Journal of Volcanology and Geothermal Research 315 (2016) 15–27

Contents lists available at ScienceDirect

Journal of Volcanology and Geothermal Research

journal homepage: www.elsevier.com/locate/jvolgeores

Pattern recognition applied to seismic signals of Llaima volcano (Chile): An evaluation of station-dependent classifiers

Millaray Curilem [a,*], Fernando Huenupan [a], Daniel Beltrán [a], Cesar San Martin [a], Gustavo Fuentealba [b], Luis Franco [c], Carlos Cardona [c], Gonzalo Acuña [d], Max Chacón [d], M. Salman Khan [e], Nestor Becerra Yoma [f]

[a] Department of Electrical Engineering, Universidad de La Frontera, Temuco, Chile
[b] Department of Physics, Universidad de La Frontera, Temuco, Chile
[c] Observatorio Vulcanológico de los Andes Sur, Temuco, Chile
[d] Department of Computer Engineering, Universidad de Santiago de Chile, Santiago, Chile
[e] Department of Electrical Engineering, University of Engineering and Technology, Peshawar, Pakistan
[f] Department of Electrical Engineering, Universidad de Chile, Santiago, Chile

ARTICLE INFO

Article history:
Received 15 September 2015
Accepted 8 February 2016
Available online 18 February 2016

Keywords:
Seismic discrimination
Volcano monitoring
Pattern recognition
Station-dependent classifiers
Support vector machines

ABSTRACT

Automatic pattern recognition applied to seismic signals from volcanoes may assist seismic monitoring by reducing the workload of analysts, allowing them to focus on more challenging activities, such as producing reports, implementing models, and understanding volcanic behaviour. In a previous work, we proposed a structure for automatic classification of seismic events in Llaima volcano, one of the most active volcanoes in the Southern Andes, located in the Araucanía Region of Chile. A database of events taken from three monitoring stations on the volcano was used to create a classification structure, independent of which station provided the signal. The database included three types of volcanic events: tremor, long period, and volcano–tectonic and a contrast group which contains other types of seismic signals. In the present work, we maintain the same classification scheme, but we consider separately the stations information in order to assess whether the complementary information provided by different stations improves the performance of the classifier in recognising seismic patterns. This paper proposes two strategies for combining the information from the stations: i) combining the features extracted from the signals from each station and ii) combining the classifiers of each station. In the first case, the features extracted from the signals from each station are combined forming the input for a single classification structure. In the second, a decision stage combines the results of the classifiers for each station to give a unique output. The results confirm that the station-dependent strategies that combine the features and the classifiers from several stations improves the classification performance, and that the combination of the features provides the best performance. The results show an average improvement of 9% in the classification accuracy when compared with the station-independent method.

1. Introduction

Automatic detection and classification of volcano-seismic activity has gained importance because of the growing need to monitor and complement the existing seismic networks on active volcanoes. In Chile, this is even more challenging due to the large number of active volcanoes and increased economic activity around them, such as housing, tourism, agriculture, and mining. Volcano seismic patterns help in identifying processes that occur inside the volcano and could potentially be used as precursors of an eruption (Lahr et al., 1994; Chouet, 1996; Zobin, 2012). Tectonic seismicity, for example, although it may not be an indication of an eruption, could be used in setting different alert levels for the safety of citizens and authorities.

The Southern Andes Volcano Observatory (OVDAS) belongs to the National Service of Geology and Mining (SERNAGEOMIN) and is responsible for volcano monitoring in Chile. OVDAS personnel have designed and installed a network of instruments for monitoring more than 40 active volcanoes. OVDAS needs to process signals from volcanoes to identify on the one hand changes in base-level activity, and on the other patterns to develop models in order to identify precursors of eruptive stages. This requires a large number of analysts and a multidisciplinary group of experts in the areas of volcanology, seismology, physics, and geology, to assess and interpret volcanic seismic signals continuously. Precursor signals of volcanic eruptions are difficult to identify; however, experts can detect situations in which alerts can be activated.

