

Article

Proposal for a System Model for Offline Seismic Event Detection in Colombia

Julián Miranda ¹, Angélica Flórez ^{1,*}, Gustavo Ospina ², Ciro Gamboa ³, Carlos Flórez ⁴
and Miguel Altuve ³

¹ Faculty of Systems and Informatics Engineering, Pontifical Bolivarian University, Bucaramanga 680006, Colombia; julian.miranda@upb.edu.co

² Faculty of Civil Engineering, Pontifical Bolivarian University, Bucaramanga 680006, Colombia; gustavo.ospina@upb.edu.co

³ Faculty of Electrical and Electronic Engineering, Pontifical Bolivarian University, Bucaramanga 680006, Colombia; ciro.gamboa.2013@upb.edu.co (C.G.); miguel.altuve@ieee.org (M.A.)

⁴ Faculty of Civil Engineering, Francisco de Paula Santander University, Cúcuta 540003, Colombia; carloshumbertofg@ufps.edu.co

* Correspondence: angelica.florez@upb.edu.co

Received: 12 November 2020; Accepted: 30 November 2020; Published: 18 December 2020



Abstract: This paper presents an integrated model for seismic events detection in Colombia using machine learning techniques. Machine learning is used to identify P-wave windows in historic records and hence detect seismic events. The proposed model has five modules that group the basic detection system procedures: the seeking, gathering, and storage seismic data module, the reading of seismic records module, the analysis of seismological stations module, the sample selection module, and the classification process module. An explanation of each module is given in conjunction with practical recommendations for its implementation. The resulting model allows understanding the integration of the phases required for the design and development of an offline seismic event detection system.

Keywords: seismic event detection; detection model; seismology; classification

1. Introduction

Earthquakes have been one major concern to societies around the world. Earthquakes are a consequence of earth tectonics, which cause intercontinental plate drifts. Deformation energy is stored along the plates. Once one or more fault lines exhaust their elastic deformation capacity, rupture occurs and the stored energy is released as seismic waves, propagating along with the earth's crust. Depending on the amount of energy released and the depth of rupture, seismic waves can hit civil infrastructure, causing major impacts. Such events as in Sumatra (Indonesia, 2004), Haiti (2010), and Tohoku (Japan, 2011) are proof of how devastating can earthquakes be over human infrastructure and society as well. In Colombia, the Armenia earthquake (1999, Mw 6.2) is referenced as the worst seismic event for the country, which forced the government to call for an updating of the existing design and construction code. The resulting document was introduced into the Colombian legislation, making it a mandatory practice among civil infrastructure designers and constructors [1,2].

Earthquake engineering is a branch of engineering born to reduce the effects of earth seismicity. The approaches that earthquake engineering take can be seen from two perspectives: a study of the seismic phenomena and a study of the structural response after the seismic event. Research in earthquake engineering has increased in depth as new materials and computational power have been conceived. Simple techniques for the characterization of earthquake events can be used intensively in an attempt

to formulate methodologies that provide people with a time frame to evacuate civil infrastructure during significant seismic events. However, current methodologies involving computational power deal with limitations in storage capacity (storage of seismic traces and raw data), processing capabilities (multichannel seismic acquisition), and lack of compatibility and integration of software resources, adding difficulties for the implementation of a successful seismic event detection technique [3–5].

Few academic research groups in Colombia, including the Colombian Geological Service, dedicate their efforts to boost techniques and methodologies focused on the seismic phenomena. Most of the research efforts aim toward a better understanding of both site seismic and structural responses, while some research has been carried on the understanding of local seismicity. Countries such as Mexico and Chile, both with similar seismic characteristics as Colombia, dedicate their research efforts to improve cities' structural resilience and to improve the social response during seismic events as well. The proposal of this research paper is a system model for offline seismic event detection in Colombia, where a set of integrated modules for reading and processing historical seismic raw data deals with the reduction of the computational costs for the successful detection of seismic events.

This article structures the proposal of a machine learning-based model for the detection of seismic events as follows: (a) problem statement (seismology and seismic data recording), (b) earthquake detection methodologies (traditional vs. current approaches), (c) seismic detection model proposal (model architecture and modules description), and (d) article conclusions.

2. Problem Statement

Countries over the Pacific coast of Southern America have a long history of catastrophic earthquake events. According to the US Geological Survey, five of the top 20 largest earthquake events occurred since the earlier 1900s, including the largest, have occurred along the fault line traced by the borders of the Nazca Plate that subducts below the South American Plate [6]. The tectonic environment in Colombia can be described by its two main fault zones: (a) Romeral zone (intraplate seismic zone that runs from north to south of the country's Pacific coast with an approximate length of 1200 km) and (b) the frontal fault of Eastern Cordillera (fault system that divides the Colombian Andean territory from its eastern great plains, most likely a southern border of the Caribbean Plate) [7]. One local seismic zone on the Colombian northeast of great activity is the Bucaramanga Seismic Nest, where at least eight events with a magnitude $M_w > 4.7$ occur each year [8].

Typical geological and seismic observation services such as SGC in Colombia provide seismic analysis in a two-stage fashion: first, by acquiring and storing seismic records (which can include up to three spatial components of accelerations, velocities, and ground displacements) and second, by performing seismic event recognition by looking for particular seismic characteristics within the stored data. Then, the geological service within a short time frame reports the occurrence of a seismic event, which is usually information that commonly contains the event magnitude and its approximate geolocation. [9–11]. This two-step procedure is complex, since it involves algorithms to read, synchronize, and process seismic data information. Geological services usually rely on black-box software that performs these tasks, closing the door to monitor sub-stages and therefore not letting the user integrate alternative algorithms that could eventually improve and/or fit specific site characteristics to the seismic data analysis sub-stage [12,13].

To establish methods for the detection and analysis of seismic events, the disposition of a set of historical seismic records that can be stored, read, and processed is essential to develop an accurate detection of future events (classic approach of learning from data). However, it is difficult to find a seismic dataset that fits the requirements for later processing stages, and when retrieved, the seismic files are not easy to interpret, as they include specific seismic parameters contained in legible formats that only specialized software such as SEISAN and SeisComP can process [14,15].

Moreover, several factors make an integral analysis of these Colombian historical records unfeasible: (a) limitations on the storage capacity, (b) limitations on the compatibility between current software resources, (c) limitations on the processing power required, (d) use of techniques that are not integrated

within the detection models, and (e) low flexibility of the existing tools for modifications and adequations of storage and processing algorithms [16,17].

Furthermore, the online detection of coming earthquakes can be done by picking the seismic phases manually or by establishing a fixed-threshold approach; these techniques are statistically earthquake-proven for significant earthquakes and signals with low sampling rates and few numbers of components, given a higher signal-to-noise ratio. When high samples rates are considered from multiple three-dimensional seismological stations, the phases may be picked differently, introducing bias into the detection [18].

In this sense, a system model for offline seismic event detection for the Colombian region is proposed, which allows the identification of patterns and dynamics in historical records, using machine learning techniques.

The following sections present a theoretical basis and the description of the model for seismic event detection.

3. Earthquake Detection Methodologies

Seismic detection algorithms are used by public and/or private services dedicated to monitor and study seismic activity. Several agencies dedicate efforts to maintain an updated database of information that can help scientists and engineers analyze any activity that could represent a hazard to the infrastructure and population, including volcanic and seismic activities [19]. Data collected include ground motion records (accelerations, velocities, and/or displacements), which are used by detection algorithms as input data.

Several approaches have been conceived to perform seismic events detection. In the seismic signal, amplitude, shape, power, or several other time-domain characteristics can be used to formulate a detection procedure [19,20], depending on the desired purpose of the outcome. In practical terms, seismic signals are identified by monitoring isolated ground vibrations, which under changes in amplitude, frequency content, or motion direction indicate the arrival of seismic waves [20]. Current developments on earthquake signals monitoring aim to provide faster and more reliable detection algorithms for warning systems [21].

