A study on the effect of input data length on deep learning based magnitude classifier

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

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

Abstract

The rapid characterisation of earthquake parameters such as its magnitude is at the heart of Earthquake Early Warning (EEW). In traditional EEW methods the robustness in the estimation of earthquake parameters have been observed to increase with the length of input data. Since time is a crucial factor in EEW applications, in this paper we propose a deep learning based magnitude classifier and, further we investigate the effect of using five different durations of seismic waveform data after first P-wave arrival– 1s, 3s, 10s, 20s and 30s. This is accomplished by testing the performance of the proposed model that combines Convolution and Bidirectional Long-Short Term Memory units to classify waveforms based on their magnitude into three classes– "noise", "low-magnitude events" and "high-magnitude events". Herein, any earthquake signal with magnitude equal to or above 5.0 is labelled as high-magnitude. We show that the variation in the results produced by changing the length of the data, is no more than the inherent randomness in the trained models, due to their initialisation.

1 Introduction

The earthquake magnitude, defined as a logarithmic measure of the relative strength of an earthquake, is one of the most fundamental parameters in its characterisation[1]. The complex nature of the geophysical processes affecting earthquakes makes it very difficult to have a single reliable measure for its size [2] and hence, magnitude values measured in different scales often differ by more than 1 unit. This is especially true for larger events due to saturation effects [3, 4]. Owing to above-mentioned reasons and the empirical nature of majority of the magnitude scales, it is one of the most difficult parameters to estimate [5, 6].

Some of the classical approaches to obtain first estimates of earthquake magnitude have used empirical relations for parameters such as predominant period τ_p^{max} [7, 8], effective average period τ_c [9, 10] in the frequency domain and parameters such as peak displacement (P_d) [10, 11] in the amplitude domain calculated from the initial 1-3 seconds of P-waves. These relations form the basis of existing Earthquake Early Warning (EEW) systems in Japan, California, Taiwan etc. ([12] and the references therein). The accuracy of such estimates have been shown to increase with the duration of data used to calculate them [13].

The recent developments in the area of deep learning [14], combined with the availability of affordable high-end computational power through GPUs, have led to state-of-the-art results in image recognition [15, 16], speech recognition [17, 18] and natural language processing [19, 20]. In fields such as seismology, where the volume of available data has increased exponentially over the last decades [21], deep learning has achieved great success in tasks such as seismic phase picking [22–24], event detection [25–27], magnitude estimation [1], event location characterisation [28–30], and first motion polarity detection [31].

Considering that timeliness is of the essence in rapid earthquake characterisation, it becomes important to find an optimum duration for the input data, that can provide a reliable and statistically significant estimate for various earthquake parameters while using minimum amount of P-wave data. In this study, we present a deep learning model to perform time-series multiclass classification [32, 33] that classifies seismic waveforms as – "noise, "low-magnitude" or "high-magnitude". Here a local magnitude of 5.0 is taken to be the boundary between the low-magnitude and high-magnitude classes. We further investigate the effect of using different lengths of data on the model performance. Please note, that the boundary of 5.0 is arbitrarily chosen, and can be modified depending on the purpose of the model and the local geology (which influences the correlation between earthquake magnitude and intensity). The boundary in itself does not influence the model performance. Unlike [34], which uses data from three seismic station to characterise different earthquake parameters, the model discussed in this paper only uses three-component data from a single station.

2 Methodology

2.1 Data Used

We use data from the STanford EArthquake Dataset (STEAD) [35] to train and test our model. STEAD is a high-quality bench-marked dataset created for machine learning and deep learning applications and contains seismic event and noise waveforms of duration 1 minute recorded by over 2,500 seismic stations across the globe. The waveforms have been detrended and filtered with a bandpass filter between 1.0 to 40.0 Hz, followed by a resampling at 100Hz. A metadata consisting of 35 attributes for earthquake traces and 8 attributes for noise traces is provided by the authors.

To ensure consistency in magnitude we only use traces for which the magnitude is provided in ‘ml’ scale (as this is the case for most of the traces in the dataset). We also discard traces with signal-to-noise ratio less than 10dB for quality control. We divide the noise and earthquake traces into training, validation and test sets in the ratio 60:10:30. Care is taken to make sure that the three aforementioned datasets are non-overlapping. This means, that traces corresponding to a particular earthquake (represented by the ‘source_id’ attribute) but recorded at different stations are included in only one of the three sets. For noise traces, recordings from a particular seismic station are included in only one of the three sets.

In this paper, we propose a classifier model for rapid earthquake characterisation. Furthermore, we investigate the effect of using different lengths of data after the first P-arrival (1s, 3s, 10s, 20s and 30s) on the performance of this classifier model. In each case the P-wave data is preceded by 2.8-3.0 seconds of pre-signal noise, so the model can learn the noise characteristics of the station [36]. The data labels 0, 1, and 2 are used to denote the classes noise, low-magnitude and high-magnitude, respectively.