* Corresponding author.
E-mail addresses: millaray.curilem@ufrontera.cl (M. Curilem), fernando.huenupan@ufrontera.cl (F. Huenupan), cesar.sanmartin@ufrontera.cl (C. San Martin), gustavo.fuentealba@ufrontera.cl (G. Fuentealba), luis.franco@sernageomin.cl (L. Franco), carlos.cardona@sernageomin.cl (C. Cardona), gonzalo.acuna@usach.cl (G. Acuña), max.chacon@usach.cl (M. Chacón), salmankhan@uetpeshawar.edu.pk (M.S. Khan), nbecerra@ing.uchile.cl (N. Becerra Yoma).

http://dx.doi.org/10.1016/j.jvolgeores.2016.02.006
0377-0273/

Table 1
Description of seismic stations used in this study.

St	Cod	Lat (°) S	Long (°) W	Dist. crater
Motín	MOT	38.675	71.784	3 (north)
Laguna Verde	LAV	38.701	71.651	7 (west)
Llaima	LLA	38.784	71.695	10 (south)

In the literature, many methods have been proposed for automatic detection and classification of seismic events, to improve response times (Carniel, 2014). The results are generally encouraging, but only in limited experiments excluding on-line processes in which discrimination tends to be more complex. Generally, a classic pattern recognition approach has been adopted in which a database of events is created, data are pre-processed, features are extracted and, finally, the events are classified, that is, a classifier assigns a label (class) to the event. The information from a single station or from several stations is normally used to generate a single database to design and validate the classifiers. This typically leads to a classifying process which does not consider the information of which station recorded the event.

Data from a single station were used to design different classification approaches in Masotti et al. (2006), Beyreuther et al. (2008), Rouland et al. (2009), Langer et al. (2011), and Bicego et al. (2015). In other works, the information of many stations was considered but they were merged to form a single database, as in Álvarez et al. (2012), Esposito et al. (2013), Carniel et al. (2013), Cortés et al. (2014, 2015), Curilem et al. (2014a, 2014b), and Bicego et al. (2015). A large variety of classification tools were used in these works, but in all of them, the classification of events was station-independent, since even when more than one station was considered, a single database was created to train and validate the classification models. This procedure is used despite the fact that most of the volcanoes are monitored by multiple seismic stations; not much research has been carried out into how to exploit the complementary information contained in the signals received from different stations to enhance classification performance. Based on the location of the stations and the source of the seismic activity, the signals will suffer different types of distortion, depending on the type of sensor, its proximity to the crater, the distance from the sources that generated the seismic activity, or the type of terrain that waves travel through. Moreover, the transducer response may differ from one station to another. Hence the need arises for a classification system that incorporates the additional information contained in signals observed by multiple stations.

There are few works which explore the alternatives offered by the use of multiple stations. In Scarpetta et al. (2005), the authors considered four stations and designed a specialized classifier for each station. They performed station-dependent classification of two classes: volcano-tectonic earthquakes and other kinds of event (thunder, quarry blast, and explosions). The authors employed data from short-period analogue stations and digital broadband stations in the Mt. Vesuvius (Italy) monitoring network. Although the results show good individual performances, the station-dependent classifiers were not combined. Ibáñez et al. (2009) performed a portability test on the classification system of different stations on two active volcanoes, Stromboli and Etna (Italy), to verify whether a classifier trained with the database from one station and one volcano could be employed to recognize signals from other stations or volcanoes. In both cases, the same type of short-period instruments was used. Two types of events were considered: explosions and noise for Stromboli, and tremor bursts and background tremor for Etna. Hidden Markov models were used to perform the classification. The results confirmed the robustness of the proposal, subject to the accuracy of the manual segmentation to train the models. In Duin et al. (2010), the authors investigated whether combining signals from different stations would improve the classification performance or not, and whether a classifier trained with one station could be tested with data from another station. They considered earthquakes, icequakes, and long-term tremor events obtained from five seismic stations on Nevado del Ruiz (Colombia) and used subspectra quadratic classifiers. The results showed that individual stations had significantly different performances; also that combining signals of different stations

