

Artificial Neural Networks as Emerging Tools for Earthquake Detection

Otilio Rojas^{1,2}, Beatriz Otero³, Leonardo Alvarado^{4,2}, Sergi Mus³, Rubén Tous³

¹ Barcelona Supercomputing Center (BSC), Barcelona, Spain

² Universidad Central de Venezuela (UCV), Caracas, Venezuela

³ Universitat Politècnica de Catalunya (UPC), Barcelona, Spain

⁴ Fundación Venezolana de Investigaciones Sismológicas (Funvisis), Caracas, Venezuela

otilio.rojas@bsc.es, botero@ac.upc.edu

Abstract. As seismic networks continue to spread and monitoring sensors become more efficient, the abundance of data highly surpasses the processing capabilities of earthquake interpretation analysts. Earthquake catalogs are fundamental for fault system studies, event modellings, seismic hazard assessment, forecasting, and ultimately, for mitigating the seismic risk. These have fueled the research for the automation of interpretation tasks such as event detection, event identification, hypocenter location, and source mechanism analysis. Over the last forty years, traditional algorithms based on quantitative analyses of seismic traces in the time or frequency domain, have been developed to assist interpretation. Alternatively, recent advances are related to the application of Artificial Neural Networks (ANNs), a subset of machine learning techniques that is pushing the state-of-the-art forward in many areas. Appropriated trained ANN can mimic the interpretation abilities of best human analysts, avoiding the individual weaknesses of most traditional algorithms, and spending modest computational resources at the operational stage. In this paper, we will survey the latest ANN applications to the automatic interpretation of seismic data, with a special focus on earthquake detection, and the estimation of onset times. For a comparative framework, we give an insight into the labor of human interpreters, who may face uncertainties in the case of small magnitude earthquakes.

Keywords. P and S seismic waves, earthquake hypocenters, supervised, unsupervised and semisupervised, deep and convolutional neural networks, training and testing data sets.

1 Introduction

Large-magnitude earthquakes may induce strong ground shaking, surface ruptures, landslides, liquefaction and Tsunamis, being any of these direct or indirect causes of human deaths and severe economic losses. Fortunately, earthquake magnitude and frequency are inversely proportional, so events of large and moderate magnitudes are scarce on seismic catalogs compared to earthquakes of low magnitudes. However, human or automatic detection by traditional techniques may fail for low-magnitude events, in case of traces with poor signal-to-noise ratios or recording overlapping events. Even though, small-magnitude earthquakes are not usually harmful for lives and civil structures, they must be also recorded in seismic catalogs. These catalogs are used in seismic hazard assessment to map the expected ground motion under a given earthquake

magnitude, and must influence building codes. In addition, the detection of small earthquakes in an aftershock sequence left by a big event, are used to study the distribution of the fault system, total rupture area, and the evolution of the stress configuration. This information, in combination to additional geological studies, allow assessing the destructive potential of that given fault.

The detection problem of local earthquake events consists of estimating the arrival times of the primary (P) and secondary (S) waves to a recording seismic station. These onset times feed subsequent processing analyses to estimate hypocenter location, focal source mechanism and important spectral properties, being this information recorded in earthquake catalogs afterward. Detection and processing analyses are routinely performed by a human analyst, who must estimate arrival times based on his own knowledge and experience. This estimation requires a significant amount of pattern recognition, and is tied to the identification problem of individual P and S arrivals from their amplitude, propagation speed, and induced signal polarization. Nowadays, the need of automatic processing tools grows rapidly with the amount of seismic data delivered by expanding monitoring networks, so detection and identification of P and S waves can be performed in a faster, robust and objective way.

The automation of seismic phase detection and identification must deal with characteristic features of seismic records. As commonly used for data analysis, signals are processed on time windows. In these cases, those windows may contain a lot of noise and high redundancy, so true P and S waves can be placed anywhere in the time span and polluted by alternative signals. For instance, wave conversions that depend on underground structures and site conditions, or even surface waves, may be also present on long time windows. Also, these windows may contain P and S waves triggered by simultaneous or overlapping seismic events. In many cases, seismic stations record data in several channels, allowing either a separate or combined processing of horizontal, vertical and transverse components of soil kinematics.

All these features make difficult and laborious the recognition of similar patterns on seismic signals,

and the P and S detection and identification from one or various events. Filtering, polarization analysis, and additional signal processing techniques have been of primary help for automation.

Artificial Neural Network (ANN) represent a natural implementation framework for P and S wave detection and identification because of the huge amount of pattern recognition involved. The pioneer applications in [16] and [17] have been followed by multiple and diverse networks, using either assisted, or semisupervised or completely unsupervised learning. In the supervised family, we find several generic and convolutional networks, and even some scatter cases of recurrent and residual networks. In the unassisted family, there are applications based on generative adversarial networks and others exploiting autoencoders.

Only few cases are built upon ensemble learning, but their potentials are worthy of mentioning. Now, all these ANN applications must deal with the processing difficulties of time windows of seismic records, mentioned above. These difficulties translate into more complex heuristic rules to estimate the number of neurons per each network level, and the size of the learning set, if it is required. This paper reviews all these implementation aspects, as well as the detection and identification performance achieved by ANN.

Alternative machine learning (ML) techniques can also offer advanced pattern-matching abilities and well serve as automated earthquake monitoring algorithms. Moreover, the expanding ML applications to seismology have reached areas such as earthquake early warnings, ground motion estimations, and seismic tomography and inversion, and we here also briefly comment on prominent works, for the sake of completeness. However, we refer the reader to the following complementary survey papers [18, 23, 29], Kong2019, for additional insights.

2 Picking of Seismic Phases by a Trained Analyst

Modern data acquisition systems use sample rates of more than 100 Hz, and this limits the estimation accuracy of P and S arrival times, among

other factors such as signal-to-noise conditions at monitoring stations [11]. In addition, hand-picked data are human dependent and errors are difficult to estimate, since this activity is performed by several analysts. Different trained analysts may give different onset times for the same signal phase, and sometimes the same analyst may give different interpretations after some time has passed.

For events at epicentral distances not larger than 10° (~ 1000 Km), recorded seismic waves have propagated to the crust and along the Moho discontinuity. At those stations, these waves are called crustal waves, being triggered by local and regional earthquakes. On the contrary, teleseismic events refer to earthquakes that occur at distances larger than 10° from the measurement site. The next description follows [32] and it is focused on crustal waves.

In general, a seismic phase is identified by two main features. A signal amplitude increase that exceeds the background noise in the case of P waves, or exceeds the coda of earlier phases in the case of S and later waves. A change of the dominant frequency that is often much more difficult to visualize and quantify [21]. Once the phase is identified, we have to determine the precise wavelet onset or arrival time, and the precise phase type. That is, a Pg phase or a direct P wave, a Pn that is the refracted P wave along the uppermost mantle, a Sg phase or a direct S wave, among other possibilities. The visual examination of amplitude and frequency changes can be rather subjective and would depend on the chosen width of time windows. As illustrative examples, figures 1 and 2 show the P and S phases identified at the CUMV and CACV stations of the Venezuelan Funvisis seismic network (2019/01/16 13:42, 4.0 Mw). In figure 1, both P and S phases are identified through amplitude changes of the recorded seismogram. In figure 2, identification of the P wave is straightforward from a clear change of the frequency content after its arrival, with respect to the background noise.