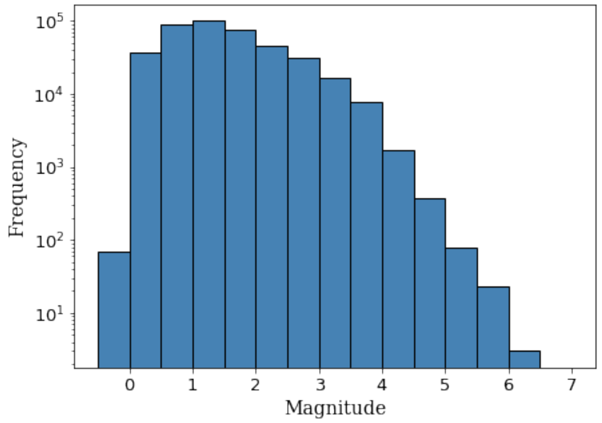


Figure 1: Original distribution of local magnitudes in the chunk of STEAD [35] data used for training.

As mentioned earlier, we take a local magnitude 5.0 to be the decision boundary between high-magnitude and low-magnitude events. However, the training dataset originally has a magnitude distribution as shown in Figure 1; this would lead to a high imbalance between the low-magnitude and high-magnitude classes (a ratio of nearly 3300:1). It is widely agreed by the Machine Learning community that most classifiers assume an equal distribution between the different classes.[37] Although examples from some domains where models perform reasonably well even in highly imbalanced datasets, show that there are other factors at play, imbalanced datasets not only are a major hindrance in the development of good classifiers but can also lead to misleading evaluations of the accuracy of the model [37]. To tackle this *imbalance problem* we apply resampling of the data [38] as follows:

- Events with magnitude equal to or above 5.0 are represented 20 times in the dataset, by using a *shifting window* starting from 300 samples to 280 samples before the first P-arrival sample, the window being shifted by 2 samples for each representation. Each of these traces are also *flipped*, i.e. their polarity is reversed, since it does not affect the magnitude information of the data. Such data augmentation techniques used for images have also been found to be useful for time series data [37, 39].
- For low-magnitude events the following strategy of random-undersampling is adopted:
  1. All events with magnitude between 4.5 and 5.0 are used.
  2. 1/3rd of events with magnitude between 4.0 and 4.5 are used.
  3. 1/50th of events with magnitude between 2.0 and 4.5 are used.
  4. 1/100th of events with magnitude less than 2.0 are used.
- 1/25th of the available noise traces are used.

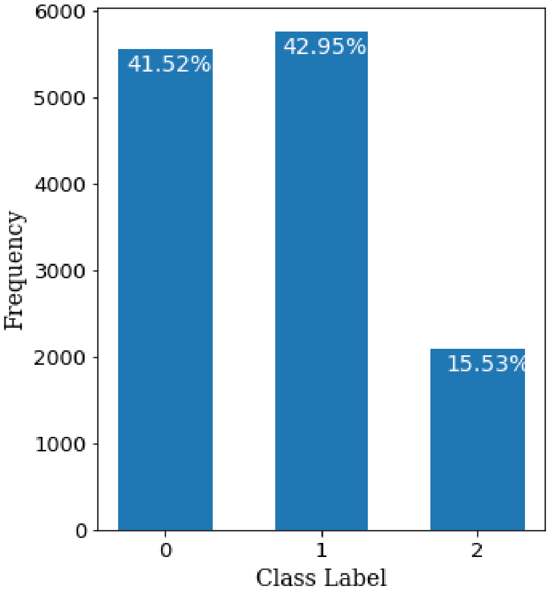


Figure 2: The distribution of classes in the training dataset obtained by under-sampling noise and low-magnitude data and applying data-augmentation to high-magnitude events. Classes 0,1 and 2 represent *'noise'*, *'low-magnitude'* and *'high-magnitude'* data, respectively. A similar distribution of classes is seen in the validation and test datasets as well.

Note that special care is taken to include more events close to the decision boundary, so that the model can learn to differentiate between events of magnitude say, 4.0 to 5.0 which is more difficult compared to differentiating between events of magnitude say, 2.0 and 5.0. The corresponding distribution of the different classes is shown in Figure 2. The validation and test datasets follow a similar distribution. As one can see, in spite of the resampling techniques employed, the high-magnitude class is still under-respresented in the dataset, as

