Geophysical Journal International

Geophys. J. Int. (2024) **236,** 1728–1742
Advance Access publication 2024 January 11
GJI Seismology

https://doi.org/10.1093/gji/ggae018

Earthquake or blast? Classification of local-distance seismic events in Sweden using fully connected neural networks

Gunnar Eggertsson, Björn Lund, Michael Roth and Peter Schmidt

Department of Earth Sciences, Uppsala University, SE-75236, Uppsala, Sweden. E-mail: gunnar.eggertsson@geo.uu.se

Accepted 2024 January 5. Received 2023 December 7; in original form 2023 May 26

SUMMARY

Distinguishing between different types of seismic events is a task typically performed manually by expert analysts and can thus be both time and resource expensive. Analysts at the Swedish National Seismic Network (SNSN) use four different event types in the routine analysis: natural (tectonic) earthquakes, blasts (e.g. from mines, quarries and construction) and two different types of mining-induced events associated with large, underground mines. In order to aid manual event classification and to classify automatic event definitions, we have used fully connected neural networks to implement classification models which distinguish between the four event types. For each event, we bandpass filter the waveform data in 20 narrow-frequency bands before dividing each component into four non-overlapping time windows, corresponding to the *P* phase, *P* coda, *S* phase and *S* coda. In each window, we compute the root-mean-square amplitude and the resulting array of amplitudes is then used as the neural network inputs. We compare results achieved using a station-specific approach, where individual models are trained for each seismic station, to a regional approach where a single model is trained for the whole study area. An extension of the models, which distinguishes spurious phase associations from real seismic events in automatic event definitions, has also been implemented. When applying our models to evaluation data distinguishing between earthquakes and blasts, we achieve an accuracy of about 98 per cent for automatic events and 99 per cent for manually analysed events. In areas located close to large underground mines, where all four event types are observed, the corresponding accuracy is about 90 and 96 per cent, respectively. The accuracy when distinguishing spurious events from real seismic events is about 95 per cent. We find that the majority of erroneous classifications can be traced back to uncertainties in automatic phase picks and location estimates. The models are already in use at the SNSN, both for preliminary type predictions of automatic events and for reviewing manually analysed events.

Key words: Machine learning; Neural networks, fuzzy logic; Statistical methods; Time-series analysis; Seismicity and tectonics.

1 INTRODUCTION

Seismic waves produced by small earthquakes and blasts (e.g. mining or quarry blasts) can share similar characteristics. The task of distinguishing between different types of seismic events in regions where both earthquakes and blasts occur is important for a variety of reasons, for example, for understanding the drivers of natural seismicity and for seismic hazard assessment. Finding seismic discriminants capable of distinguishing between earthquakes and blasts has long been an active research topic. Many of the methods developed for that purpose have exploited the amplitude or spectral ratios of different seismic phases. Examples for seismic events at regional distances include the *P*/*S* ratio (Baumgardt & Young 1990), and the *P*/*Lg* ratio (Dysart & Pulli 1990; Baumgardt & Young 1990; Kim *et al.* 1993). At local distances, examples include the *P*/*S* ratio (O'Rourke *et al.* 2016) and the *Rg*/*Sg* ratio (Tibi *et al.* 2018; Kintner *et al.* 2020). Methods based on amplitude or spectral ratios have, in general, performed well on the task of seismic event classification.

In recent years, advances in machine learning have, in conjunction with a widespread availability of large data sets, been a catalyst for progress in method development for seismic event classification. Several recent studies have demonstrated the suitability of deep-learning based methods, in particular convolutional neural networks (CNNs), a subclass of artificial neural networks, for the task of seismic event classification (Linville *et al.* 2019; Tibi *et al.* 2019; Kong *et al.* 2021, 2022; Hourcade *et al.* 2023). These studies

use images of event spectrograms as inputs into CNNs which are capable of automatic feature extraction and expert-level classification of earthquakes and explosive sources.

Linville *et al.* (2019) applied a form of recurrent neural network; Long-Short-Term-Memory, and CNNs on images of event spectrograms to distinguish between explosive and tectonic sources at local distances. They found that both methods were able to correctly predict event labels in over 99 per cent of cases, once label errors assigned by analysts had been accounted for. They also reported classification accuracy of over 98 per cent for shallow earthquakes, indicating limited effects of depth on event classification. Hourcade *et al.* (2023) expanded on the work of Linville *et al.* (2019) by increasing the frequency content of the used spectrograms and thereby implementing a tool capable of classifying low-magnitude events. They also demonstrate the versatility of their approach by successfully applying it in different geographical areas. Kong *et al.* (2022) combined physics-based features with a deep-learning approach, operating on waveforms and spectrograms, to distinguish between earthquakes and explosions. Their selected physics-based features were high-frequency *P*/*S* amplitude ratios and the difference between local and coda duration magnitudes. They found that the combined method outperformed models developed exclusively with deep learning when applied to new regions, that is, the combination improved generalization performance.

Other machine-learning-based methods for classifying seismic events, using manually selected features, have also been successful. For example, Kortström *et al.* (2016) developed an algorithm to automatically distinguish between earthquakes and explosive sources within a regional seismic network located in Finland. Their method applied a supervised pattern recognition technique, support vector machine, on computed waveform features, filtered in different frequency bands. With evaluation data consisting of 5435 automatic event detections, their method correctly predicted the source type for 94 per cent of the events, with a 3 per cent risk of falsely identifying an event as an earthquake. Another example is the work of Miao *et al.* (2020), who use an artificial neural network to discriminate local and regional earthquakes from quarry and mining blasts in eastern Kentucky with high accuracy. After conducting tests with several types of input features, model architectures and hyperparameters, a trained model containing the authors' preferred combination of input features and model architecture achieves a high precision (>97 per cent) when identifying both earthquakes and blasts, using single-station recordings.

In Sweden, natural (tectonic) earthquakes, man-made blasts at construction sites, quarries and in mines, and mining-induced events frequently occur in close spatial proximity. Located in a stable continental region in northern Europe, the large majority of seismic events detected in Sweden are man-made, that is, either blasts or mining-induced events. The natural earthquakes detected in Sweden by the Swedish National Seismic Network (SNSN) are mostly low magnitude. The network detects earthquakes down to local magnitudes below −1 with an estimated magnitude of completeness of about 0.5 and an average annual detection rate of about seventeen earthquakes with local magnitude 2 or higher (Lund *et al.* 2021). The task of distinguishing between different event types is typically performed by expert analysts and can thus be both time and resource expensive. At SNSN, analysts use four different event types in the routine analysis; earthquakes, blasts and two different types of mining-induced events associated with large, underground mines. The main objective of this study is to develop an algorithm capable of reliable automatic event classification for the four event event types detected in Sweden. Inspired by the study of Kortström *et al.* (2016) for a seismic network located in geological settings comparable to Sweden, we have opted to implement a classification model based on similar, manually selected features. The classification is implemented as traditional, feed-forward, fully connected neural networks; a machine-learning-based approach widely used for classification tasks. As a first approach, we implement the classification on a station-specific basis, training separate models for each seismic station used in the study, with data recorded at the station. Subsequently, we train a larger, regional model with data from all stations and compare its performance to the station-specific approach.

In Section 2, we describe our method, starting with the neural networks' architecture, followed by the data processing steps and a description of how station-specific predictions are combined to form a final prediction about an event's class. In Section 3, we introduce the SNSN, the data used in this study and the seismicity of Sweden. In Section 4, we present our results, both individually on a station basis and on a multistation basis for an evaluation data set, spanning six months of seismic events. We also present the results of different applications of our method to seismic data. Finally, in Section 5, we discuss the general implications of our results, compare with other similar studies and discuss the impact of various factors on our results.

2 METHOD

We use fully connected artificial neural network models to classify the source types of seismic events in Sweden. The seismic stations operated by the SNSN are spread over a large geographical area, with notable differences in the types of seismic events and ambient noise levels recorded at the different stations. For this reason, as a first approach, we develop station-specific models for each SNSN station, using data recorded at the stations.

The neural network models consist of three hidden and two dropout layers to reduce the risk of overfitting (Srivastava *et al.* 2014). Each hidden layer consists of 256 nodes and a rectified linear unit activation function. The dropout layers have a dropout rate of 0.5. The number and size of the hidden layers along with other neural network hyperparameters were decided through a grid-search-based evaluation process where different combinations of parameters were ranked in terms of their effect on the models' performance. For the model training, the label, that is, source type of the event, has been manually assigned. On most stations, the classification problem is treated as binary, that is, all events are assumed to be either a natural (tectonic) earthquake or a blast. Here, the neural network output consists of one node with a sigmoid activation function. The node contains a value between 0 and 1 which represents the probability of the input event being an earthquake. On six selected stations, the presence of mining-induced events increases the number of possible classes from two to four. Here, the output layer contains a number of nodes equal to the number of possible classes, that is, four, with a softmax activation function. In this case, each node will have a value between 0 and 1 with the individual values summing to 1. Each node value then represents the probability of the input event belonging to the corresponding class (Goodfellow *et al.* 2016). All the analysis is implemented in the Python programming language. Data processing is done using NumPy (Harris *et al.* 2020) and pandas (McKinney 2010). The models were constructed using the Keras deep learning API (Chollet *et al.* 2015), running on top of the machine learning platform TensorFlow (Abadi *et al.* 2015). Evaluation metrics for model performance are computed

using sklearn (Pedregosa *et al.* 2011). Figures were generated using Matplotlib (Hunter 2007) and The Generic Mapping Tools (Wessel *et al.* 2019).

For each station used in the study, we start by compiling a list of manually revised seismic events with both *P*- and *S*-phase arrivals picked at the station, to use for model training. For each event we correct the associated waveforms for instrument response and rotate the horizontal components into radial and transverse coordinates. Each component is subsequently processed in the same way as suggested by Kortström *et al.* (2016). First, it is bandpass filtered in a narrow-frequency band. Then, it is divided into four non-overlapping time windows corresponding to *P* phase, *P* coda, *S* phase and *S* coda, using the manually picked *P*- and *S*-phase arrivals (see Fig. 1). All time windows have equal length, corresponding to half the differential time between the *P*- and *S*-phase arrivals. Finally, the root-mean-square (RMS) amplitude is computed in each time window. This process is repeated for a total of 20 narrow-frequency bands, resulting in a total of $3 \times 4 \times 20 = 240$ RMS amplitudes which represent our neural network inputs. The passbands used are the ones suggested by Kortström *et al.* (2016), ranging from 1 to 41 Hz with a bandwidth of 3 and 1 Hz overlap between adjacent bands, that is, 1–3, 2–5, 4–7, . . . , 36–39, 38–41 Hz.