Signal filtering can help to identify the phase arrivals by improving the signal-to-noise ratio. However, filtering can also introduce shape changes on the waveform, and even a phase

shift. This time shift could modify the phase arrivals in the order of tenth of seconds [21]. Figure 3 shows the vertical component of ground acceleration recorded at the BAUV station of the Funvisis network, in Cojedes, Venezuela. This station is located 198 km away from the epicenter of the event (2018/04/02 16:24, 2.0 Mw). Later, this seismogram is passband filtered (5-10 Hz) and resulting signal is depicted in figure 4. Here, we can see how the P wave is enhanced, and it can be easily identified by the first notorious change in amplitude.

There are software available to assist the human analyst that could even be used for automatic processing, most of them being created by university groups or seismological institutions. Some packages widely used are SHM, SEISAN, SeisComp3, Earthworm and ObsPy. These processing software were developed and published under a free, or partially free license. Software libraries for routinely analysis and additional research work, can be downloaded from web sites of the United State Geological Survey (USGS: <https://earthquake.usgs.gov/research/software/>), and of the Incorporated Research Institutes for Seismology (https://www.iris.edu/hq/data_and_software/).

3 Machine Learning Algorithms and Deep Neural Networks

The fundamentals and pivotal developments on automatic pattern recognition and ML techniques can be reviewed on reference textbooks, such as [10], [41], and [25], as well as on some comprehensive online courses, as for instance, the "Neural Networks for Machine Learning" from G. Hinton¹ and the "Machine Learning" from A. Ng². This section highly lacks of such generality, and only presents a basic ML classification based on the learning strategy, that later serves to introduce ML and ANN applications to geophysical data, in next sections. Thus, we recommend

¹http://www.cs.toronto.edu/~hinton/course_slides.html

²<https://www.coursera.org/learn/machine-learning>

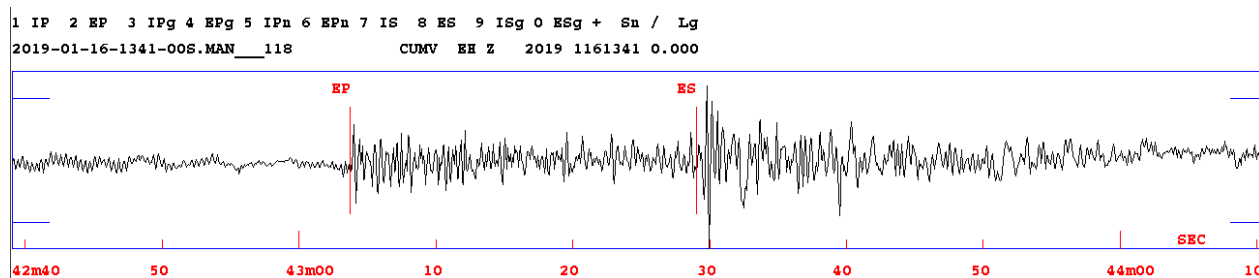


Fig. 1. Seismogram recorded at CUMV station in Sucre, Venezuela (2019/01/16 13:42 4.1 Mw), and used for P and S wave identification based on the change of signal amplitude

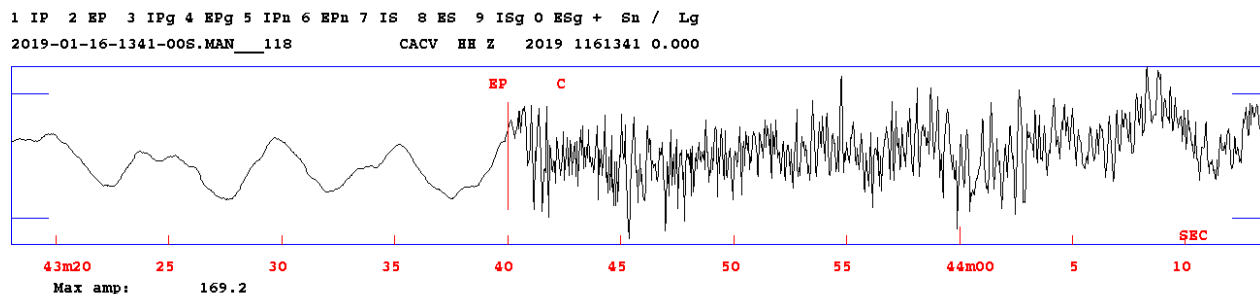


Fig. 2. P-wave identification based on the change of the frequency content for the same event in Figure 1, but recorded at CACV station in Bolivar, Venezuela

aforementioned books and materials for a broader presentation.

The rich variety of ML algorithms can be grouped according to the type of data used during training, a stage where optimization strategies allow tuning the algorithm parameters usually by minimizing a cost function. To this purpose, the training dataset can be fully or partly labelled, and even it may present no labels at all. Coarsely speaking, this dataset property frames the learning algorithm and also defines its potential applications. Supervised learning employs labelled datasets and develops models used for either quantitative prediction or categorical classification. In this case, the training and evaluation stages are time consuming because of the labelling and data preparation. Alternatively, unsupervised learning works on unlabelled datasets aiming at clustering the input data into groups based on similarity measures, or at reducing the input data dimensions. Because of labelling is not required, the preparation of the training and evaluation datasets takes less

effort, which allows using bigger amount of data. Finally, semisupervised learning uses a hybrid dataset for training and evaluation, typically with a small fraction of labelled data, compared to a larger amount of unlabelled data. These different machine learning algorithms are shown in the categorization figure 5 along with their typical applications that vary from data classification, quantitative regression, grouping or clustering, and reduction of data dimensionality. An additional application pursuing data grouping by following implicit association rules to discover relations among variables in large databases, is also considered in this figure. The specific algorithms referred in the technical literature are also shown for each category at the bottom.

Figure 5 places ANN into the classification hierarchy of supervised learning algorithms, delineating also their possible applications when processing earthquake data. In fact, specific ANN implementations on earthquake detection are mentioned in next sections, along with interesting

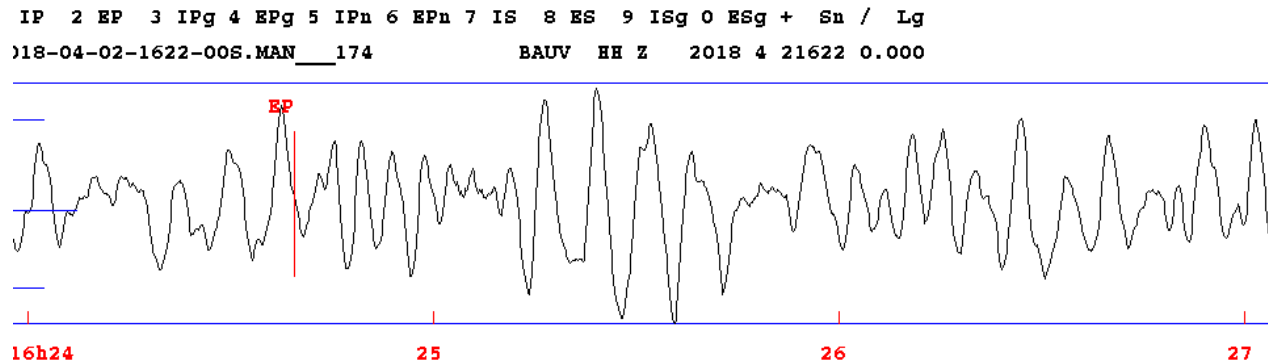


Fig. 3. P wave onset on the vertical acceleration seismogram recorded at the Funvisis BAUV station in Cojedes, Venezuela, during a local earthquake (2018/04/02 16:24 Mw=2,0)