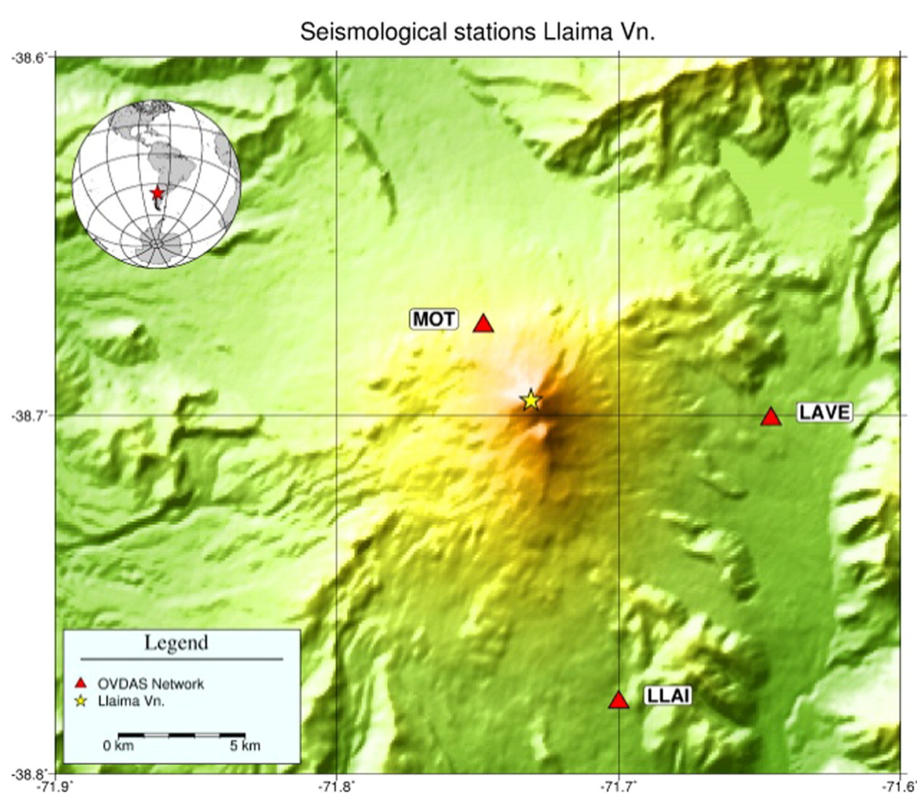

Fig. 1. Location of the seismic stations considered in the study.

