

# PolarCAP – A deep learning approach for first motion polarity classification of earthquake waveforms



Megha Chakraborty<sup>a,b</sup>, Claudia Quinteros Cartaya<sup>a</sup>, Wei Li<sup>a</sup>, Johannes Faber<sup>a,c</sup>, Georg Rümpker<sup>a,b</sup>, Horst Stoecker<sup>a,c,d,e</sup>, Nishtha Srivastava<sup>a,b,\*</sup>

<sup>a</sup> Frankfurt Institute for Advanced Studies, 60438, Frankfurt am Main, Germany

<sup>b</sup> Institute of Geosciences, Goethe-University Frankfurt, 60438, Frankfurt am Main, Germany

<sup>c</sup> Institute for Theoretical Physics, Goethe-University Frankfurt, 60438, Frankfurt am Main, Germany

<sup>d</sup> Xidian-FIAS International Joint Research Center, Giersch Science Center, 60438, Frankfurt am Main, Germany

<sup>e</sup> GSI Helmholtzzentrum für Schwerionenforschung GmbH, 64291, Darmstadt, Germany

## ARTICLE INFO

### Keywords:

First-motion polarity  
Earthquake waveforms  
Convolutional

## 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 earth-quake 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.

## 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 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, (i) require 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 and 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 (Chen, 2018; Zhu and Beroza, 2019; Li et al., 2021, 2022), magnitude characterization (Mousavi and Beroza, 2020; Chakraborty et al., 2021a,b, 2022). 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

\* Corresponding author. Frankfurt Institute for Advanced Studies, 60438, Frankfurt am Main, Germany.

E-mail address: [srivastava@fias.uni-frankfurt.de](mailto:srivastava@fias.uni-frankfurt.de) (N. Srivastava).

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 and 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. The dataset also contains traces for which multiple detections are retrieved using EQTransformer (Mousavi et al., 2020); this is indicated by the metadata parameter trace\_EQT\_number\_detections. For simplicity we use only those traces for which the number of detections is 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 10 dB. Thus, we ignore traces with SNR less than 10 dB 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 Uchide (2020), 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 and 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.

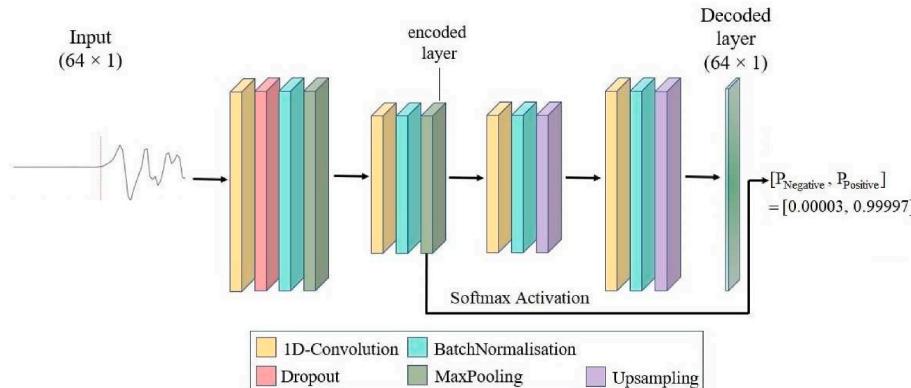
### 2.2. Model architecture and training

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 Fig. 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 and BaAdam, 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 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 10 dB to ensure a good quality of training data, so we test it separately on data with SNR above and below 10 dB. 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) (Southern California earthquake data center, 2013) 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 Fig. 2. As one can see, the accuracy for traces with SNR above 10 dB 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 Fig. 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