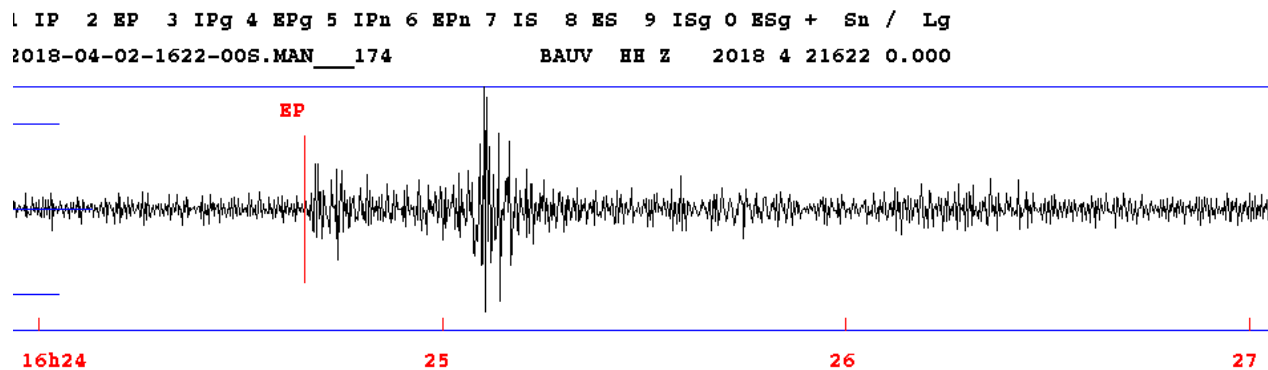


Fig. 4. The new signal after applying a Butterworth filter in the band 510 Hz to the seismogram in figure 3.

additional works based on autoencoder algorithms, whose applications also facilitate this interpretation task.

To emphasize the focus of our study on these ML algorithms, we use grey boxes in figure 5. Most of those ANN applications are based on deep learning, as a way to achieve high accuracy by the incorporation of several neuron layers, and therefore heuristic rules. The ANN prediction accuracy would partly depend on a previous good training, but also on the number of network layers that allow a parallel and multi-level detection of important data features (also referred as abstraction levels).

The presence of two or more neuron layers is defined as *deep learning*, while *very deep learning* is referred to using more than ten layers. Nowadays, applications based on very

deep learning are scarce (because of complexity of the network structure), as also those using a shallow ANN with only one hidden layer (because of the limited learning capability).

It is worthy of mentioning that deep architectures could first use unsupervised learning, and then appeal to fine-tuning in a final stage of supervised training, aiming at a superior performance [49].

4 Machine Learning Applications to Earthquake Data

The physics of earthquakes is very complex and seismologists rely on processing and interpreting massive data sets in search for insights. ANN, among other ML strategies, have the potential to find unseen patterns and new features, both with

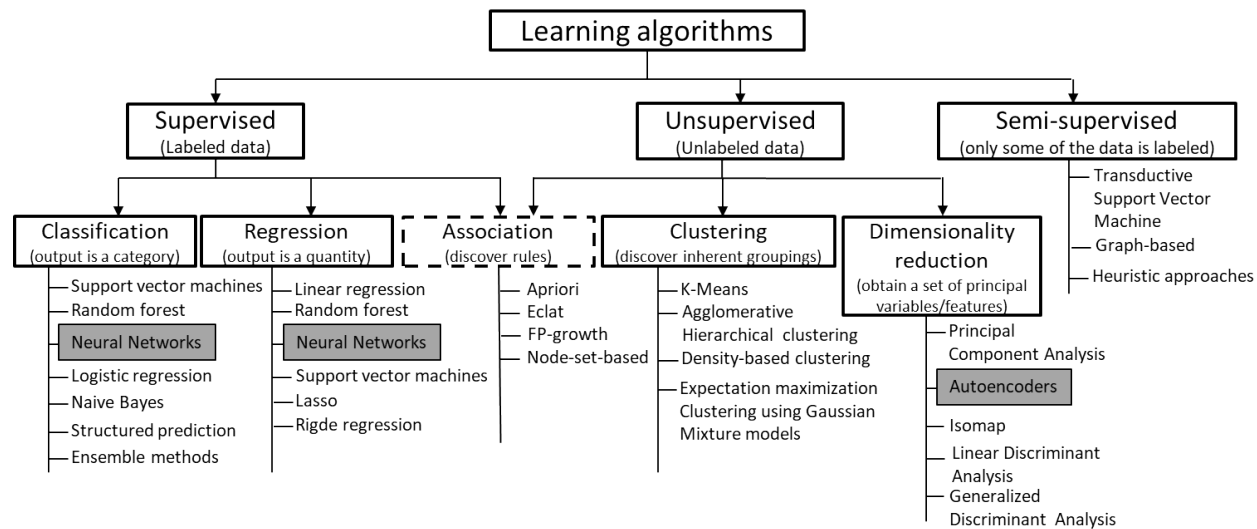


Fig. 5. Supervised, unsupervised and semisupervised algorithms for ML

physical consistency, in available datasets. Thus, seismologists and computer scientists have been actively developing ML tools to process earthquake data and assist interpretation. Numerous applications have focused on earthquake detection and phase picking, exploiting the ML proficiency of data classification and regression when processing real seismic traces.

Recent and interesting ML applications to earthquake detection are the classification Markov model proposed in [6], the image-based search for waveform matching presented in [56] and the seismic arrival picking based on fuzzy clustering developed in [14]. In next section, we review additional and prominent works in this area, but focus on contributions based on ANN. However, we before comment on three main additional ML applications to seismology, which are under current expansion, and relate to earthquake early warnings, ground motion estimations, and seismic tomography and inversion.

Earthquake early-warning systems intend to identify a seismic event onset at remote stations, and exploits the faster propagation of P waves compared to more destructive S and surface waves, to trigger alerts at local sites, seconds or few minutes before the strongest shaking happens. Earthquake early warning require tools

for earthquake detection, and some works based on ANN can be found in [26, 13, 30]. Alternatively, similar applications have been also addressed by the ML approaches in [15, 12, 30].