3.1. Traditional Approaches

Detection algorithms for earthquake detection assume that seismic signals correspond to ground vibrations isolated from human activity. Only stationary background noise is registered prior to earthquake waves' arrival. To automate the process of identifying the arrival of earthquake waves, specialized detection algorithms are required. These algorithms deal with the task of effectively discriminating background noise from seismic events, to avoid the recording of unnecessary data or the loss of actual seismic signals. Then, detection algorithms require having a high rate of positive event identification, which is easily achieved when strong motions occur (e.g., triggers such as signal's threshold can discriminate noise from strong seismic signals). However, if a seismic event is detected far from the causative fault, a decrease in the signal's amplitude can be expected, making it harder for the algorithm to perform a positive event detection.

The simplest approach for an earthquake detection system is a front detection system, which consists of the direct monitoring of a given seismic source. Monitoring the signal's amplitude allows a central managing system to perform pre-defined tasks such as shutting down the power on certain areas or generating alerts of populated areas far enough from the strong ground motion epicenter. Mexico's earthquake monitoring system (SASMEX) implements front detection for this purpose. The front detection approach requires the analysis of most of the seismic signal to validate the trigger, dismissing valuable time that can be used to alert a wider area in case of a strong ground motion event. To deal with this, further approaches attempt to wider the alert time window by analyzing a shorter segment of the seismic signal, requiring more elaborated metrics that can be positively correlated to a significant earthquake event. Some of those metrics include the average noise level, predominant signal period,

or cumulative energy [22]. However, these algorithms have a high rate of false alarms when dealing with weak-motion earthquake events [23]. Detection triggers are also specialized to work with frequency-domain data. In this case, signal energy metrics are used as thresholds (e.g., average power). Transform methods such as Fourier or Walsh and signal filtering have been used to provide faster and more reliable detection algorithms [5]. Time-frequency domain techniques such as the wavelet transform have been used to track the initiation of ground motion [24,25]. A technique found more reliable and widely used is based on the short-time average through the long-time average ratio (STA/LTA). The technique is based on the fact that when seismic events occur, the current signal average (STA) is different from the long-term signal average (LTA) where no events occurred [5,26].

Figure 1 shows the implementation of the STA/LTA algorithm over a strong-motion record. The seismic record (top figure) shows the arrival of P-waves in the interval 5–10 s. P-waves (P for primary) travel across the earth’s mantle in tension–compression mode. Rocks have their highest stiffness (force to deformation ratio) for compression forces, and thus, compression waves can travel the fastest across the earth’s mantle, arriving at the surface prior to secondary waves. After 10 s, the seismogram on the top figure shows the strongest acceleration recorded by the seismic station for the event. Secondary and surface waves arrive at the seismic station seconds later than P-waves. Cities with poor seismic resilient infrastructure usually take the highest toll on human and economic losses when they experience strong ground accelerations.

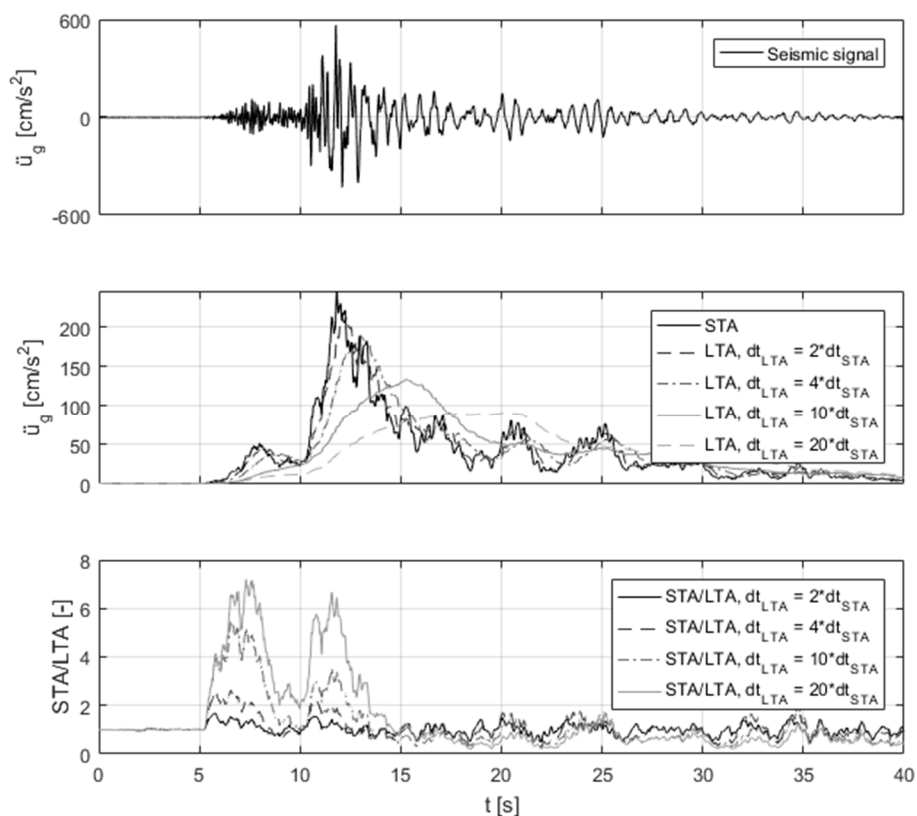


Figure 1. Short-time average through the long-time average ratio (STA/LTA) algorithm.

Figure 1 (middle) shows the STA and LTA parameters. It can be seen that the short-time average is useful to indicate the arrival of the seismic event as it suddenly arises from a very low value (assumed to be noise). The long-time average is less sensitive to the arrival of the seismic signal but keeps track of the signal’s duration. The STA/LTA ratio (Figure 1, bottom) points out the location of the seismic event’s start point, which is one of the most important features required by any detection methodology.

The STA/LTA algorithm has shown to be very effective due to its simplicity [5,27,28], but it requires the optimization of user-defined parameters to obtain a high rate of positive-event detection.

Parameters such as the sampling rate, detection threshold, or even, pre-event and post-event parameters are required to achieve a desired positive detection with the algorithm.

3.2. Current Approaches

Actual developments of detection algorithms take advantage of current technological advances that allow the capture, processing, and storage of data with high resolution. A large amount of available data today is used for advanced and still rarely used techniques: local similarity (quantifies consistency of data between neighboring stations) [29], probability (parameters such as distance to the seismogenic zone or signal phases are treated as random variables with an associated probability) [30], data mining (establishes a fingerprint of seismic waves for later comparison) [31], neural networks (neural network-based algorithms are trained to identify several characteristics of seismic waveforms) [19], and social sensing (based on trending hashtags or key words on social media, algorithms can trigger responses on alert systems) [32–34]. These approaches are all computationally powerful and help identify waveforms on large historic seismic arrays that were not processed so far or that could have been processed by techniques with less accuracy. The cost of these techniques is the computational time. In terms of computational efficiency, the STA/LTA concept [26,35] arises as a traditional and yet highly efficient parameter for earthquake detection.

4. Seismic Detection Model Proposal

This research paper proposes the model presented in Figure 2—a set of modules in which seismic signals can be processed—from the seismic data collection to the detection mechanism expressed in the classification process module. The applicability of the model is directly related to the selection of the geographical zone whose seismicity is to be studied—in this case, the northeastern region of Colombian. The size of the study area, the rate of occurrence of events, and the homogeneity of the subsoil are some of the variables that directly influence the number of observations to be analyzed and the performance of each of the modules, which is why a careful selection of the region of interest must be first carried out. Applicability of the model on a different geographical region would require historic seismic arrays specific to the location.

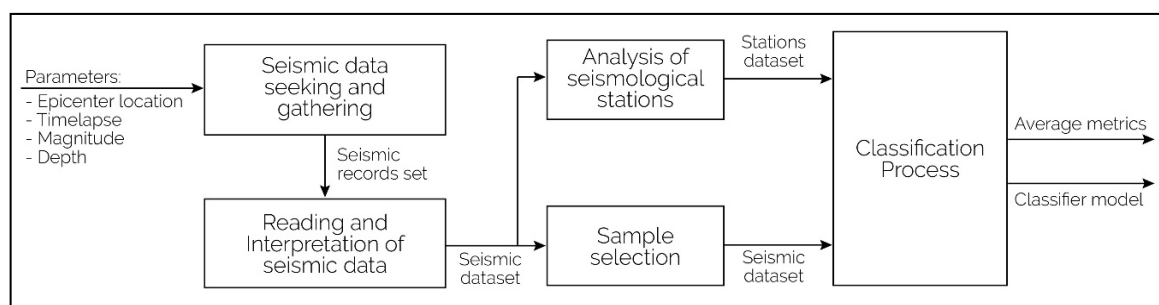


Figure 2. Model for offline seismic event detection.