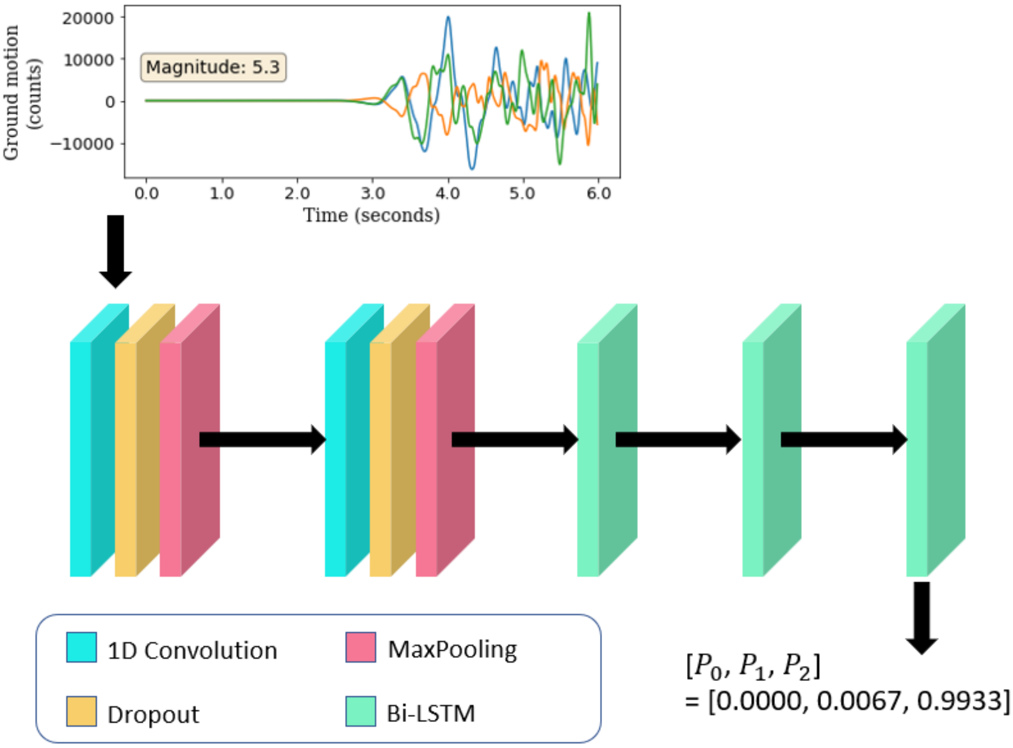


Figure 3: The architecture of the model used to perform the 3-class classification. The input to the model is 3-component seismic waveform data from a single station. The example shown here corresponds to the case where 3 seconds of P-wave data is used (the total length of data is, thus, 6 seconds). The 1D Convolution layers have a kernel size of 4 and 8 filters each; the drop rate for each Dropout layer is 0.2, and each MaxPooling layer reduces the size of the data by a factor of 4; the Bi-LSTM layers have dimensions of 256, 256 and 128, respectively. The final layer is a Softmax layer, that outputs the probability of the trace belonging to classes 0 (noise), 1 (low-magnitude) and 2 (high-magnitude), represented here as P_0, P_1 and P_2 respectively. In this case a probability of 0.9933 is assigned to class 2, for an event with magnitude 5.3; thus, this is a case of correct classification.

compared to the other two classes. So we apply a class-weight [38] of 1:1:10 (chosen, experimentally) for classes 0,1 and 2 while training the model. The data is used without instrument response removal. Unlike [40] we do not normalise the data. Only the waveform information is provided to the model.

2.2 Model Architecture and Model Training

The model architecture [41] consists of two sets of 1D Convolution [42], Dropout [43] and MaxPooling[44], followed by three bidirectional Long-Short Term Memory (LSTM) [45] layers; the final layer is a Softmax layer [46] which gives a three-element array of the form $[P_0, P_1, P_2]$, where P_i is the probability of the waveform belonging to the class i (Figure 3).

The model is trained using Adam optimiser [47], Categorical Crossentropy [48] loss and a batch size of 256. Early stopping [49] is used to prevent overfitting, whereby the validation loss is monitored and the training stops when there is no reduction in it for 20 consecutive epochs. We start with a learning rate of 10^{-3} and reduce it by a factor of 10 if the validation loss does not reduce for 15 consecutive epochs until it reaches 10^{-6}. The model for the epoch corresponding to the lowest validation loss is retained.

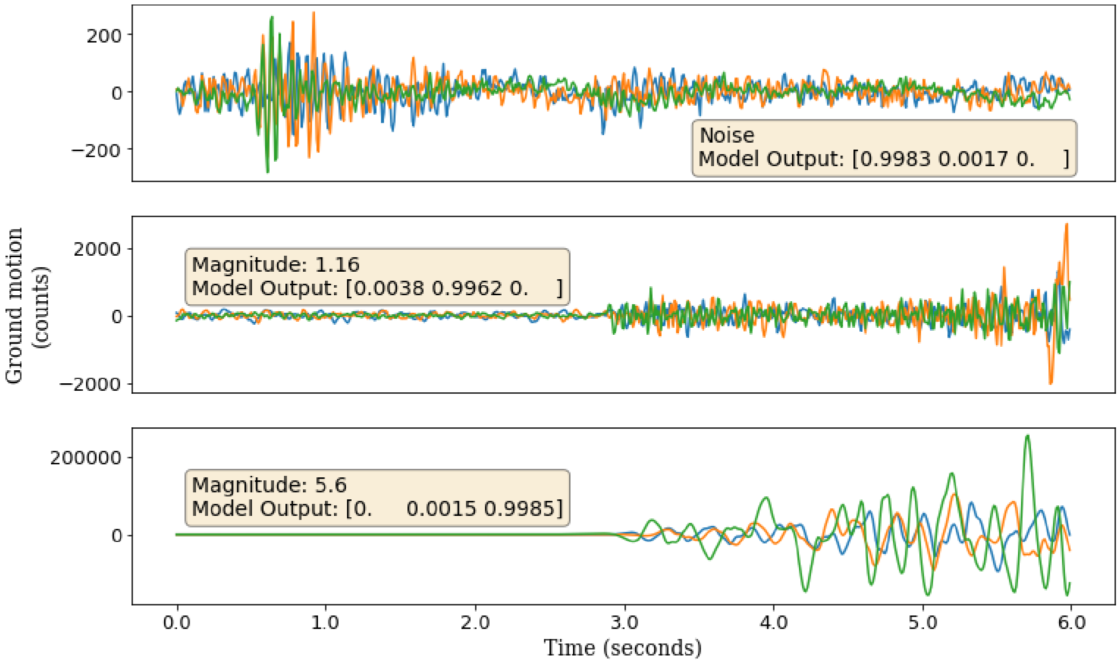


Figure 4: Examples of waveforms that have been correctly classified. In each case the highest probability corresponds to the respective class.

