PolarCAP – A Deep Learning approach for First Motion Polarity Classification of Earthquake Waveforms

Megha Chakraborty[1,2], Claudia Quinteros Cartaya[1], Wei Li[1], Johannes Faber[1,3], Horst Stoecker[1,3,4,5], Georg Rümpker[1,2], Nishtha Srivastava[1,2]

[1]Frankfurt Institute for Advanced Studies, 60438 Frankfurt am Main, Germany
[2]Institute of Geosciences, Goethe-University Frankfurt, 60438 Frankfurt am Main, Germany
[3]Institute for Theoretical Physics, Goethe Universität, 60438 Frankfurt am Main, Germany
[4]Xidian-FIAS international Joint Research Center, Giersch Science Center, 60438 Frankfurt am Main, Germany
[5]GSI Helmholtzzentrum für Schwerionenforschung GmbH, 64291 Darmstadt, Germany

Key Points:

- We present PolarCAP, a deep learning model that can classify the polarity of a waveform with known P-arrival time with a 98% accuracy.
- The first-motion polarity of seismograms is a useful information, but its manual determination can be laborious and imprecise.
- We demonstrate that in several cases the model can assign trace polarity more accurately than a human analyst.

Corresponding author: Nishtha Srivastava, `srivastava@fias.uni-frankfurt.de`

Abstract

The polarity of first P-wave arrivals plays a significant role in the effective determination of focal mechanisms specially for smaller earthquakes. Manual estimation of polarities is not only time-consuming but also prone to human errors. This warrants a need for an automated algorithm for first motion polarity determination. We present a deep learning model - PolarCAP that uses an autoencoder architecture to identify first-motion polarities of earthquake waveforms. PolarCAP is trained in a supervised fashion using more than 130,000 labelled traces from the Italian seismic dataset (INSTANCE) and is cross-validated on ∼22,000 traces to choose the most optimal set of hyperparameters. We obtain an accuracy of 0.98 on a completely unseen test dataset of almost 33,000 traces. Furthermore, We check the model generalizability by testing it on the datasets provided by previous works and show that our model achieves a higher recall on both positive and negative polarities.

Plain Language Summary

The polarity of an earthquake waveform describes the vertical direction of ground motion when the seismic waves first arrive at a station. This information is used to determine the mechanism of the faulting responsible for the earthquake. Traditionally polarity is determined manually by looking at the waveforms. However this is not just time consuming but often suffers from human errors. Hence there is a need for an algorithm that can perform this task automatically and reliably. In this paper we present a deep learning model PolarCAP (Polarity determining Convolutional Auto-encoder for first motion P-waves) which leverages the manually assigned polarity information of over 130,000 earthquake traces provided by the Italian seismic dataset (INSTANCE) for machine learning to learn the polarity estimation of any seismogram with an accuracy above 98%. We also demonstrate that the predictions from the model are more reliable than human-assigned labels.

1 Introduction

The first motion polarity of earthquake waveforms is an important parameter in determining focal mechanisms, particularly for smaller earthquakes. Traditionally the first-motion polarity is assigned manually by expert analysts. However, based on their observation on a dataset from the Northridge region, Hardebeck and Shearer (2002) reported that the picked polarities are inconsistent with the true polarities about 10% (for impulsive onset where polarities are more easily determined) to 20% (for emergent onset where polarities are more ambiguous) of the times. This, coupled with the growing volumes of seismological data, warrants the need for a faster, more precise and efficient method for the picking of polarities.

An automated polarity picking algorithm proposed by C. Chen and Holland (2016) is based on comparing the signal amplitude with the background noise and checking whether it crosses a user-defined threshold. Pugh et al. (2016) presented a Bayesian inference approach to polarity determination. Such numerical approaches, however, require (i) intensive human involvement, (ii) are heavily dependent on a limited number of parameters, and (iii) fail to account for the complex nature of seismograms; and hence cannot compete with manual picks (Ross et al., 2018).

Data-driven computer vision techniques, such as convolutional neural networks have been shown to be capable of analysing spatially independent information by mimicking the perception of images by the human brain (Voulodimos et al., 2018; Lundervold & Lundervold, 2019; Brachmann et al., 2017). Like in most research fields, deep learning has been successfully applied to seismology for tasks such as event detection & location (Perol et al., 2018), seismic phase identification & picking (Y. Chen, 2018; Zhu & Beroza, 2019; Li et al., 2021, 2022), magnitude characterization (Mousavi & Beroza, 2020; Chakraborty et al., 2022, 2021, 2004). The applicability of simple Convolutional Neural Networks (CNNs) in

the picking of first-motion polarities has been demonstrated by Ross et al. (2018), Hara et al. (2019) and Uchide (2020).

In this study we use an autoencoder model for determining the first-motion polarities. Unlike Mousavi et al. (2019) who also use autoencoders for polarity classification in an unsupervised fashion, we adopt a supervised approach and leverage the polarity information provided in the metadata of the INSTANCE dataset (Michelini et al., 2021). Since there are several algorithms that are capable of picking P-arrival times with an accuracy of the order of 0.01s (Mousavi et al., 2020; Li et al., 2021; Zhu & Beroza, 2019; Liao et al., 2021) we focus solely on classification of polarities and not the picking of P-arrival times. Therefore, we assume the P-arrival time to be an a-priori knowledge and use data windows of fixed length centred around the known P-arrival sample. We perform extensive analysis on our results to investigate its potential at outperforming human analysts and to investigate the possible scenarios that can lead to an error in assignment of polarity by the model.

2 Methodology

2.1 Data

Nearly 1.2 million waveforms recorded primarily by the Italian National Seismic Network between January 2005 and January 2020 and obtained from the INSTANCE dataset (Michelini et al., 2021) are used to train and validate our model. For simplicity we leave out those traces for which the number of detected P-wave arrivals, as reported in the metadata, is not equal to 1. It has been observed by Ross et al. (2018) that the accuracy in the prediction of polarity falls exponentially with signal-to-noise ratio (SNR), when the SNR is below 10dB. Thus, we ignore traces with SNR less than 10dB to ensure minimum discrepancy in the training data. This leaves us with 443,002 traces out of which 109,748 have polarity information identified by expert analysts and made available through catalogs. These 109,748 traces are divided in the ratio 60:10:30 for training, validation and testing purposes.

Each trace in the training and validation sets is included twice in the respective set – once in its original form and once by flipping it (i.e. multiplying it by -1). This augmentation technique, previously used by Hara et al. (2019), not only doubles the volume of training data but also helps in balancing out the two classes ('positive' and 'negative' polarity), which is known to benefit the learning of a classifier (Batista et al., 2004). No such augmentation is applied to the test dataset to retain the original distribution of classes. We do not apply any pre-processing steps, other than normalising each waveform by dividing it with its maximum absolute value. As mentioned earlier, we work with the assumption that the first P-arrival time is already known to us. We use a fixed data window centred around the P-arrival sample. The length of the window was chosen experimentally to be 64 samples after exploring several values between 32 to 1024 samples. Note that only powers of 2 were used since the auto-encoder reduces the data dimensionality by a factor of 2 at each step.

2.2 Model architecture and training

We use an autoencoder model (Rumelhart et al., 1986) whose architecture is shown in Figure 1. It uses two sets of 1D Convolution (Kiranyaz et al., 2015) and Maxpooling (Nagi et al., 2011) layers to map the data into a 16-dimensional latent space (encoded layer). The decoder for reconstructing the data consists of two sets of Convolutional and Upsampling layers. Further details on the hyperparameters used can be found on the caption for Figure 1. A softmax function is applied to the encoded layer to perform the classification. The model is implemented using Keras (Charles, 2013) and trained and tested on an NVIDIA A100 GPU. An Adam Optimiser (Kingma & Ba, 2014) is used for backpropagation. The loss function is a weighted sum of the reconstruction and classification losses with weights of 1 and 200 respectively. Since, we are more interested in the classification performance than the reconstruction performance, and the reconstruction is only used to facilitate the

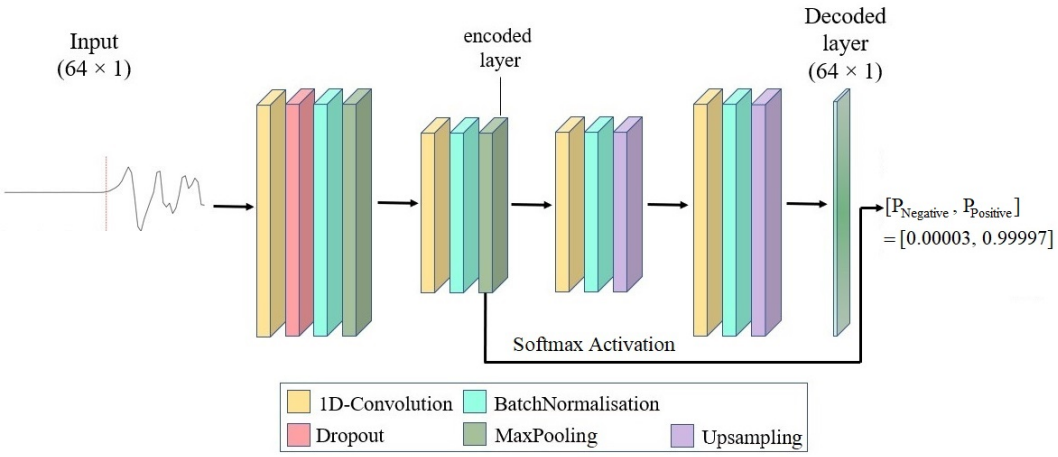


Figure 1. The autoencoder architecture used for our study. The 1D convolutional layers in the encoder use 'relu' activation and have 32 and 8 filters respectively and kernel sizes of 32 and 16, respectively. Each Maxpooling layer reduces the data dimension by 2. The drop out rate used is 0.3. The convolutional layers in the decoder use 'tanh' and 'relu' activations respectively and 8 and 32 filters and kernel sizes of 16 and 32, respectively. The final decoder layer has a 'tanh' activation. A softmax function is applied on the encoded layer to get the probability of the waveform having a 'negative' and a 'positive' polarity respectively.