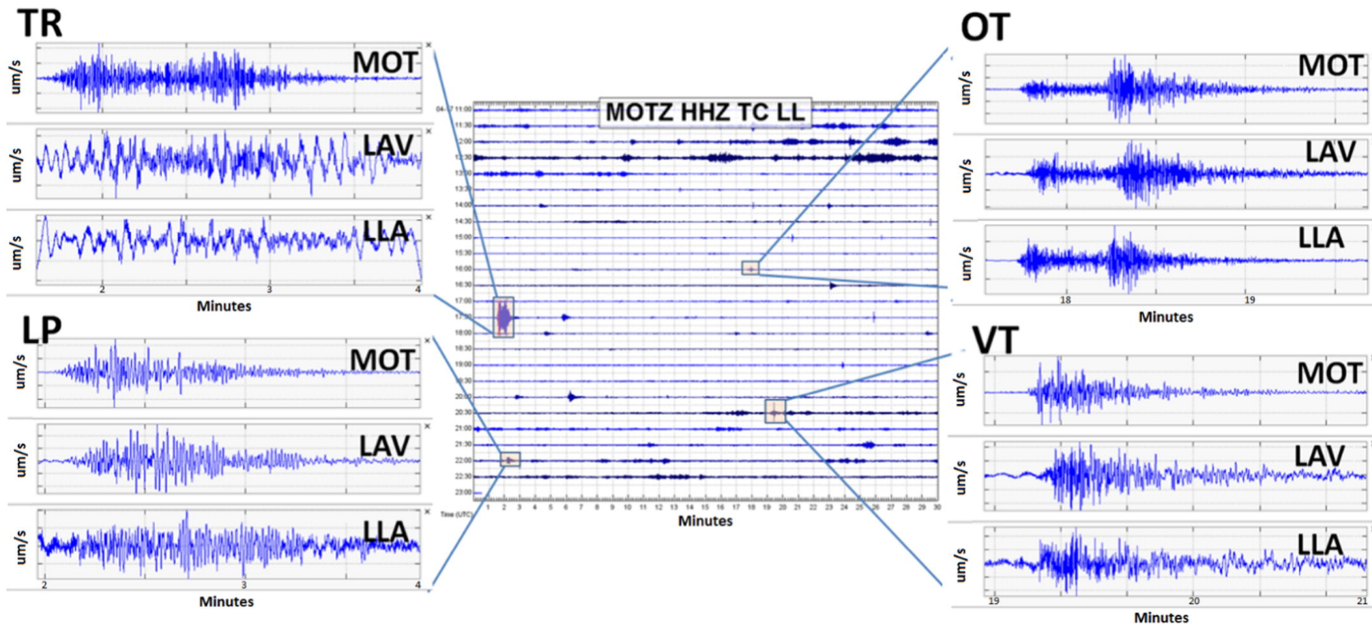

Fig. 2. Time representations of events of the four classes (TR, LP, OT, VT) recorded at the different stations (MOT, LAV, LLA) using the Seismic Wave Analysis and Real-time Monitoring (SWARM, 2011) tool.

improved the classification results of individual station classifiers. Finally, the authors showed that combining classifiers of different stations did not improve the results of individual station classifiers; however, merging the signals led to an improvement in the accuracy of classification, showing that although the signals measured by the different stations were similar, they contain useful additional information. The results of these preliminary works motivated us to apply merging strategies and evaluate if they are able to improve the capacity for automatic classification of our classifying structures for the Llaima volcano.

Distortions resulting from the location and quality of the stations affect the records. This is not generally taken into account by the automatic classification structures of literature, although it is interesting to note that in practice, human analysts consider information from several stations when labelling the seismic events. In the present work, we analyse how information from different stations could be combined to improve discrimination of the seismic events considered. The case study is Llaima volcano, and we included three of its stations placed strategically with respect to the crater. The general classification structure presented here is inspired by Curilem et al. (2014a); however, in this work, we propose different strategies for combining the information from the stations to generate station-dependent classifiers. The objective is to assess whether the procedure carried out by a human analyst, who analyses several stations in order to take a decision, can be applied to automatic classification to improve its performance. The following sections describe the data used in this study, the classification structures applied, the combination strategies proposed, and the principal results obtained.