During an earthquake, ground motion experienced at a given surface location would depend on several parameters of the seismic source, material properties along wave propagation paths and local site conditions. The statistical prediction of ground motion is based on regression models that combines these parameters, and strives for reducing uncertainties by increasing observational data and model complexity. Ground motion estimates are key inputs to probabilistic analyses of seismic hazard, that are sensitive to these prediction uncertainties [4].

Alternatively, Authors in [19, 20, 28] employ ANN for ground motion estimation using as input, key parameters of the seismic source and the velocity model. In particular, results in [28] are more accurate than those provided by the Central and Eastern North America attenuation model (CENA) developed in [8]. Alternative ML approaches have been also used in similar applications, as discussed in [45, 48].

Travel-time tomography and full-waveform inversion are subsurface imaging procedures based on the minimization of certain discrepancies

between seismic data and simulations. Given a reference model, the former seeks the reduction of source-receiver travel times differences by relocating material interfaces, while the latter updates medium properties to reduce mismatches between data and synthetic waveforms. The convergence of both methods to a realistic structural model is time consuming, while tomography may require continuous intervention of a human expert, full-waveform inversion is computationally demanding and requires a good initial model.

Recent ANN applications to subsurface imaging claim to overcome these deficiencies, using seismic data as input to identify important structures. In particular, [39] uses earthquake data to accurately predict 1-D velocity models, and applications in [5, 34, 27, 53] employ data collected in seismic surveys for structural model building with interest in hydrocarbon exploration. In addition, the study in [37] applies a ANN to infer the prior distribution of acoustic properties of a geological model, that is later improved by full waveform inversion. This work includes promising applications to a synthetic reservoir-scale dataset of channel bodies. Lastly, we would like to note on alternative ML approaches that use data-driven dictionaries to adaptively capture complex profiles of geophysical parameters, and used to improve the convergence of seismic tomography [9] or inversion [58].

5 Neural Networks for Earthquake Detection

According to the figure 5, the identification of P and S waves on seismic traces is a classification problem, while the estimation of onset arrival times, as required for complete detection, corresponds to a regression problem. Either processing objective will be highlighted below when reviewing ANN solution approaches developed during the last thirty years, and this discussion will also include a ANN feature suitable to achieving it. The network input data can be the direct (probably processed) seismograms, or some previously extracted (primarily statistical) features from these signals.

Full traces provide to the ANN with the whole information carried by the seismic signal, but due to its high dimensionality, the mapping of learning examples to the feature space might be highly sparse, making the ANN prone to wrong convergences at the production stage. Alternatively, a more compact feature space results from a network learning process using only relevant signal features as input data, that will translate into high operational accuracy. However, if some important features are omitted, the ANN might performs poorly. In the following, we only focus on earthquake detection ANN under supervised and unsupervised learning given their very active development, and enough contributions to motivate this survey paper.

5.1 Supervised Networks

Supervised ANN fall into four main categories: Feed forward (FFNN), recurrent (RcNN), convolutional (CNN) and residual (RsNN). FFNN present a simple structure and are commonly used for data classification. RcNN are usually applied to time series analysis to identify sequence patterns, while CNN are typically employed for image recognition. Finally, among all supervised ANN, RsNN present a more complex structure with skipped connections or shortcuts, that enable a fast learning and give better performance than a plain network.

5.1.1 FFNN

For picking seismic phases on local earthquake data, deep FFNN using the direct full waveforms as input are employed in [16, 17, 22, 57]. Network in [16] uses the amplitude of the three-component input seismograms, and achieves successful results on more of the 90% of the testing data for both P and S phases. Same Authors in [17] employ a single component of the input data, and attain P and S detection performances mostly above 80% in tested cases, with strong variations according to the chosen signal component, given the influences of the ray path and source position. The input traces to the three-layer ANN in [22] were band-pass filtered for a better arrival time prediction of P waves. The detection performance is assessed in terms of different

noise levels, that consistently affects prediction accuracy as graphically shown by Authors. The technique in [57] behaves successfully for the phase identification of P waves on 95% of a testing set, that comprises 1254 seismograms of the IRIS network. The estimated onset times present an error less than 0.5 sec in 80% of the cases.

The parallel application in [51] for P wave detection implements two FFNN, which employ trace features as input data. The first one uses an extension of the standard STA/LTA triggering algorithm, developed in [2, 3], where input are given by the ratios of short- and long-term averages (STA and LTA, respectively) of normalized trace windows. The rate of correct phase identification was nearly 92%. The second network operates with spectrograms of moving trace windows, which typically present a different behavior of those from non-earthquake signals. For the same testing data set, 98% of correct detection is observed. Same Authors in [50], develop a S-phase identification and picking FFNN that includes as input attributes, autoregressive model coefficients and measures of the signal power and polarization angles. On testing results, 86% correct rate of phase identification was achieved, and 74% of them were picked with onset time errors less than 0.1 sec.

Networks developed in all aforementioned works have been trained on and applied to local and regional earthquake data. FFNN's introduced in [46] use a training data base comprising P-wave signals of 193 teleseismic events, and operate on input data consisting of STA/LTA values computed in seven frequency bands. After training, these ANN detectors found 25% more events of the official event bulletin, compared to the reference Murdock-Hutt detector [40]. Appropriate ANN weight tuning led to a reduction of the false alarm rate. In addition, [46] explores the detection capabilities of Elman-Jordan RcNN, but found poor performances.

An important contribution to the automatic onset time picking of P and S waves is the FFNN proposed in [24], a neural tree called IUANT2, that presents a problem adaptive structure. The best structure of IUANT2 is inferred during the training phase, while allowing for noise

filtering. The three-component input seismograms are preprocessed by removing trace mean and pass-band filtering, followed by the calculation of the vector magnitude of both horizontal components.

The estimation of P and S onset times are performed on statistical features, such variance, absolute skewness, kurtosis and combinations of these statistics, which are altogether computed on time windows of preprocessed traces. The testing dataset consists of more than 300 local earthquakes recorded by 23 different stations, and the picking accuracy, quantified through the normal distribution of time differences relative to manual picks, have standard deviations of 0.06 sec and 0.1 sec, for P and S waves, respectively. The accuracy of IUANT2 is then measured as the inverse of these deviation values, even though manual picks could also be erroneous.

5.1.2 RcNN

A RcNN for the real time detection of small magnitude (below 2.5 Mw) earthquakes is presented in [52]. In this case, the distribution of seismic stations in populated areas yields data with significant levels of noise. This requires a special detection method capable of recognizing small events without using preprocessing based on standard pass-band filtering. Instead, [52] applies a filter bank of STA/LTA ratios of the vertical component of seismic waveforms, with an elaborate design to keep all important frequencies of the signal, including the highest range, where disturbances are less significant. Thus, this RcNN is low prone to false detection occurrences. The initial training set comprises about 170 events including regional and teleseismic earthquakes, but the full operability was achieved after progressive tuning for real time P and S phase detection, on data from various monitoring stations. Testing was carried out during different periods of time in 2009, and relative to the standard STA/LTA method, this RcNN misses a few percentage of real events, but behaves much less sensitive to false detections, as expected. Authors found that the number of triggering stations is an important

parameter to adjust, so it is crucial for potential applications in new regions.