learning, higher weight is assigned to classification loss. To calculate the reconstruction and classification losses we use the Keras (Charles, 2013) inbuilt loss functions for mean squared error and huber loss with delta = 0.5, respectively. We use early stopping (Prechelt, 2012) to prevent overfitting, whereby the training stops automatically if the validation loss does not decrease for 15 consecutive epochs, and the best set of learned model weights (one with lowest validation loss) is saved iteratively. We also use *ReduceLROnPlateau* function to reduce the learning rate by a factor of 10 if the validation loss does not decrease for 10 epochs, starting with a learning rate of 10^{-3} and letting it have a minimum value of 10^{-6}. Using these conditions, the model trains for 75 epochs with an average training time of 1s per epoch.

3 Results and Discussion

As stated previously, PolarCAP was trained only on data with signal-to-noise ratio (SNR) above 10dB to ensure a good quality of training data, so we test it separately on data with SNR above and below 10dB. This testing was performed on two independent datasets: (i) the portion of the INSTANCE dataset (Michelini et al., 2021) not used for training or validation and (ii) test dataset from Southern California Seismic Network (SCSN) used in Ross et al. (2018). For the latter dataset, however, we ignored the waveforms for which the polarity was 'undetermined'. The results have been summarised in the first half of Table 1 and the corresponding confusion matrices (Ting, 2017) can be found in Figure 2. As one can see, the accuracy for traces with SNR above 10dB is around 98% for both the datasets which means for about 98% of the traces the polarity labels determined by our model agrees with the polarity labels assigned through manual analysis. A few examples of such traces are shown in Figure 3a. As expected, the accuracy is lower for smaller SNR since a higher noise level makes the polarity information ambiguous and hence difficult to determine either manually or using a deep learning model.

Table 1. Summary of Model Performance when trained with and without data augmentation on the two datasets for traces with SNR above and below 10dB

	Test Dataset		Accuracy (%)	Precision (%)		Recall (%)	
				Positive	Negative	Positive	Negative
Without Data Augmentation	**INSTANCE** [1]	SNR $\geq$ 10dB	98.19	99.06	96.49	98.22	98.12
		SNR $<$ 10dB	96.22	98.18	93.63	95.30	97.50
	SCSN [2]	SNR $\geq$ 10dB	97.53	98.89	94.98	97.35	97.87
		SNR $<$ 10dB	89.65	96.11	77.67	88.87	91.48
With Data Augmentation	**INSTANCE** [1]	SNR $\geq$ 10dB	97.65	98.61	95.77	97.86	97.23
		SNR $<$ 10dB	94.24	97.68	89.98	92.36	96.90
	SCSN [2]	SNR $\geq$ 10dB	97.78	98.99	95.50	97.64	98.05
		SNR $<$ 10dB	92.46	96.86	83.55	92.26	92.93

[1]Michelini et al. (2021); [2]Ross et al. (2018)

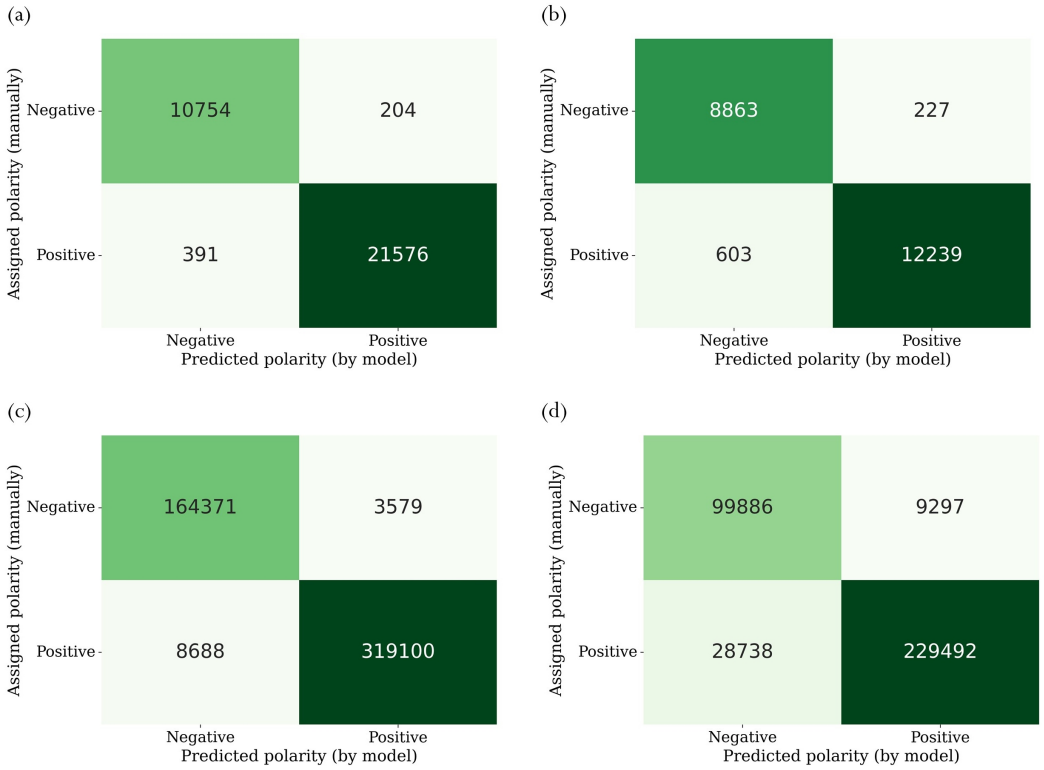


Figure 2. Confusion matrices for testing model on INSTANCE data (Michelini et al., 2021) with (a) SNR $\geq$ 10dB (b) SNR $<$ 10dB and SCSN data used by Ross et al. (2018) with (c) SNR $\geq$ 10dB (d) SNR $<$ 10dB.

3.1 Manual inspection of traces with mismatched assigned and predicted polarity

We further investigate the 595 traces from INSTANCE dataset with SNR above 10dB for which the polarity predicted by PolarCAP and the manually assigned polarities were in disagreement (refer to Figure 2). We could classify these cases into one of the following three categories:

- The polarity assigned by the analyst was correct.
- The polarity predicted by the model was correct.
- The polarity information was ambiguous due to high noise levels, or incorrect P-picking.

Based on our analysis we observed that for 40.8% of the cases the polarity predicted by the model was correct whereas in 27.6% of the cases the model predicted incorrect polarities. In the remaining 31.6% cases the polarity information was hard to determine through manual inspection, mostly due to high noise levels. Some examples of the former case can be found in Figure 3b. We further found that incorrect classification of traces by the model, was either due to incorrect picking of the first P-arrival (although one can see from Figure 3b that some degree of time shift is accounted for by the model based on what is encountered in the training data) or the P-arrivals being emergent in nature. We went on to inspect the traces where we identified the P-arrival sample to be incorrectly picked, and picked the P-arrivals ourselves using the EPick model (Li et al., 2021). Figure 4a shows five traces where the P-arrival samples determined by EPick seemed more accurate than those provided in the metadata. We then fed 64 sample windows centred around the picked P-phases and in each of these cases the polarities predicted by the model now matched with the assigned polarities. Figure 4b shows some examples of emergent arrivals. It is also observed that the probability of prediction for emergent onsets are usually lower than that for impulsive onsets.

3.2 Factors affecting model accuracy

We looked at the distribution of the incorrectly classified traces in terms of signal-to-noise ratio, magnitude, focal depth, and epicentral distances. Signal-to-noise ratio (SNR) can influence the ease with which the first P-arrival is picked and hence the first motion polarity can be determined, and magnitude can be correlated with the SNR. However, we did not find any observable correlation between signal-to-noise ratio, magnitude and model accuracy, as shown in Figure 5a, which means the model is capable of performing polarity determination across a wide range of SNR and magnitudes. One can also see from Figure 5b that the incorrect classifications are restricted to shallower events (<80 km) even though these are most represented in the training data. This is likely to be because deeper earthquakes tend to have a more impulsive nature as compared to shallow earthquakes of similar magnitude (Bormann et al., 2014).

We also looked at the fraction of incorrect classification for different site conditions as characterised by the average shear wave velocity of the top 30 meters of subsurface (V_{S30}). We find that the tendency for incorrect polarity classification is slightly higher for $V_{S30} < 420ms^{-1}$, although for the 11 events recorded at stations with $V_{S30} < 240ms^{-1}$ there is no observed misclassification (Figure 5c).