Fig. 1 shows example waveform segments where the vertical components of seismic event records have been divided into the prescribed time windows. The upper panel is from an earthquake which occurred at a depth of approximately 20 km, while the lower panel is from a near-surface quarry blast. *Rg* waves are often prominent features on seismograms of shallow events of all types, including both blasts and earthquakes, at the local distances used in this study (Båth 1975). The lower panel of Fig. 1 shows how the *Rg* wave generated by the local-distance near-surface event is most prominent within the time window we associate with *S* coda. *Lg* waves are often prominent features in high-frequency seismograms at regional distances beyond 150 km (Furumura *et al.* 2014). At the local distances used in this study, the relative amplitude and the time separation between the *Lg* and *S* phases are typically too small to reliably distinguish the *Lg* phase which effectively merges into the *S* phase (Press & Ewing 1952).

For the model training, available data are randomly split with a 50, 25 and 25 per cent ratio into training, validation and test data, respectively. On most stations the number of blasts is much larger than the number of earthquakes in the data, which leads to class imbalance. To limit prediction bias caused by the class imbalance, we apply the Synthetic Minority Over-Sampling Technique (Chawla *et al.* 2002) on stations where the number of earthquakes in the data is less than 80 per cent of the number of blasts. We use an oversampling ratio of 0.8, that is, we supplement the earthquake class with synthetic examples until the ratio between earthquakes and blasts reaches 0.8. The oversampling ratio was determined through an evaluation process where different ratios were applied to model training on all stations. The selected ratio of 0.8 is the one which resulted in the highest average validation set accuracy over all stations. When fitting the model we use a batch size of 64. This means that 64 events from the training data are passed through the network on each iteration, before model parameters are updated. When all training data have passed once through the network, it is termed an epoch. We continue training for a maximum of 500 epochs, that is, the network will work through the entire training data no more than 500 times. At the end of each epoch the model's classification accuracy on the validation set is estimated. We stop training once the validation accuracy has not improved over 20 epochs. We select the final model parameters as the ones corresponding to the epoch where the highest validation accuracy was achieved. Depending on the size of the training data, our average training time per epoch ranges from 0.2 to 0.9 s. This is when using an 11th Gen Intel(R) Core(TM) i7-11700 @ 2.50GHz with 8 cores and 32 GB of RAM. For an average-sized training set containing about 3000 events, the average training time per epoch is approximately 0.6 s.

To make a final prediction about the class of a new seismic event, we first generate individual class predictions from all stations located within a 10–200 km distance of the event's epicentre that have at least one phase pick available from either manual analysis or automatic phase association. The 10 km distance limit ensures that the time windows are long enough to contain meaningful information while the 200 km limit reflects the typical range of manual phase picks available for the data. For missing *P*- or *S*-phase picks we compute synthetic arrival times based on the local velocity model of the SNSN, the same velocity model as used for manual analysis and automatic event detections. Given that we have predictions from at least two stations, we compute a final probability distribution from the individual station predictions according to eq. (1), as the the arithmetic means of the individual station distributions.

$$\begin{aligned}\mathbf{P} &= [P(c_1), P(c_2), P(c_3), P(c_4)] \\ &= \frac{1}{n}\left[\sum_{j=1}^{n} P(c_{1,j}), \sum_{j=1}^{n} P(c_{2,j}), \sum_{j=1}^{n} P(c_{3,j}), \sum_{j=1}^{n} P(c_{4,j})\right], \\ n &\geq 2 \end{aligned} \tag{1}$$

where $P(c_i)$ is the final probability of an event belonging to class i, n is the total number of stations which generated individual distributions and $P(c_{i,j})$ is the probability of an event belonging to class i, computed at station j. On stations where the classification problem is binary, $P(c_{3,j}) = P(c_{4,j}) = 0$. The predicted class is taken be the one corresponding to the highest probability in the final distribution. The prediction is assigned an integer quality factor (QF) between 0 and 99 which represents the highest probability in the final distribution and the number of stations used to generate the prediction. It is designed to rank prediction reliability, as we associate higher station count with higher reliability. The QF is computed according to eq. (2).

$$\mathrm{QF} = \mathrm{floor}\left(\left(\max(\mathbf{P}) - \frac{1}{n^2}\right) \times 100\right). \tag{2}$$

We evaluate the performance of our models using several metrics. For each class, we adopt the following notation: true positives (TP): number of events correctly predicted to belong to the class. False positives (FP): number of events erroneously predicted to belong to the class. True negatives (TN): number of events correctly predicted to belong to a different class. False negatives (FN): number of events erroneously predicted to belong to a different class. We define the evaluation metrics in a traditional way (Burkov 2019). Accuracy represents the ratio between the number of correct event predictions and the total number of event predictions for all the data; accuracy = (TP+TN)/(TP+FP+TN+FN). It provides an overall summary of the model's performance across all classes. For a given class, precision represents how many, out of all events predicted to belong to the class, are correctly predicted; precision = TP/(TP+FP). Recall represents how many, out of all events which truly belong to the class, are correctly predicted; recall = TP/(TP+FN). The F1-score is defined as a harmonic mean of precision and recall

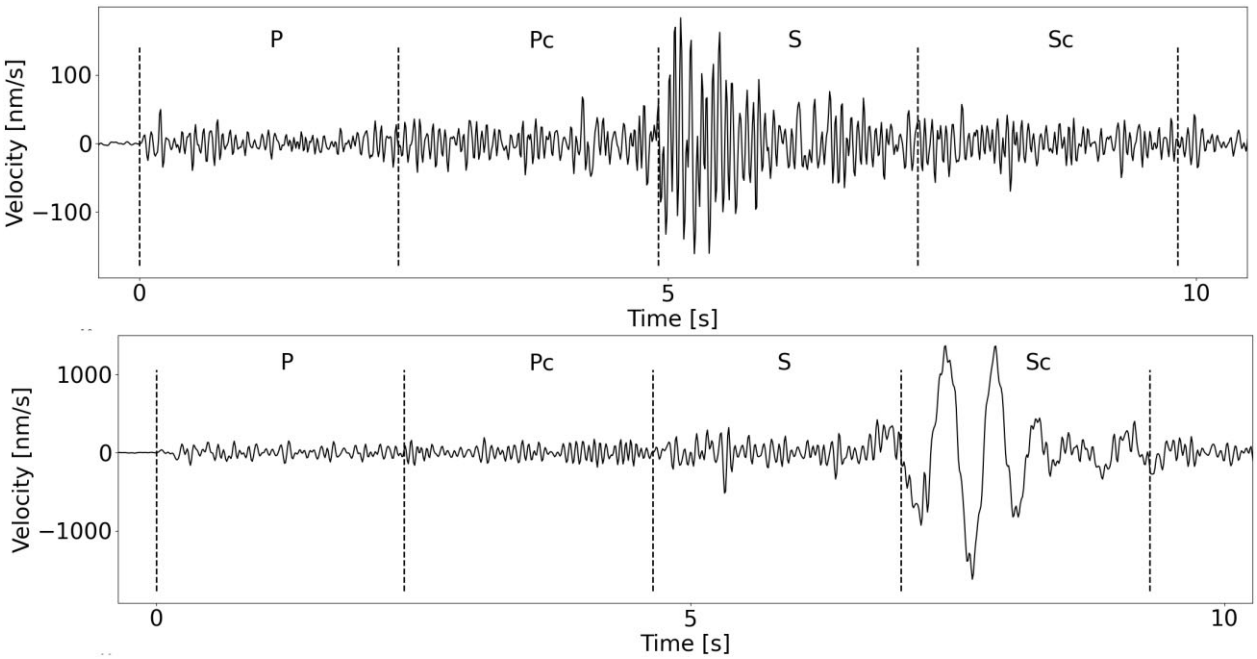

Figure 1. Vertical component of seismic records for two seismic events in Sweden. The letters represent the four non-overlapping time windows. *P*: *P* phase, *Pc*: *P* coda, *S*: *S* phase and *Sc*: *S* coda. The upper record is from an earthquake with local magnitude 0.7, recorded at an epicentral distance of 36 km. The lower record is from a quarry blast with local magnitude 1.5, recorded at an epicentral distance of 38 km. It contains a (relatively) long-period *Rg* wave starting in the *S*-phase window and most prominent in the *S*-coda window.

(Sasaki *et al.* 2007):

$$F_1 = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}. \quad (3)$$

The F1-score can take values ranging from 0 to 1. An F1-score of 0 for a given class indicates that either precision or recall for the class is equal to 0, whereas an F1-score of 1 indicates both precision and recall are equal to a maximum value of 1 (Sasaki *et al.* 2007). We use the F1-score as our primary evaluation metric for the station-specific models because it provides a balanced measure of the models' performance by considering both precision and recall for each class.

In addition to classifying real seismic events, we also train another set of station-specific classification models, with the aim of distinguishing real seismic events from spurious phase associations. The network architecture of these event-or-not models is identical to the event classification models but the length and positioning of the time windows is slightly adjusted. We extend and shift the time windows to include a noise window before the first *P*-phase arrival to include an estimate of signal-to-noise ratio, one *P*-phase window, one *S*-phase window and one post-*S*-coda window. The length of each window equals the differential time between the *P*- and *S*-phase arrivals. The training data for the event-or-not models contains a mixture of real seismic events and spurious phase associations with a label indicating whether or not an event has been deemed to be a real seismic event during manual analysis. The classification problem is thus formulated as a binary problem; for a given automatic event detection, is it a real seismic event (irrespective of event class) or spurious phase associations? The models are applied in an identical way to the event classification models, with an arithmetic mean of station-specific predictions used to generate the final prediction.

3 DATA

The SNSN (Lund *et al.* 2021) is the primary monitoring institution for seismic activity in Sweden. The network currently comprises 67 permanent seismic stations, each equipped with a broadband instrument providing continuous real-time data, with a sampling rate of 100 samples per second. Each station is connected via cellular network to a central computer in Uppsala (Lund *et al.* 2021). Fig. 2 shows the locations of SNSN stations in operation in 2022 August.