Recent applications of RcNN to earthquake data are given in [26, 7] with the purpose of earthquake prediction. The common architecture of a RcNN presents a set of Long Short-Term Memory (LSTM) layers, combined to dropout layers to avoid overfitting, and supply an efficient framework for detection and inference of temporal patterns. In both works, a series of past earthquakes is used for network training, that is used to predict a next trend of known events, to allow testing. Results are encouraging, although earthquake forecast remains as a highly controversial subject.

5.1.3 CNN

Among the state-of-the-art earthquake detection methods, the most popular deep CNN are probably ConvNetQuake [44], the generalized phase detection (GPD) [42], and PhaseNet [59]. However, we next also comment on a few additional and prominent works. Figure 6 illustrates the schematic structure of an automatic detection seismic system based on a fully connected ANN for classification and regression.

The raw seismic traces are usually subjected to waveform normalization (mean removal and amplitude scaling), band-pass filtering to isolate seismic phases, time splitting and windowing. Then, preprocessed data is used for network learning and validation, probably preceded by data labeling to allow supervised learning. In many cases, the network outputs reduce to class probabilities that associate window traces with P waves, S waves, and simple noise. The actual architecture of CNN mentioned below are somewhat more complex, and some cases are also capable of locating earthquakes.

ConvNetQuake is based on collection of nonlinear filters, and operates on time windows of single-station three-channel seismograms for P-wave detection and event location. Its main application has been the detection of natural or human-induced (related to waste water injection) low-magnitude earthquakes in Central Oklahoma, USA. Data preprocessing splits traces into monthly streams, applies mean subtraction and divides

each trace component by the absolute peak amplitude, to finally defines appropriate windows. For training, near to two thousand cataloged seismic events occurred during 2014-2016 were employed, and 209 were later used for testing.

Here, ConvNetQuake displays a precision of 94.8%, measured as the fraction of detected events that are true events, with 74.5% of them being correctly located on a coarse Voronoi geographic partition of the study region. Also, the recall was 100%, defined as the fraction of true events correctly detected. On an independent testing set comprising 21 earthquakes with magnitude below 4.1 Mw that occurred in Northern California, ConvNetQuake achieves 100% detection, confirmed by autocorrelation, and 74.6% location correctness.

In seismology, nonetheless, an earthquake's hypocenter corresponds to be the physical location of the starting point of the rupture process, where stored strain energy in the rock is first released. Commonly, hypocenter coordinates are given in terms of longitude, latitude and depth below surface, and those define the earthquake location. Mapping the hypocenters of foreshocks, main event, and aftershocks allows picturing the three-dimensional movement of the fault system. Thus, the event association to a cluster of different geographical areas, as performed by ConvNetQuake, represents an incomplete information to seismologists. Recent CNN based approaches aim at actually solving the location problem.

The network in [31], handles three component seismic records of multiple stations, and after training the first convolutional layer becomes sensitive to characteristic features of seismic waveforms. Thus, this layer can behave as an event detector by itself. Training employs an earthquake swarm of 2000 events recorded by nine local stations. Later, during testing, this network successfully locates 908 earthquakes with standard deviations nearly of 56 m, 124 m and 136 m, along east-west, north-south and vertical directions, respectively. Alternatively, the network in [54] operates on single-station waveforms located in Oklahoma, USA, where earthquakes

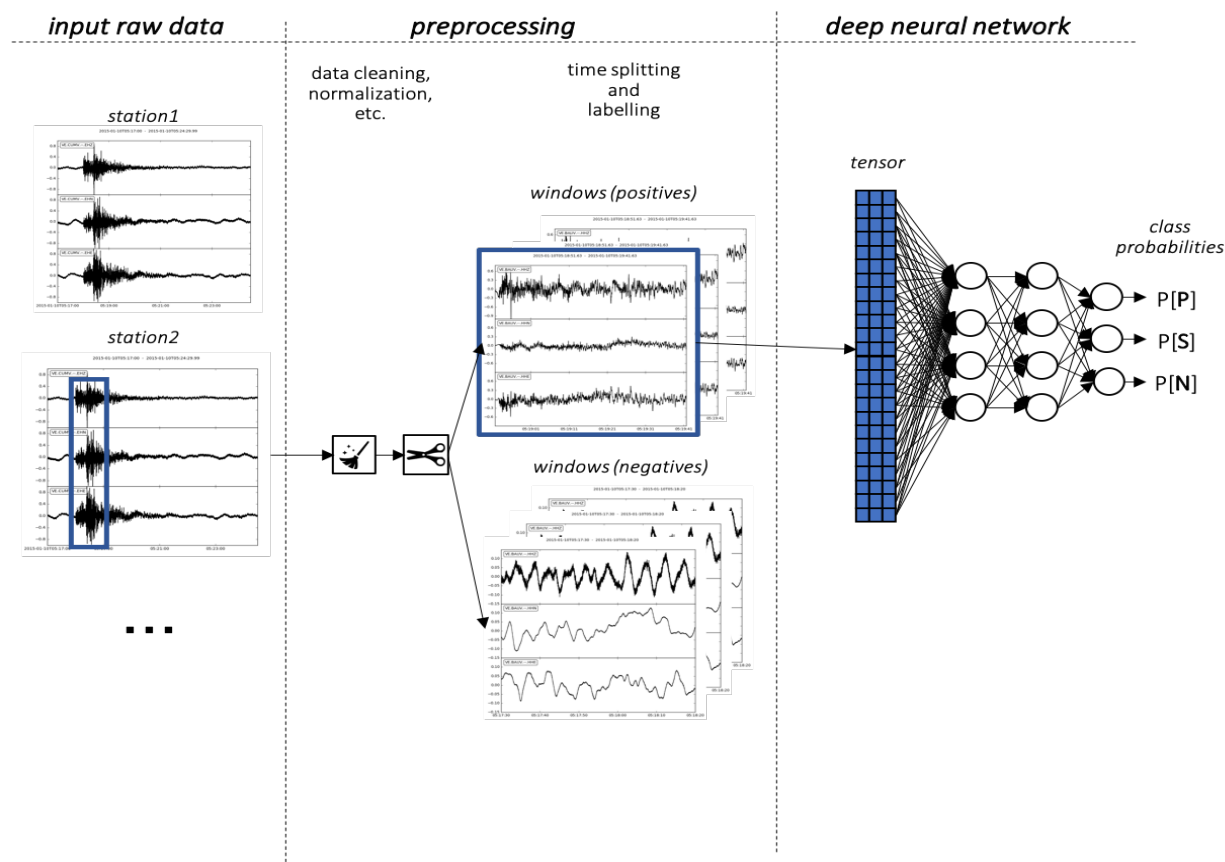


Fig. 6. Schematic structure of an automatic detection system of seismic phases based on a fully connected ANN for classification and regression

have been induced by hydrocarbon (oil and gas) production.

A training dataset consists of 1,013 historic events, and the output is a 3D probability volume of location likelihood inside the Earth. Testing was carried out using 194 earthquakes, and results present errors of approximately 4.9 km to the epicenter and 1.0 km to the hypocentral depth, on average, based on data from 30 network stations.