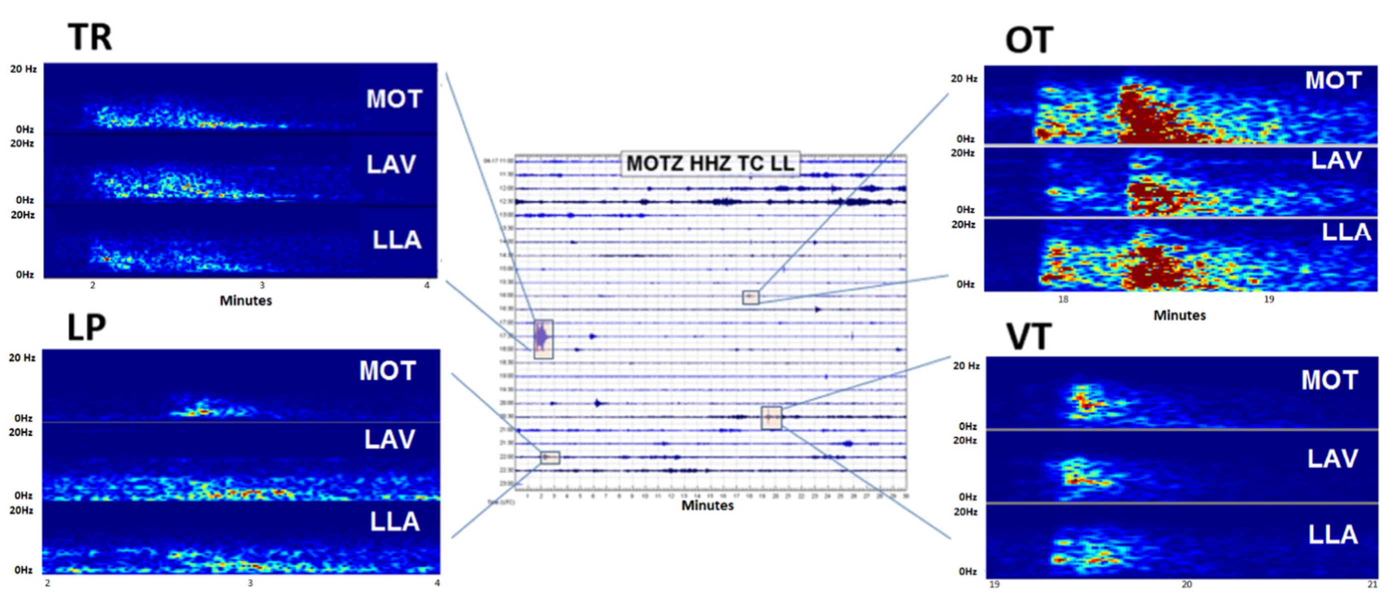

Fig. 3. Spectral representations of the events of the four classes (TR, LP, OT, VT) recorded at the different stations (MOT, LAV, LLA) using the Seismic Wave Analysis and Real-time Monitoring (SWARM, 2011) tool. In the graphs, the x-axis is time and the y-axis is frequency.

Table 2
Number of events of each class recorded at the three seismic stations.

	Station		
Class	MOT	LAV	LLA
OT	166	166	166
LP	296	296	296
VT	134	134	134
TR	173	173	173
Event/station	769	769	769
Total		2307	

2. Case study: Llaima volcano and its monitoring system

2.1. The volcano

Llaima, one of the biggest volcanoes in Chile by volume and one of the most active in South America, is a compound strato-volcano of basaltic to andesite–basaltic composition, located in the volcanic zone of the Southern Andes (Stern, 2004; Mora-Stock et al., 2012). It has produced major eruptions, with a historical record of around 50 documented eruptions between 1640 and 2009 (Naranjo & Moreno, 2005), the biggest of which was in 1640. Its latest eruption cycle (2007–2010) gave rise to two large eruptions in January 2008 and April 2009 with VEI – volcanic explosivity index (Newhall & Self, 1982) – of 3 and 2, respectively. In general, the ascent and going up of magmatic materials through the volcanic conduits generate a wide variety of signals which are recorded by sensors placed around the volcanic edifice. Although instrumental monitoring has been carried on since 1996 with temporary seismic stations, the National Volcanic Vigilance Programme, started in 2009, decided to complement and expand monitoring of the volcano with an instrumental network. Today, Llaima volcano has a network equipped with seismic sensors, GPS, tilt-meters, scan DOAS (differential optical absorption) and cameras for visual observation. The data are transmitted in real time to the national monitoring and processing centre in the city of Temuco, 70 km west of the volcano. Llaima volcano has 9 seismic stations, which together with the other instruments are used to monitor its activity. In the present study, only three stations were considered: Motín (MOT), Laguna Verde (LAV), and Llaima (LLA), Table 1.

MOT is the reference station for the volcano (seismic station where all measurements from seismic events are recorded). Historically, these stations present good stability over time, while their different distances from the crater, as shown in Fig. 1 and Table 1, ensure adequate variability of the shape and other characteristics of seismic records. The analysed data come from broadband seismic stations (Güralp 6TD 30s). Only the Z component was considered for all stations because it provides a better signal to noise ratio in most events.

The seismic volcanic events considered were Long-Period (LP), Tremor (TR), and Volcano–Tectonic (VT); they were recorded from January 2010 to December 2013. All the classified events meet the classification threshold defined by OVDAS (which depends on the distance between seismic source and stations) i.e. their amplitudes are superior to 2 μm/s and their signal-to-noise ratio (SNR) are superior to 1.5 (at the reference MOT station). The seismic record observed in the volcano presents predominantly LP events and to a lesser degree TR; these signals are typically associated with magmatic and hydrothermal fluids (Chouet, 1996). The latest eruptive period included continuous tremor and discrete tremor signals with durations longer than 5 min.

Fig. 4. General structure of the proposed pattern recognition system.