Currently, all major seismic events occurring in Sweden are subject to routine manual analysis at SNSN. The analysis includes tasks such as phase picking, event location and classification, magnitude estimation and focal mechanism generation. Historically, analysts at SNSN have have distinguished between three different source types for seismic events occurring in Sweden. These are natural earthquakes, man-made blasts, typically related to industry, and mining-induced events. Since 2020, the class of mining-induced events has been further sub-divided into two separate classes, owing to a difference in source mechanics and frequency content. Fig. 3 shows the revised locations of all blasts (a) and earthquakes (b) in Sweden and neighbouring countries registered and classified by SNSN between the years 2010 and 2021. Events outside of Sweden are generally not classified as earthquakes or blasts unless in collaboration with neighbouring seismologists (Lund *et al.* 2021). The figure shows that blast locations are widely distributed over the country. The two biggest sources of blasts are the large underground iron ore mines in Kiruna and Malmberget in northern Sweden (see Fig. 3a). In the southern part of the country, most blasting events originate from construction work and quarry blasting (Lund *et al.* 2021). The earthquake locations are mostly concentrated around postglacial faults, major continuous fault scarps in northern Sweden where fault movement took place at or near the time when the ice receded at the end of the latest glaciation, ca 9500 yr ago (Juhlin & Lund 2011; Lindblom *et al.* 2015), and around lake Vänern,

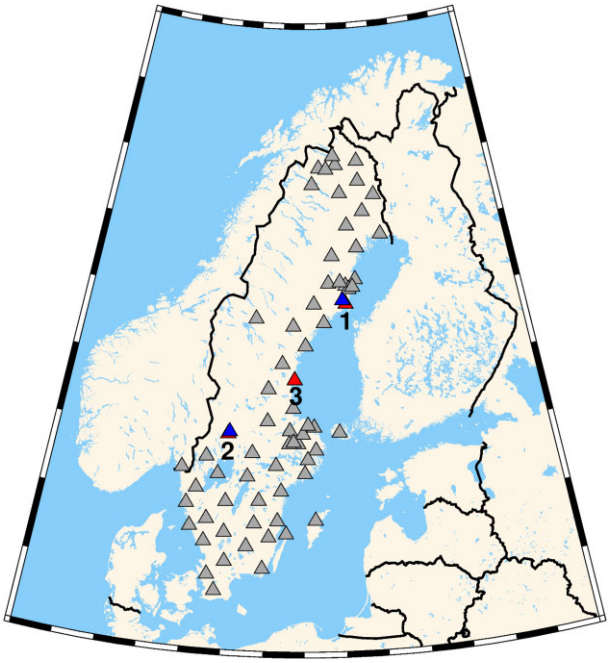

Figure 2. SNSN station locations, 2022 August, shown in triangles. (1) stations UMAU and BLMU (closely located), (2) stations UDD and HFB1 (closely located) and (3) station ARNU.

the large lake in southwestern Sweden. The two most seismically active postglacial faults in Sweden currently are the Burträsk and Pärvie faults (see Fig. 3(b). Mining-induced events are exclusively observed in the vicinity of large underground mines, most notably Kiruna and Malmberget.

The primary system used for automatic detection and event definition at SNSN is the South Iceland Lowland (SIL) system (Böðvarsson *et al.* 1999; Lund *et al.* 2021). On average, the system defines 100 to 200 automatic events per day, whereof about 30–40 are typically confirmed to be real earthquakes, blasts or mining-induced events via routine manual analysis (Lund *et al.* 2021). SNSN has collected and manually revised digital seismic data continuously since 2000 August. Between the years 2000 and 2020, approximately 171 000 analyst-reviewed seismic events were recorded by SNSN, whereof about 11 000 have been classified as earthquakes and about 160 000 as man-made (blasts or mining-induced, Lund *et al.* 2021). In some areas the network detects earthquakes to below local magnitude −1 but its magnitude of completeness is approximately 0.5. The detected earthquakes are generally low magnitude, with an average annual rate of 17 events with local magnitude M 2+ (Lund *et al.* 2021). This extensive availability of labelled training data makes SNSN well positioned to take advantage of supervised learning based techniques to automate parts of its routine analysis.

The data used for model training in this study are waveform data from manually analysed seismic events, recorded in the years 2010–2021, by all permanent seismic stations in operation by SNSN at the start of 2022 (SNSN 1904). Separate models were developed for each SNSN station, using event data recorded at the station. The event classification problem is treated as binary on most stations, with two possible classes of events observed; earthquakes and blasts. On stations within a 70 km distance radius from one of the mines in Kiruna and Malmberget, the number of possible classes increases to four due to the presence of two types of mining-induced events in addition to earthquakes and blasts. Fig. 4 shows the number of events in the data set used for each station, before oversampling and capping. An event is included in a station's data set if it has been located within a 10–200 km distance from the station and both *P*- and *S*-phase picks are available at the station from manual analysis. For computational efficiency reasons, the number of blasts and mining-induced events from the Kiruna and Malmberget mines in the relevant stations' data set are capped at 1000 events per class from each mine. The figure shows that the data for all stations south of station ARNU (marked with a dashed line, see also Fig. 2) contain significant class imbalance where the ratio between number of blasts and number of earthquakes in the data is high. This is consistent with Fig. 3 which show that the majority of earthquakes recorded in Sweden by SNSN occur north of station ARNU.

The majority of automatic event detections by the SIL system at SNSN comes from spurious phase associations (Lund *et al.* 2021). The spurious phase associations are typically distinguished from real earthquakes, blasts and mining-induced events via routine manual analysis. It is therefore of practical value to SNSN for the automatic event classification to also have the capability of distinguishing spurious phase associations from real seismic events. For this event-or-not classification, the data used for model training consist of spurious phase associations from the SIL system in conjunction with real, manually analysed seismic events, recorded in the years 2010–2021. Automatic phase picks from the SIL system are used for time window selection for the spurious phase associations and manual phase picks are used for the real seismic events. An event is included in a station's data set if it has been located (automatically or manually) within a 10–200 km distance from the station and both *P*- and *S*-phase picks are available at the station from automatic phase association (spurious phase associations) or manual analysis (real seismic events). Each station's event-or-not data set contains two event classes; spurious phase associations and real seismic events. Due to the large numbers of spurious phase associations available for some SNSN stations, the size of the event-or-not data sets is capped at 1000 events from each class and thus each station's event-or-not data set contains at most 2000 events.

4 RESULTS

This section presents the results from the method described in Section 2 and different applications on seismic data.

4.1 Station-specific models

Fig. 5 shows the F1-score achieved on the test data (25 per cent of each station's data set) for each individual station. The displayed score is computed as a macro average, that is, an unweighted mean of the F1-score for each individual class. The figure shows that all stations within the network achieve a good general performance with an F1-score of over 0.95 on all stations and a mean score of approximately 0.98 over the whole network. For stations trained to distinguish between four different classes the mean score is approximately 0.96 but for stations with only two classes it is 0.99. The high F1-score achieved on all stations indicates that the station models' predicted classes agree with analyst assigned classes on the large majority of events from all classes. The figure also shows that most of the southern stations maintain a relatively stable F1-score, whereas some of the northern stations display a quasi-linear trend of increasing F1-score with decreased latitude. In addition to some

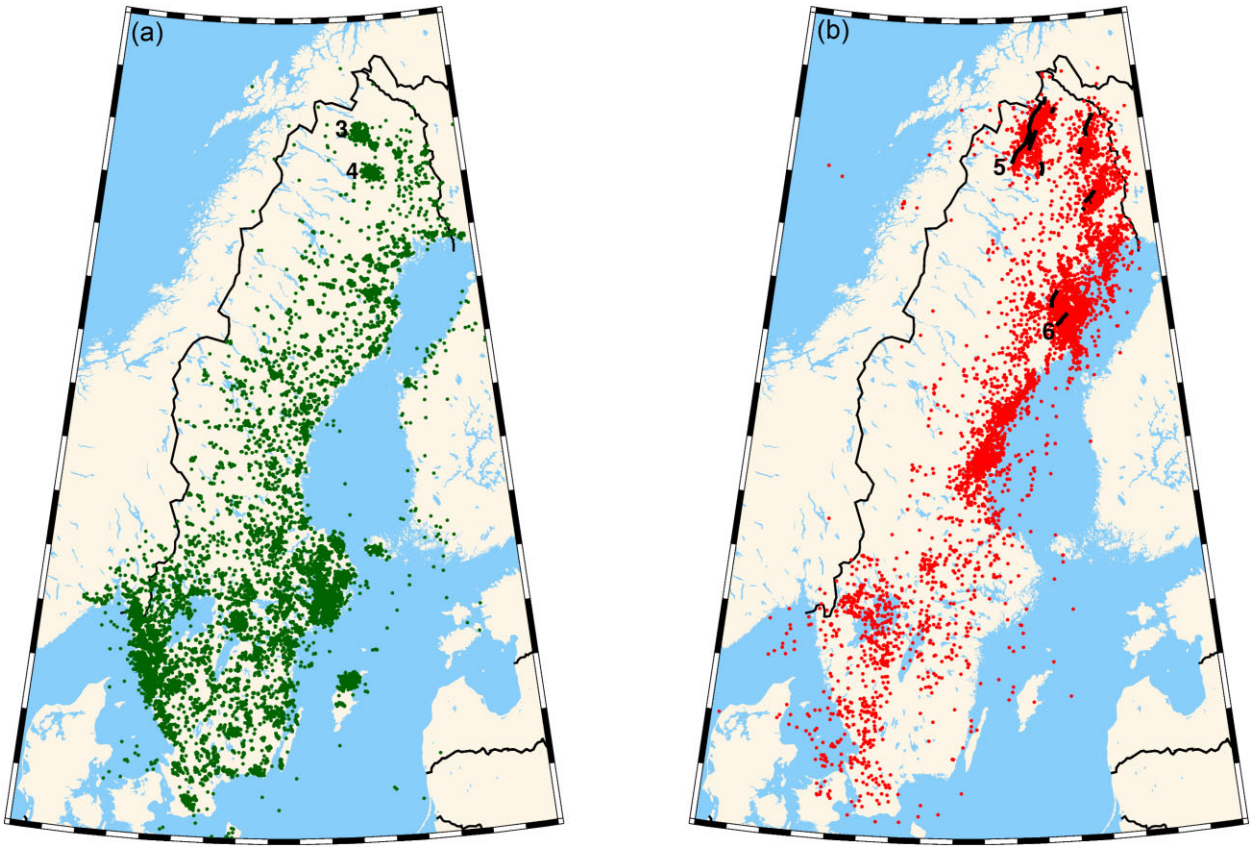

Figure 3. Locations of blasts (a) and earthquakes (b) recorded by SNSN in Sweden and neighbouring countries, 2010–2021. Mines: (3) Kiruna, (4) Malmberget. Postglacial faults: (5) Pärvie and (6) Burträsk.