GPD employs a feature extraction system from seismic data, that are later used as input to a fully connected ANN, and finally outputs a classification as P waves, S waves, or just noise. The feature extraction system is a collection of preprocessing layers where seismic data is sequentially convolved with a set of digital filters

for characteristic recognition, and then decimated for down-sampling and trace evaluation at different length scales.

Both processes combined to an activation function, allows the identification of seismic phases anywhere in a seismogram, regardless their duration and amplitude. The final stage of ANN classification outputs probabilities of the likelihood of each possible class (P, S or noise). For training and validation, GPD employs 4.5 millions of four second windows, where P and S waves correspond to events of magnitude below 5.7 Mw, recorded by the Southern California Seismic Network. The validation precision is nearly 99% for both phases and various detection probability thresholds, and recall is somewhat lower, between

96% and 99% for most threshold choices, suggesting a minor number of misclassifications of seismic phases as noise. GDP detection has been also validated on the 2016 Bombay Beach, California swarm of small and moderate (≤ 4.8 Mw) earthquakes, and with data of the 2016 Mw 7.0 Kumamoto earthquake recorded at multiple stations, within 100 km of the hypocenter. Highly successful results were shown for both cases.

PhaseNet operates on three-component unfiltered seismograms, and estimates P- and S-wave arrival times by generating probability time series that quantify the likelihood of P- and S-wave onsets. Its exhaustive training and validation make use of the extremely large dataset of analyst-labelled P and S arrival times from the Northern California Earthquake Data Center. Specifically, Authors employ stratified sampling based on stations and divide data into a training, validation and test sets, with more than 600K, 77K and 78K waveform samples, respectively.

A minimum preprocessing is applied to the training data, where each data component is normalized by removing its mean and dividing it by the standard deviation. Testing results are compared to those obtained by the standard *AR picker* algorithm [1], and PhaseNet outperforms this scheme with significant improvements on both phases, particularly for the S waves.

A densely connected CNN to capture laboratory slip events of different durations is given in [53]. This network presents a cascaded architecture to generate multi-scale slip proposals and detect events with various lengths. It also exploits atrous convolutions with different dilation rates to enrich feature extraction.

Training, validation and testing proceed using acoustic data acquired at the Rock and Sediment Mechanics Laboratory of Penn State University, that presents a thousand manually picked events, and 800 are taken for training. Relative to a well established template matching algorithm, Authors report a detection accuracy significantly higher, especially for highly different subsequent events.

5.1.4 RsNN

Residual networks are usually a combination of the three previous types of supervised ANN, and have achieved higher accuracy when solving same problems already tackled by one of them, or even dealing with more complex problems. For earthquake detection, the RsNN proposed in [38] and named CRED, first employs a CNN for feature extraction, that are later used by a RcNN for recognition of temporal patterns. The next and last stage of this network is a fully connected FFNN used for phase identification. Training and validation is based on a dataset of 550K three component seismograms recorded by 889 broadband and short-period stations in North California, and P-wave and S-wave arrivals have been labelled by manual picking. Traces of seismic noise are also records from the same network stations.

A first test on 50K waveform samples, yields detection precision higher than 96% and a recall above 99%, regardless the threshold choice on the output probabilities. Further testing for microseismic detection was undertaken in Central Arkansas, with a substantially different crustal structure, and events of lower magnitude with shorter epicentral distances. Without additional training, CRED finds 3 orders of magnitude more events compared with STA/LTA results. Comparisons against the FAST algorithm [55], reveal that the CRED detection rate is much lower, but it takes less than a hundredth of the computation time spent by FAST (non-parallel version).

5.2 Unsupervised

In this final section, we discuss few applications of unsupervised ANN to seismic event detection. In general, ANN of this kind are classified as autoencoders (AE) and generative adversarial networks (GAN).

5.2.1 AE

An autoencoder is a deep learning approach with noticeable advantages over most other ML techniques, especially when processing unlabelled data. This unsupervised technique is able to directly process the input raw data, with no need for preprocessing and previous labeling. A main application of AE is data compression, aiming at extracting representative features of the input and removing redundancies, so the overall size of the data may be highly reduced [33], [36]. This concept is called Dimensionality Reduction. Denoising AE (DAE) are also trained locally to mitigate noise on corrupted versions of the input data, with a consecutive step of reconstruction a cleaner version of the data, even under the presence of high noise levels. These networks are known as stacked DAE's (SDAE) [49]. In [43], a SDAE is developed to mitigate background noise on seismic traces, to later pick the P wave onset times on the cleaner waveforms. For evaluation, both synthetic and field seismic data are employed, and the SDAE algorithm accurately picks the onset time of 94.1% for 407 of the field seismograms, with a standard deviation error of 0.10 s. Results also indicates that this algorithm can make precise arrival time inferences on data with SNR as lows as -14 dB.

A hybrid network, that combines a SAE and a deep-belief ANN, has been applied to classify seven different classes of volcano-seismic events in [47]. Classification results validate the efficient capturing of complex relationships on volcano-seismic data by this new network, and compared to those delivered by other ML algorithms, the performance is higher under faster convergence.

5.2.2 GANs

The GAN proposed in [35] is trained to learn the characteristics of both earthquake first arrival P waves and background noise, resulting in a discriminator that mitigates false triggering. Training employs 300K seismic records from southern California and Japan to act as an automatic feature extractor, and then a Random Forest classifier is subjected to a secondary

learning on a large set of earthquake signals and noise waveforms. When tested, the recognition performance was about 99.2% for P waves, and 98.4% for signals of pure noise. This excellent performance is accompanied by a very low sensitivity to false triggers from local impulsive noise.

6 Discussion and Conclusions

Detection of earthquake signals is fundamental for observational seismology. Main features of a reliable detection algorithm of seismic waves must include: high sensitivity to small magnitude events with variable trace patterns, low sensitivity to ambient noise and non-earthquake signals, enough flexibility to account for data from one or multiple seismic stations, and high computational efficiency for fast real time assistance or large dataset processing. These features are naturally fulfilled by Artificial Neural Networks (ANN). ANN, among few other machine learning techniques, offer advanced pattern recognition abilities that allow matching the detection performance of experienced human analysts, after being appropriate trained.

During the mid-late 90s, training of pioneer ANN proceeds on subsets of hundreds or few thousands of seismic traces, but the paramount advances of computational platforms and ANN implementation frameworks afterwards, have allowed using regional-wide earthquake catalogs with hundreds of thousands or even millions traces, during the ANN learning over last years. Thus, training of state-of-the-art ANN really accounts for multiple station information, seismic and noise signals with different shapes, local and non local earthquake data, and labeling (in case it is required) from various human experts. Because of different analysts may pick seismic phases in a dissimilar way, such a modern well-trained ANN become practically free of any human bias or likely errors.

Inspired on the basic time and frequency processing of seismic traces undertaken by human analysts, usually assisted by interpretation software, the input data to most of detecting ANN are not raw seismograms. A big variety of ANN operate on preprocessed traces, after a standard normalization (with optional filtering)

and windowing, but other specimens only treat previously extracted statistical features, or even more specialized data representations such as spectrograms.