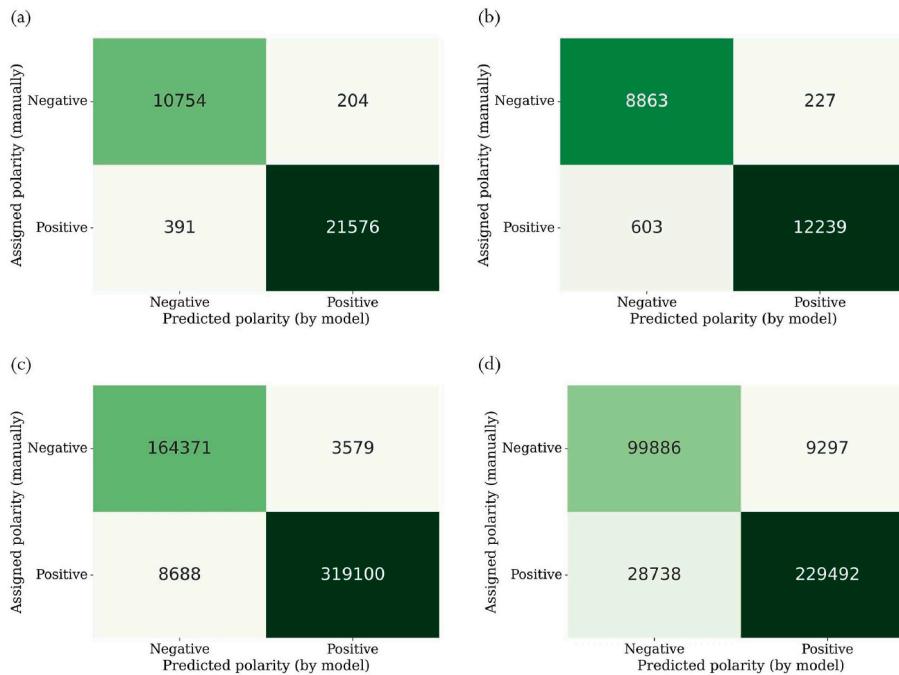


**Fig. 1.** The autoencoder architecture used for our study. The 1D convolutional layers in the encoder use 'relu' activation and have 32 and 8 filters respectively and kernel sizes of 32 and 16, respectively. Each Maxpooling layer reduces the data dimension by 2. The drop out rate used is 0.3. The convolutional 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.

**Table 1**

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

	Test Dataset		Accuracy (%)	Precision (%)		Recall (%)	
				Positive	Negative	Positive	Negative
Data without Augmentation	INSTANCE <sup>a</sup>	SNR $\geq$ 10 dB	98.19	99.06	96.49	98.22	98.12
		SNR < 10 dB	96.22	98.18	93.63	95.3	97.5
	SCSN <sup>b</sup>	SNR $\geq$ 10 dB	97.53	98.89	94.98	97.35	97.87
		SNR < 10 dB	89.65	96.11	77.67	88.87	91.48
Data with Augmentation	INSTANCE <sup>a</sup>	SNR $\geq$ 10 dB	97.65	98.61	95.77	97.86	97.23
		SNR < 10 dB	94.24	97.68	89.98	92.36	96.9
	SCSN <sup>b</sup>	SNR $\geq$ 10 dB	97.78	98.99	95.5	97.64	98.05
		SNR < 10 dB	92.46	96.86	83.55	92.26	92.93

<sup>a</sup> (Michelini et al., 2021).<sup>b</sup> (Southern California earthquake data center, 2013).**Fig. 2.** Confusion matrices for testing model on INSTANCE data (Michelini et al., 2021) with (a) SNR 10 dB (b) SNR < 10 dB and SCSN data used by (Ross et al., 2018) with (c) SNR 10 dB (d) SNR.

model. It is also worth noting that although the model was trained only on data from the Italian National Seismic Network it can perform reasonably well on data from Southern California Seismic Network. This demonstrates the model generalizability.

### 3.1. Comparison with previously published models

In this section we compare the performance of PolarCAP with models described in previous works such as Ross et al. (2018); Uchide (2020); Hara et al. (2019). To do this, we train each model on the same traces, while using the same model architecture, training parameters and input data length specified by the respective authors. We then test the models on both test datasets, with SNR above 10 dB for a reliable comparison. Our observations are summed up in Table 2. As one can see PolarCAP outperforms the other models.

