the lowest loss of 0.068097 with an optimized batch size of 256 and a learning rate of 10^{-2}.

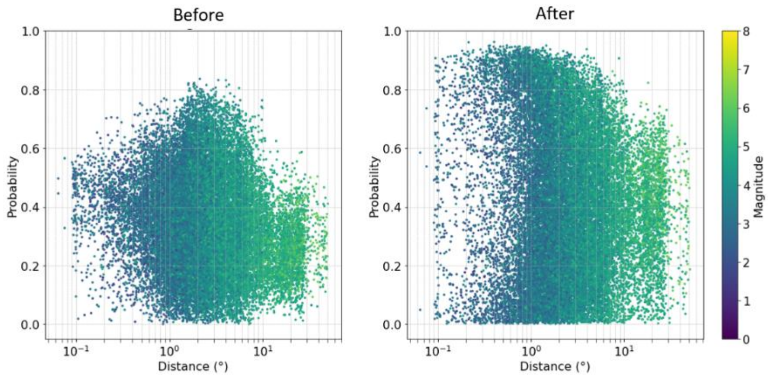

Figure 4. Probability distribution of the model.

After retraining, the model showed improved confidence in classifying the BMKG dataset, with more evenly distributed high-probability detections within the catalog and reduced detections outside it. A detailed probability distribution plot (Figure 4) revealed consistent phase detection across various probabilities, though a decrease in detection probability was noted for events at longer distances. Despite these improvements, the retrained model had a 1.4% reduction in detection quantity within the catalog and a slight decrease in onset time accuracy, with increased RMS error and MAE. The transfer learning enhanced the model's performance in key areas but introduced minor trade-offs in detection quantity and accuracy.

4.3. Phase associaition

The phase association process aims to link detected phases from the retrained model into coherent earthquake events using the GaMMA method. After retraining on the BMKG dataset, the model identified 31,988,962 earthquake phases from continuous waveform data recorded in March 2023. To minimize false positives, a threshold was applied based on precision using the F0.3 score formula, which weights precision three times as heavily as recall. The optimal probability threshold of 0.2 resulted in a precision of 96.45% and a recall of 73.79%. Using the GaMMA software, 29,769 of the filtered phase picks were successfully associated with 2,488 earthquake events, applying criteria such as a minimum of five phase picks per event and maximum phase time residual of 2.5 seconds, with a 40-second maximum time between neighboring picks.

4.4. Catalog comparison

The comparison between the new earthquake catalog generated by GaMMA and the BMKG catalog revealed significant improvements in earthquake detection and characterization. GaMMA identified 2,488 earthquake events in March 2023, which is about a 2.3 times increase over the 1,088 events recorded by BMKG. GaMMA's catalog demonstrated more frequent seismic detections and higher precision, with an RMS time residual error consistently lower than BMKG's, indicating improved accuracy.

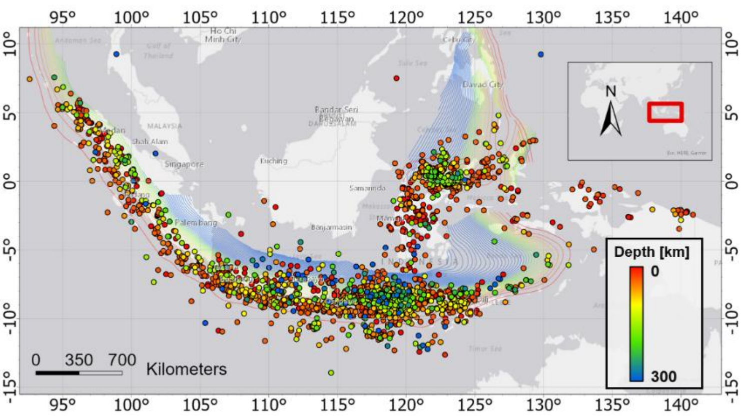

Figure 5. Spatial distribution map of the GaMMA catalog.

The spatial distribution of earthquakes in the GaMMA catalog aligned with known seismic regions, as shown in Figure 5. However, due to sparse seismic network coverage, GaMMA's detection was limited in some areas, such as the Philippines and eastern Indonesia. The depth distribution showed that GaMMA provided a more defined representation of earthquake depths, especially in subduction zones. Matching events between GaMMA and BMKG showed that GaMMA had a 56.52% recall rate, successfully identifying 615 out of 1,088 BMKG events, though it missed several earthquakes, particularly in eastern Indonesia. While GaMMA offered substantial improvements in monitoring and identifying earthquakes in Indonesia, further work is needed to address coverage stations gaps and refine detection parameters for comprehensive seismic monitoring.