The modules are described below as follows: Seismic data seeking and gathering, Reading and interpretation of the seismic data, Analysis of seismological stations, Sample selection, and Classification process.

4.1. Seismic Data Seeking and Gathering

To analyze seismic signals by algorithmic means, it is necessary to have a set of records that can be manipulated. In this first step, a download of the historical seismic records is made, filtering by the chosen region. A set of seismological records of each station is obtained.

Depending on the region of interest, it is possible to obtain seismological records through web services, such as the United States Geological Survey (USGS), the European-Mediterranean

Seismological Centre (EMSC), the National Earthquake Information Center (NEIC), the National Institute of Seismology, Volcanology, Meteorology, and Hydrology of Guatemala (INSIVUMEH), the Mexican National Seismological Service (SSN), the National Seismological Center of Chile (CSN), or the Colombian National Seismological Network (RSNC by its Spanish acronym), among others. These web services allow downloading data of seismic events one by one, although it is usually necessary to provide searching filters, generally concerning magnitudes, depths, and dates.

The seismic records are usually of public access, so it is possible to request the set of desired data directly to the seismology agency responsible for its storage. Another easier way to do this data-gathering process is to use automated interaction web tools to download the requested files, such as the creation of web snippets using the web-scraping technique, which allows interacting with the web resources of the web service. This technique is usually legally authorized by the RSNC and other services, since all the information downloaded is public access and no intrusion for non-authorized domains or web resources is made. Although the scraping procedure is legally accepted, it is recommended to inform the geological services about this practice when executed.

The technological infrastructure needed to download and to store seismic files depends on the volume of data to be processed. In Colombia, the RSNC gathers the seismic records into two categories:

- Trace files (Waveforms), which contain the seismic samples taken by all seismological stations available around the region of interest.
- Parameter files (Sfiles), which provide detailed information about the seismic events, such as the longitude and latitude of the epicenter and the P-wave and the S-wave arrival times, among others.

The seismic traces recorded in the Waveform files are usually a large size because they record non-event samples that occur before and after the seismic event picking. Their content is dependent on the duration of the recorded earthquake. Each trace file can have a storage size from approximately 5 MB if it corresponds to a microseism that has been registered by one or few seismological stations, or a specific seismic event registered by a couple of stations, and up to approximately 120 MB, if it is registered by most seismological stations with a duration close to five to ten minutes. The RSNC registers up to 10,000 seismic events per year, with an average of 60 MB of storage per seismic record.

The seismic records may not be stored completely. The storage of all the records allows faster access for later processing stages; however, as has already been shown, the computational load applied to the data storage is high. On the other hand, storing portions of data that are processed and then erasing the unrequired portions, or processing the records one by one so that the results are stored and the records are erased are two recommended procedures to save storage space. Nevertheless, if the data are processed one by one, any subsequent processes to be done or corrections to previous processes will force to access the data sources again, which will hinder processing.

4.2. Reading and Interpretation of the Seismic Records

Once the seismic records are obtained, it is necessary to understand the format in which the data are presented, to establish the mechanisms by which they will be read. After reading the data, a sample selection is proposed, which depends on the characteristics found, and a set of memory instances that represent the trace files and read parameters is obtained as an output.

Among the most common international formats for the Sfiles [36] are HYPO71, HYPOINVERSE, and Nordic formats. The most used international formats for Waveform files are SEED, miniSEED, and SimpleASCII [37]. There are comprehensive seismological analysis tools capable of reading a wide range of formats of these two types of files, such as SEISAN and SeisComP. It is also possible to read the files through native programming languages or using libraries linked to these languages, such as SEISPP for C++ [38], Obspy for Python [39], or the open-source tools made available by the USGS for Java [40].

The implementation of specific architectures for reading multiple seismic records is of utmost importance. It is vital to analyze the computational capacity in terms of volatile memory, mainly

by considering the techniques exposed and the data stored during the seeking and gathering of seismic records. Therefore, a sub-module can be incorporated for balancing the computational load, which includes techniques for transforming volatile information into non-volatile (hard disk storage), as well as considering parallel processing mechanisms to facilitate the processing of large sets of seismic records.

Additionally, it is pertinent to include file selection algorithms that can discard repeated files or easily identifiable irregularities in both parameter files and trace files to prevent their storage. This process is called data wrangling, in which a data-cleaning procedure is required. Among the irregularities found in the seismic records from the RSNC are inconsistencies in the format, absence of trace files that correspond to existing parameter files, lack of start and end times of the event in the records, as well as non-existent P-wave and S-wave arrival times in some of the records recorded by the seismological stations. A key process to clean the seismic data from the RSNC is shown in Figure 3.

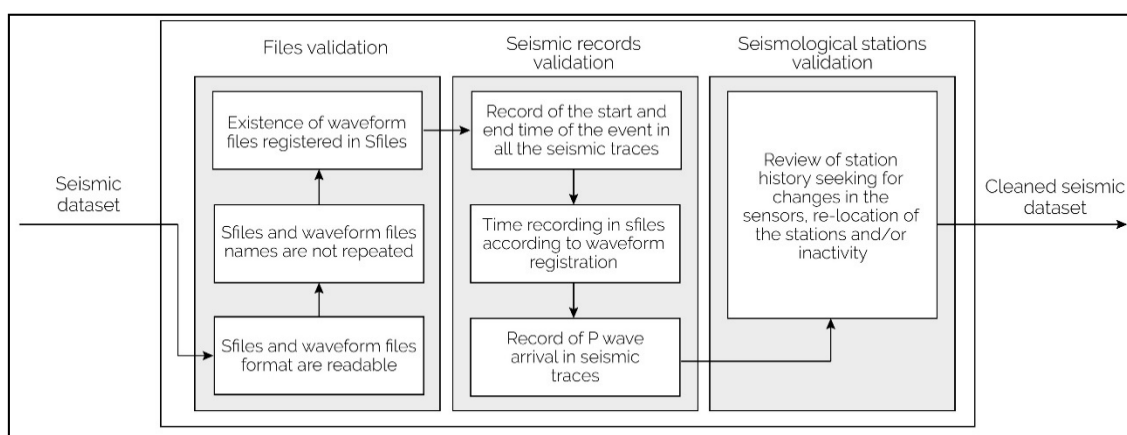


Figure 3. Seismic data cleaning process for RSNC files.

The data cleaning process is executed from three approaches: (1) the validation of the sought and gathered files, (2) the validation of the seismic samples in these files, and (3) the validation of the stations that register these samples.

In the first stage of this cleaning process in the file validation approach, it is checked that the sfile and waveform files can be read correctly for processing. First, if the files cannot be read, the inconvenience may be the source of the data. Secondly, it is verified that there are no repeated files, because the complexity of processing grows and there is no compensation to the investigation for processing the same data more than once. Third, it is verified that the general sfiles data files have a correct record of the corresponding waveform file. If in any of these stages there are inconsistencies, it is recommended that the file(s) are discarded, as they may create a bias in the general seismic analysis.

In the second stage of the cleaning process that focuses on the validation of the samples, it is convenient to check first if there is a record of the start and end time of the events that is homogeneous between each sfile and waveform files, secondly, that the P-waves have been recorded in both files.

Finally, in the third stage of the cleaning process, changes in the seismological station sensors, their relocation, or periods in which they have stopped operating must be considered. This permits the definition of a time interval in which the analysis will be executed, with the certainty that the dynamics of the waves will not be altered by external changes that do not concern the merely seismological field.

4.3. Analysis of Seismological Stations

Once the historical seismic archives are read, an analysis of the seismological stations that have recorded the events of interest is carried out, so that those that best represent the events and allow a reduction of the computational load in the processing of the data can be selected. The selected seismological stations are obtained as an output of this process.

Each seismic station provides a specific recording pattern that is dependent on several factors:

- The distance from the hypocenter and the epicenter (hypocenter and epicenter distances) to the geographic position of the station defines the amount of attenuation of the seismic wave.
- The geomorphology to which the seismic waves are exposed on the way to the station defines the propagation pattern and the attenuation of the seismic waves.
- The natural and artificial noise sources demean the seismic records due to the loss of quality regarding the content associated with seismic information, adding sources of information that concern other events that are not from a seismic nature.
- The technical parameters of the stations such as measurement channels, signal-to-noise ratio, analog-to-digital conversion, sampling rates, sensitivity, and dynamic range define how the seismic event is perceived from an analog source to a digital environment.