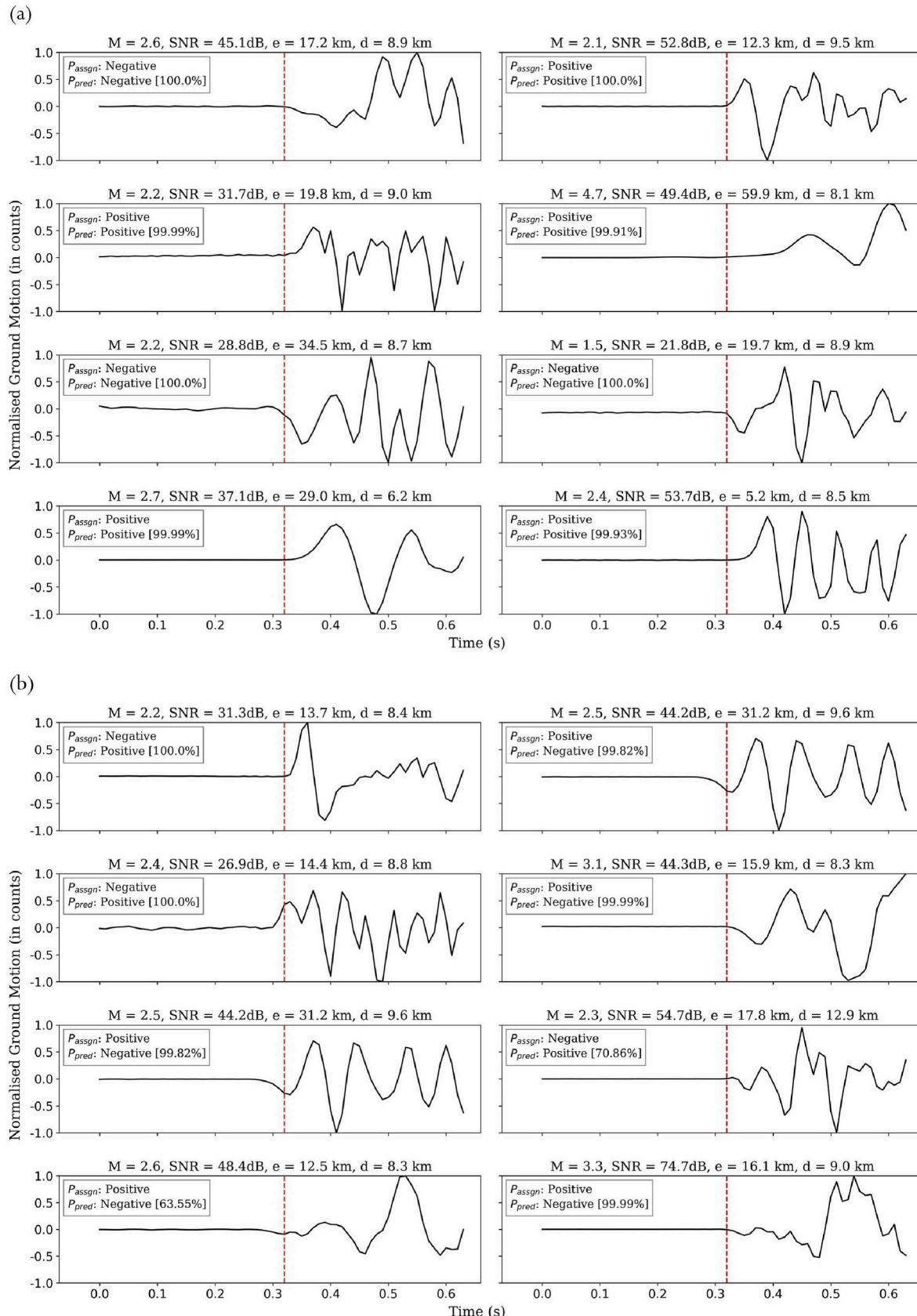
### 3.2. Manual inspection of traces with mismatched assigned and predicted polarity