3.3 Further improvement through augmentation

As outlined above, the incorrect determination of polarity was caused by incorrect picking of P-wave arrival time or emergent nature of P-onsets. In order to tackle the first issue, we used data augmentation to add a time shift to some of the traces in the training data. This augmentation technique was also explored in Uchide (2020). Since we did not

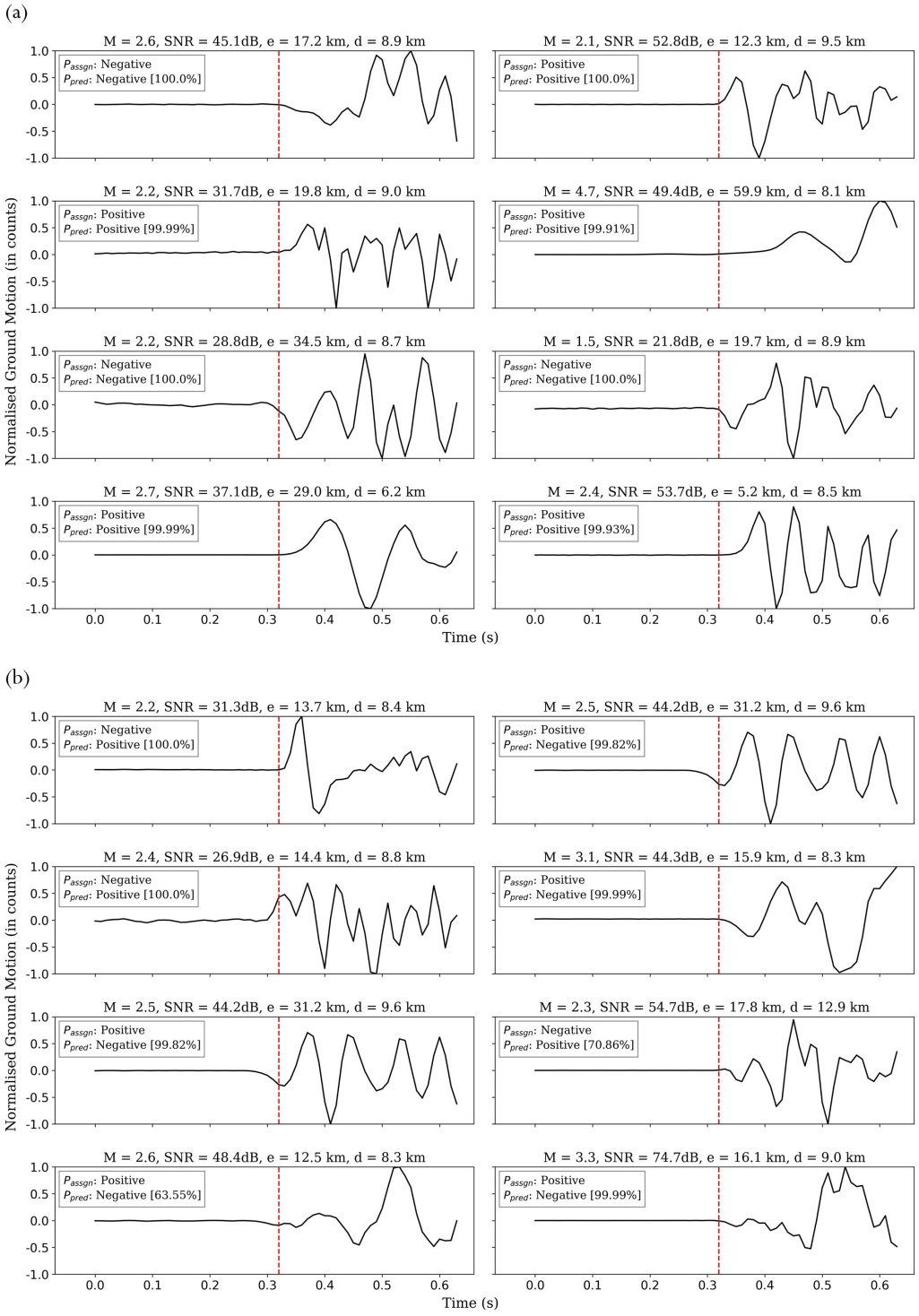


Figure 3. Examples of traces where (a) the polarity predicted by the model matches with the polarity assigned by human analysts, (b) the model predicts the correct polarity as opposed to manually assigned polarity. The red dashed line shows the P-arrival sample provided in the metadata. In the figure titles, M stands for event magnitude, SNR for signal-to-noise ratio, e for epicentral distance and d for focal depth. P_{assgn} and P_{pred} represent the assigned and predicted polarities respectively and the percentages in square brackets represent probabilities corresponding to the predicted polarity.

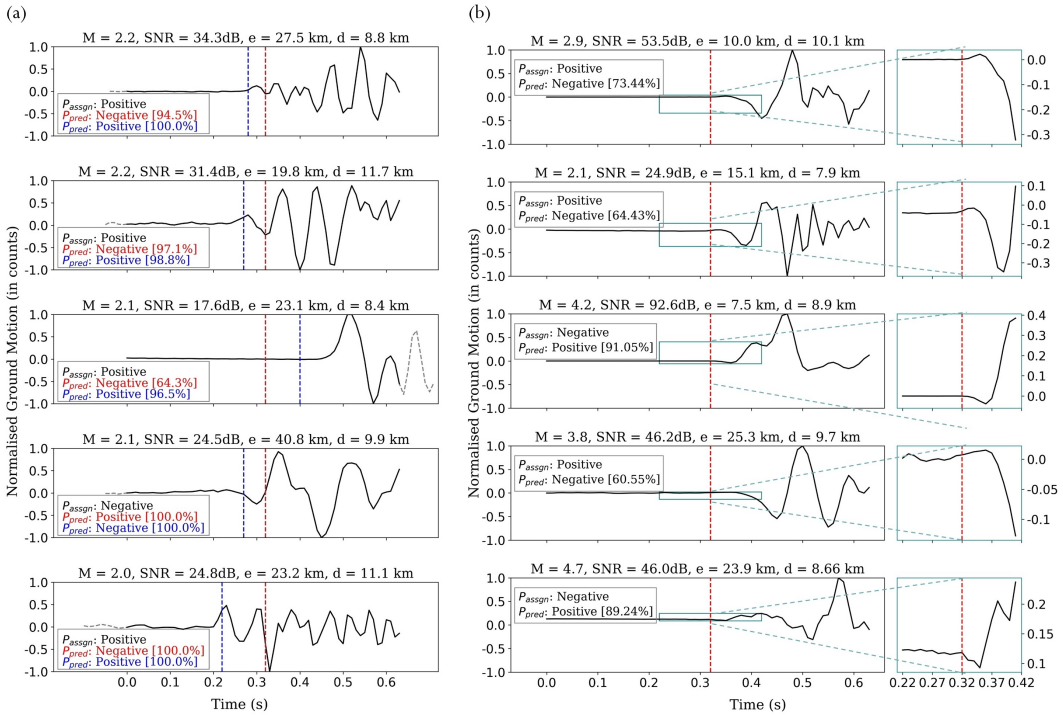


Figure 4. (a) Examples of traces where the polarity prediction was incorrect due to an error in the P-arrival time provided in the metadata (showed with red-dashed line). The P-arrival sample picked by the model from Li et al. (2021) is shown with a dotted blue line. The dotted gray line shows the new time window. The predicted polarities in red and blue correspond to the arrival times provided in the metadata and those picked using Li et al. (2021) respectively (b) Examples of traces where the polarity was predicted incorrectly by the model due to the emergent nature of the P-arrival. The panel on the right shows a magnified plot of just 10 samples around the P-arrival sample (provided in the metadata). In the figure titles, M stands for event magnitude, SNR for signal-to-noise ratio, e for epicentral distance and d for focal depth. P_{assgn} and P_{pred} represent the assigned and predicted polarities respectively and the percentages in square brackets represent probabilities corresponding to the predicted polarity.