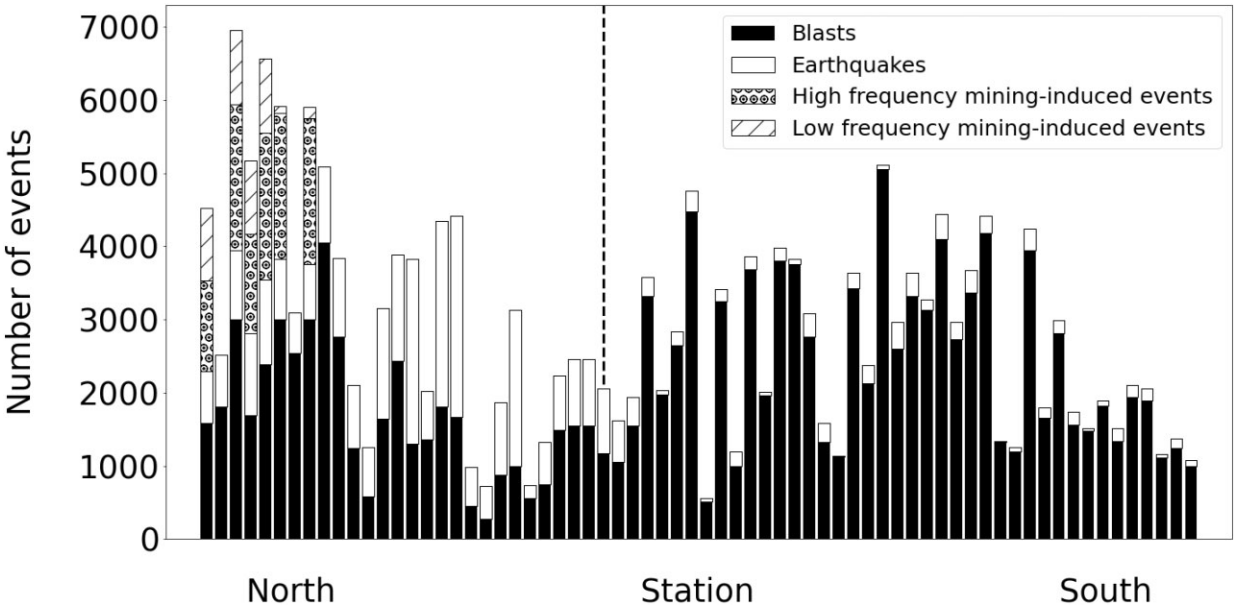

Figure 4. Number of events from each class used for model training. Each bar represents the class distribution on a single station. The dashed vertical line represents station ARNU, south of which significant class imbalance is observed on all stations. Station ordering is from north on the left to south on the right.

northern stations being trained to distinguish between four event classes instead of two, we also observe positive correlation between the stations' F1-score and the ratio between number of blasts to earthquakes in their training data. This suggests that stations with more class imbalance perform slightly better overall than stations where the class distribution is more balanced. We believe this to be a result of stations with significant class imbalance having higher numbers of synthetic, statistically similar earthquakes in their test

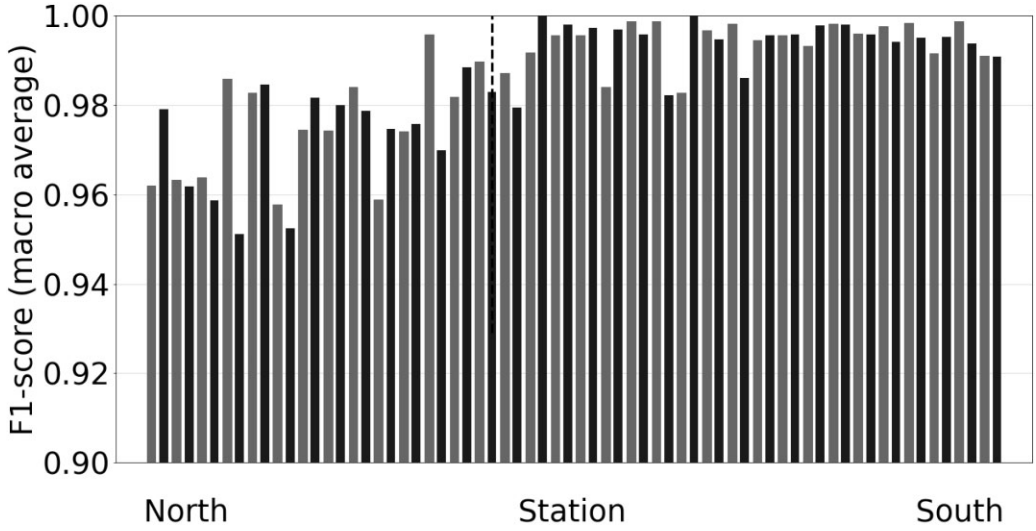

Figure 5. F1-score at individual stations. Each bar represents the F1-score computed at a single station. The score is a macro average, that is, an unweighted mean of the calculated F1-score for each individual class. The dashed vertical line represents station ARNU, south of which significant class imbalance is observed on all stations. Station ordering is from north on the left to south on the right. Note that the vertical axis starts at 0.9.

data, and thus less feature variability, resulting in better classification performance.

4.2 Multistation predictions

To evaluate the generalization performance of our multistation classification models we applied them to all seismic events that occurred in Sweden and were analysed by the SNSN, from 2022 January to June. For each event we used the relevant station-specific models to generate class probability distributions, computed a final prediction for the event and compared the result with the analyst assigned class. The total number of events was 6220. From these, 4933 events were detected automatically by the SIL system, for which automatic phase picks could be used for time window selection. This allowed for evaluating predictions both of automatic and manually analysed events.

Table 1 (left) shows the results for all manually analysed events located at least 25 km away from the mines in Kiruna and Malmberget; a total of 2707 events. In general, the classification problem here is binary, with only two classes of events observed; blasts and earthquakes. Occasional exceptions occur at other underground mines, as demonstrated by the four mining-induced events listed in the table, but their numbers were not sufficient to include in the training of the station-specific models. Overall, the predicted class agrees with the analyst assigned class for 2690 events, resulting in a classification accuracy of 99.4 per cent. The F1-score exceeds 0.98 for both blasts and earthquakes. The seven earthquakes erroneously predicted to be blasts have local magnitudes ranging from 0.1 to 0.7, with an average of 0.4.

Table 1 (right) shows the results for events located less than 25 km away from either Kiruna or Malmberget mine; a total of 3513 events. This is where the two classes of mining-induced events are primarily observed and thus instances of all four event classes are observed, although only one natural earthquake was detected during the evaluation period. Most erroneous classifications involve blasts predicted as high-frequency mining-induced events and vice-versa. Overall, the predicted class agrees with the analyst assigned class for 3374 events, resulting in a classification accuracy of 96.0 per cent. With the exception of earthquake class, the F1-score exceeds 0.95 for each individual event class.

Table 2 shows the corresponding results for the automatic SIL detections. Of the 1847 events located at least 25 km away from the mines in Kiruna and Malmberget (left), the predicted class agrees with the analyst assigned class on 1811, resulting in a classification accuracy of 98.1 per cent. Both the blast and earthquake classes receive an F1-score above 0.95. Of the 3086 events located less than 25 km away from the mines in Kiruna and Malmberget (right), the predicted class agrees with the analyst assigned class on 2770, resulting in a classification accuracy of 89.8 per cent. The F1-score for each individual class exceeds 0.90 with the exception of the earthquake class which has a relatively high number of FPs and only one TP, resulting in low precision and, consequently, low F1-score. Again, the majority of erroneous classifications are associated with the high-frequency mining-induced event class. Fig. 6 summarizes the classification accuracy achieved for the automatic and manually analysed events.

4.3 Quality factor threshold

One way of taking the final predictions' QF (eq. 2) into consideration for the event classification is to define a threshold value and prioritize predictions with QF above the threshold. Fig. 7 shows how the classification accuracy and event retention, defined as the proportion of events which prediction QF exceeds a given threshold value, change as a function of QF threshold when using manual phase picks (dashed curves) and automatic SIL phase picks (solid curves) for time window selection. For a given QF threshold, a point on the dark curves represents the classification accuracy for the subset of events which prediction quality is at least as high as the threshold. A point on the lighter curves represents the size of the subset, relative to the total number of events. We find that a QF threshold of about 70 provides a decent trade off between accuracy and event retention, with classification accuracy of about 99 per cent for 93 per cent of the events when using manual phase picks and about 98 per cent for 80 per cent of the events when using automatic SIL phase picks.

Table 1. Generalization performance of the event classification for manually analysed events at SNSN between 2022 January and June. Left: events located at least 25 km away from the mines in Kiruna and Malmberget. Right: events located less than 25 km away from the mines in Kiruna and Malmberget.

	Predicted class						Predicted class				
Analyst	B	E	H	L	**Re**	**Analyst**	B	E	H	L	**Re**
B	2297	6	0	0	0.997	B	1437	0	47	0	0.968
E	7	393	0	0	0.983	E	0	1	0	0	1
H	4	0	0	0	0	H	77	1	1584	8	0.949
L	0	0	0	0	–	L	0	0	7	351	0.980
Pr	0.995	0.985	–	–		**Pr**	0.949	0.5	0.967	0.978	
F1	0.996	0.984	–	–		**F1**	0.959	0.667	0.958	0.979	

Notes. Event classes, B: blasts, E: earthquakes, H: high-frequency mining-induced events and L: low-frequency mining-induced events. Metrics, Re: recall, Pr: precision and F1: F1-score.

Table 2. Generalization performance of the event classification for automatic event detections by the SIL system at SNSN between 2022 January and June. Left: events located at least 25 km away from the mines in Kiruna and Malmberget. Right: events located less than 25 km away from the mines in Kiruna and Malmberget.