Then, these factors influence the composition and patterns of the traces that are transmitted to the monitoring site and stored for further offline processing. Each Colombian station records the seismic events individually by considering the named factors. The more stations that detect the event, the more data that can be transmitted, processed, and stored.

When a microseism occurs, usually few stations record it, since it can be a noise event due to a local disturbance on the surface or a seismic event of very low magnitude and/or considerable depth. In this case, the amount of data that contains useful information is not extensive and can be analyzed quickly and stored without major physical space costs.

However, when there are long-term, shallow, or large-scale seismic events that are perceived in various regions, the computational capacity for analysis and storage is high. Therefore, selecting a set of stations to analyze the traces that are recorded allows the dimensionality reduction of the data and reduction of the computational load.

Defining the stations to be studied is a process that requires a strong seismological criterion; however, the use of algorithms for statistical analysis facilitates the discernment between station selection criteria. For example, using libraries for the geographical mapping of stations and seismic epicenters, clustering and sampling the data exposed in the parameter files, among other procedures, allow contrasting the information stored and decide about the stations that best represent the events analyzed.

Figure 4 proposes a general procedure for the selection of the seismological stations used to identify the presence of seismic events in the Colombian monitored signals.

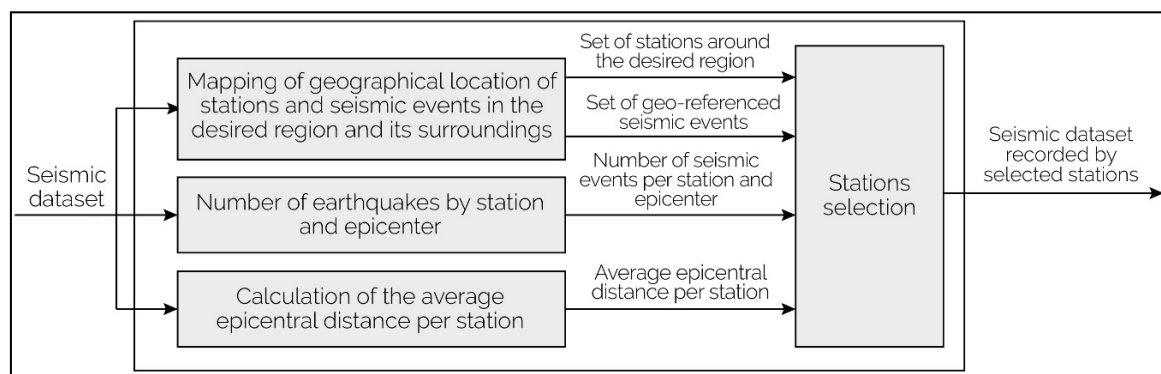


Figure 4. Seismological station selection process.

At first sight, it is necessary to identify the geo-referenced position of the stations, so that there is a spatial perspective of their distribution. Depending on the area under monitoring, the stations of interest will be those closer to this area. This is because the dynamics of local earthquakes are more marked and detectable than the dynamics of regional earthquakes and teleseisms, which have

attenuated and difficult to model features. It is recommended that more than two stations be selected, as events that have a non-seismic nature can be detected as such if only the activity is monitored in one or two stations.

In the second stage, it becomes clear to calculate the identification rate of seismic events labeled by each station, according to the monitoring region. This supports the selection of the stations concerning their geo-referenced position, as it is an indicator of the proximity of the stations to the epicenters (epicentral distance) and the signal attenuation index when arriving at the stations.

The stations with the highest identification rate of seismic events must be checked against their epicentral distance. A good relationship for the choice, as the third stage in this process, is to select the stations that have identified the most earthquakes, with a short epicentral distance. It is advisable to include the processing capacity as a third attribute in the selection process, since an additional station can signify the processing of 210,000 additional samples, on average.

4.4. Sample Selection

The sample selection process can be carried out in parallel to the station analysis process, since these two processes are independent of each other. There are several factors to consider when selecting a Colombian seismic event sample, as it was shown in Section 4.2:

- Inconsistencies in the file formats: There are different formats in which a seismic file can be structured, as SEED and miniSEED. During the processing and storage stages, the data are susceptible to be modified or lost, since there are multiple sources of information. Sometimes, these modifications alter the file formats, making them inconsistent. The files that present inconsistencies in the format and cannot be read correctly must be discarded.
- Absence of trace files that correspond to SEED and SAC existing parameter files: As part of the data storage process, the seismic information extracted from the seismic events (Sfiles) and the seismic samples (Waveforms) are recorded in separate files, as described in previous sections. Some of them are stored as part of the dataset without being associated. In this way, cases in which seismic information is recorded and samples were lost and vice versa can be found. Those files where the description data do not correspond to the seismic traces must be discarded.
- Lack of start and end times and/or inexistence of P-wave and S-wave arrival times in the events recorded: when a seismic event is recorded, some variables are measured, among which are the start time and end time of the event and the P-wave and S-wave arrival times. These values are very important to train classification algorithms, as some specific samples can be extracted from the seismograms, knowing when the earthquake began and when it finished. Unfortunately, some files can be well stored but lacking one or more of these four key parameters. In this case, it should be analyzed whether it is possible to determine the start or end date of the event by processing the seismic traces. If this is not possible, the files must be discarded.

It is also important to consider the structural changes of the stations, such as changes in the sampling frequency, sensors, digitizers, and number of spatial components, among others. These variations, although some of them are subtle, represent substantial alterations in the seismic patterns that might not be detected, since the detecting algorithms learn from specific patterns shown in the learning stages.

The inconsistencies in the files may be a consequence of the wrong acquisition, processing, and storage processes that are sometimes attributable to the algorithms that execute those processes for the seismology entities of each region or country.

In Colombia, between 20% and 30% on average of the selected files within the initial population of seismic events are prone to be discarded due to these inconsistencies, although some regions that are affected by very strong seismic events have accurate and well-stored files [41]. Other less frequent irregularities that may occur are (a) recording of the seismic data from stations that are inactive, (b) recording of the seismic data from components that the seismological station does not have

(e.g., the registration of three components for stations with monoaxial sensors), (c) recording seismic traces with a different sampling frequency from that described in the sensor datasheet, (d) recording of the seismic traces with different sampling frequencies among the components from the same station, either by components or by events, (e) the annotation of the P-wave in the traces is outside the measurement period of the seismic events, and (f) recording of all relevant seismic attributes described of the seismic signals with a magnitude of zero.

4.5. Classification Process

The analysis of seismological stations and sample selection processes provide the input datasets to the classification process, regarding seismological stations and seismic record datasets, respectively. With these inputs, the classification process implements supervised learning strategies using the selected seismological stations to detect the seismic event. The outputs of the classification process are (a) the average performance metrics of the classification approach and (b) the classification model, which is trained and validated, as described in Figure 5. The performance metrics are related to the ability of the classifier to differentiate between a seismic event and a non-seismic event, i.e., a binary classification, and the classification model corresponds to the implementation of the classifier, and it is able to classify new signals and provide the event detection output.

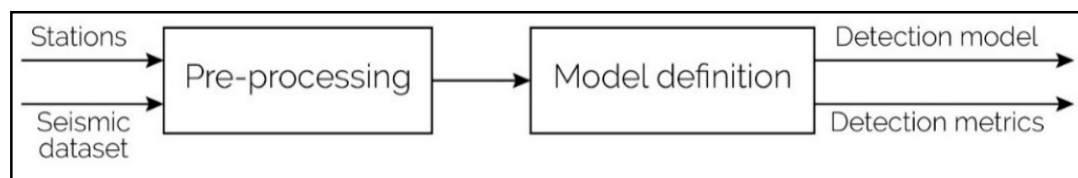


Figure 5. Block diagram of the classification process.