We further investigate the 595 traces from INSTANCE dataset with SNR above 10 dB for which the polarity predicted by PolarCAP and the manually assigned polarities were in disagreement (refer to Fig. 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 Fig. 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 Fig. 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). Fig. 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



**Fig. 3.** Examples of traces where (a) the polarity predicted by the model matches with the polarity assigned by human analysts, (b<sup>9</sup>) 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.  $P_{\text{assign}}$  and  $P_{\text{pred}}$  represent the assigned and predicted polarities respectively and the percentages in square brackets represent probabilities corresponding to the predicted polarity.

**Table 2**

Comparison between different models. As one can see from the table, PolarCAP outperforms the other models for both the datasets.

Test Dataset	Model	Accuracy (%)	Precision (%)		Recall (%)	
			Positive	Negative	Positive	Negative
INSTANCE <sup>a</sup>	PolarCAP	98.19	99.06	96.49	98.22	98.12
	Ross et al. (2018) <sup>b</sup>	97.2	94.82	98.42	96.88	97.36
	Hara et al. (2019) <sup>c</sup>	97.06	94.7	98.26	96.56	97.31
	Uchide (2020) <sup>d</sup>	98.01	96.08	99.01	98.03	98.01
SCSN <sup>e</sup>	PolarCAP	97.53	98.89	94.98	97.35	97.87
	Ross et al. (2018) <sup>b</sup>	96.18	91.71	98.69	97.53	95.48
	Hara et al. (2019) <sup>c</sup>	97.27	94.37	98.85	97.79	97.01
	Uchide (2020) <sup>d</sup>	97.23	94.04	98.96	98.02	96.82

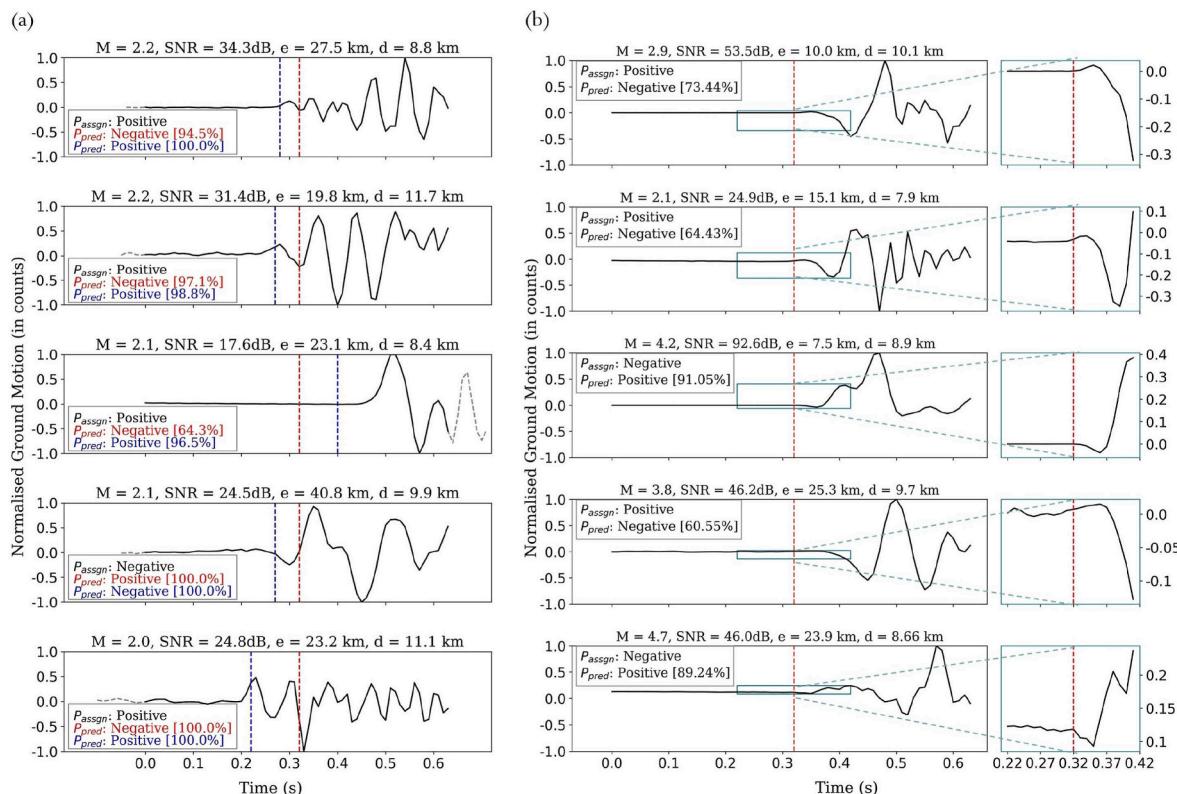
<sup>a</sup> (Michelini et al., 2021).