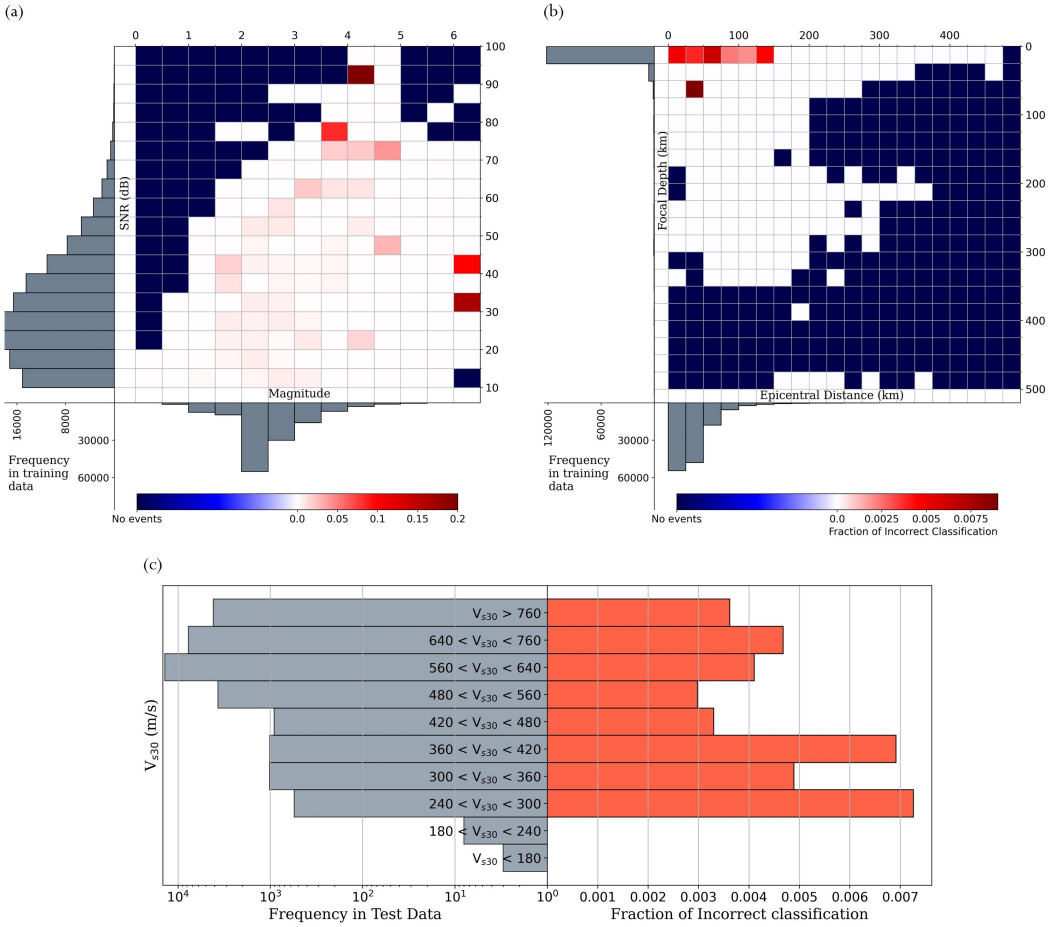


Figure 5. Distribution of incorrectly classified traces as a fraction of total number of traces, with respect to (a) Signal-to-noise ratio and magnitude (b) Focal depth and epicentral distance (c) Average shear wave velocity of the top 30 meters of subsurface (V_{S30}).

have any information on the quality of P-wave onsets in the metadata, it was not possible to apply augmentation to increase the amount of emergent traces in the training data. The time shift was applied to 1/5th of the traces (which were then added back to the dataset); the amount of time shift was chosen from a normal distribution with mean 0 and standard deviation of 10 samples. A maximum shift of 30 samples was allowed to ensure that the first P-arrival is included in the window.

After re-training the model on the augmented data, we tested it again on the same test sets. The corresponding evaluation of the model performance is shown in the second half of Table 1. As one can see, the use of data augmentation resulted in lower accuracy on the INSTANCE dataset (for both low and high SNR) This could be because in the dataset, more often than not the traces had correctly picked P-arrival times and hence the augmentation was not needed. In case of the test data from Ross et al. (2018), on the other hand, the accuracy increased to 0.25% for high SNR data and 2.81% for low SNR data upon using data augmentation.

4 Conclusion

In this study, we explored the potential of a deep learning model - PolarCAP to determine first-motion polarity of earthquake waveforms when the P-arrival information is available, faster and more accurately as compared to human analysts. We show that when tested on unseen traces, the polarity predicted by the model, matches the ones assigned by human analysts over 98% of the times. We observed that the major reasons behind incorrect assignment of polarity by the model were incorrect P-arrival picks and emergent arrivals; to that end we also found the event depth to indirectly affect the fraction of incorrect predictions by affecting the quality of P-wave arrival. However in spite of these hurdles we find that when the polarities predicted by the model differed from those presented in the metadata, it was usually the model that was correct (almost 41% of the times, while the model was wrong only about 28% of the times), thus demonstrating its ability to overcome human errors.

5 Open Research

The seismic waveforms used in our research are a part of two datasets – INSTANCE (Michelini et al., 2021) which was downloaded from `https://doi.org/10.13127/instance` (last accessed January 2022) and the dataset described in Ross et al. (2018) which can be downloaded from `https://scedc.caltech.edu/data/deeplearning.html#picking_polarity` (last accessed April 2022).

Acknowledgments
This research is supported by the "KI-Nachwuchswissenschaftlerinnen" - grant SAI 01IS20059 by the Bundesministerium für Bildung und Forschung - BMBF. Calculations were performed at the Frankfurt Institute for Advanced Studies' new GPU cluster, funded by BMBF for the project Seismologie und Artifizielle Intelligenz (SAI). The authors are also thankful to Darius Fenner and Jonas Köhler for their kind suggestions. H.St. gratefully acknowledges the Judah M. Eisenberg Laureatus - Professur at Fachbereich Physik, Goethe Universität Frankfurt, funded by the Walter Greiner Gesellschaft zur Förderung der physikalischen Grundlagenforschung e.V.

References

Batista, G. E. A. P. A., Prati, R. C., & Monard, M. C. (2004, June). A study of the behavior of several methods for balancing machine learning training data. *SIGKDD Explorations Newsletter*, *6*(1), 20–29. Retrieved from `https://doi.org/10.1145/`

1007730.1007735 doi: 10.1145/1007730.1007735

Bormann, P., Klinge, K., & Wendt, S. (2014). Data analysis and seismogram interpretation. In *Bormann, p. (ed.), new manual of seismological observatory practice 2 (nmsop-2), potsdam : Deutsches geoforschungszentrum gfz* (p. 1-126).

Brachmann, A., Barth, E., & Redies, C. (2017). Using cnn features to better understand what makes visual artworks special. *Frontiers in Psychology*, *8*. Retrieved from https://www.frontiersin.org/article/10.3389/fpsyg.2017.00830 doi: 10.3389/fpsyg.2017.00830

Chakraborty, M., Fenner, D., Li, W., Faber, J., Zhou, K., Rümpker, G., ... Srivastava, N. (2022). Creime – a convolutional recurrent model for earthquake identification and magnitude estimation. *Earth and Space Science Open Archive*, 49. Retrieved from https://doi.org/10.1002/essoar.10511140.1 doi: 10.1002/essoar.10511140.1

Chakraborty, M., Li, W., Faber, J., Ruempker, G., Stoecker, H., & Srivastava, N. (2021). *A study on the effect of input data length on deep learning based magnitude classifier.* arXiv. Retrieved from https://arxiv.org/abs/2112.07551 doi: 10.48550/ARXIV.2112.07551

Chakraborty, M., Rümpker, G., Stöcker, H., Li, W., Faber, J., Fenner, D., ... Srivastava, N. (2004, June). Real time magnitude classification of earthquake waveforms using deep learning. *SIGKDD Explor. Newsl.*, *6*(1), 20–29. Retrieved from https://doi.org/10.1145/1007730.1007735 doi: 10.1145/1007730.1007735

Charles, P. (2013). *Project title.* GitHub. Retrieved from https://github.com/charlespwd/project-title

Chen, C., & Holland, A. A. (2016). Phasepapy: A robust pure python package for automatic identification of seismic phases. *Seismological Research Letters*, *87*(6), 1384–1396.

Chen, Y. (2018). Automatic microseismic event picking via unsupervised machine learning. *Geophys. J. Int.*, *212*(1), 88–102.

Hara, S., Fukahata, Y., & Iio, Y. (2019). P-wave first-motion polarity determination of waveform data in western Japan using deep learning. *Earth Planets Space*, *71*(127). doi: https://doi.org/10.1186/s40623-019-1111-x

Hardebeck, J. L., & Shearer, P. M. (2002). A new method for determining first-motion focal mechanisms. *Bulletin of the Seismological Society of America*, *92*(6), 2264–2276.

Kingma, D. P., & Ba, J. (2014, 12). *Adam: A method for stochastic optimization.*

Kiranyaz, S., Ince, T., Hamila, R., & Gabbouj, M. (2015). Convolutional neural networks for patient-specific ecg classification. , 2608-2611. doi: 10.1109/EMBC.2015.7318926

Li, W., Chakraborty, M., Fenner, D., Faber, J., Zhou, K., Rümpker, G., ... Srivastava, N. (2021). *Epick: Multi-class attention-based u-shaped neural network for earthquake detection and seismic phase picking.* Retrieved from https://arxiv.org/abs/2109.02567

Li, W., Sha, Y., Zhou, K., Faber, J., Ruempker, G., Stoecker, H., & Srivastava, N. (2022). *Deep learning-based small magnitude earthquake detection and seismic phase classification.* arXiv. Retrieved from https://arxiv.org/abs/2204.02870 doi: 10.48550/ARXIV.2204.02870

Liao, W., Lee, E., Mu, D., Chen, P., & Rau, R. (2021, 03). ARRU Phase Picker: Attention Recurrent-Residual U-Net for Picking Seismic P- and S-Phase Arrivals. *Seismol. Res. Lett.*, *92*(4), 2410-2428. Retrieved from https://doi.org/10.1785/0220200382 doi: 10.1785/0220200382

Lundervold, A. S., & Lundervold, A. (2019). An overview of deep learning in medical imaging focusing on mri. *Zeitschrift für Medizinische Physik*, *29*(2), 102-127. Retrieved from https://www.sciencedirect.com/science/article/pii/S0939388918301181 (Special Issue: Deep Learning in Medical Physics) doi: https://doi.org/10.1016/j.zemedi.2018.11.002

Michelini, A., Cianetti, S., Gaviano, S., Giunchi, C., Jozinović, D., & Lauciani, V. (2021). Instance the italian seismic dataset for machine learning, seismic waveforms and associated metadata. *Istituto Nazionale di Geofisica e Vulcanologia (INGV)*. Retrieved from https://doi.org/10.13127/instance