3 Results

To analyse the effect of different lengths of data on the performance of the classifier model, we use the metrics listed below to evaluate the model performance. The metrics are calculated in terms of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN).

- **Accuracy**: The accuracy of a classifier is the proportion of testing samples that are correctly classified. Mathematically, it can be defined as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \tag{1}$$

- **Precision**: This is the ratio of the number of times the model *correctly* predicts a class to the total number of times it predicts that class. Mathematically it is defined as:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2}$$

- **Recall**: This is the ratio of the number of times the model correctly predicts a class to the total number occurences of that class in the dataset. Mathematically it is defined as:

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3}$$

Figure 4 shows three waveforms, (one from each class) that has been correctly classified. The softmax probalities, as described in section 2.2, are also shown. In each case the highest probability is predicted for the corresponding class.

Figure 5 shows the softmax probalities, predicted by the model for different lengths of the same waveform. Although the waveform is correctly classified in each case, the predicted probabilities are different and show no dependence on the length of input data. Figure 6a shows the variation in the model performance with the duration of P-wave data used. We also look at the randomness in the performance when the model is trained on the same data five times (Figure 6b), as we do not tune a random seed during model training [50, 51]. Thus, we can see that the variation in the results caused by changing the length of data is comparable to the randomness in the results due to random-initialisation upon re-training the model

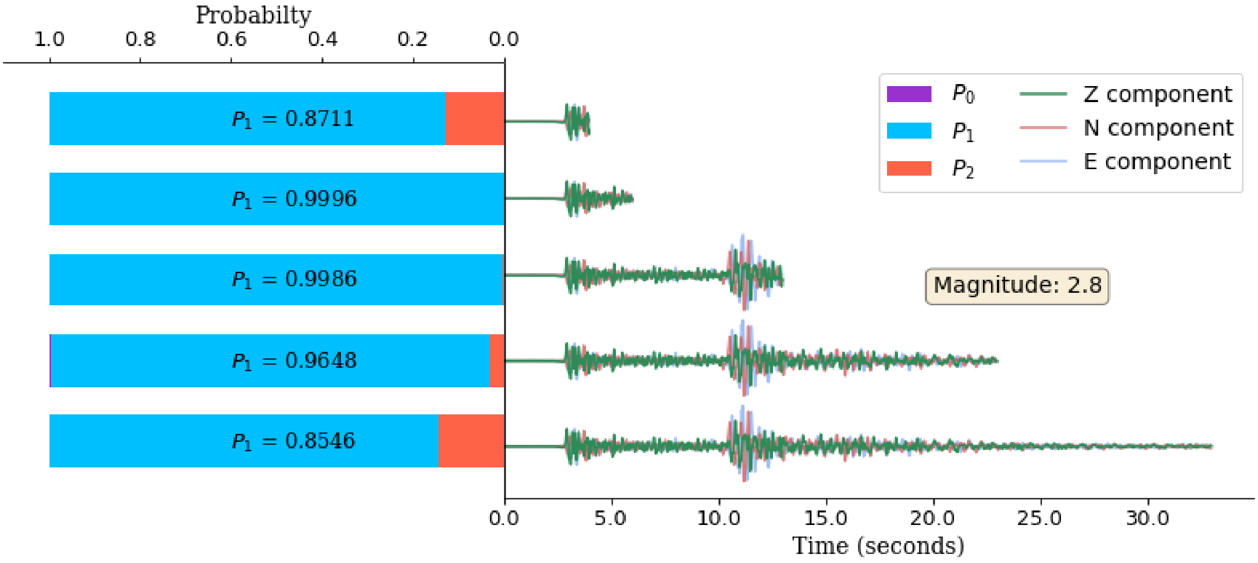
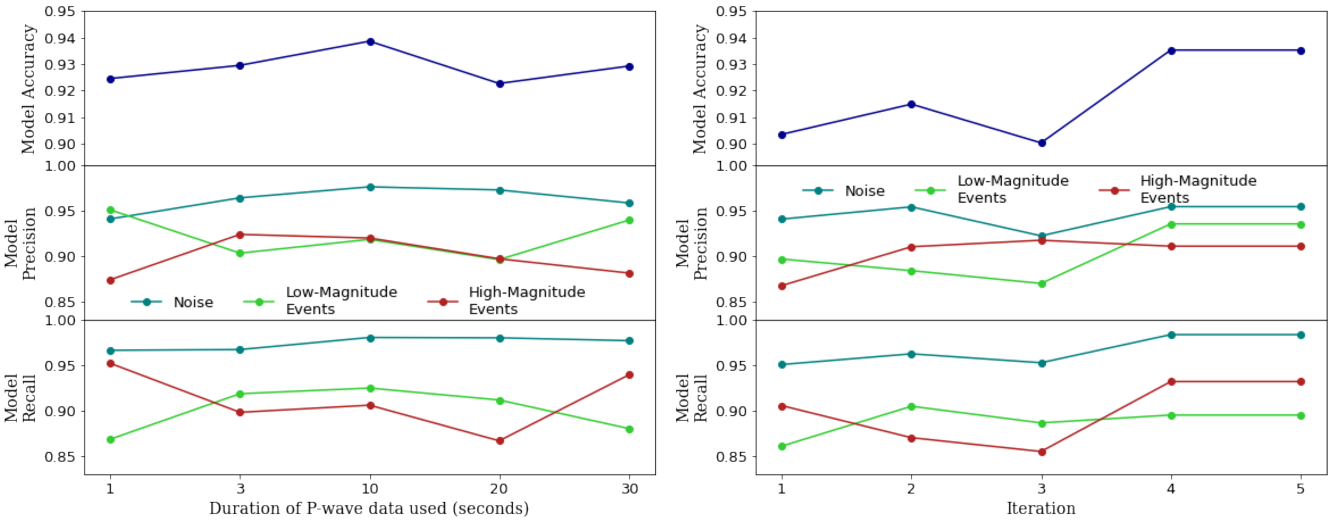