	Predicted class						Predicted class				
Analyst	B	E	H	L	**Re**	**Analyst**	B	E	H	L	**Re**
B	1482	22	0	0	0.985	B	1087	14	49	0	0.945
E	10	329	0	0	0.971	E	0	1	0	0	1
H	4	0	0	0	0	H	132	80	1396	22	0.856
L	0	0	0	0	–	L	3	2	14	286	0.938
Pr	0.991	0.937	–	–		**Pr**	0.890	0.010	0.957	0.929	
F1	0.988	0.954	–	–		**F1**	0.917	0.020	0.904	0.933	

Notes. Event classes, B: blasts, E: earthquakes, H: high-frequency mining-induced events, L: low-frequency mining-induced events. Metrics, Re: recall, Pr: precision and F1: F1-score.

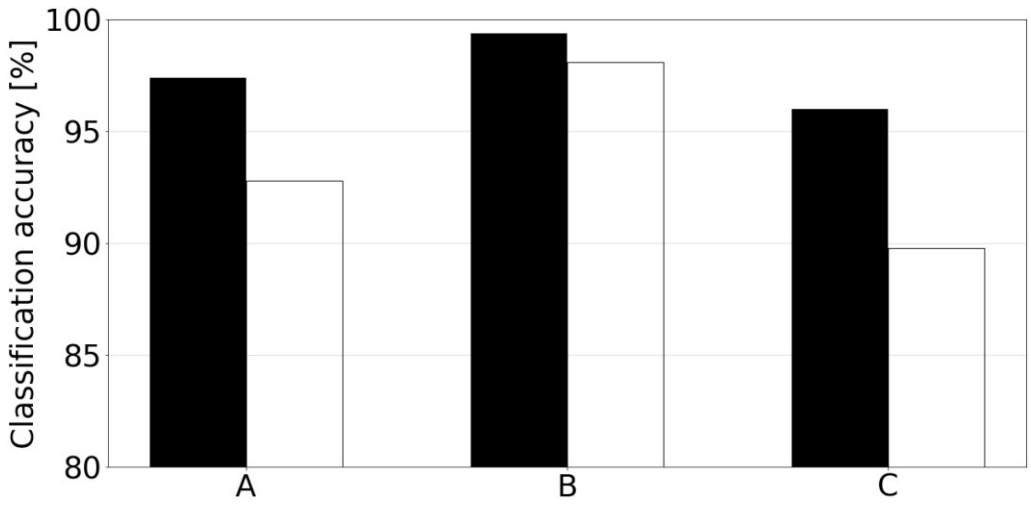

Figure 6. Classification accuracy for seismic events in Sweden analysed by SNSN, 2022 January to June. The vertical axis depicts accuracy; the proportion of events where the predicted class agrees with analyst assigned class. Filled bars indicate events where manually assigned phase picks were used to divide the seismic records into the prescribed time windows. Unfilled bars indicate events where automatic phase picks from the SIL system were used. Event categories, A: all events, B: events located at least 25 km away from the mines in Kiruna and Malmberget and C: events located less than 25 km away from the mines in Kiruna and Malmberget.

4.4 Shallow earthquakes

Shallow earthquakes, here defined as earthquakes occurring at depths above 5 km, represent a case study for our classification model, due to the potential for generating significant surface wave energy. For the seismic events recorded by SNSN, significant surface wave energy is more typically observed from blasting events close to the surface than earthquakes. Reliable depth estimates for shallow earthquakes are also typically challenging to establish at regional networks, due to the lack of station coverage sufficiently close to the event. To evaluate the model performance when predicting the event class of shallow earthquakes we compiled evaluation data consisting of 20 such events, with origin times outside the period used for model training, generated class predictions and compared with analyst assigned classes. We found that 19 out of the 20 earthquakes generated predictions matching manually assigned classes, indicating that the method is relatively robust to predicting the class of shallow earthquakes and that depth-sensitive effects alone do not play a decisive role in the classification.

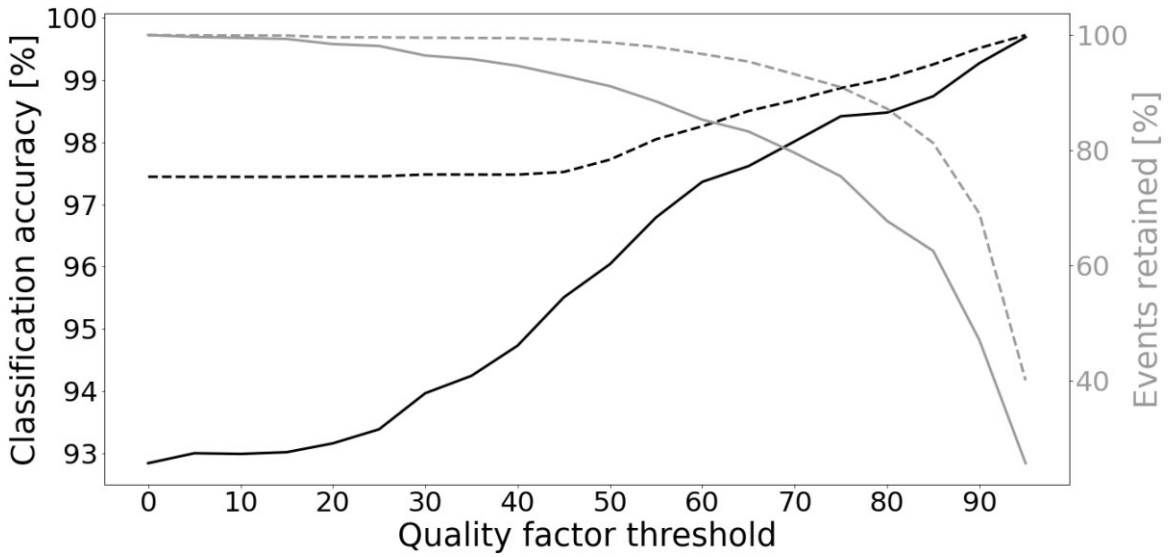

Figure 7. Effect of different QF threshold values on classification accuracy and event retention for all events during the evaluation period. Dashed curves represent events where manual phase picks are used for time window selection and solid curves represent events where automatic SIL picks are used. Dark curves represent classification accuracy and lighter curves represent event retention, that is, the proportion of events with final prediction QF above the corresponding threshold value.

4.5 Feature importance

The neural networks used in this study contain 240 manually selected input features. The features represent RMS amplitudes computed in different time windows, filtered in different frequency bands. One way of ranking individual features in terms of their relative importance to the model predictions is through a method referred to as permutation importance, first described by Breiman (2001) for random forests. In essence, the method estimates how a model's performance score, for example, classification accuracy, is affected when a single feature is randomly shuffled. We implemented permutation importance on all the station-specific models and computed the importance of each feature as a mean decrease of validation-set accuracy when the feature was shuffled. Fig. 8 shows the mean and standard deviation of the feature importance for each model input feature, over all SNSN stations.

We find that, in general, features corresponding to *P*-phase and *P*-coda time windows at intermediate to high frequencies (12–41 Hz) are of high importance. For most frequency bands, their importance significantly outweighs the importance of the *S*-phase and *S*-coda windows in the same bands. The reason for the high relative importance of these features is unclear to us and may be of interest to further investigate. Towards the lower end of the spectrum, three features stand out in terms of importance. Two of these features correspond to the *S*-coda time windows on the vertical and radial components in the lowest passband, corresponding to 1–3 Hz. The average importance of these features is among the highest of all the features. We interpret the importance of these features as an indication that the presence of *Rg* surface waves plays an important role in the classification on many stations. The models' apparent robustness when predicting shallow earthquakes does, however, suggest that the presence of significant *Rg* wave energy, by itself, is not decisive to the final classification. The third important feature on the lower end of the spectrum corresponds to the *S*-phase time window on the transverse component, also in the lowest passband. A possible explanation for the importance of this feature is the presence of *Lg* surface waves. It should be noted that, overall, the standard deviations of the computed feature importances are high compared to their mean values, as displayed in Fig. 8. In other words, the importance of each individual feature varies considerably from one station model to another.

4.6 Weighting of station predictions

During method development we tested different weighting schemes, where individual station predictions were weighted based on station-event distance before computing the final class prediction for an event. Fig. 9 shows the mean classification accuracy as a function of station-event distance in left-adjusted bins of 10 km width, computed over all stations used in the study for events in the evaluation data, 2022 January to June. The figure shows that beyond epicentral distances of 120 km, the mean accuracy starts decreasing after remaining relatively stable for shorter distances. We believe that this result can partly be attributed to more pronounced path effects at greater station-event distances, making the different classes more challenging to distinguish.

With Fig. 9 in mind, we tested weighting schemes where station predictions at station-event distances beyond 120 km were downweighted before computing final predictions. For example, we tested a weighting scheme where all predictions up to 120 km station-event distance were equally weighted and then downweighted predictions linearly with distance beyond 120 km. In general, we found the difference in classification accuracy to be minimal with and without weighting. While applying the different weighting schemes did increase the average QF of the final predictions, it did not increase the overall number of predictions matching analyst assigned classes. Conversely, we found that for three events in the evaluation data, assigning the weights led to final predictions which disagreed with manually assigned classes where they previously had agreed, prior to assigning the weights. The common factor shared by these 'flipped' events was that they only had one or two stations within a 120 km station-event distance. When one of these nearby stations produced a false class prediction, the downweighing of more distant stations

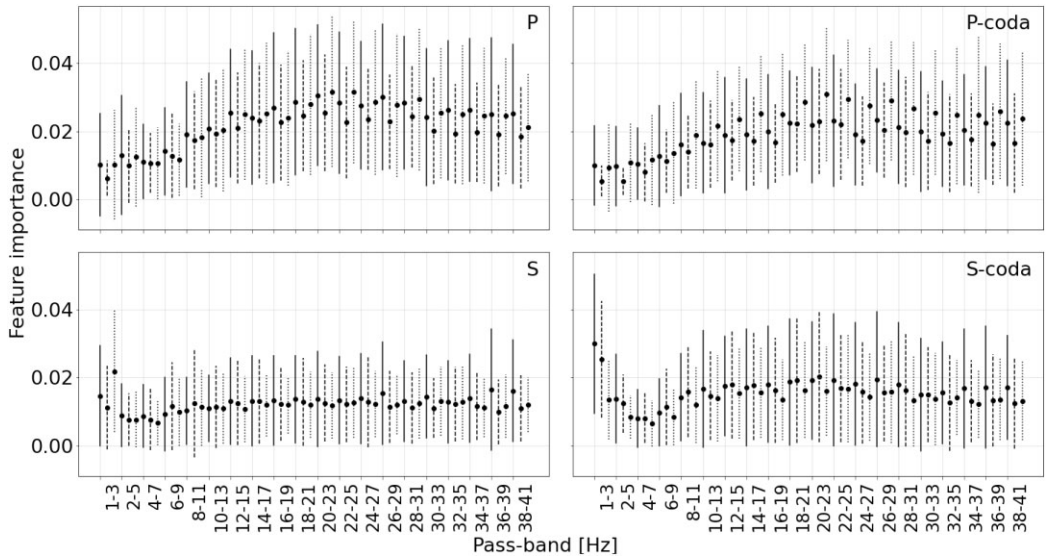

Figure 8. Mean importance of model input features over all station-specific models. Each point represents the importance of a single input feature, computed as the mean decrease of validation-set accuracy when the feature was randomly shuffled. Vertical lines from each point represent standard deviations. Different subfigures denote different time windows; top left: *P* phase, top right: *P* coda, bottom left: *S* phase and bottom right: *S* coda. The labels on the horizontal axis indicate passband of the filter applied to compute the feature. The three points shown for each passband correspond to the sensor components; vertical (solid line), radial (dashed line) and transverse (dotted line).