Mousavi, S. M., & Beroza, G. C. (2020). A machine-learning approach for earthquake magnitude estimation. *Geophys. Res. Lett.*, *47, e2019GL085976.*. Retrieved from https://doi.org/10.1029/2019GL085976

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

Mousavi, S. M., Zhu, W., Ellsworth, W., & Beroza, G. (2019). Unsupervised clustering of seismic signals using deep convolutional autoencoders. *IEEE Geoscience and Remote Sensing Letters*, *16*(11), 1693-1697. doi: 10.1109/LGRS.2019.2909218

Nagi, J., Ducatelle, F., Di Caro, G. A., Cireşan, D., Meier, U., Giusti, A., ... Gambardella, L. M. (2011). Max-pooling convolutional neural networks for vision-based hand gesture recognition. In *2011 ieee international conference on signal and image processing applications (icsipa)* (p. 342-347). doi: 10.1109/ICSIPA.2011.6144164

Perol, T., Gharbi, M., & Denolle, M. (2018). Convolutional neural network for earthquake detection and location. *Sci. Adv.*, *4*(2), e1700578. doi: 10.1126/sciadv.1700578

Prechelt, L. (2012). Early stopping — but when? In *Neural networks: Tricks of the trade: Second edition* (pp. 53–67). Springer Berlin Heidelberg. Retrieved from https://doi.org/10.1007/978-3-642-35289-8_5 doi: 10.1007/978-3-642-35289-8_5

Pugh, D., White, R., & Christie, P. (2016). Automatic bayesian polarity determination. *Geophysical Journal International*, *206*(1), 275–291.

Ross, Z. E., Meier, M.-A., & Hauksson, E. (2018). P wave arrival picking and first-motion polarity determination with deep learning. *Journal of Geophysical Research: Solid Earth*, *123*(6), 5120-5129. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2017JB015251 doi: https://doi.org/10.1029/2017JB015251

Rumelhart, D., Hinton, G., & Williams, R. (1986). Learning internal representations by error propagation. In *Explorations in the microstructure of cognition* (Vol. 1, p. 318–362). MIT Press, Cambridge, MA, USA. Retrieved from http://dl.acm.org/citation.cfm?id=104279.104293

Ting, K. M. (2017). Confusion matrix. In *Encyclopedia of machine learning and data mining* (pp. 260–260). Boston, MA: Springer US. Retrieved from https://doi.org/10.1007/978-1-4899-7687-1_50 doi: 10.1007/978-1-4899-7687-1_50

Uchide, T. (2020, 08). Focal mechanisms of small earthquakes beneath the Japanese islands based on first-motion polarities picked using deep learning. *Geophysical Journal International*, *223*(3), 1658-1671. doi: 10.1093/gji/ggaa401

Voulodimos, A., Doulamis, N., Doulamis, A., & Protopapadakis, E. (2018). Deep learning for computer vision: A brief review. *Computational Intelligence and Neuroscience*, *2018*(7068349). doi: https://doi.org/10.1155/2018/7068349

Zhu, W., & Beroza, G. C. (2019). Phasenet: a deep-neural-network-based seismic arrival-time picking method. *Geophys. J. Int.*, *216*(1), 261-273. Retrieved from https://doi.org/10.1093/gji/ggy423

Figure_1.jpg.

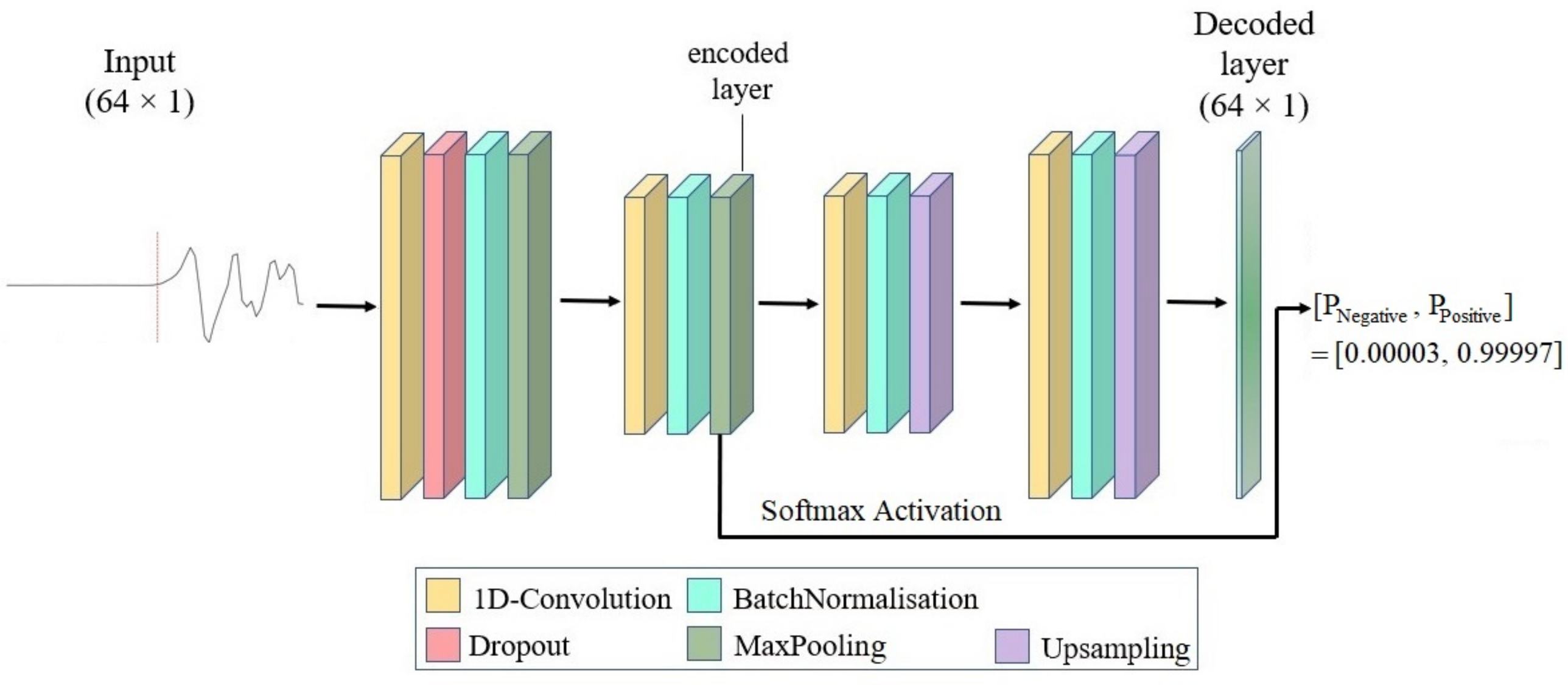
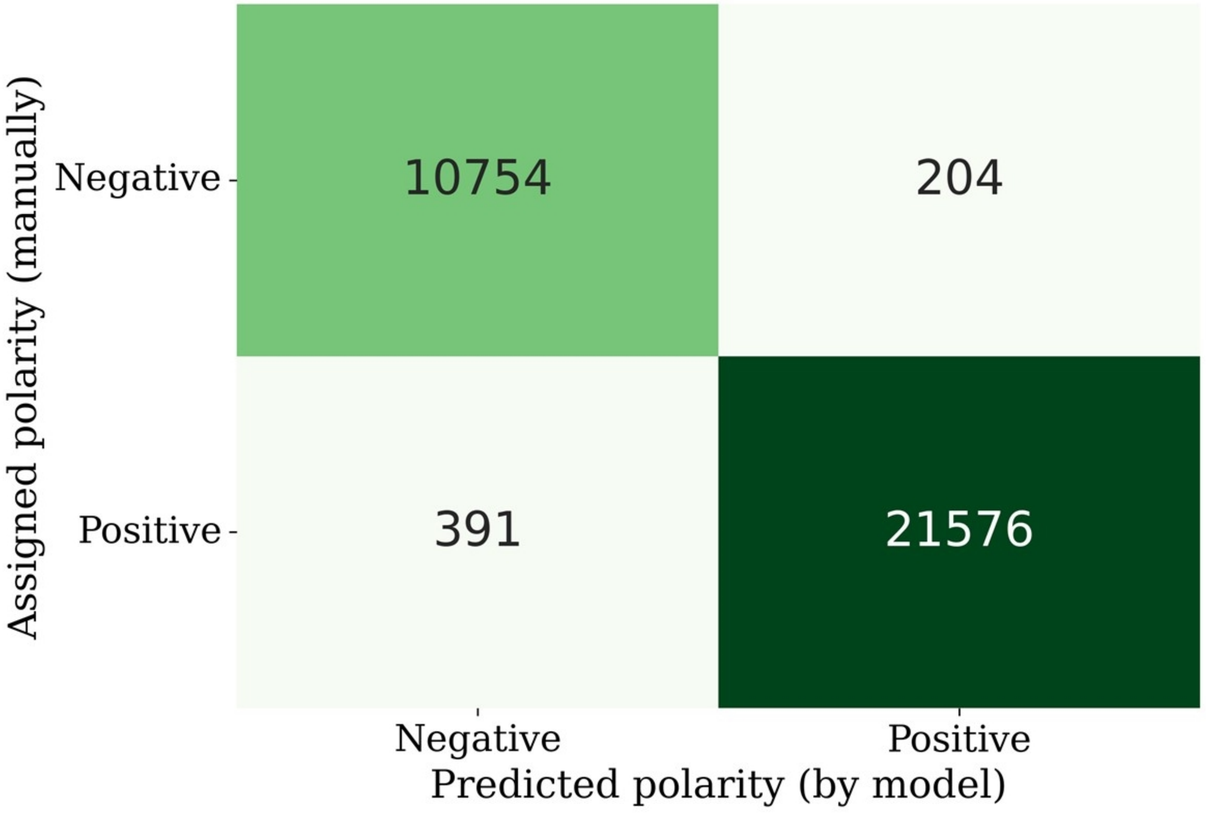
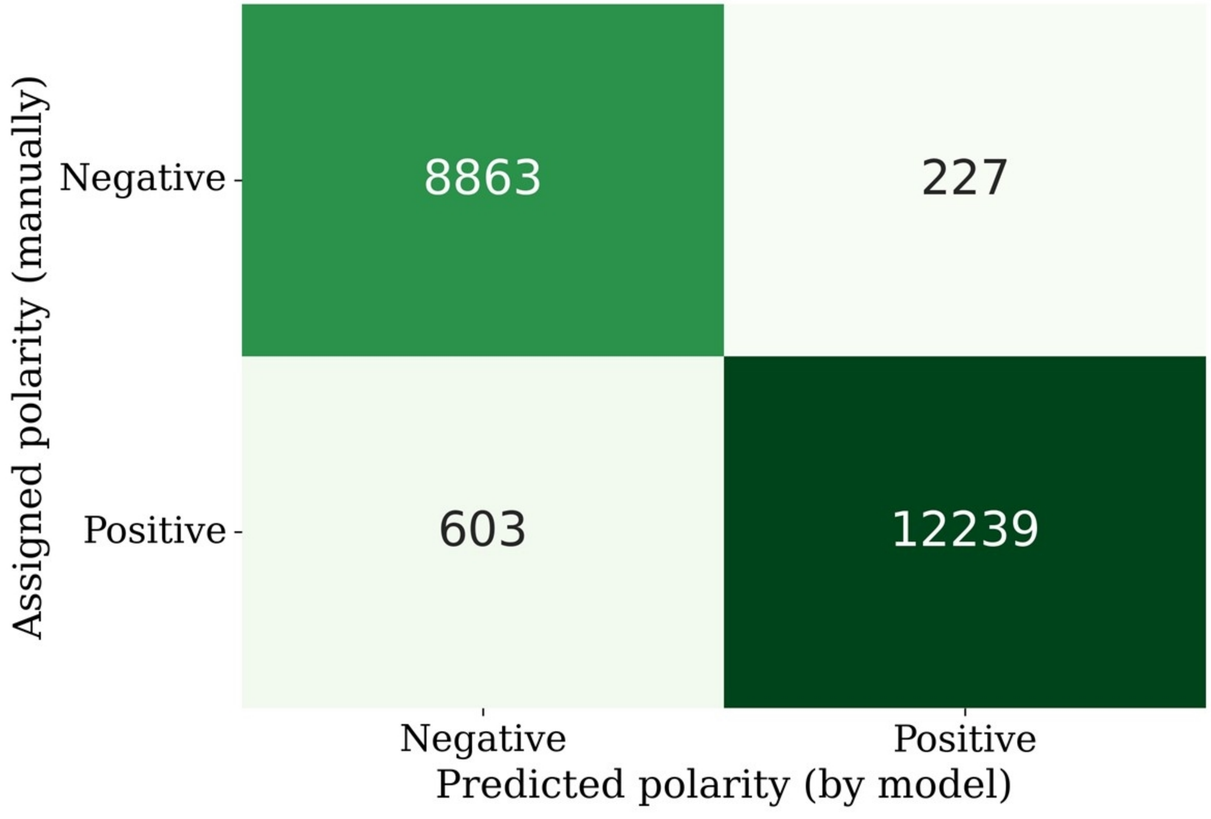