Figure 5: Softmax probabilities for different input lengths of the same , predicted by the models trained on the corresponding lengths of data. The waveform used here corresponds to an event of magnitude 2.8, although the maximum probability corresponds to class 1, the values of these probabilities are different for different data lengths, and there is no clear dependence between the length of the data and this probability.



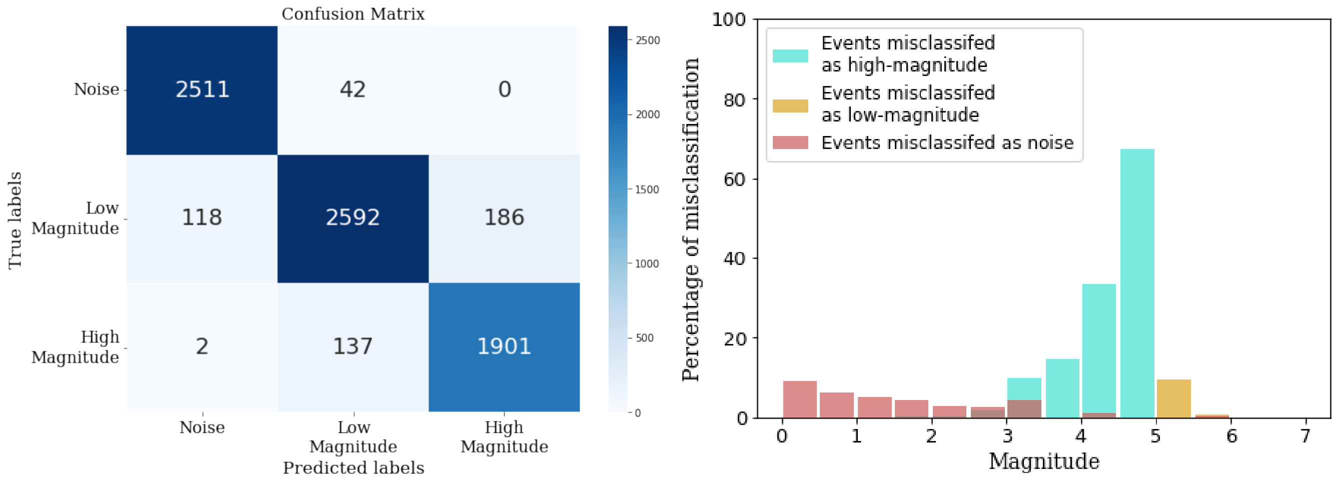
(a)

(b)

Figure 6: (a) Variation in classifier model performance when different duration of P-wave data are used; (b) Variation in the classifier model performance when the same model is re-trained on the same data (in this case 3 seconds of P-wave data used) five times. This shows that the variation in the two cases are comparable.

on the same data.

Figure 7 shows the classification statistics for one of the iterations of the model trained on the 3 second data. One can see that the events classified as noise tend to be of low magnitude, while the mis-classification of low-magnitude events as high-magnitude and vice-versa, is most pronounced at the decision boundary of 5.0. Another important observation is that the degree of misclassification of low-magnitude events is much higher than the reverse case; approximately 65% of events with magnitude between 4.5 and 5.0 and 35% of events with magnitude between 4.0 and 4.5 get classified as high magnitude, while less than 10% of events with magnitude between 5.0 and 5.5 are classified as low-magnitude; this is intentional as a missed alarm is considered more dangerous than a false alarm in this context [52] and is achieved by giving the high-magnitude class more weight during



(a)

(b)

Figure 7: The classification results for a model trained on the 3 second data. (a) The confusion matrix [53] for a model trained and tested on the 3 second data. (b) The mis-classification statistics for the same model, for different magnitude values. Note how the highest degree of mis-classification happens close to the decision boundary; the percentage of low-magnitude events classified as high-magnitude is much higher than the percentage of high-magnitude events classified as low-magnitude; this is a result of the class-weights we used while training the model.

model training.