In the present work, tremor events lasting up to 10 min were included, with a general average of approximately 2 min. The tremor was distinguished from distant regional seismic events by analysing the energy variation in the frequency band 0.5–15 Hz. VT seismicity, related with the rupture of fragile material (Chouet and Matoza, 2013), is less recurrent in Llaima volcano, and most frequently found in the crater and 15 km south of the crater. A fourth group called OT (other types) was defined to cluster all the signals that did not correspond to any of the first three events (OT principally groups non-volcanic events). The OT set is mainly composed of background noise that is saved by analysts twice a day to keep track of the background signal recorded by the station. The OT set also contains ice-quake or ice breaking events, Local Tectonic (principally the Liquiñe-Ofqui Fault Zone), Regional Tectonic (Nazca-South America subduction zone), Distant Tectonic, avalanches caused by slippage of snow and ice, and other events detected by the analysts the origin of which is unknown and which cannot be classified in any of the above established classes. The purpose of creating the OT class was to offer a contrast group of signals to train the classifiers and thus increase their discrimination performance. This set is very important for discriminating signals that are not related to volcanic activity.

2.2. The database of seismic events

The database was generated by supervised selection of the events from recorded signals. As mentioned before, the signals in this work were treated separately according to their station. After recognising the signal as a volcanic event, an OVDAS analyst performed the segmentation (start and end recognition) from the digital file in which seismic signals are kept after manual classification (labelling) of each event, for all the stations, following the procedures used in the observatory. We only considered events present at least in two of the three stations. If an event was not recorded in one station, its start and end values were set by replicating the start and end values of the nearest station in which the event was recorded, considering that on Llaima the distances between seismic sources and stations are small (<10 km) and estimating the velocities of seismic phases (>3 km/s). The record ought to be included in the time window which we set (>2 min). The segmented signals were exported to the Matlab environment. A volcano seismologist also analysed the data to confirm that the events were classified correctly, that is, they followed the same procedure used by OVDAS. It should be noted that the events were segmented manually, thus their duration is variable.

Fig. 2 shows examples of events of the different classes recorded at different stations. Fig. 3 shows the time–frequency plots (spectrograms) of the events of Fig. 2. Each spectrogram is obtained by normalising the event with the maximum amplitude of its signal, and then dividing it into segments, each one being 10% of the total length of the signal with 86% overlap. Each window is then transformed to frequency domain using the FFT function in Matlab to represent a slice in the

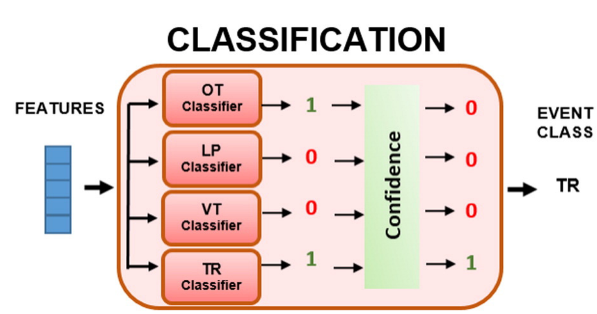

Fig. 5. Classification structure: "one versus all" structure of the classifiers with the confidence step – the stage at which it is decided which of the two "1" values prevails – for the output final classification.

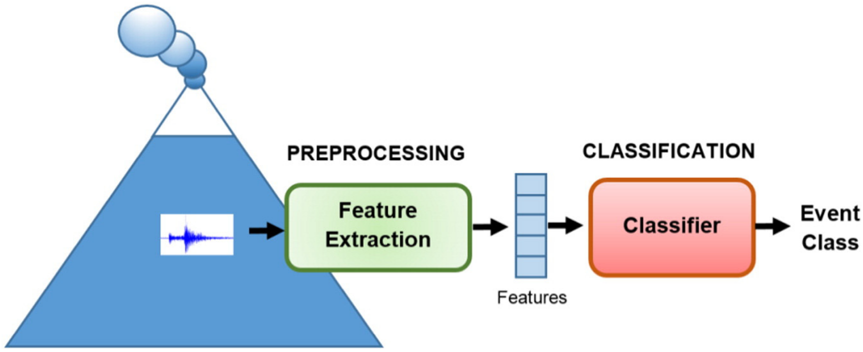

Fig. 6. Station-independent classification: In this approach, a single classifying structure is used to classify the signals coming from any station of Llaima volcano.

spectrogram that covers the entire duration of a window (Oppenheim et al., 1999). Table 2 shows the distribution of the classes for each station.