The identification and classification of seismic events can be done using different techniques, as described in Section 3. For instance, the phase picking of seismic waves is widely used for real-time monitoring, detection, and localization with the P-wave picking is the main method for detection in early warning systems. Several algorithms for P-wave picking have been proposed in the time-domain [42], whereas the STA/LTA and its variations are the most implemented algorithm in observation and detection networks [43]. In Colombia, the SeisComp3 software is currently used by the RSNC for the acquisition of seismic data from stations located throughout the national territory. With the STA/LTA-based AutoPick module, the P-wave is detected by a SeisComp3 implementation [44,45].

Traditional approaches such as STA/LTA are suitable options for the classification algorithm. Nevertheless, these approaches have limitations regarding their adaptability to the behavior of seismic waves [5,46,47]. As [48] state, the automatic picking of seismic waves can remove the ambiguity derived from the lack of synchronization between channels and signals proceeding from the seismological stations. Furthermore, [48–50] have shown that STA/LTA and cross-correlation approaches have a high rate of Types I and II errors (namely false negative and false positive) due to excessive noise that cannot be removed from the source and very low-frequency components that might not be enhanced. These factors can be handled accurately (as far as possible) using machine learning techniques.

Considering that the dynamic behavior of seismic signals recorded by sensors is subject to many factors that influence the signal, as denoted in Section 4.2, machine learning algorithms represent an appropriate alternative for the development of classification models, since they enable the abstraction of attributes associated with the signals, based on the modeling of large training datasets. Machine learning algorithms rely on the quantity and quality of data and, as described in Sections 3.2 and 4.1, current seismological services can provide huge amounts of data that can be processed to obtain datasets to classify seismic records such as the ones provided as output in the Sample Selection process.

As [51] state, machine learning algorithms are particularly well-used in seismology due to their facility to model complex relationships of a wide range of variables. Since the majority of tasks in

this context are normally targeting classification problems, machine learning drives a well-structured solution, since it can build a predictive environment in which a model is trained over sample data and tested over unseen data, guaranteeing the generalization of the solution against the data and the context. Sometimes, this procedure can be harmful if there is not enough data to proceed with the training procedure or when the representative descriptors (features or covariates) are almost the same as the number of samples. In these cases, additional machine learning approaches can be implemented, such as feature selection using forward or backward procedures [52].

With the selection of a machine learning algorithm as the classification algorithm, the classification process can be configured with a set of subprocesses depending on the specific machine learning algorithm requirements, which may be feature-based or time-series-based. The proposed set of subprocesses for the classification process is shown in the subprocesses diagram of Figure 6. The sequence of subprocesses can fulfill the classification capabilities for attribute-based or time series machine learning algorithms.

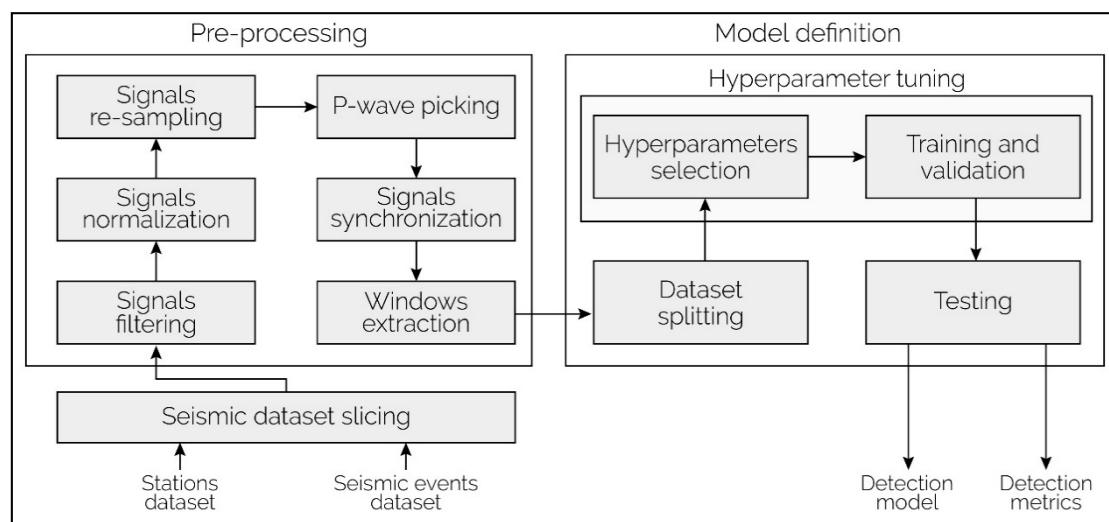


Figure 6. Classification process.

To set up the subprocesses, it is necessary to determine the number of classification models desired to represent the dynamics perceived by the stations. A single classification model might represent the events linking all the stations throughout a centralized fusion, or multiple classification models might correspond to each station separately throughout decentralized functions (such as the ones proposed by bagging techniques). In the case of decentralized fusion, each station has its classification model; hence, the Pre-processing and the Model definition stages need to be done separately for each selected station. The results provided by the per-station models can be composed into a combined result in the Testing subprocess using logical functions or more complex integration functions, producing a single classification model based on individual station analysis.

When the selected stations and the Seismic events dataset enter the classification process, the first subprocess that manipulates the datasets is the *Seismic dataset slicing* subprocess. This subprocess oversees linking the events to the selected stations and splits the Seismic events dataset into datasets per station if a decentralized schema is going to be used. In the *Pre-processing* stage, six subprocesses oversee formatting the dataset into a scheme ready to be used as input for a learning algorithm in the *Model definition* stage. Each subprocess in the *Pre-processing* stage is applied to every component that the signal may contain.

After the *Seismic dataset slicing*, the seismic signals associated with each event in the dataset are filtered in the *Signals filtering* subprocess. This subprocess considers the influence of noise sources that affect the seismic signals. The considered noise sources are the soil vibrations produced by natural causes and the instrumental noise associated with the measurement equipment hardware [41]. The filter

choice is crucial for the classification model performance, since the quality of the signal depends largely on the quality of the filter; therefore, the ability of the learning algorithm to generalize P-wave dynamics relies on the filter. It is recommended to make a frequency analysis to obtain the frequency components of the signals and the noise. The suggested frequency band is between 1 and 12 Hz [53], and different filter techniques can be applied to this subprocess [54].

When the *Signals filtering* subprocess is performed successfully, the dataset is relocated to the *Signals normalization* subprocess, whose objective is to standardize the scale of the signals to the $[-1, 1]$ interval according to the minimum and maximum value of each signal and remove the direct current component that is commonly added by the instrumentation. Then, the normalized signals are re-sampled in the *Signals re-sampling* subprocess. The definition of a common sampling rate for the signals is necessary due to the variety of the sampling rates that the signals may have, which are the outcome of diverse sampling properties that the acquisition devices and the digitalization algorithms present in the seismological stations.

The seismic signals associated with each record are commonly signals that contain information before the arrival of the P-wave and after the arrival of the S wave. To focus on the P-wave dynamics, in the *P-wave picking* subprocess, the identification of the time where the P-wave arrived is performed. This picking time annotation is commonly found in the Sfiles. Then, in the *Signal synchronization* subprocess, the picking time obtained in the *P-wave picking* is used to determine the exact sample where the P-wave arrived, relying on the defined sampling rate used in the *Signals re-sampling* subprocess, and a standard amount of samples is selected for a homogeneous duration time for each signal component. Therefore, all the signals representing an event start at the same time according to their P-wave time and end at the same time depending on the selected duration time.

The last *Pre-processing* subprocess is the *Windows extraction* subprocess whose purpose is to extract segments from the signals where the P-wave dynamics are contained and segments where there is no P-wave. The window length in samples is subject to the P-wave picking time and the window duration. A 2-s window is a recommended length [55]. It is necessary to obtain non-P-wave windows as well, since the classification algorithm learns to distinguish between the attributes of a P-wave and the attributes of a non-P-wave signal. Therefore, it is recommended to have the same number of windows concerning P-waves and non-P-waves to avoid class imbalance issues. With the *Window extraction* subprocess, the dataset of filtered, re-sampled, and synchronized signals associated with the events turns into a windows dataset with two classes, P-wave and non-P-wave, which is a common approach for the binary classification of seismic events.

With the window dataset, it is possible to generate a feature dataset that is commonly used by some feature-oriented machine learning techniques. A feature is a description of a record; the seismic events can be statistically described in terms of time, frequency, and non-linearity, among others. The features selection has a huge impact on the classification model's metrics, since they define how the classification algorithm perceives the seismic events. Each event is represented by a set of features in a matrix, and each feature acts as an input to the classification algorithm.