4 Conclusion

In this study, we present a deep learning model that classifies seismic waveform into three-classes: *noise*, *low-magnitude* events and *high-magnitude* events, with events having local magnitude equal to or above 5.0 categorised as 'high-magnitude'. We investigate the effect of using different duration of P-wave data to perform the said task and demonstrate that changing the length of the waveform has no significant effect on the model performance. We also find that the model classifies most the data above a magnitude of 4.5 as high-magnitude, even though the decision boundary is chosen at 5.0, due to the higher class weight assigned to high-magnitude events. We obtain an overall accuracy ranging between 90.04% and 93.86% (which is comparable to the magnitude classification accuracy of 93.67% achieved by [34] using data from three seismic stations).

5 Acknowledgement

This research is supported by the "KINachwuchswissenschaftlerinnen" - 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).

References

[1] S. M. Mousavi and G. C. Beroza. "A machine-learning approach for earthquake magnitude estimation." In: *Geophysical Research Letters* 47, e2019GL085976. (2020). URL: `https://doi.org/10.1029/2019GL085976`.

[2] H. Kanamori and G. S. Stewart. "Seismological aspects of the Guatemala Earthquake of February 4, 1976". In: *Journal of Geophysical Research: Solid Earth* 83.B7 (1978), pp. 3427–3434. DOI: `https://doi.org/10.1029/JB083iB07p03427`.

[3] B.F. Howell Jr. "On the saturation of earthquake magnitudes." In: *Bulletin of the Seismological Society of America* 71(5) (1981), pp. 1401–1422. URL: https://doi.org/10.1785/BSSA0710051401.

[4] H. Kanamori. "Magnitude scale and quantification of earthquakes." In: *Tectonophysics* 93(3-4) (1983), pp. 185–199. URL: https://doi.org/10.1016/0040-1951(83)90273-1.

[5] D. H. Chung and D. L. Bernreuter. "Regional relationships among earthquake magnitude scales". In: *Reviews of Geophysics* 19.4 (1981), pp. 649–663. DOI: https://doi.org/10.1029/RG019i004p00649. eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/RG019i004p00649. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/RG019i004p00649.

[6] G. Ekström and A. Dziewonski. "Evidence of bias in estimations of earthquake size." In: *Nature* 332 (1988), pp. 319–323. DOI: https://doi.org/10.1038/332319a0.

[7] Y. Nakamura. "On the Urgent Earthquake Detection and Alarm System (UrEDAS)". In: *9th world conference on earthquake engineering* VII.B7 (1988), pp. 673–678.

[8] R. Allen and H. Kanamori. "The Potential for Earthquake Early Warning in Southern California". In: *Science (New York, N.Y.)* 300 (May 2003), pp. 786–789. DOI: 10.1126/science.1080912.

[9] H. Kanamori. "REAL-TIME SEISMOLOGY AND EARTHQUAKE DAMAGE MITIGATION". In: *Annual Review of Earth and Planetary Sciences* 33.1 (2005), pp. 195–214. DOI: 10.1146/annurev.earth.33.092203.122626.

[10] X. Jin, H. Zhang, J. Li, Y. Wei, and Q. Ma. "Earthquake magnitude estimation using the τ_c and P_d method for earthquake early warning systems". In: *Earthquake Science* 26.es-26-1-23 (2013), pp. 23–31. ISSN: 1674-4519. DOI: 10.1007/s11589-013-0005-4.

[11] Y.M. Wu and L. Zhao. "Magnitude estimation using the first three seconds P-wave amplitude in earthquake early warning". In: *Geophysical Research Letters* 33.16 (2006). URL: https://doi.org/10.1029/2006GL026871.

[12] R. Allen, P. Gasparini, O. Kamigaichi, and M. Böse. "The Status of Earthquake Early Warning around the World: An Introductory Overview". In: *Seismological Research Letters* 80 (2009), pp. 682–693. DOI: 10.1785/gssrl.80.5.682.

[13] A. Ziv. "New frequency-based real-time magnitude proxy for earthquake early warning". In: *Geophysical Research Letters* 41.16 (2014), pp. 7035–7040. URL: https://doi.org/10.1002/2014GL061564.

[14] Y. LeCun, Y. Bengio, and G. Hinton. "Deep learning." In: *Nature* 521 (2015), pp. 436–444. URL: https://doi.org/10.1038/nature14539.

[15] A. Krizhevsky, I. Sutskever, and G. E. Hinton. "ImageNet Classification with Deep Convolutional Neural Networks". In: *Communications of the ACM* 60.6 (2017), pp. 84–90. ISSN: 0001-0782. URL: https://doi.org/10.1145/3065386.

[16] K. He, S. Ren, J. Sun, and X. Zhang. "Deep Residual Learning for Image Recognition". In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016), pp. 770–778. DOI: 10.1109/CVPR.2016.90.

[17] T. Mikolov, A. Deoras, D. Povey, L. Burget, and J. Černocký. "Strategies for training large scale neural network language models". In: *2011 IEEE Workshop on Automatic Speech Recognition Understanding.* 2011, pp. 196–201. DOI: 10.1109/ASRU.2011.6163930.