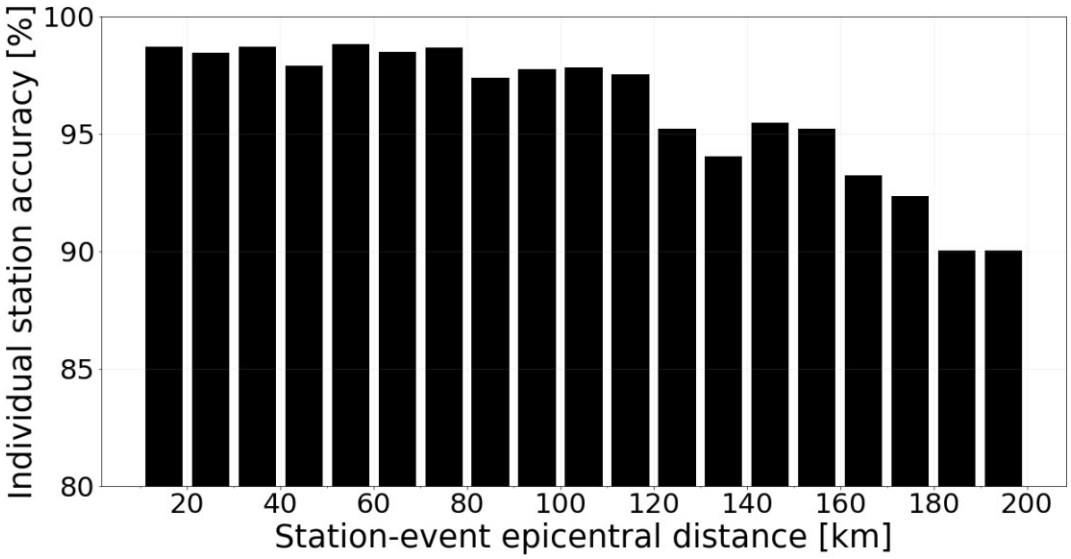

Figure 9. Mean classification accuracy as a function of station-event epicentral distance for individual station predictions during the evaluation period. Each vertical bar represents the mean classification accuracy over all individual station predictions from stations within a 10 km station-event epicentral distance range. Note that the vertical axis starts at 80 per cent.

resulted in erroneous final predictions. Due to the minimal observed difference and for the sake of simplicity we have opted to assign uniform weights and compute final predictions as arithmetic means of all individual station predictions.

4.7 Transferability of model parameters

During 2022, two permanent SNSN stations, Uddeholm (UDD) and Umeå (UMAU), were decommissioned and replaced with nearby stations from the Hagfors array (HFB1), operated by the Swedish Defence Research Agency, and Bullmark (BLMU), respectively (see Fig. 2). Since our classification models are station specific, when a new station is introduced into the network, we would typically need to wait for the station to collect enough data to train its own prediction model before having it available for class predictions. To test the transferability of station-specific model parameters we conducted tests where the classification models from the decommissioned stations, UDD and UMAU, were used to generate predictions for events recorded at stations HFB1 and BLMU, without applying any additional model training. We found that for events recorded on station HFB1 in the first two months of 2023, predictions made using the model trained for station UDD generated class predictions with 95 per cent accuracy. For events recorded at station BLMU, in the same period, the model trained for station

UMAU generated class predictions with 94 per cent accuracy. For comparison, the classification accuracy of station UMAU on its own test data was approximately 98 per cent.

These results offer encouragement that models trained for specific stations can be applied to data recorded on other nearby stations whilst maintaining relatively high classification accuracy. To further study the transferability of specific station model parameters we compiled a list of 120 events recorded at station BLMU during 2023 January and February and used station models from all other SNSN stations to generate class predictions. Fig. 10 shows the resulting accuracy results. We find that the classification accuracy ranges from about 60 to 96 per cent for the different station models, with an average of 83 per cent. In general, we observe a negative correlation between accuracy and distance from BLMU with some outliers, including the model which achieved the highest accuracy of about 96 per cent, from a station located about 400 km away from BLMU. Notably, the model which achieved the lowest accuracy of about 60 per cent also comes from a station located about 400 km away from BLMU.

These results encouraged us to test the feasibility of training one large classification model for the whole region as an alternative to the station-specific model approach. For this purpose, we started with the binary classification case and trained one model with all the blasts and earthquakes used in the original station-specific training data, excluding blasts associated with the mines in Kiruna and Malmberget. We subsequently applied the model, in the same way as the station-specific models, to all manually analysed blasts and earthquakes from the same evaluation period as before, 2022 January to June, excluding events located less than 25 km away from the mines in Kiruna and Malmberget. Of these 2703 events (see Table 1), 2690 were correctly classified by the model, the exact same number as were correctly classified by the station-specific models, yielding an accuracy of 99.3 per cent. Of the 13 events mis-classified by the single model, ten were also mis-classified by the station-specific models. When we apply the single model to data collected by a new station (BLMU), which was not part of the training data, we find that the single model achieves a classification accuracy of 97.4 per cent, outperforming all the station-specific models (see Fig. 10). This suggests that the single model is superior to the station-specific models when introducing new stations in the station network.

As a second test, we repeated the one model approach, this time including the two mining-induced event classes and trained the model using all the training data used for the station-specific approach, including events associated with the Kiruna and Malmberget mines. This time the overall classification accuracy during the evaluation period was 90.7 per cent, a significant decrease from the 97.4 per cent accuracy achieved using the station-specific models (see Fig. 6). The decrease in accuracy was most significant for events located at least 25 km away from the mines in Kiruna and Malmberget where the accuracy dropped from 99.3 to 86.4 per cent with the addition of the mining-induced event classes to the model. For events located less than 25 km away from the mines the accuracy drop off was smaller, going from 96.0 to 94.5 per cent.

4.8 Event-or-not classification

During 2022 June, the SIL system produced 1554 automatic event detections in Sweden with locations more than 50 km away from the mines in Kiruna and Malmberget. From these, 1221 were deemed to be spurious phase associations by analysts and 333 were real seismic events. Table 3 shows the results from comparing the output of the event-or-not classification to analyst assigned labels. The classification agreed with the analyst on 1173 of the spurious phase associations and 300 of the real seismic events, resulting in a classification accuracy of 94.8. The F1-scores are 0.881 for the real seismic event class and 0.967 for the spurious phase associations. As demonstrated by the high recall achieved for the spurious event class, the classification is able to correctly identify over 97 per cent of the spurious events. The SIL system also detects a large number of events, both spurious associations and real events, close to the mines in Kiruna and Malmberget. Due to their high frequency of occurrence, many of the smaller mining-induced events are typically not manually analysed and thus the accuracy of the event-or-not classification has not been determined close to the mines.

5 DISCUSSION

The results demonstrate that station-specific classification models based on fully connected neural networks are well suited to the task of classifying the seismic events detected in Sweden. The models are already in use at SNSN and have proven to be a useful addition to automatic event analysis, providing reliable preliminary class predictions to automatic event detections. They are also applied as a revision tool after manual analysis has been completed, highlighting potential mistakes and helping to classify ambiguous events. The event-or-not classification will be used to screen out the majority of spurious event detections prior to manual revision, thus reducing analyst workload.

5.1 Impact of various factors on predictions

5.1.1 Class imbalance

The models' high-performance scores involving the earthquake class, under-represented in most stations' training data, are an important achievement. They demonstrates the positive effect of oversampling the training data, as the models are able to correctly identify a large majority of the earthquakes in the evaluation period (high recall). Before applying actions to compensate for the class imbalance, for example, oversampling, many of the station models trained on the imbalanced data would predict almost all new events to belong to the over-represented class (blasts), leading to low-performance scores for the earthquake class. We tested alternative methods to compensate for the class imbalance, including undersampling the blast class and assigning different weights to events, proportional to the relative size of their class. All the tested methods offered significant improvement to the performance scores of the stations with imbalanced training data. We found that in our case, oversampling the earthquake class offered the most overall improvement in performance scores, particularly for stations with very few earthquakes (< 30) in their training data.