Developments before 2010 were mainly FFNN, mostly applied to local and regional earthquake data, although few cases target teleseismic events. For historical reasons, this survey paper cited few prominent works in that period, but the contribution in [24] pushed forward the feature-based trace processing, detection for single or multi station information, and the performance assessment as well. In current decade RcNN, with an inherent efficiency for recognition and inference of temporal patterns, have been used for real-time small-event monitoring under poor signal-to-noise ratios, and even have found successful applications in early warning systems and earthquake forecasts. During last two years, CNN and RsNN have taken the automatic seismic phase identification and onset estimation to a new level of regional-wide training and operation, with few cases also capable of seismological hypocentral location. With special emphasis, we here discussed contributions in [38], [44], [42] and [59].

A performance assessment among different algorithms for automatic phase detection and picking is rather difficult because of the specialized target application, the variety of accuracy metrics used, and the case-dependent testing dataset employed by each algorithm. For a systematic evaluation of the accuracy in phase onset pickings, for instance in [24].

Authors use the standard deviation of the time differences between the automatic and manual picks. This metric tends to be insensitive to the size of the testing dataset. They also employ the metrics of *recall* and *precision* to measure the ability of the system to detect the correct onset of seismic waves and to reject false alarms, considering the total number of reference picks. These metrics were rapidly adopted by others studies, and few others emerged. Because of the highly active developments of ANN, among others ML and general algorithms, in this area, there is a pressing need for establishing standardized benchmark datasets, of several sizes and SNR

levels, to facilitate full assessment with clear quality measures.

Acknowledgements

This work is partially supported by the Spanish Ministry of Economy and Competitiveness under contract TIN2015-65316-P, by the SGR programmes (2014-SGR-1051 and 2017-SGR- 962) of the Catalan Government and has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 777778 (MATHROCKS).

We thank FUNVISIS for providing the seismic data subject of our current studies.

References

1. **Akazawa, T. (2004).** *A technique for automatic detection of onset time of P and S phases in strong motion records.*
2. **Allen, R. (1978).** Automatic earthquake recognition and timing from single traces. *Bulletin of the Seismological Society of America*, Vol. 68, No. 5, pp. 1521–1532.
3. **Allen, R. (1982).** Automatic phase pickers: Their present use and future prospects. *Bulletin of the Seismological Society of America*, Vol. 72, No. 6B, pp. S225–S242.
4. **Anderson, J. & Brune, J. (1999).** Probabilistic seismic hazard analysis without the ergodic assumption. *Seismological Research Letters*, Vol. 70, No. 1, pp. 19–28.
5. **Araya, M., Jennings, J., Adler, A., & Dahlke, T. (2018).** Deep-learning tomography. *The Leading Edge*, Vol. 37, pp. 58–66.
6. **Beyreuther, M., Hammer, C., Wassermann, J., Ohrnberger, M., & Megies, T. (2012).** Constructing a hidden Markov model based earthquake detector: Application to induced seismicity. *Geophysical Journal International*, Vol. 189, No. 1, pp. 602–610.
7. **Bhandarkar, T., Vardaan, K., Satish, N., Sridhar, S., Sivakumar, R., & Ghosh, S. (2019).** Earthquake trend prediction using long short-term memory RNN. *International Journal of Electrical and Computer Engineering (IJECE)*, Vol. 9, pp. 1304–1312.

8. **B.Hassani & Atkinson, G. (2015).** Referenced empirical ground-motion model for Eastern North America. *Seismological Research Letters*, Vol. 86, pp. 477–491.
9. **Bianco, M. J. & Gerstoft, P. (2018).** Travel time tomography with adaptive dictionaries. *IEEE Transactions on Computational Imaging*, Vol. 4, No. 4, pp. 499–511.
10. **Bishop, C. M. (2006).** *Pattern Recognition and Machine Learning*. Springer Verlag, New York.
11. **Bormann, P., Klinge, K., & Wendt, S. (2014).** Data analysis and seismogram interpretation. In **Bormann, P.**, editor, *New Manual of Seismological Observatory Practice 2 (NMSOP-2)*. Deutsches GeoForschungsZentrum GFZ, Potsdam, pp. 1–126.
12. **Böse, M., Heaton, T. H., & Hauksson, E. (2012).** Real-time finite fault rupture detector (FinDer) for large earthquakes. *Geophysical Journal International*, Vol. 191, No. 2, pp. 803–812.
13. **Bose, M., Wenzel, F., & Erdik, M. (2008).** PreSEIS: A neural network-based approach to earthquake early warning for finite faults. *Bulletin of the Seismological Society of America*, Vol. 98.
14. **Chen, Y. (2018).** Automatic microseismic event picking via unsupervised machine learning. *Geophysical Journal International*, Vol. 212, No. 1, pp. 88–102.
15. **Cua, G. & Heaton, T. (2007).** The virtual seismologist (vs) method: A bayesian approach to earthquake early warning. In **Gasparini, P., Manfredi, G., & Zschau, J.**, editors, *Earthquake Early Warning Systems*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 97–132.
16. **Dai, H. & MacBeth, C. (1995).** Automatic picking of seismic arrivals in local earthquake data using an artificial neural network. *Geophysical Journal International*, Vol. 128, No. 3, pp. 758–774.
17. **Dai, H. & MacBeth, C. (1997).** The application of back-propagation neural network to automatic picking seismic arrivals from single-component recordings. *Journal of Geophysical Research: Solid Earth*, Vol. 102, No. B7, pp. 15105–15113.
18. **der Baan, M. V. & Jutten, C. (2000).** Neural networks in geophysical applications. *Geophysics*, Vol. 65, No. 4, pp. 1032–1047.
19. **Derras, B., Bard, P., & Cotton, F. (2014).** Towards fully data driven ground-motion prediction models for Europe. *Bulletin of Earthquake Engineering*, Vol. 12, No. 1, pp. 495–516.
20. **Derras, B., Bard, P., Cotton, F., & Bekkouche, A. (2012).** Adapting the neural network approach to PGA prediction: An example based on the KiK-net data. *Bulletin of the Seismological Society of America*, Vol. 102, pp. 1446–1461.
21. **Diehl, T., E, K., & Bormann, P. (2012).** Tutorial for consistent phase picking at local to regional distances. In *New Manual of Seismological Observatory Practice 2 (NMSOP-2)*. Deutsches GeoForschungsZentrum GFZ, pp. 1–21.
22. **Enescu, N. (1996).** Seismic data processing using nonlinear prediction and neural networks. *IEEE NORSIG Symposium, Espoo, Finland*.
23. **Florida, E., Aznarte, J., Morales-Esteban, A., & MartÁnez-Álvarez, F. (2016).** Earthquake magnitude prediction based on artificial neural networks: A survey. *Croatian Operational Research CROORS*, Vol. 7, No. 2, pp. 159–169.
24. **Gentili, S. & Michelini, A. (2006).** Automatic picking of p and s phases using a neural tree. *Journal of Seismology*, Vol. 10, pp. 39–63.
25. **Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016).** *Deep Learning*. MIT Press, Cambridge, Massachusetts.
26. **Ibrahim, M. A., Park, J., & Athens, N. (2018).** Earthquake warning system: Detecting earthquake precursor signals using deep neural networks.
27. **Iturrarán-Viveros, U., Munoz-García, A. M., Parra, J. O., & Josué, T. (2018).** Validated artificial neural networks in determining petrophysical properties: A case study from colombia. *Interpretation*, Vol. 6, No. 4, pp. T1067–T1080.
28. **Khosravikia, F., Zeinali, Y., Nagy, Z., Clayton, P., & Rathje, E. (2018).** Neural network-based equations for predicting pga and pgv in texas, oklahoma, and kansas. pp. 1–12.
29. **Kislov, K. & Gravirov, V. (2018).** Deep artificial neural networks as a tool for the analysis of seismic data. *Seismic Instruments*, Vol. 54, No. 1, pp. 8–16.
30. **Kong, Q., Allen, R., Schreier, L., & Kwon, Y.-W. (2016).** Myshake: A smartphone seismic network for earthquake early warning and beyond. *Science Advances*, Vol. e1501055, No. 2, pp. 1–8.
31. **Kriegerowski, M., Petersen, G., Vasyura-Bathke, H., & Ohrnberger, M. (2018).** A deep convolutional neural network for localization of clustered earthquakes based on multistation full waveforms. *Seismological Research Letters*.