[18] G. Hinton et al. "Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups". In: *IEEE Signal Processing Magazine* 29.6 (2012), pp. 82–97. DOI: 10.1109/MSP.2012.2205597.

[19] M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer. "Deep contextualized word representations". In: (June 2018), pp. 2227–2237. DOI: 10.18653/v1/N18-1202. URL: https://aclanthology.org/N18-1202.

[20] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa. "Natural language processing (almost) from scratch". In: *Journal of machine learning research* 12.ARTICLE (2011), pp. 2493–2537. DOI: 10.5555/1953048.2078186.

[21] Q. Kong, D. T. Trugman, Z. E. Ross, M. J. Bianco, B. J. Meade, and P. Gerstoft. "Machine Learning in Seismology: Turning Data into Insights". In: *Seismological Research Letters* 90.1 (2018), pp. 3–14. URL: https://doi.org/10.1785/0220180259.

[22] W. Zhu and G. C. Beroza. "PhaseNet: a deep-neural-network-based seismic arrival-time picking method". In: *Geophysical Journal International* 216.1 (2019), pp. 261–273. URL: https://doi.org/10.1093/gji/ggy423.

[23] W.Y. Liao, E.J. Lee, D. Mu, P. Chen, and R.J. Rau. "ARRU Phase Picker: Attention Recurrent-Residual U-Net for Picking Seismic P- and S-Phase Arrivals". In: *Seismological Research Letters* 92.4 (Mar. 2021), pp. 2410–2428. ISSN: 0895-0695. DOI: 10.1785/0220200382. eprint: https://pubs.geoscienceworld.org/ssa/srl/article-pdf/92/4/2410/5351037/srl-2020382.1.pdf. URL: https://doi.org/10.1785/0220200382.

[24] W. Li, M. Chakraborty, D. Fenner, J. Faber, K. Zhou, G. Ruempker, H. Stoecker, and N. Srivastava. *EPick: Multi-Class Attention-based U-shaped Neural Network for Earthquake Detection and Seismic Phase Picking.* 2021. URL: https://arxiv.org/abs/2109.02567.

[25] J. Wang and T.L. Teng. "Artificial neural network-based seismic detector". In: *Bulletin of the Seismological Society of America* 85.1 (Feb. 1995), pp. 308–319. URL: https://doi.org/10.1785/BSSA0850010308.

[26] S.M. Mousavi, W.L. Ellsworth, W. Zhu, L.Y. Chuang, and G.C. Beroza. "Earthquake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking". In: *Nature Communications* 11.3952 (2020). URL: https://doi.org/10.1038/s41467-020-17591-w.

[27] M.A. Meier, Z. E. Ross, A. Ramachandran, A. Balakrishna, S. Nair, P. Kundzicz, Z. Li, J. Andrews, E. Hauksson, and Y. Yue. "Reliable Real-Time Seismic Signal/Noise Discrimination With Machine Learning". In: *Journal of Geophysical Research: Solid Earth* 124.1 (2019), pp. 788–800. DOI: https://doi.org/10.1029/2018JB016661. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018JB016661.

[28] T. Perol, M. Gharbi, and M. Denolle. "Convolutional neural network for earthquake detection and location". In: *Science Advances* 4.2 (2018), e1700578. DOI: 10.1126/sciadv.1700578.

[29] A. Panakkat and H. Adeli. "Recurrent neural network for approximate earthquake time and location prediction using multiple seismicity indicators". In: *Computer-Aided Civil and Infrastructure Engineering* 24.4 (2009), pp. 280–292. URL: https://doi.org/10.1111/j.1467-8667.2009.00595.x.

[30] H. S. Kuyuk and O. Susumu. "Real-time classification of earthquake using deep learning". In: *Proc. Comput. Sci* 140 (2018), pp. 298–305. URL: https://doi.org/10.1016/j.procs.2018.10.316.

[31] Z. E. Ross, M.A. Meier, and E. Hauksson. "P wave arrival picking and first-motion polarity determination with deep learning". In: *Journal of Geophysical Research: Solid Earth*

123.6 (2018), pp. 5120–5129. URL: https://doi.org/10.1029/2017JB015251.

[32] H. Ismail Fawaz, G. Forestier, J. Weber, L. Idoumghar, and P.A. Muller. "Deep learning for time series classification: a review." In: *Data Mining and Knowledge Discovery* 33 (2019), pp. 917–963. DOI: https://doi.org/10.1007/s10618-019-00619-1.

[33] M. Aly. "Survey on multiclass classification methods". In: *Neural Netw* 19 (2005), pp. 1–9.

[34] O. M. Saad, A. G. Hafez, and M. S. Soliman. "Deep Learning Approach for Earthquake Parameters Classification in Earthquake Early Warning System." In: *IEEE Geoscience and Remote Sensing Letters* (2020), pp. 1–5. DOI: 10.1109/LGRS.2020.2998580.

[35] S. M. Mousavi, Y. Sheng, W. Zhu, and G. C. Beroza. "STanford EArthquake Dataset (STEAD): A Global Data Set of Seismic Signals for AI." In: *IEEE Access* 7 (2019), pp. 179464–179476. URL: https://doi.org/10.1109/ACCESS.2019.2947848.