5. CONCLUSIONS

This study highlights the significant advancements in seismic phase detection and earthquake catalog generation achieved through machine learning (ML) models. Evaluation of 20 pretrained ML models on the BMKG dataset identified EQTransformer (INSTANCE), GPD (INSTANCE), and PhaseNet (NEIC) as the top performers, with PhaseNet-NEIC showing exceptional results after threshold optimization. The PhaseNet-NEIC model demonstrated a strong balance with 89% precision and 85% recall, proving reliable for automated phase detection and minimizing false positives.

The transfer learning process, which adapted PhaseNet to the BMKG dataset, enhanced phase detection confidence and probability distribution, though it slightly reduced detection quantity and onset time accuracy. Using the GaMMA method, 29,769 filtered phases were successfully associated with 2,488 earthquake events for March 2023. Compared to the BMKG catalog, GaMMA identified 120% more events that aligned with known seismic regions, especially in western and central Indonesia, and provided a more defined depth distribution, particularly in subduction zones.

Despite these improvements, limitations were noted in areas with sparse seismic network coverage and deep earthquake detection. The study underscores the potential of ML-enhanced methods to advance earthquake monitoring in Indonesia, emphasizing the importance of adapting ML models to local data and optimizing phase association parameters. Future research should focus on refining these models, incorporating additional datasets, and exploring real-time processing to further enhance seismic monitoring and hazard assessment.

ACKNOWLEDGEMENTS

My heartfelt appreciation extends to my thesis supervisor, Dr. Masumi Yamada; my advisors, Dr. Takumi Hayashida and Dr. Saeko Kita; the faculty and staff of IISEE/BRI; and the dedicated officers at JICA and GRIPS.

REFERENCES

Allen, R. V. (1978). Automatic earthquake recognition and timing from single traces. *Bulletin of the seismological society of America*, *68*(5), 1521-1532.

Beyreuther, M., Barsch, R., Krischer, L., Megies, T., Behr, Y., & Wassermann, J. (2010). ObsPy: A Python toolbox for seismology. *Seismological Research Letters*, *81*(3), 530-533.

Kennett, B., & Engdahl, E. (1991). Traveltimes for global earthquake location and phase identification. *Geophysical Journal International*, *105*(2), 429-465.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Michelini, A., Cianetti, S., Gaviano, S., Giunchi, C., Jozinović, D., & Lauciani, V. (2021). INSTANCE– the Italian seismic dataset for machine learning. *Earth System Science Data*, *13*(12), 5509-5544.

Moritz, P., Nishihara, R., Wang, S., Tumanov, A., Liaw, R., Liang, E.,…Jordan, M. I. (2018). Ray: A distributed framework for emerging {AI} applications. 13th USENIX symposium on operating systems design and implementation (OSDI 18),

Mousavi, S. M., Ellsworth, W. L., Zhu, W., Chuang, L. Y., & Beroza, G. C. (2020). Earthquake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking. *Nature communications*, *11*(1), 3952.

Mousavi, S. M., Sheng, Y., Zhu, W., & Beroza, G. C. (2019). STanford EArthquake Dataset (STEAD): A global data set of seismic signals for AI. *IEEE Access*, *7*, 179464-179476.

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

Woollam, J., Münchmeyer, J., Tilmann, F., Rietbrock, A., Lange, D., Bornstein, T.,…Jozinović, D. (2022). SeisBench—A toolbox for machine learning in seismology. *Seismological Society of America*, *93*(3), 1695-1709.

Woollam, J., Rietbrock, A., Bueno, A., & De Angelis, S. (2019). Convolutional neural network for seismic phase classification, performance demonstration over a local seismic network. *Seismological Research Letters*, *90*(2A), 491-502.

Yeck, W. L., Patton, J. M., Ross, Z. E., Hayes, G. P., Guy, M. R., Ambruz, N. B.,⋯Earle, P. S. (2021). Leveraging deep learning in global 24/7 real - time earthquake monitoring at the National Earthquake Information Center. *Seismological Society of America*, *92*(1), 469-480.

Zhu, W., & Beroza, G. C. (2019). PhaseNet: a deep-neural-network-based seismic arrival-time picking method. *Geophysical Journal International*, *216*(1), 261-273.

Zhu, W., McBrearty, I. W., Mousavi, S. M., Ellsworth, W. L., & Beroza, G. C. (2022). Earthquake phase association using a Bayesian Gaussian mixture model. *Journal of Geophysical Research: Solid Earth*, *127*(5), e2021JB023249.