The *Pre-processing* stage differs from feature-based learning algorithms and time-series learning algorithms. The *Signals filtering*, *Signals re-sampling*, and *Signals normalization* subprocesses may be skipped depending on the time series technique, and some extra processes must be carried out, such as checks for stationarity, correlation, and autocorrelation. It is necessary to indicate the location of the P-wave in the signals (*P-wave picking*) to synchronize the signal's arrival and duration throughout the dataset (*Signals synchronization*) and to determine the characteristics of the moving window (*Window extraction*).

The input dataset for a feature-based technique is a set of single values that describes the original P-waves and non-P-waves in the seismic signals. Among the most commonly used machine learning feature-based algorithms applied to P-wave detection are Hidden Markov Models [56], Bayesian Networks [57], Support Vector Machines [58–60], Logistic Regression [53,61,62] and Artificial Neural Networks (ANN) [41,53,63–67]. Conversely, the input dataset for a time-series technique is a set of

signals split into P-wave signals and non-P-wave signals. Among the most used time-series forecasting techniques (TTF) for P-wave detection are Autoregressive Integrated Moving Average (ARIMA) [68], Seasonal ARIMA (SARIMA) [69], and ARIMA with Exogenous Regressors (ARIMAX) [70]. Some time-series forecasting recent methods used for P-wave detection include Pure Linear Neural Networks (PLNN) [71] and Polynomial Neural Networks (PNN) [72].

By following the suggested set of subprocesses and using a machine learning algorithm as the ones previously described, the classification process outputs a classification model and a set of metrics associated with that classification model. The produced model is suitable for performing the offline detection of seismic signals through the identification of P-waves. The mentioned metrics indicate the performance of the classification algorithm with the best set of hyperparameters scored on the test set records. The classification model, as denoted in the graph of the resulting classification model (Figure 7), can be interpreted as a system with testing preprocessed signals as the input to a function that contains the resultant algorithm responsible for the event detection. The output of this algorithm is a value that indicates whether the input signal contains a P-wave or not, in case of being a binary classification process.

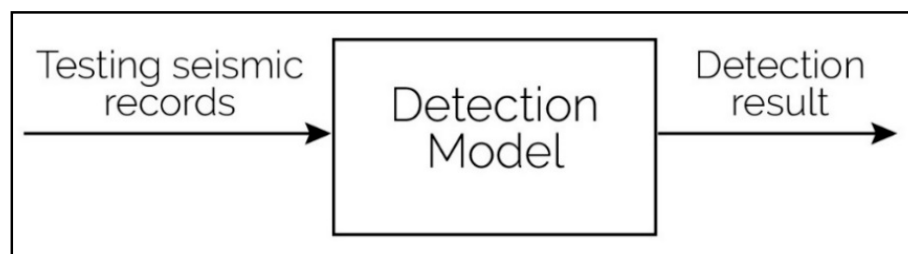


Figure 7. Classification model.

The classification model is the result of the System Model for Offline detection. It is recommended to generate several classification models by testing different approaches. In this sense, depending on the performance of the classification model reflected on the metrics, a change in the selection of classification process parameters, such as the filter type and length of the window, among others, may be considered to improve the classification performance. In the same way, other parameters that belong to macro processes, such as the number of stations in the Analysis of seismological stations or filtering the seismic records by magnitude range in Seismic data seeking and gathering, may have a huge influence on the classification performance.

Some studies have been carried out using the proposed System Model to perform the identification of P-waves in Colombia. In [41], a dataset of seismological records of events with epicenter in the department of Santander, Colombia between 2010 and 2017 was selected in the Seismic data seeking and gathering process. In the Reading and interpretation of the seismic records process, the records were described in the Nordic format. Then, in the Analysis of seismological stations, four stations were selected according to the epicenter distance. In the Sample Selection process, 20% of the downloaded records gathered were discarded due to inconsistencies in seismic attributes. Finally, in the Classification Process, the selected classification algorithm was logistic regression. The classification model was composed of four logistic regressors, one per each station, and a decentralized voting function that applied a logical function to the output of each regressor, to produce a binary output (P-wave or not P-wave). The obtained classification model produced an accuracy of 98.26% for the detection of the P-wave.

Similarly, in [41], the use of the System Model was applied to a dataset of events with an epicenter in Santander, using four stations. In the binary classification process, the selected classification algorithm was a feed-forward back-propagation Artificial Neural Network (ANN) after being cleaned and the missing values handled appropriately. Unlike the voting function applied in [53], the classification algorithm relied on the behavior associated with all the stations, which were analyzed as a group in a

centralized function represented by the ANN. The degree of polarization, the ratio of vertical power to total power, skewness, and kurtosis of the three-component seismic data for each station were used as the selected features to feed the training process of the named ANN binary classifier. All the input features were extracted from observations whose classes were balanced and equally distributed in the datasets. With the described settings, the obtained classification model produces an accuracy of 99.24% for the detection of the P-wave.

5. Conclusions

The proposed five modules of the seismic detection model facilitated the comprehension of the integration of the phases of an offline detection system. A set of historical seismic records is first downloaded and read (depending on the format of the data). The data can be filtered to make it easy to process by the subsequent phases. A selection of seismological stations that recorded an event of interest may allow the reduction of the computational load. This selection can be made based on the distances, geomorphology, noise sources, and technical parameters of the stations. The identification of a seismic event is a binary classification task, i.e., the presence/absence of a P-wave on the seismic signal.

The proposed model allows specifying detection and classification tasks for seismic events, which is applicable to the Colombian region. However, it can be extrapolated to other regions, as the detailed procedures are general enough to be applied in the local seismological networks that have monitoring stations with data formats and equivalent measurements.

Author Contributions: J.M. and C.G. contributed to the design of the seismic detection model. A.F. and M.A. supervised the research activities, and then carried out the coordination activities and the analysis of results. G.O. and C.F. conceived and wrote the earthquake detection methodologies. J.M., A.F., G.O., C.G. and M.A. contributed to the final version of the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: This research was supported by the Pontifical Bolivarian University at Bucaramanga, the Colombian National Seismological Network, and the Francisco de Paula Santander University. Thanks to Maria A. Claro and Diana T. Gómez, who were actively involved in the research project in which this paper is framed, and the seismological expert Edwar Zambrano, who provided insight and expertise that greatly assisted the research.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. El Tiempo, C.E.E. ¿Cuán Vulnerable es Colombia Ante un Sismo? Available online: <https://www.eltiempo.com/archivo/documento/CMS-16571309> (accessed on 5 November 2019).
2. El Tiempo, C.E.E. Los Peores Terremotos en la Historia de Colombia. Available online: <https://www.eltiempo.com/colombia/otras-ciudades/terremotos-mas-fuertes-de-colombia-155006> (accessed on 18 February 2020).
3. Wen-xiang, J.; Hai-ying, Y.; Li, L. A Robust Algorithm for Earthquake Detector. Available online: [/paper/A-Robust-Algorithm-for-Earthquake-Detector-Wen-xiang-Hai-ying/fa34661c1689cc9ee9a871ae4e1740bf323a54d2](https://arxiv.org/abs/2002.07402) (accessed on 18 February 2020).
4. Cuéllar, A.; Suárez, G.; Espinosa-Aranda, J.M. Performance Evaluation of the Earthquake Detection and Classification Algorithm 2(t_S-t_P) of the Seismic Alert System of Mexico (SASMEX). *Bull. Seismol. Soc. Am.* **2017**, *107*, 1451–1463. [CrossRef]
5. Sharma, B.K.; Kumar, A.; Murthy, V.M. Evaluation of seismic events detection algorithms. *J. Geol. Soc. India* **2010**, *75*, 533–538. [CrossRef]
6. USGS 20 Largest Earthquakes in the World. Available online: https://www.usgs.gov/natural-hazards/earthquake-hazards/science/20-largest-earthquakes-world?qt-science_center_objects=0#qt-science_center_objects (accessed on 27 November 2020).
7. Sarria Molina, A. *Ingeniería Sísmica*; Ediciones Uniandes, Ecoe Ediciones: Bogotá, Colombia, 1995; ISBN 958-9057-49-7.