3. Classification strategies

In a previous work (Curilem et al., 2014a), a station-independent classifying structure was proposed to classify the signals recorded at the volcano by any seismic station. In this paper, station-dependent event classification is performed using two different approaches: combination of the features and combination of the classifiers. We started with an assumption that the OVDAS analyst applies when identifying and classifying seismic signals: the more stations reviewed, the greater the degree of certainty in labelling an event. This section will first present the classifying structure proposed in the previous work, and then show how this structure was applied to the two merging approaches proposed in the present paper.

3.1. General method for the design of the station-independent classifiers

The classical pattern recognition structure shown in Fig. 4 was applied in Curilem et al. (2014a) to design a classifying structure for LP, VT, TR, and OT events. It can be divided in two main steps: pre-processing and classification. The pre-processing step transforms the signal into a feature vector that provides a discriminative parameterization of the event. The classification step is formed by a recognizer structure, the task of which is to assign one of the four class labels to a particular event, according to its features. The input of the classifying structure is a seismic event and the output is its label (OT, LP, VT, and TR). These steps are described in detail below.

3.1.1. Pre-processing step

Since the most important information from volcano events is below 25 Hz, signals coming from the sensors were down-sampled from 100 to 50 Hz and filtered with a 4th order Butterworth band-pass filter between [0.5 Hz, 24 Hz]. This band was used because most of the seismic energy for the events considered is contained in this range.

The signals coming from different seismic stations were inspected by an analyst who performed a manual segmentation (definition of the beginning and end) of the events. Thus, the events extracted from the continuous records of many sensors have variable length and they are stored in a unique database. Five features were extracted from the variable length events:

— Three features from the time domain: the mean, the median, and the maximum value of the event. These features were calculated for each segmented event, directly from the time samples, with the exception of the mean that is calculated over the absolute values of the samples.
— One feature from the frequency domain: the dominant frequency. This feature was calculated by the mean of the five highest frequency peaks of the Fourier transform of each event.
— One from the time–frequency domain: the energy in 4th band ([1.56–3.13] Hz) of the wavelet transform. This feature was calculated as the ratio between the sum of the components of the 4th wavelet band over the sum of all the wavelet components (in all the bands). It is expressed in percentage. A Daubechies mother wavelet type five was used with five decomposition levels.

These features were selected by a feature selection process performed in previous works (Curilem et al., 2009, 2014a), to find which combination of the features improved the performance of the

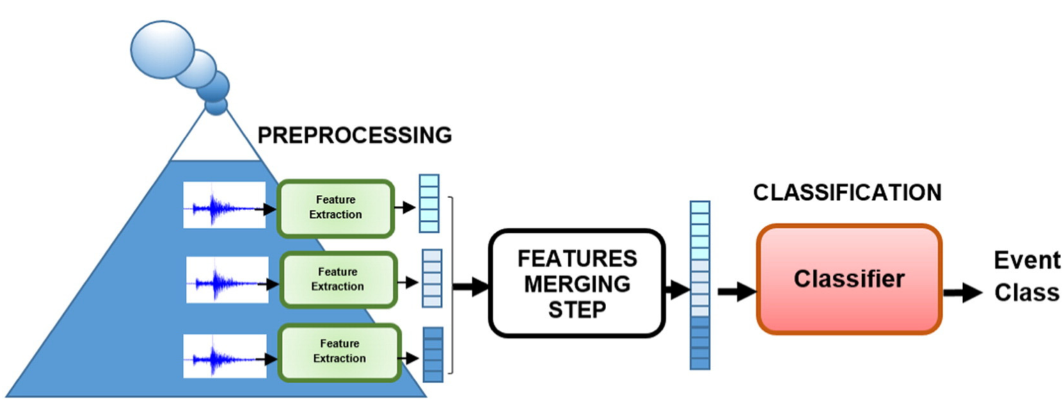

Fig. 7. Station-dependent classification: Combination of the features per station. In this approach, only the pre-processing step is performed by station. The feature extraction produces a feature vector of 5 features × 3 stations long. In the classifying step, this new feature vector is the input of the classifying structure.