<sup>b</sup> (Ross et al., 2018).

<sup>c</sup> (Hara et al., 2019).

<sup>d</sup> (Uchide, 2020).

<sup>e</sup> (Southern California earthquake data center, 2013).



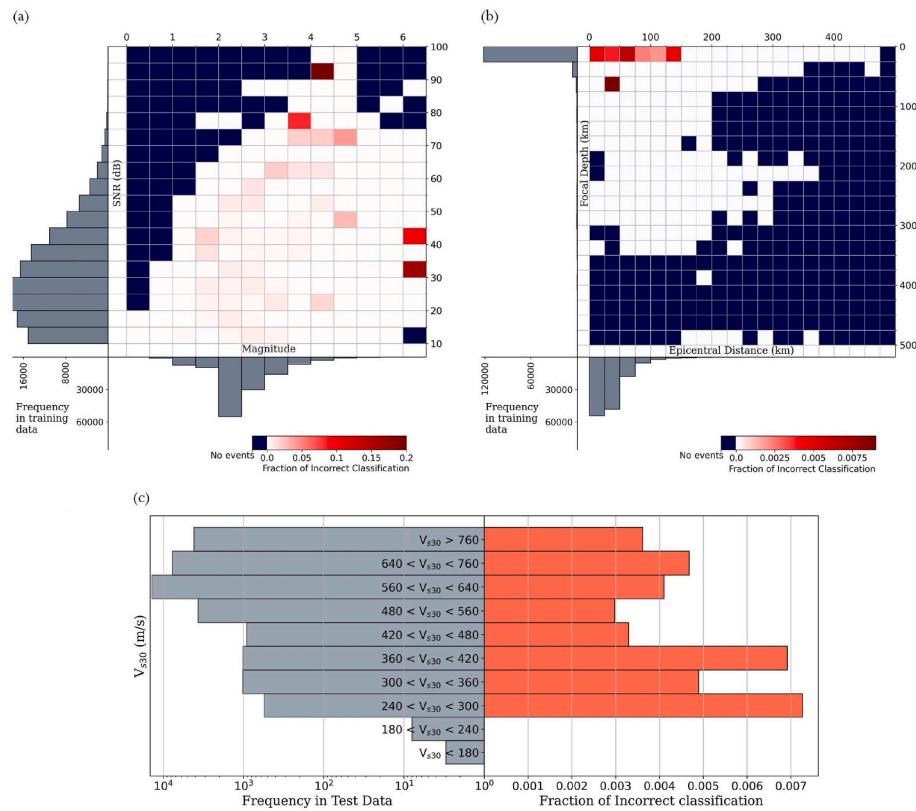
**Fig. 4.** (a) Examples of traces where the polarity prediction was incorrect due to an error in the P-arrival time provided in the metadata (showed with red-dashed line). The P-arrival sample picked by the model from (Li et al., 2021) is shown with a dotted blue line. The dotted gray line shows the 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.  $P_{\text{assign}}$  and  $P_{\text{pred}}$  represent the assigned and predicted polarities respectively and the percentages in square brackets represent probabilities corresponding to the predicted polarity.

picked P-phases and in each of these cases the polarities predicted by the model now matched with the assigned polarities. Fig. 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.