Input
(64 × 1)
encoded
layer
Decoded
layer
(64 × 1)
[P_Negative , P_Positive]
= [0.00003, 0.99997]
Softmax Activation
1D-Convolution
BatchNormalisation
Dropout
MaxPooling
Upsampling

Figure_2.jpg.

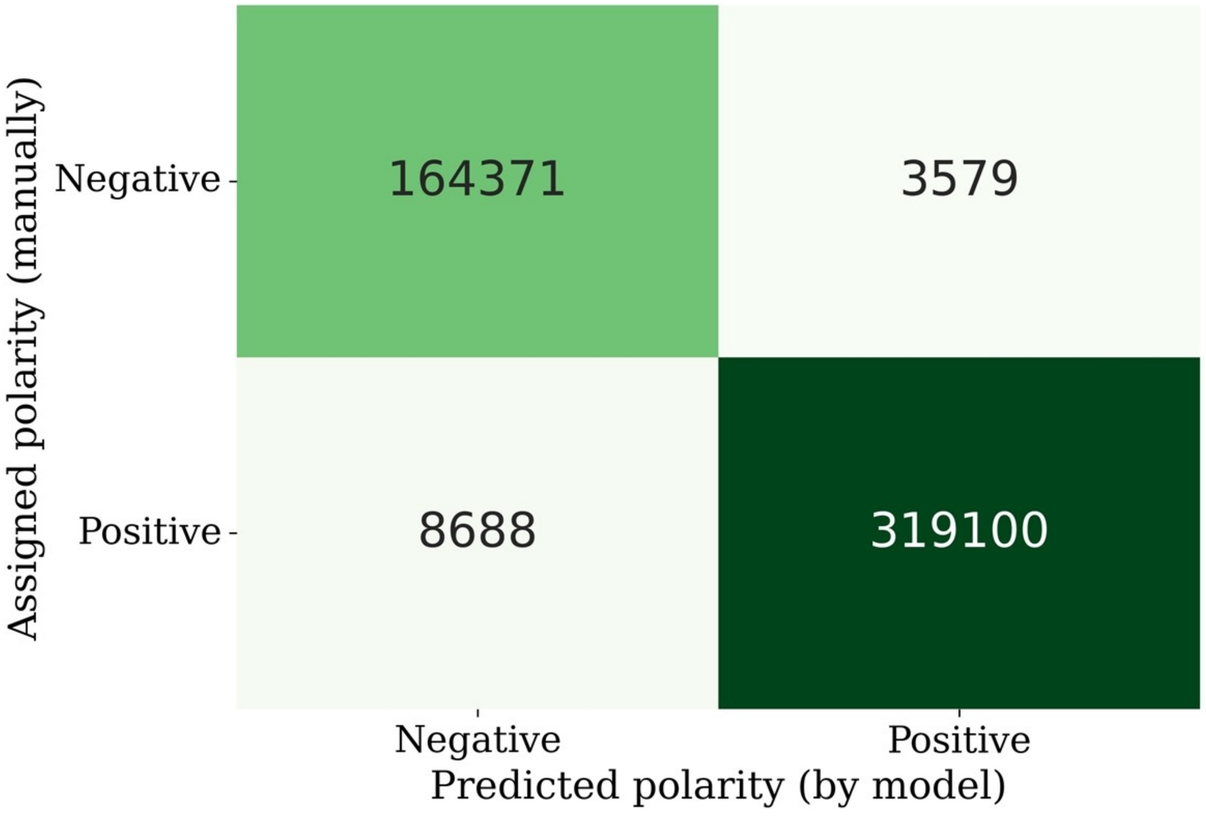
(a)


Assigned polarity (manually)
Negative
Positive
10754
204
391
21576
Negative
Positive
Predicted polarity (by model)

(b)


Assigned polarity (manually)
Negative
Positive
8863
227
603
12239
Negative
Positive
Predicted polarity (by model)

(c)


Assigned polarity (manually)
Negative
Positive
164371
3579
8688
319100
Negative
Positive
Predicted polarity (by model)

(d)