5.1.2 Accuracy of phase picks and event locations

The classification accuracy is slightly lower when automatic phase picks are used for the time window selection than when manual phase picks are used. This can partly be attributed to a higher number of phase picks being available after manual revision and thus more stations being available to generate predictions. Another

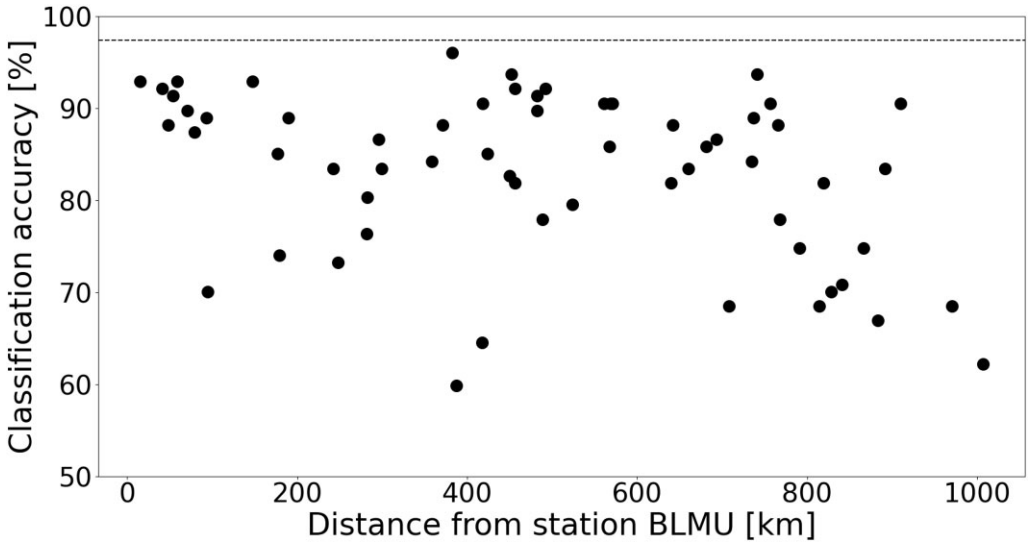

Figure 10. Classification accuracy for data recorded at station BLMU, using station models trained for other permanent SNSN stations. The dashed line indicates the corresponding accuracy when using one model trained with all the blasts and earthquakes used in the original station-specific training data, excluding blasts associated with the mines in Kiruna and Malmberget.

Table 3. Generalization performance of the event-or-not classification for automatic SIL event detections, 2022 June.

	Predicted class		
Analyst	Real	Spurious	**Re**
Real	300	48	0.862
Spurious	33	1173	0.973
Pr	0.901	0.961	
F1	0.881	0.967	

Notes. Metrics, Re: recall, Pr: precision and F1: F1-score.

reason is the higher uncertainty of the automatic phase picks compared to manual ones, which leads to less accurate time window selections. Automatic event locations generally have higher uncertainty than manually revised ones and thus calculations of theoretical phase arrival times and rotation of components also become less accurate. We have found that the main limiting factor of our method, when using automatic phase picks, is its sensitivity to the timing of the phase picks being accurate. Most events, for which the models generate a final class prediction which contradicts the later manually assigned class, share the common feature that the timing of their phase picks is inaccurate, leading to an inaccurate time window selection and erroneous class predictions. This can manifest itself in different ways. For example, if an automatic *P*- or *S*-phase pick is missing at a given station for an event which epicenter, depth and/or origin time are poorly constrained, the missing phase arrival needs to be computed theoretically and can lead to significant inaccuracies. Another source of inaccurate phase picks occurs when automatic phase association mixes up *P* and *S* phases, leading to erroneous time window selections. Fig. 11 shows an example of a seismic trace from a mining-induced event in the Kiruna mine where the *P*-phase arrival has been automatically picked as an *S*-phase arrival. Once the *P*-phase arrival has been theoretically computed, based on the erroneous *S*-phase pick, the consequent time window selection leads to an erroneous class prediction. Moving forward, we see increasing accuracy of automatic phase picks as one way of further improving the event classification. The accuracy results for the manually analysed events serve as a benchmark to the classification's potential when pick quality is optimal.

In general, the models described in this study are capable of classifying seismic events in Sweden to a high degree of accuracy, irrespective of their location. The accuracy is, however, lower for events located close to the mines in Kiruna and Malmberget, compared to events at further distances away from the mines. This can primarily be attributed to the station models having four classes to distinguish instead of two. Another reason for lower accuracy close to the mines when using automatic phase picks for the time window selection is a higher relative number of events with inaccurate phase picks/association in the SIL system compared to events further away from the mines. Finally, the waveform features from events associated with the mines, particularly blasts and high-frequency mining-induced events often share similar features, making them hard to distinguish, even for human analysts. In practice, analysts often rely on an event's origin time to distinguish between blasts and mining-induced events in Kiruna and Malmberget, since the blasts typically occur during the same specific hours each day.

5.1.3 Mislabelled events in the training data

In this study, we have not explicitly estimated the effects of mislabelled events in the training data on the results. We have however identified several events in the training data, where our model produces a final predicted class with a high QF which disagrees with the analyst assigned class. For many of these events, closer manual inspection has led us to conclude that the model is correct and a mistake was made during the manual analysis. Such examples confirm the presence of mislabelled events in the training data although their extent and effect on the final results are difficult to quantify. The examples also highlight the capability of the model to identify classification mistakes in the data archives and provide the opportunity to review specific events. Currently at SNSN, the model is run for each calendar month, after manual analysis has been completed. On average, it detects 1–2 events every month where an event was mislabelled during manual analysis. The application of the model

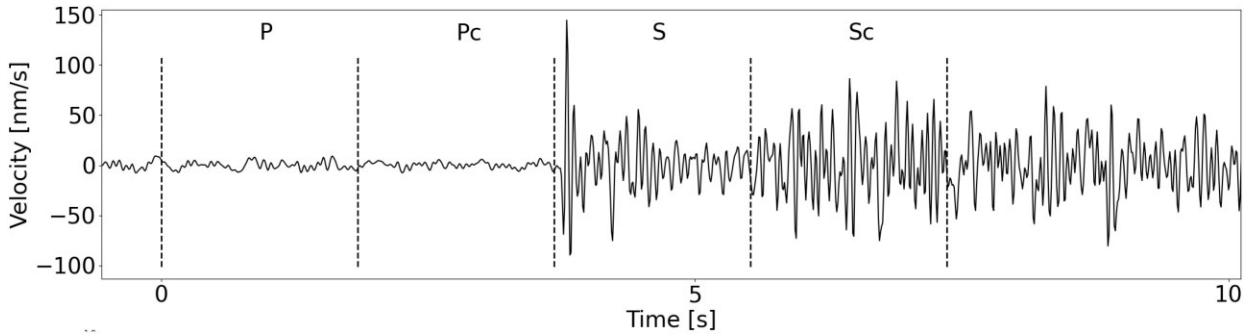

Figure 11. Vertical component from a mining-induced event in the Kiruna mine where an automatic *S*-phase pick has been erroneously assigned to the arrival of the *P* phase, leading to erroneous time window selection.

thus allows us to fix mistakes and reduce the number of errors in the event catalogue.

5.1.4 Rotation of horizontal components

One step in our data processing is the rotation of the horizontal data components into radial and transverse coordinates. The idea behind applying this rotation is to better capture individual seismic phases on the horizontal components, such as the *Rg* phase on the radial component and the *S*/*Lg* phase on the transverse component. Rotating the horizontal components also means that the computed waveform features are less sensitive to the direction from which seismic phases arrive at a station and thus helps with transferability of model parameters from one station to another. However, rotating the horizontal components requires an estimate of an event's location to compute the backazimuth. For automatic event detections, the uncertainty in an event's location can be high, leading to wrong rotations. We have seen that some of the erroneous classifications made by our model, particularly for events located close to the mines in Kiruna and Malmberget, can be traced to erroneous location estimates and subsequent rotations rather than low-phase pick accuracy.

5.1.5 Station-specific models or not?

Our results on parameter transferability (Fig. 10) show that models trained on data recorded at a specific station have the capacity to be applied to data recorded on a different station and maintain high classification accuracy. Sweden is part of the Fennoscandian Shield, an area of low seismicity in northern Europe (Gregersen *et al.* 2021). It is feasible that the relatively uniform and stable Swedish bedrock, with its low levels of attenuation, contributes to favourable conditions for model parameter transferability between different stations. However, Fig. 10 also shows that the while the accuracy achieved on different stations tends to decrease with distance, it is not a simple linear relationship. Both the lowest and highest accuracy scores (about 60 and 96 per cent) are achieved at stations located about 400 km away from the recording station.

Our results from testing the feasibility of using one model for the whole region (Section 4.7), as opposed to the station-specific approach, suggest that both approaches represent feasible alternatives for seismic event classification in Sweden. Our primary reasons for starting with station-specific models were the large geographical spread of the seismic stations operated by the SNSN and the variability in the types of seismic events and ambient noise levels recorded by the different stations. In northern Sweden, the primary source of seismic events classified by the SNSN as 'blasts' are mining blasts originating from the underground mines in Kiruna and Malmberget. In contrast, the majority of 'blasts' recorded in other parts of the country result from quarry blasting and construction work. Most earthquakes recorded in northern Sweden can be associated with known postglacial faults whereas the sources of earthquakes recorded in southern Sweden are typically not as clear. We hypothesized that station-specific models would allow us to better capture the individual characteristics of each seismic station, thus making for a more reliable class prediction.

Our results suggest that when events associated with the two mines are included, where four event classes are involved, the station-specific approach remains favourable to the one model approach in terms of prediction accuracy. For the binary classification of blasts and earthquakes, using the one model approach has proved to be equally effective as the station-specific approach and thus both approaches represent feasible alternatives. Arguments for selecting the one model approach might include less work managing model updates and more seamless integration of new stations into the classification. With the good performance achieved by the one model approach, we have recently started testing the application of the model to data from neighbouring stations outside of the SNSN. The results look generally promising, especially in regions with geological structure similar to Sweden. Fig. 10 demonstrates that in the case of BLMU, a new station which was added to the SNSN, applying the one model approach to data collected by the station outperforms all the individual, station-specific models. We hypothesize that this difference in performance can primarily be attributed to the significantly larger training data set involved in the one-model approach compared to any station-specific model, allowing for more sophisticated learning of the model features. A potential way of further developing the methodology presented in this study may be to attempt to extract the best from both scenarios by adapting ideas from the field of transfer learning, a subfield of machine learning. With the knowledge learned from the one model approach we may improve the station-specific models by using the large model as a base to fine-tune the station-specific models.

5.2 Comparison to other methods

The accuracy achieved in our study for the binary classification of blasts and natural earthquakes (98 per cent for automatic event detections and 99 per cent for manually analysed events) is comparable to other recent studies devoted to the same purpose in different regions. In contrast to many of the more recent studies on seismic event classification (Hourcade *et al.* 2023; Kong *et al.* 2022, 2021;

Linville *et al.* 2019; Tibi *et al.* 2019) our method does not rely on images of event spectrograms and thus does not require the application of CNNs. Whether the method generalizes as well to different geographical areas as, for example, the method proposed by Hourcade *et al.* (2023) has not yet been tested. Our method also depends on manual feature selection. The feature selection is inspired by the work of Kortström *et al.* (2016), for a seismic network located in geological settings comparable to Sweden. Our accuracy results for distinguishing between blasts and earthquakes are somewhat better than Kortström *et al.* (2016) which suggests the suitability of the selected features for event classification, at least in the stable continental region of the Fennoscandian Shield. Our accuracy involving four event classes (90 per cent for automatic event detections and 96 per cent for manually analysed events) is more difficult to compare to other studies but suggests that the method is also capable of performing event classification involving multiple event classes to a high accuracy.