### 3.3. 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 Fig. 5a, which means the model is capable of performing polarity determination across a wide range of SNR and magnitudes. One can also see from Fig. 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



**Fig. 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 m of subsurface ( $V_{S30}$ ).

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 m of subsurface ( $V_{S30}$ ). We find that the tendency for incorrect polarity classification is slightly higher for  $V_{S30} < 420 \text{ ms}^{-1}$ , although for the 11 events recorded at stations with  $V_{S30} < 240 \text{ ms}^{-1}$  there is no observed misclassification (Fig. 5c).

#### 3.4. 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 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 Southern California Earthquake Data Center (2013), 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. As one might expect, the improvement was more significant for low SNR cases. This could indicate that the augmentation helps in better representing the traces in the SCSN dataset (Southern California Earthquake Data Center, 2013).

#### 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.

#### Declaration of competing interest

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

#### 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 Artificielle Intelligenz (SAI). The authors are also thankful to Darius Fenner and Jonas Köhler for their kind suggestions. Prof. Dr. Horst Stoecker 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. A study of the behavior of several methods for balancing machine learning training data. SIGKDD Explorations Newsletter 6, 20–29. <https://doi.org/10.1145/1007730.1007735>.
- Bermann, P., Klinge, K., Wendt, S., 2014. Data Analysis and Seismogram interpretation, pp. 1–126.
- Brachmann, A., Barth, E., Redies, C., 2017. Using cnn features to better understand what makes visual art works special. Frontiers in Psychology 8. <https://doi.org/10.3389/fpsyg.2017.00830>.
- Chakraborty, M., Fenner, D., Li, W., Faber, J., Zhou, K., Rümpker, G., et al., 2022. CREIME—a convolutional recurrent model for earthquake identification and magnitude estimation. J. Geophys. Res. Solid Earth 127, e2022JB024595. <https://doi.org/10.1029/2022JB024595>.
- Chakraborty, M., Li, W., Faber, J., Ruempker, G., Stoecker, H., Srivastava, N., 2021a. A Study on the Effect of Input Data Length on Deep Learning Based Magnitude Classification. <https://doi.org/10.48550/ARXIV.2112.07551>. <https://arxiv.org/abs/2112.07551>.
- Chakraborty, M., Rümpker, G., Stöcker, H., Li, W., Faber, J., Fenner, D., Zhou, K., Srivastava, N., 2021b. Real Time Magnitude Classification of Earthquake Waveforms using Deep Learning, EGU General Assembly 2021, online, EGU21-15941. <https://doi.org/10.5194/egusphere-egu21-15941>.
- Charles, P., 2013. Project title. <https://github.com/charlespwd/project-title>.
- Chen, Y., 2018. Automatic microseismic event picking via unsupervised machine learning. Geophys. J. Int. 212, 88–102.
- Chen, C., Holland, A.A., 2016. Phasepapy: a robust pure python package for automatic identification of seismic phases. Seismological Research Letters 87, 1384–1396.
- 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. <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, 2264–2276.
- Kingma, D.P., Ba, J., Adam, 2014. A Method for Stochastic Optimization.
- Kiranyaz, S., Ince, T., Hamila, R., Gabbouj, M., 2015. Convolutional Neural Networks for Patient-specific Ecg Classification, pp. 2608–2611. <https://doi.org/10.1109/EMBC.2015.7318926>.