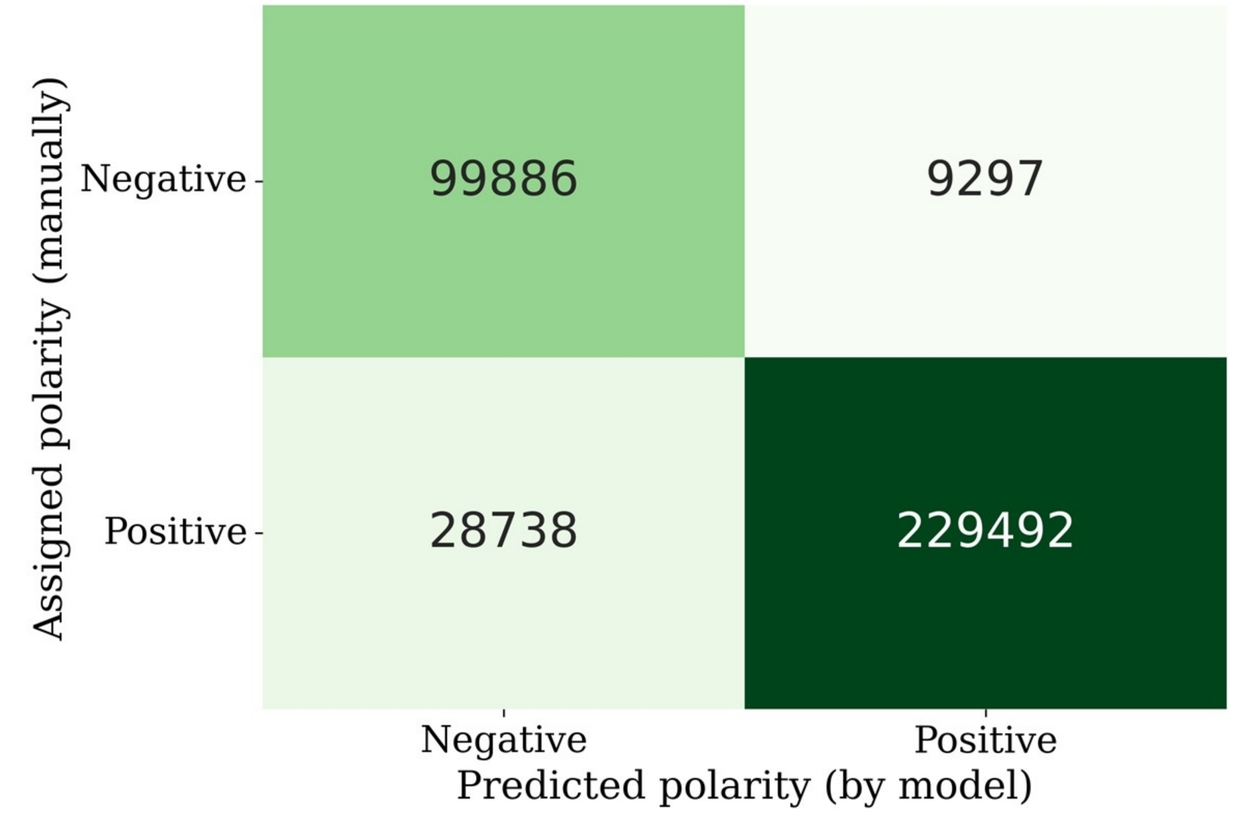

Assigned polarity (manually)
Negative
Positive
99886
9297
28738
229492
Negative
Positive
Predicted polarity (by model)

Figure_3.jpg.

(a)

M = 2.6, SNR = 45.1dB, e = 17.2 km, d = 8.9 km
P_{assgn}: Negative
P_{pred}: Negative [100.0%]

M = 2.1, SNR = 52.8dB, e = 12.3 km, d = 9.5 km
P_{assgn}: Positive
P_{pred}: Positive [100.0%]

M = 2.2, SNR = 31.7dB, e = 19.8 km, d = 9.0 km
P_{assgn}: Positive
P_{pred}: Positive [99.99%]

M = 4.7, SNR = 49.4dB, e = 59.9 km, d = 8.1 km
P_{assgn}: Positive
P_{pred}: Positive [99.91%]

M = 2.2, SNR = 28.8dB, e = 34.5 km, d = 8.7 km
P_{assgn}: Negative
P_{pred}: Negative [100.0%]

M = 1.5, SNR = 21.8dB, e = 19.7 km, d = 8.9 km
P_{assgn}: Negative
P_{pred}: Negative [100.0%]

M = 2.7, SNR = 37.1dB, e = 29.0 km, d = 6.2 km
P_{assgn}: Positive
P_{pred}: Positive [99.99%]

M = 2.4, SNR = 53.7dB, e = 5.2 km, d = 8.5 km
P_{assgn}: Positive
P_{pred}: Positive [99.93%]

Normalised Ground Motion (in counts)

Time (s)

(b)

M = 2.2, SNR = 31.3dB, e = 13.7 km, d = 8.4 km
P_{assgn}: Negative
P_{pred}: Positive [100.0%]

M = 2.5, SNR = 44.2dB, e = 31.2 km, d = 9.6 km
P_{assgn}: Positive
P_{pred}: Negative [99.82%]

M = 2.4, SNR = 26.9dB, e = 14.4 km, d = 8.8 km
P_{assgn}: Negative
P_{pred}: Positive [100.0%]

M = 3.1, SNR = 44.3dB, e = 15.9 km, d = 8.3 km
P_{assgn}: Positive
P_{pred}: Negative [99.99%]

M = 2.5, SNR = 44.2dB, e = 31.2 km, d = 9.6 km
P_{assgn}: Positive
P_{pred}: Negative [99.82%]

M = 2.3, SNR = 54.7dB, e = 17.8 km, d = 12.9 km
P_{assgn}: Negative
P_{pred}: Positive [70.86%]

M = 2.6, SNR = 48.4dB, e = 12.5 km, d = 8.3 km
P_{assgn}: Positive
P_{pred}: Negative [63.55%]

M = 3.3, SNR = 74.7dB, e = 16.1 km, d = 9.0 km
P_{assgn}: Positive
P_{pred}: Negative [99.99%]

Normalised Ground Motion (in counts)

Time (s)

Figure_4.jpg.

(a)

(b)

M = 2.2, SNR = 34.3dB, e = 27.5 km, d = 8.8 km

P_{assgn}: Positive
P_{pred}: Negative [94.5%]
P_{pred}: Positive [100.0%]

M = 2.2, SNR = 31.4dB, e = 19.8 km, d = 11.7 km

P_{assgn}: Positive
P_{pred}: Negative [97.1%]
P_{pred}: Positive [98.8%]

M = 2.1, SNR = 17.6dB, e = 23.1 km, d = 8.4 km

P_{assgn}: Positive
P_{pred}: Negative [64.3%]
P_{pred}: Positive [96.5%]

M = 2.1, SNR = 24.5dB, e = 40.8 km, d = 9.9 km

P_{assgn}: Negative
P_{pred}: Positive [100.0%]
P_{pred}: Negative [100.0%]

M = 2.0, SNR = 24.8dB, e = 23.2 km, d = 11.1 km

P_{assgn}: Positive
P_{pred}: Negative [100.0%]
P_{pred}: Positive [100.0%]

Normalised Ground Motion (in counts)

Time (s)

M = 2.9, SNR = 53.5dB, e = 10.0 km, d = 10.1 km

P_{assgn}: Positive
P_{pred}: Negative [73.44%]

M = 2.1, SNR = 24.9dB, e = 15.1 km, d = 7.9 km

P_{assgn}: Positive
P_{pred}: Negative [64.43%]

M = 4.2, SNR = 92.6dB, e = 7.5 km, d = 8.9 km

P_{assgn}: Negative
P_{pred}: Positive [91.05%]

M = 3.8, SNR = 46.2dB, e = 25.3 km, d = 9.7 km

P_{assgn}: Positive
P_{pred}: Negative [60.55%]

M = 4.7, SNR = 46.0dB, e = 23.9 km, d = 8.66 km

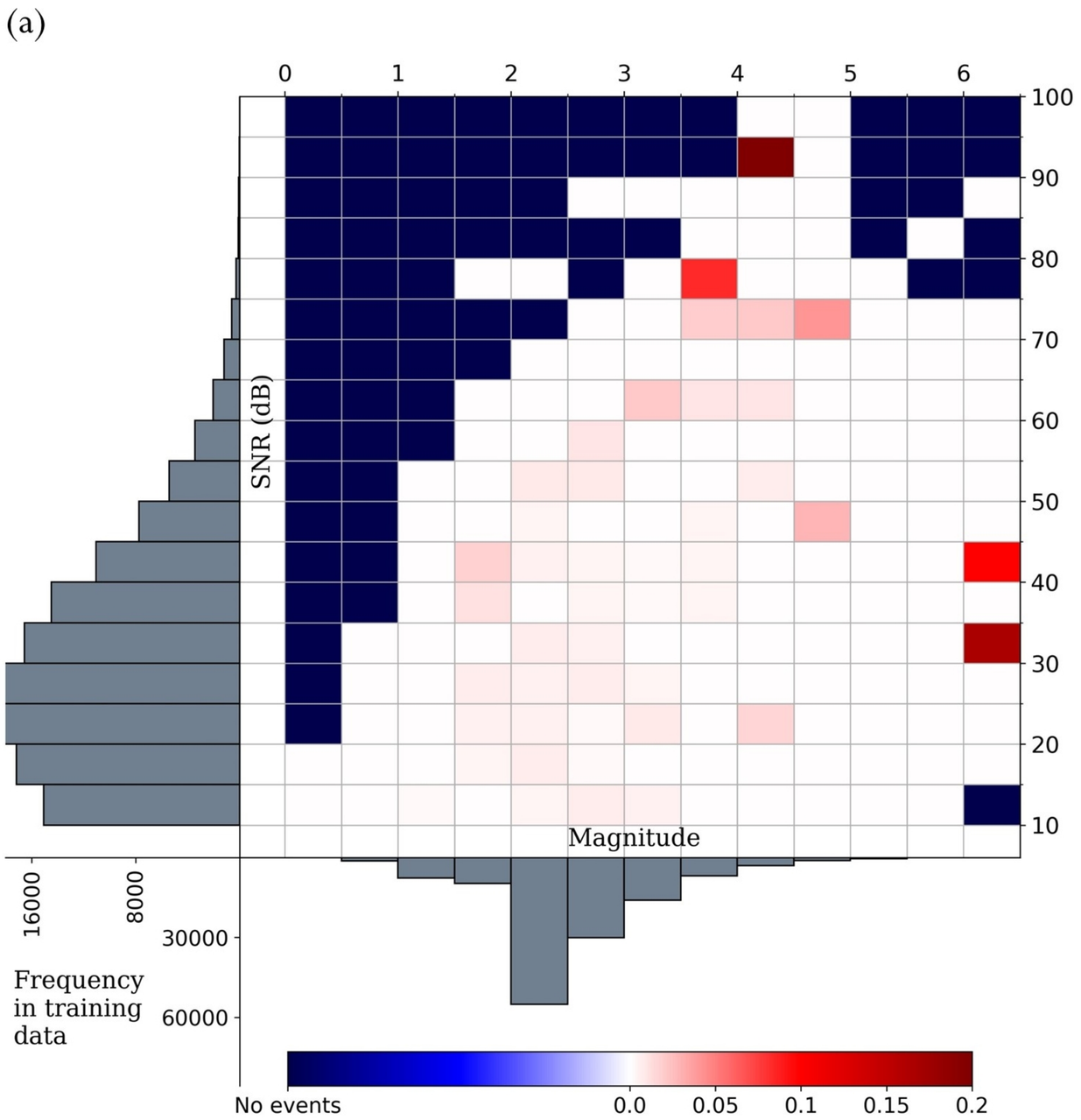
P_{assgn}: Negative
P_{pred}: Positive [89.24%]