6 CONCLUSIONS

The main objective of this study is to develop an algorithm capable of reliably predicting the source types of seismic events detected in Sweden. Such an algorithm will allow a reduction in the time spent by analysts establishing the source type of automatically detected seismic events. We have developed station-specific event classification models using traditional fully connected artificial neural networks. The models distinguish between four different event classes representative of analysis routines at the SNSN; natural (tectonic) earthquakes, blasts (from e.g. quarries, mines and construction) and two types of mining-induced events, with a high degree of accuracy. An extension of the models is capable of distinguishing the majority of spurious phase associations from real seismic events of interest.

The classification accuracy achieved by the models when distinguishing between earthquakes and blasts is 98 per cent for automatic event detections and 99 per cent for manually analysed events, comparable to other studies dedicated to the same purpose in different geographical areas. The models also maintain high classification accuracy in areas close to the underground mines in Kiruna and Malmberget, with four different event classes to distinguish, 90 per cent for automatic event detections and 96 per cent for manually analysed events. Most erroneous classifications for automatic event detections can be traced back to uncertainties in automatic phase picks and/or association, leading to erroneous time window selections. Other reasons for erroneous classifications include high automatic location estimate uncertainties, leading to erroneous rotations of the horizontal data components. The accuracy of the event-or-not classification, distinguishing spurious phase associations from real seismic events, is about 95 per cent for events located more than 50 km away from the mines in Kiruna and Malmberget.

The most important features, on average, for the model predictions are computed from time windows associated with *P* phase and *P* coda at intermediate to high frequencies (12–41 Hz) as well as *S* phase and *S* coda at low frequencies (1–3 Hz), corresponding to time windows where surface wave energy (*Rg*/*Lg*) is strong. Models trained on data recorded at specific stations have the capacity to be applied to data recorded on a different station and maintain high classification accuracy. Training one model for the whole region results in equally high accuracy when distinguishing blasts and earthquakes as training station-specific models does. With the addition of the mining-induced event classes, the station-specific approach performs better. The models are already in use at the SNSN, where they serve as a tool to assign preliminary type predictions to automatic event detections and to review manually analysed events, identifying potential mistakes. The models will also be used to evaluate which events to subject to manual analysis in the future.

ACKNOWLEDGMENTS

The work presented in this study was made possible by the thorough manual analysis work undertaken by analysts at the SNSN in the years 2010–2022. The authors wish to thank the Swedish Defence Research Agency for use of data from the Hagfors array, and LKAB for information on events in the mines. The suggestions of an anonymous reviewer and the associate editor improved the manuscript, which we thank them for.

DATA AVAILABILITY

This study uses manual and automatic event definitions, and waveform data, from 80 000 seismic events recorded by the SNSN (1904). Subsets of the data are available upon request to the SNSN: https://doi.org/10.18159/SNSN

REFERENCES

Abadi, M. *et al.*, 2015. *TensorFlow: Large-scale machine learning on heterogeneous systems, Software available from tensorflow.org,* last accessed 9 January 2024.

Båth, M., 1975. Short-period rayleigh waves from near-surface events, *Phys. Earth planet. Inter.,* **10**(4), 369–376.

Baumgardt, D.R. & Young, G.B., 1990. Regional seismic waveform discriminants and case-based event identification using regional arrays, *Bull. seism. Soc. Am.,* **80**(6B), 1874–1892.

Böðvarsson, R., Rögnvaldsson, S.T., Slunga, R. & Kjartansson, E., 1999. The sil data acquisition system—at present and beyond year 2000, *Phys. Earth planet. Inter.,* **113**(1–4), 89–101.

Breiman, L., 2001. Random forests, *Mach. Learn.,* **45,** 5–32.

Burkov, A., 2019. *The Hundred-page Machine Learning Book,* Vol. **1,** Andriy Burkov, Quebec City, QC, Canada.

Chawla, N.V., Bowyer, K.W., Hall, L.O. & Kegelmeyer, W.P., 2002. Smote: synthetic minority over-sampling technique, *J. Artif. Intell. Res.,* **16,** 321–357.

Chollet, F. *et al.*, 2015. *Keras,* https://keras.io., last accessed 9 January 2024.

Dysart, P.S. & Pulli, J.J., 1990. Regional seismic event classification at the noress array: seismological measurements and the use of trained neural networks, *Bull. seism. Soc. Am.,* **80**(6B), 1910–1933.

Furumura, T., Hong, T.-K. & Kennett, B.L., 2014. Lg wave propagation in the area around japan: observations and simulations, *Prog. Earth planet. Sci.,* **1,** 1–20.

Goodfellow, I., Bengio, Y. & Courville, A., 2016. *Deep Learning,* MIT Press, http://www.deeplearningbook.org.

Gregersen, S., Lindholm, C., Korja, A., Lund, B., Uski, M., Oinonen, K., Voss, P.H. & Keiding, M., 2021. *Seismicity and Sources of Stress in Fennoscandia,* pp. 177–197, Cambridge University Press, doi:10.1017/9781108779906.

Harris, C.R. *et al.*, 2020. Array programming with NumPy, *Nature,* **585**(7825), 357–362.

Hourcade, C., Bonnin, M. & Beucler, É., 2023. New cnn-based tool to discriminate anthropogenic from natural low magnitude seismic events, *Geophys. J. Int.,* **232**(3), 2119–2132.

Hunter, J.D., 2007. Matplotlib: A 2d graphics environment, *Comput. Sci. Eng.,* **9**(3), 90–95.

Juhlin, C. & Lund, B., 2011. Reflection seismic studies over the end-glacial burträsk fault, skellefteå, sweden, *Solid Earth,* **2**(1), 9–16.

Kim, W.-Y., Simpson, D. & Richards, P.G., 1993. Discrimination of earthquakes and explosions in the eastern united states using regional high-frequency data, *Geophys. Res. Lett.,* **20**(14), 1507–1510.

Kintner, J.A., Michael Cleveland, K., Ammon, C.J. & Nyblade, A., 2020. Testing a local-distance r g/s g discriminant using observations from the bighorn region, wyoming, *Bull. seism. Soc. Am.,* **110**(2), 727–741.

Kong, Q., Chiang, A., Aguiar, A.C., Fernández-Godino, M.G., Myers, S.C. & Lucas, D.D., 2021. Deep convolutional autoencoders as generic feature extractors in seismological applications, *Artif. Intell. Geosci.,* **2,** 96–106.

Kong, Q., Wang, R., Walter, W.R., Pyle, M., Koper, K. & Schmandt, B., 2022. Combining deep learning with physics based features in explosion-earthquake discrimination, *Geophys. Res. Lett.,* **49**(13), e2022GL098645. https://doi.org/10.1029/2022GL098645.

Kortström, J., Uski, M. & Tiira, T., 2016. Automatic classification of seismic events within a regional seismograph network, *Comput. Geosci.,* **87,** 22–30.

Lindblom, E., Lund, B., Tryggvason, A., Uski, M., Bödvarsson, R., Juhlin, C. & Roberts, R., 2015. Microearthquakes illuminate the deep structure of the endglacial pärvie fault, Northern Sweden, *Geophys. J. Int.,* **201**(3), 1704–1716.

Linville, L., Pankow, K. & Draelos, T., 2019. Deep learning models augment analyst decisions for event discrimination, *Geophys. Res. Lett.,* **46**(7), 3643–3651.

Lund, B., Schmidt, P., Hossein Shomali, Z. & Roth, M., 2021. The Modern Swedish National Seismic Network: two decades of intraplate microseismic observation, *Seismol. Res. Lett.,* **92**(3), 1747–1758.

McKinney, W. 2010. Data structures for statistical computing in python, in *Proceedings of the 9th Python in Science Conference,* pp. 56–61.

Miao, F., Carpenter, N.S., Wang, Z., Holcomb, A.S. & Woolery, E.W., 2020. High-accuracy discrimination of blasts and earthquakes using neural networks with multiwindow spectral data, *Seismol. Res. Lett.,* **91**(3), 1646–1659.

O'Rourke, C.T., Baker, G.E. & Sheehan, A.F., 2016. Using p/s amplitude ratios for seismic discrimination at local distancesusing p/s amplitude ratios for seismic discrimination at local distances, *Bull. seism. Soc. Am.,* **106**(5), 2320–2331.

Pedregosa, F. *et al.*, 2011. Scikit-learn: machine learning in Python, *J. Mach. Learn. Res.,* **12,** 2825–2830.

Press, F. & Ewing, M., 1952. Two slow surface waves across North America, *Bull. seism. Soc. Am.,* **42**(3), 219–228.

Sasaki, Y. *et al.*, 2007. The truth of the f-measure, *Teach Tutor Mater,* **1**(5), 1–5.

SNSN, 1904. *Swedish National Seismic Network,* Uppsala University, Uppsala, Sweden. Other/Seismic network, https://doi.org/10.18159/SNSN.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting, *J. Mach. Learn. Res.,* **15**(1), 1929–1958.

Tibi, R., Koper, K.D., Pankow, K.L. & Young, C.J., 2018. Depth discrimination using rg-to-sg spectral amplitude ratios for seismic events in Utah recorded at local distancesdepth discrimination using rg-to-sg spectral amplitude ratios for seismic events in utah, *Bull. seism. Soc. Am.,* **108**(3A), 1355–1368.

Tibi, R., Linville, L., Young, C. & Brogan, R., 2019. Classification of local seismic events in the utah region: a comparison of amplitude ratio methods with a spectrogram-based machine learning approach, *Bull. seism. Soc. Am.,* **109**(6), 2532–2544.

Wessel, P., Luis, J., Uieda, L., Scharroo, R., Wobbe, F., Smith, W.H. & Tian, D., 2019. The generic mapping tools version 6, *Geochem. Geophys. Geosyst.,* **20**(11), 5556–5564.