[36] J. Münchmeyer, D. Bindi, U. Leser, and F. Tilmann. "The transformer earthquake alerting model: a new versatile approach to earthquake early warning". In: *Geophysical Journal International* 225.1 (2020), pp. 646–656. DOI: 10.1093/gji/ggaa609. URL: https://doi.org/10.1093/gji/ggaa609.

[37] G. E. A. P. A. Batista, R. C. Prati, and M. C. Monard. "A Study of the Behavior of Several Methods for Balancing Machine Learning Training Data". In: *SIGKDD Explorations Newsletter* 6.1 (June 2004), pp. 20–29. ISSN: 1931-0145. DOI: 10.1145/1007730.1007735. URL: https://doi.org/10.1145/1007730.1007735.

[38] B. Krawczyk. "Learning from imbalanced data: open challenges and future directions". In: *Progress in Artificial Intelligence* 5.4 (2016), pp. 221–232. URL: https://doi.org/10.1007/s13748-016-0094-0.

[39] Q. Wen, L. Sun, F. Yang, X. Song, J. Gao, X. Wang, and H. Xu. "Time Series Data Augmentation for Deep Learning: A Survey". In: *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence* (Aug. 2021). DOI: 10.24963/ijcai.2021/631. URL: http://dx.doi.org/10.24963/ijcai.2021/631.

[40] A. Lomax, A. Michelini, and D. Jozinović. "An Investigation of Rapid Earthquake Characterization Using Single-Station Waveforms and a Convolutional Neural Network". In: *Seismological Research Letters* 90 (Feb. 2019), pp. 517–529. DOI: 10.1785/0220180311.

[41] M. Chakraborty, G. Rümpker, H. Stöcker, W. Li, J. Faber, D. Fenner, K. Zhou, and N. Srivastava. "Real Time Magnitude Classification of Earthquake Waveforms using Deep Learning". In: *EGU General Assembly 2021, online* EGU21-15941 (2021). URL: https://doi.org/10.5194/egusphere-egu21-15941.

[42] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, and D. J. Inman. "1D convolutional neural networks and applications: A survey". In: *Mechanical Systems and Signal Processing* 151 (2021), p. 107398. ISSN: 0888-3270. URL: https://doi.org/10.1016/j.ymssp.2020.107398.

[43] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. "Dropout: A Simple Way to Prevent Neural Networks from Overfitting". In: *Journal of Machine Learning Research* 15.56 (2014), pp. 1929–1958. URL: http://jmlr.org/papers/v15/srivastava14a.html.

[44] J. Nagi, F. Ducatelle, G. A. Di Caro, D. Cireşan, U. Meier, A. Giusti, F. Nagi, J. Schmidhuber, and L. M. Gambardella. "Max-pooling convolutional neural networks for vision-based hand gesture recognition". In: *2011 IEEE International Conference on Signal and Image Processing Applications (IC-*

SIPA). 2011, pp. 342–347. DOI: 10.1109/ICSIPA.2011.6144164.

[45] S. Hochreiter and J. Schmidhuber. "Long Short-Term Memory". In: *Neural Computation* 9.8 (Nov. 1997), pp. 1735–1780. ISSN: 0899-7667. DOI: 10.1162/neco.1997.9.8.1735. URL: https://doi.org/10.1162/neco.1997.9.8.1735.

[46] I. Goodfellow, Y. Bengio, and A. Courville. *Deep learning*. MIT press, 2016. URL: http://www.deeplearningbook.org.

[47] D. P. Kingma and J. Ba. *Adam: A Method for Stochastic Optimization*. Dec. 2014.

[48] K.P. Murphy. *Machine learning: a probabilistic perspective*. MIT press, 2012.

[49] L Prechelt. "Early Stopping — But When?" In: *Neural Networks: Tricks of the Trade: Second Edition*. Springer Berlin Heidelberg, 2012, pp. 53–67. ISBN: 978-3-642-35289-8. DOI: 10.1007/978-3-642-35289-8_5. URL: https://doi.org/10.1007/978-3-642-35289-8_5.

[50] Y. Bengio. "Practical Recommendations for Gradient-Based Training of Deep Architectures". In: *Neural Networks: Tricks of the Trade: Second Edition*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 437–478. ISBN: 978-3-642-35289-8. DOI: 10.1007/978-3-642-35289-8_26. URL: https://doi.org/10.1007/978-3-642-35289-8_26.

[51] P. Madhyastha and R. Jain. *On Model Stability as a Function of Random Seed*. 2019. arXiv: 1909.10447 [cs.LG].

[52] R. M. Allen and D. Melgar. "Earthquake Early Warning: Advances, Scientific Challenges, and Societal Needs". In: *Annual Review of Earth and Planetary Sciences* 47.1 (2019), pp. 361–388. DOI: 10.1146/annurev-earth-053018-060457.

[53] K. M. Ting. "Confusion Matrix". In: *Encyclopedia of Machine Learning and Data Mining*. Boston, MA: Springer US, 2017, pp. 260–260. ISBN: 978-1-4899-7687-1. DOI: 10.1007/978-1-4899-7687-1_50. URL: https://doi.org/10.1007/978-1-4899-7687-1_50.