8. Frohlich, C.; Kadinsky-Cade, K.; Davis, S.D. A reexamination of the Bucaramanga, Colombia, earthquake nest. *Bull. Seismol. Soc. Am.* **1995**, *85*, 1622–1634.
9. Prieto, G.A.; Beroza, G.C.; Barrett, S.A.; López, G.A.; Florez, M. Earthquake nests as natural laboratories for the study of intermediate-depth earthquake mechanics. *Tectonophysics* **2012**, *570*, 42–56. [[CrossRef](#)]
10. Bernal-Olaya, R.; Mann, P.; Vargas, C.A. Earthquake, tomographic, seismic reflection, and gravity evidence for a shallowly dipping subduction zone beneath the Caribbean Margin of Northwestern Colombia. In *Petroleum Geology and Potential of the Colombian Caribbean Margin*; American Association of Petroleum Geologists: Tulsa, OK, USA, 2015; pp. 247–269. [[CrossRef](#)]
11. Ochoa, L.H.; Niño, L.F.; Vargas, C.A. Fast magnitude determination using a single seismological station record implementing machine learning techniques. *Geod. Geodyn.* **2018**, *9*, 34–41. [[CrossRef](#)]
12. Kononov, A.V.; Stepnov, A.A.; Patrikeev, V.N. SEISAN software application for developing an automated seismological data analysis workstation. *Seism. Instrum.* **2012**, *48*, 270–281. [[CrossRef](#)]
13. Stepnov, A.A.; Gavrilov, A.V.; Kononov, A.V.; Ottemöller, L. New architecture of an automated system for acquisition, storage, and processing of seismic data. *Seism. Instrum.* **2014**, *50*, 67–74. [[CrossRef](#)]
14. Havskov, J.; Ottemoller, L. SEISAN earthquake analysis software. *Seismol. Res. Lett.* **1999**, *70*, 532–534. [[CrossRef](#)]
15. Weber, B.; Becker, J.; Hanka, W.; Heinloo, A.; Hoffmann, M.; Kraft, T.; Pahlke, D.; Reinhardt, J.; Thoms, H. SeisComp3—Automatic and interactive real time data processing. *Geophys. Res. Abstracts* **2007**, *9*, 09219.
16. Alfaro, A. Difficulties on psha in Colombia because of data scarcity. *Bull. Int. Inst. Seismol. Earthq. Eng.* **2012**, *46*, 145–152.
17. Mora-Páez, H.; Peláez-Gaviria, J.-R.; Diederix, H.; Bohórquez-Orozco, O.; Cardona-Piedrahita, L.; Corchuelo-Cuervo, Y.; Ramírez-Cadena, J.; Díaz-Mila, F. Space Geodesy Infrastructure in Colombia for Geodynamics Research. *Seismol. Res. Lett.* **2018**, *89*, 446–451. [[CrossRef](#)]
18. Zhu, W.; Mousavi, S.M.; Beroza, G.C. Seismic Signal Denoising and Decomposition Using Deep Neural Networks. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 9476–9488. [[CrossRef](#)]
19. Perol, T.; Gharbi, M.; Denolle, M. Convolutional neural network for earthquake detection and location. *Sci. Adv.* **2018**, *4*, e1700578. [[CrossRef](#)] [[PubMed](#)]
20. Hildyard, M.W.; Nippres, S.E.J.; Rietbrock, A. Event Detection and Phase Picking Using a Time-Domain Estimate of Predominate Period Tpd. *Bull. Seismol. Soc. Am.* **2008**, *98*, 3025–3032. [[CrossRef](#)]
21. Toledo Peña, P.A. Algoritmo de Detección de Ondas P Invariante de Escala: Caso de Réplicas del Sismo del 11 de Marzo de 2010. Ph.D. Thesis, Universidad de Chile, Santiago, Chile, 2014.
22. Allen, R.M.; Gasparini, P.; Kamigaichi, O.; Bose, M. The status of earthquake early warning around the world: An introductory overview. *Seismol. Res. Lett.* **2009**, *80*, 682–693. [[CrossRef](#)]
23. Joswig, M. Pattern recognition for earthquake detection. *Bull. Seismol. Soc. Am.* **1990**, *80*, 170–186.
24. Donoho, D.L. Nonlinear Wavelet Methods for Recovery of Signals, Densities, and Spectra from Indirect and Noisy Data. 1993. Available online: <https://statistics.stanford.edu/sites/g/files/sbiybj6031/f/EFS%20NSF%20437.pdf> (accessed on 2 December 2020).
25. Hafez, A.G.; Khan, T.A.; Kohda, T. Earthquake onset detection using spectro-ratio on multi-threshold time–frequency sub-band. *Digit. Signal Process.* **2009**, *19*, 118–126. [[CrossRef](#)]
26. Withers, M.; Aster, R.; Young, C.; Beiriger, J.; Harris, M.; Moore, S.; Trujillo, J. A comparison of select trigger algorithms for automated global seismic phase and event detection. *Bull. Seismol. Soc. Am.* **1998**, *88*, 95–106.
27. Allen, R.V. Automatic earthquake recognition and timing from single traces. *Bull. Seismol. Soc. Am.* **1978**, *68*, 1521–1532.
28. Leonard, M. Comparison of Manual and Automatic Onset Time Picking. *Bull. Seismol. Soc. Am.* **2000**, *90*, 1384–1390. [[CrossRef](#)]
29. Li, Z.; Peng, Z.; Hollis, D.; Zhu, L.; McClellan, J. High-resolution seismic event detection using local similarity for Large-N arrays. *Sci. Rep.* **2018**, *8*, 1646. [[CrossRef](#)] [[PubMed](#)]
30. Schorlemmer, D.; Woessner, J. Probability of Detecting an Earthquake. *Bull. Seismol. Soc. Am.* **2008**, *98*, 2103–2117. [[CrossRef](#)]
31. Yoon, C.E.; O’Reilly, O.; Bergen, K.J.; Beroza, G.C. Earthquake detection through computationally efficient similarity search. *Sci. Adv.* **2015**, *1*, e1501057. [[CrossRef](#)] [[PubMed](#)]