Normalised Ground Motion (in counts)

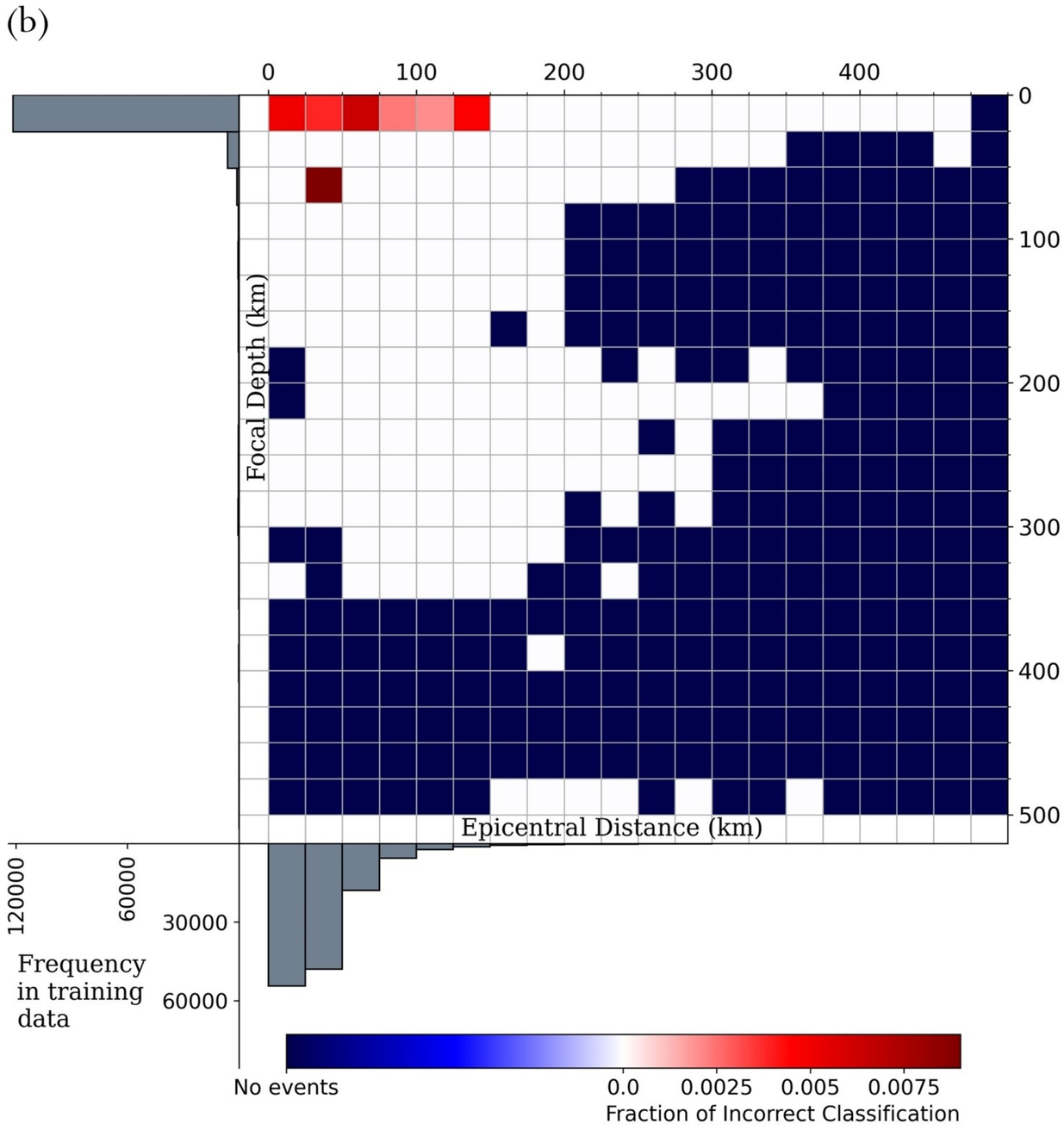
Time (s)

Figure_5.jpg.

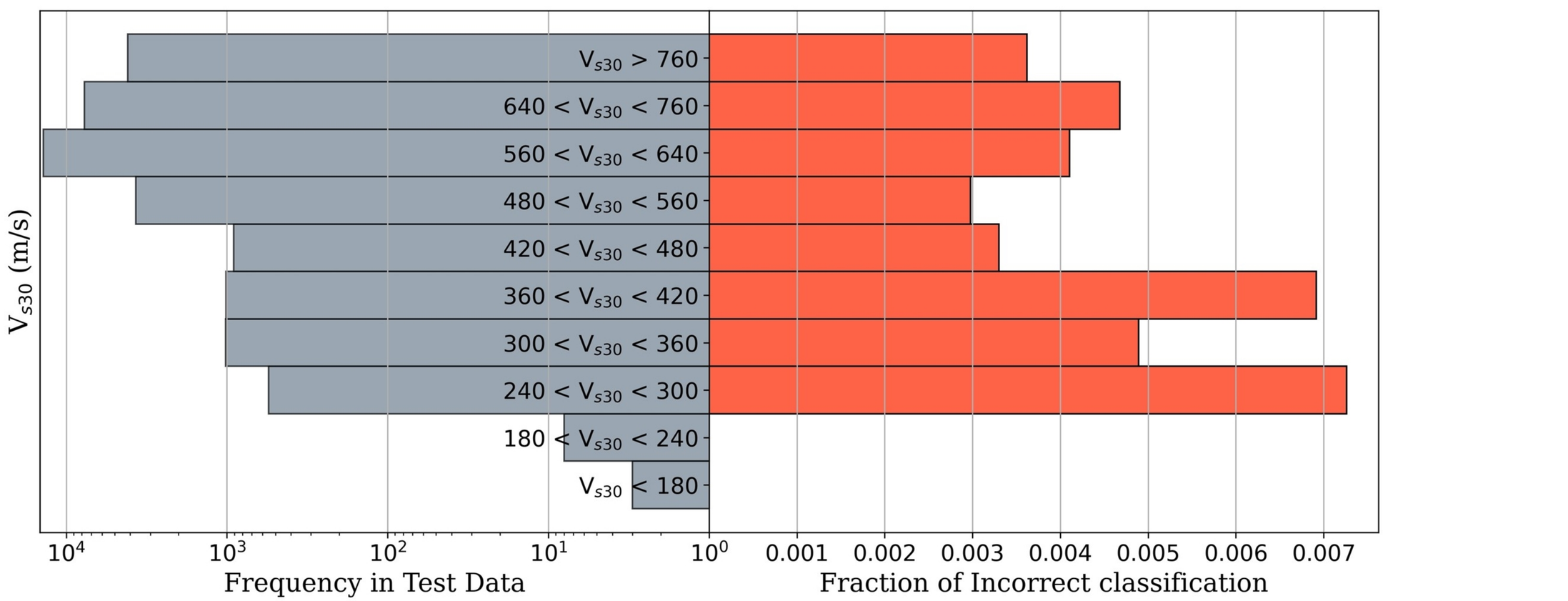
(a)


0
1
2
3
4
5
6
100
90
80
70
60
50
40
30
20
10
SNR (dB)
Magnitude
16000
8000
30000
60000
Frequency in training data
No events
0.0
0.05
0.1
0.15
0.2

(b)


0
100
200
300
400
0
100
200
300
400
500
Focal Depth (km)
Epicentral Distance (km)
120000
60000
30000
60000
Frequency in training data
No events
0.0
0.0025
0.005
0.0075
Fraction of Incorrect Classification

(c)


V_{s30} > 760
640 < V_{s30} < 760
560 < V_{s30} < 640
480 < V_{s30} < 560
420 < V_{s30} < 480
360 < V_{s30} < 420
300 < V_{s30} < 360
240 < V_{s30} < 300
180 < V_{s30} < 240
V_{s30} < 180
V_{s30} (m/s)
10^4
10^3
10^2
10^1
10^0
Frequency in Test Data
0.001
0.002
0.003
0.004
0.005
0.006
0.007
Fraction of Incorrect classification