32. **Kulhánek, O. (1990).** Anatomy of seismograms. *Developments in Solid Earth Geophysics*, Vol. 18.
33. **LeCun, Y., Bengio, Y., & Hinton, G. (2015).** Deep learning. *Nature*, Vol. 521, pp. 436–444.
34. **Lewis, W. & Vigh, D., .** Deep learning prior models from seismic images for full-waveform inversion. pp. 1512–1517.
35. **Li, Z., Meier, M., Hauksson, E., Zhan, Z., & Andrews, J. (2018).** Machine learning seismic wave discrimination: Application to earthquake early warning. *Geophysical Research Letters*, Vol. 45, No. 10, pp. 4773–4779.
36. **Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y., & Alsaadi, F. E. (2017).** A survey of deep neural network architectures and their applications. *Neurocomputing*, Vol. 234, pp. 11–26.
37. **Mosser, L., Dubrule, O., & Blunt, M. (2018).** Stochastic seismic waveform inversion using generative adversarial networks as a geological prior. *arXiv e-prints*, pp. arXiv:1806.03720.
38. **Mousavi, S. M., Zhu, W., Sheng, Y., & Beroza, G. C. (2018).** Cred: A deep residual network of convolutional and recurrent units for earthquake signal detection. *CoRR*, Vol. abs/1810.01965.
39. **Moya, A. & Irikura, K. (2010).** Inversion of a velocity model using artificial neural networks. *Computers & Geosciences*, Vol. 36, pp. 1474–1483.
40. **Murdock, J. & Hutt, C. R. (1983).** A new event detector designed for the seismic research observatories. <http://pubs.er.usgs.gov/publication/ofr83785>.
41. **Murphy, K. P. (2012).** *Machine Learning: A Probabilistic Perspective*. MIT Press, Cambridge, Massachusetts.
42. **Ross, Z. E., Meier, M.-A., Hauksson, E., & Heaton, T. H. (2018).** Generalized seismic phase detection with deep learning. *Bulletin of the Seismological Society of America*.
43. **Saad, O. M., Inoue, K., Shalaby, A., Samy, L., & Sayed, M. S. (2018).** Automatic arrival time detection for earthquakes based on stacked denoising autoencoder. *IEEE Geoscience and Remote Sensing Letters*, Vol. 15, No. 11, pp. 1687–1691.
44. **Thibaut, P., Michaél, G., & Marine, D. (2018).** Convolutional neural network for earthquake detection and location. *Science Advances*, Vol. 4, No. 2.
45. **Thomas, S., Pillai, G., Pal, K., & Jagtap, P. (2016).** Prediction of ground motion parameters using randomized anfis (ranfis). *Applied Soft Computing*, Vol. 40, No. C, pp. 624–634.
46. **Tiira, T. (1999).** Detecting teleseismic events using artificial neural networks. *Computers & Geosciences*, Vol. 25, pp. 929–938.
47. **Titos, M., Bueno, A., Garc a, L., & Ben atez, C. (2018).** A deep neural networks approach to automatic recognition systems for volcano-seismic events. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 11, No. 5, pp. 1533–1544.
48. **Trugman, D. & Shearer, P. (2018).** Strong correlation between stress drop and peak ground acceleration for recent m 1-4 earthquakes in the san francisco bay area. *Bulletin of the Seismological Society of America*, Vol. 108.
49. **Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., & Manzagol, P.-A. (2010).** Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal of Machine Learning Research*, Vol. 11, No. 12, pp. 3371–3408.
50. **Wang, J. & liang Teng, T. (1997).** Identification and picking of s phase using an artificial neural network. *Bulletin of the Seismological Society of America*, Vol. 87, No. 5, pp. 1140–1149.
51. **Wang, J. & Teng, T.-L. (1995).** Artificial neural network-based seismic detector. *Bulletin of the Seismological Society of America*, Vol. 85, No. 1, pp. 308–319.
52. **Wiszniowski, J., Plesiewicz, B. M., & Trojanowski, J. (2014).** Application of real time recurrent neural network for detection of small natural earthquakes in poland. *Acta Geophys*, Vol. 62, No. 3, pp. 469–485.
53. **Wu, Y., Lin, Y., Zhou, Z., Bolton, D. C., Liu, J., & Johnson, P. (2018).** Deepdetect: A cascaded region-based densely connected network for seismic event detection. *IEEE Geoscience and Remote Sensing Society*, pp. 1–14.
54. **Xiong, Z., Jie, Z., Congcong, Y., Sen, L., Zhibo, C., & Weiping, L. (2018).** Locating earthquakes with a network of seismic stations via a deep learning method.
55. **Yoon, C., O'Reilly, O., Bergen, K., & Beroza, G. (2015).** Earthquake detection through computationally efficient similarity search. *Science Advances*, Vol. 1, pp. e1501057–e1501057.
56. **Zhang, J., Zhang, H., Chen, E., Zheng, Y., Kuang, W., & Zhang, X. (2014).** Real-time earthquake monitoring using a search engine method. *Nature Communications*, Vol. 5664, No. 5, pp. 1–9.

57. **Zhao, Y. & Takano, K. (1999).** An artificial neural network approach for broadband seismic phase picking. *Bulletin of the Seismological Society of America*, Vol. 89, pp. 670–680.
58. **Zhu, L., Liu, E., & McClellan, J. (2017).** Sparse-promoting full-waveform inversion based on online orthonormal dictionary learning. *Geophysics*, Vol. 82, No. 2, pp. R87–R107.
59. **Zhu, W. & Beroza, G. C. (2018).** Phasenet: A deep-neural-network-based seismic arrival time picking method. <http://arxiv.org/abs/1803.03211v1>.

*Article received on 15/10/2018; accepted on 10/02/2019.
Corresponding author is Otilio Rojas.*