32. Avvenuti, M.; Cresci, S.; Marchetti, A.; Meletti, C.; Tesconi, M. EARS (Earthquake Alert and Report System): A Real Time Decision Support System for Earthquake Crisis Management. In Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 24–27 August 2014; ACM: New York, NY, USA, 2014; pp. 1749–1758.
33. Sakaki, T.; Okazaki, M.; Matsuo, Y. Earthquake shakes Twitter users: Real-time event detection by social sensors. In Proceedings of the 19th International Conference on World Wide Web, Raleigh, NC, USA, 26–30 April 2010; ACM: New York, NY, USA, 2010; pp. 851–860.
34. Earle, P.S.; Bowden, D.C.; Guy, M. Twitter earthquake detection: Earthquake monitoring in a social world. *Ann. Geophys.* **2012**, *54*. [[CrossRef](#)]
35. Liu, H.; Zhang, J.Z. STA/LTA algorithm analysis and improvement of Microseismic signal automatic detection. *Prog. Geophys.* **2014**, *29*, 1708–1714.
36. Havskov, J.; Ottemöller, L. *Seisan: The Earthquake Analysis Software for Windows, Solaris, Linux and MacOSX Version 8.1*; Institute of Solid Earth Science, University of Bergen: Bergen, Norway, 2008.
37. IRIS: Data Formats. Available online: <https://ds.iris.edu/ds/nodes/dmc/data/formats/> (accessed on 19 February 2020).
38. Pavlis, G. SEISPP Library: C++ Seismic Data Processing Library. Available online: <https://pavlab.sitehost.iu.edu/software/seispp/html/index.html> (accessed on 9 October 2020).
39. Beyreuther, M.; Barsch, R.; Krischer, L.; Megies, T.; Behr, Y.; Wassermann, J. ObsPy: A Python Toolbox for Seismology. *Seismol. Res. Lett.* **2010**, *81*, 530–533. [[CrossRef](#)]
40. United States Geological Survey Earthquake Hazards—Software. Available online: <https://www.usgs.gov/natural-hazards/earthquake-hazards/software> (accessed on 19 February 2020).
41. Miranda Calle, J.D.; Gamboa Entralgo, C.A. *Desarrollo de un Sistema para la Detección de Movimientos Sísmicos usando Redes Neuronales Artificiales*; Universidad Pontificia Bolivariana: Medellín, Colombia, 2018.
42. Chen, Z.; Stewart, R.R. A multi-window algorithm for real-time automatic detection and picking of p-phases of microseismic events. *CREWES Res. Rep.* **2006**, *18*. Available online: <https://www.crewes.org/ForOurSponsors/ResearchReports/2006/2006-15.pdf> (accessed on 2 December 2020).
43. SeisComp3 Scautopick. Makes Picks on Waveforms. Available online: <https://www.seiscomp3.org/doc/jakarta/current/apps/scautopick.html> (accessed on 19 February 2020).
44. Servicio Geológico Colombiano Consulta Catálogo Sísmico. Available online: <https://www2.sgc.gov.co/sgc/sismos/Paginas/catalogo-sismico.aspx> (accessed on 11 February 2020).
45. Catálogo Mecanismo Focal y Tensor Momento. Available online: http://bdrsnc.sgc.gov.co/sismologia1/sismologia/focal_seiscomp_3/index.html (accessed on 9 October 2020).
46. Zhang, J.; Tang, Y.; Li, H. STA/LTA Fractal Dimension Algorithm of Detecting the P-Wave Arrival. *Bull. Seismol. Soc. Am.* **2018**, *108*, 230–237. [[CrossRef](#)]
47. Akram, J.; Peter, D.; Eaton, D. A k-mean characteristic function to improve STA/LTA detection. In Proceedings of the Geoconvention, Calgary, AB, Canada, 7–11 May 2018; p. 5. Available online: https://geoconvention.com/wp-content/uploads/abstracts/2018/287_GC2018_A_k-mean_characteristic_function_to_improve_STALTA1_detection.pdf (accessed on 2 December 2020).
48. Gentili, S.; Michelini, A. Automatic picking of P and S phases using a neural tree. *J. Seismol.* **2006**, *10*, 39–63. [[CrossRef](#)]
49. Audretsch, J.M.L. Earthquake Detection using Deep Learning Based Approaches. Master’s Thesis, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia, 2020.
50. Wiszniowski, J.; Plesiewicz, B.M.; Trojanowski, J. Application of real time recurrent neural network for detection of small natural earthquakes in Poland. *Acta Geophys.* **2014**, *62*, 469–485. [[CrossRef](#)]
51. Kong, Q.; Trugman, D.T.; Ross, Z.E.; Bianco, M.J.; Meade, B.J.; Gerstoft, P. Machine Learning in Seismology: Turning Data into Insights. *Seismol. Res. Lett.* **2019**, *90*, 3–14. [[CrossRef](#)]
52. Venkatesh, B.; Anuradha, J. A Review of Feature Selection and Its Methods. *Cybern. Inf. Technol.* **2019**, *19*, 3–26. [[CrossRef](#)]
53. Miranda, J.D.; Gamboa, C.A.; Flórez, A.; Altuve, M. Voting-based seismic data classification system using logistic regression models. In Proceedings of the 2019 XXII Symposium on Image, Signal Processing and Artificial Vision (STSIVA), Bucaramanga, Colombia, 24–26 April 2019; pp. 1–5.

54. Bormann, P.; Wielandt, E. Seismic Signals and Noise. *New Man. Seismol. Obs. Pract. 2 NMSOP2* **2013**, 1–62. [[CrossRef](#)]
55. Kaur, K.; Wadhwa, M.; Park, E.K. Detection and identification of seismic P-Waves using Artificial Neural Networks. In Proceedings of the 2013 International Joint Conference on Neural Networks (IJCNN), Dallas, TX, USA, 4–9 August 2013; IEEE: Piscataway, NJ, USA, 2013; pp. 1–6.
56. Beyreuther, M.; Hammer, C.; Wassermann, J.; Ohrnberger, M.; Megies, T. Constructing a Hidden Markov Model based earthquake detector: Application to induced seismicity: Constructing a HMM based earthquake detector. *Geophys. J. Int.* **2012**, *189*, 602–610. [[CrossRef](#)]
57. Riggelsen, C.; Ohrnberger, M. A Machine Learning Approach for Improving the Detection Capabilities at 3C Seismic Stations. *Pure Appl. Geophys.* **2014**, *171*, 395–411. [[CrossRef](#)]
58. Ruano, A.E.; Madureira, G.; Barros, O.; Khosravani, H.R.; Ruano, M.G.; Ferreira, P.M. Seismic detection using support vector machines. *Neurocomputing* **2014**, *135*, 273–283. [[CrossRef](#)]
59. Asencio-Cortés, G.; Morales-Esteban, A.; Shang, X.; Martínez-Álvarez, F. Earthquake prediction in California using regression algorithms and cloud-based big data infrastructure. *Comput. Geosci.* **2018**, *115*, 198–210. [[CrossRef](#)]
60. Li, W.; Narvekar, N.; Nakshatra, N.; Raut, N.; Sirkeci, B.; Gao, J. Seismic Data Classification Using Machine Learning. In Proceedings of the 2018 IEEE Fourth International Conference on Big Data Computing Service and Applications (BigDataService), Bamberg, Germany, 26–29 March 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 56–63.
61. Amidan, B.G.; Hagedorn, D.N. *Logistic Regression Applied to Seismic Discrimination*; No. PNNL-12031; Pacific Northwest National Lab.: Richland, WA, USA, 1998.
62. Reynen, A.; Audet, P. Supervised machine learning on a network scale: Application to seismic event classification and detection. *Geophys. J. Int.* **2017**, *210*, 1394–1409. [[CrossRef](#)]
63. Ibs-von Seht, M. Detection and identification of seismic signals recorded at Krakatau volcano (Indonesia) using artificial neural networks. *J. Volcanol. Geotherm. Res.* **2008**, *176*, 448–456. [[CrossRef](#)]
64. HassanAitLaasri, E.; Akhouayri, E.-S.; Agliz, D.; Atmani, A. Seismic Signal Classification using Multi-Layer Perceptron Neural Network. *Int. J. Comput. Appl.* **2013**, *79*, 35–43. [[CrossRef](#)]
65. Giudicepietro, F.; Esposito, A.M.; Ricciolino, P. Fast Discrimination of Local Earthquakes Using a Neural Approach. *Seismol. Res. Lett.* **2017**, *88*, 1089–1096. [[CrossRef](#)]
66. Vallejos, J.A.; McKinnon, S.D. Logistic regression and neural network classification of seismic records. *Int. J. Rock Mech. Min. Sci.* **2013**, *62*, 86–95. [[CrossRef](#)]
67. Diersen, S.; Lee, E.-J.; Spears, D.; Chen, P.; Wang, L. Classification of Seismic Windows Using Artificial Neural Networks. *Procedia Comput. Sci.* **2011**, *4*, 1572–1581. [[CrossRef](#)]
68. Amei, A.; Fu, W.; Ho, C.-H. Time Series Analysis for Predicting the Occurrences of Large Scale Earthquakes. *Int. J. Appl. Sci. Technol.* **2012**, *2*, 12.
69. Adhikari, R.; Agrawal, R.K. An Introductory Study on Time series Modeling and Forecasting. *arXiv* **2013**, arXiv:1302.6613. Available online: <https://arxiv.org/abs/1302.6613> (accessed on 2 December 2020).
70. Yu, J.; Wang, R.; Liu, T.; Zhang, Z.; Wu, J.; Jiang, Y.; Sun, L.; Xia, P. Seismic energy dispersion compensation by multi-scale morphology. *Pet. Sci.* **2014**, *11*, 376–384. [[CrossRef](#)]
71. Egrioglu, E.; Khashei, M.; Aladag, C.H.; Turksen, I.B.; Yolcu, U. Advanced Time Series Forecasting Methods. Available online: <https://www.hindawi.com/journals/mpe/2015/918045/> (accessed on 9 October 2020).
72. Fong, S.; Deb, S.; Fong, S.; Deb, S. Prediction of Major Earthquakes as Rare Events Using RF-Typed Polynomial Neural Networks. In *Encyclopedia of Information Science and Technology*; IGI Global: Hershey, PN, USA, 2015; ISBN 978-1-4666-5888-2.

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).