- Li, W., Chakraborty, M., Fenner, D., Faber, J., Zhou, K., Rümpker, G., Stöcker, H., Srivastava, N., 2021. Epick: multi-class attention-based u-shaped neural network for earthquake detection and seismic phase picking. <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. <https://doi.org/10.48550/ARXIV.2204.02870>. <https://arxiv.org/abs/2204.02870>.
- Liao, W., Lee, E., Mu, D., Chen, P., Rau, R., 2021. ARRU phase picker: attention recurrent-residual U-net for picking seismic P- and S-phase Arrivals. Seismol. Res. Lett. 92, 2410–2428. <https://doi.org/10.1785/0220200382>. <https://pubs.geoscienceworld.org/ssa/srl/article-pdf/92/4/2410/5351037/s>.
- Lundervold, A.S., Lundervold, A., 2019. An overview of deep learning in medical imaging focusing on mri. Zeitschrift für Medizinische Physik 29, 102–127. <https://doi.org/10.1016/j.zemedi.2018.11.002>, special Issue. <https://www.sciencedirect.com/science/article/pii/S0939388918301181> (Deep Learning in Medical Physics).
- Michelini, A., Cianetti, S., Gaviano, S., Giunchi, C., Jozi nović, D., Lauciani, V., 2021. Instance the Italian seismic dataset for machine learning, seismic waveforms and associated metadata, Istituto Nazionale di Geofisica e Vulcanologia (INGV). <https://doi.org/10.13127/instance>.
- Mousavi, S.M., Beroza, G.C., 2020. A machine-learning approach for earth-quake magnitude estimation. Geophys. Res. Lett. 47, e2019GL085976. <https://doi.org/10.1029/2019GL085976>.
- 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, 1693–1697. <https://doi.org/10.1109/LGRS.2019.2909218>.
- Mousavi, S.M., Ellsworth, W.L., Zhu, W., Chuang, L., Beroza, G., 2020. Earth-quake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking. Nat. Commun. 11. <https://doi.org/10.1038/s41467-020-17591-w>.
- Nagi, J., Ducatelle, F., Di Caro, G.A., Cireşan, D., Meier, U., Giusti, A., Nagi, F., Schmidhuber, J., Gambardella, L.M., 2011. Max- pooling convolutional neural networks for vision-based hand gesture recognition. In: IEEE International Conference on Signal and Image Processing Applications. ICSIPA), pp. 342–347. <https://doi.org/10.1109/ICSIPA.2011.6144164>, 2011.
- Perol, T., Gharbi, M., Denolle, M., 2018. Convolutional neural network for earthquake detection and location. Sci. Adv. 4, e1700578. <https://doi.org/10.1126/sciadv.1700578>.
- Prechelt, L., 2012. Early Stopping — but when? Springer Berlin Heidelberg, pp. 53–67. [https://doi.org/10.1007/978-3-642-35289-8\\_5](https://doi.org/10.1007/978-3-642-35289-8_5).
- Pugh, D., White, R., Christie, P., 2016. Automatic bayesian polarity determination. Geophysical Journal International 206, 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, 5120–5129. <https://doi.org/10.1029/2017JB015251.arXiv>. <https://agupubs.onlinelibrary.wiley.com/doi/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. MIT Press, Cambridge, MA, USA, pp. 318–362. <http://dl.acm.org/citation.cfm?id=104279.104293>.
- Southern California Earthquake Data Center, 2013. <https://doi.org/10.7909/C3WD3xH1>.
- Ting, K.M., 2017. Confusion Matrix. Springer US, Boston, MA. [https://doi.org/10.1007/978-1-4899-7687-1\\_50](https://doi.org/10.1007/978-1-4899-7687-1_50), 260–260.
- Uchide, T., 2020. Focal mechanisms of small earthquakes beneath the Japanese islands based on first-motion polarities picked using deep learning. Geophysical Journal International 223, 1658–1671. <https://doi.org/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. <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, 261–273. <https://doi.org/10.1093/gji/ggy423>.