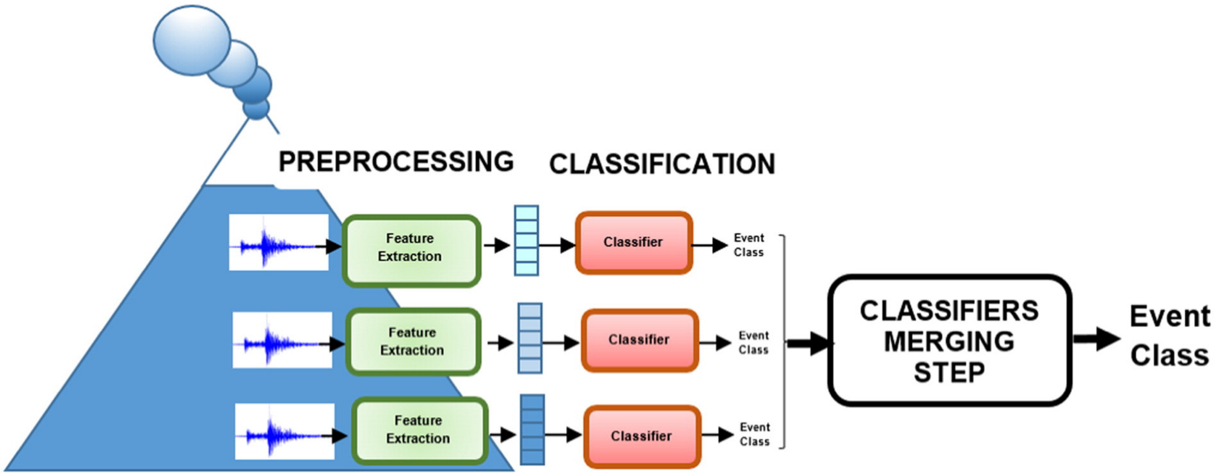

Fig. 8. Station-dependent classification: Combination of the classifiers per station. In this approach, the whole pattern recognition structure is implemented for each station. A classifiers-merging step decides the final output of the classification, according to the information of the individual classifying structures.

classifiers. All the extracted features were linearly normalized between $[-1, 1]$.

3.1.2. Classifier design

Support vector machines (SVM) were used to implement the classifying step. SVM is an approach which creates a linear decision hyperplane to separate two classes by applying a quadratic programming optimisation process (Vapnik, 1995). When the classes are not linearly separable, the input features are projected to an appropriately chosen kernel-induced feature space of higher dimension. The kernel function performs the projection of the data from its original input space to the feature space where the two classes can be linearly separable. In this work, we used one of the most applied kernel functions, the Gaussian or Radial Basis Function (RBF-kernel). The decision hyperplane is calculated given c, a regularization parameter that penalizes the errors and σ, a parameter that controls the standard deviation of the Gaussian kernel function (Hamel, 2009). The tuning of these parameters was performed by a grid search: the parameters c and σ were increased by powers of 2^X, where the exponents X took values in a discrete range (Hsu et al., 2003). The SVM functions of the bioinformatics toolbox of Matlab were used to solve the optimization problem.

Since SVM are binary classifiers, they have to be combined to handle multiclass problems (Hamel, 2009). The two-class approach described above can be extended to multiple class classification problems by adopting methods such as the "one versus all" configuration (Giacco et al., 2009). This is a simple combination in which each classifier discriminates one class from all the others. Fig. 5 shows the resulting classification structure.

Each of the outputs of the classification structure corresponds to a class; however, this output codification presents the limitation that, when a classifier is wrong, a disagreement situation may occur: none or more than one classifiers may be activated for a single event. An example is presented in Fig. 5 where the OT and TR classifiers are activated at the same time. To solve the disagreement, a confidence measure was proposed at the output of the classification structure, as shown in Fig. 5. The Bayes-based confidence measure (BBCM) was used (Yoma et al., 2005). As mentioned before, the SVM is a two-class classifier that finds a maximal margin decision hyperplane of the two classes, based on a kernel function. In this proposal, a score is assigned to the distance between the input sample and the hyperplane. If the sample is closer to the hyperplane, the decision of the classifier is less reliable and vice versa. Each classifier output has a score value assigned to it, to decide which output is the most reliable. If more than one classifier is activated, this measure retrieves the most confident positive output (the one with the highest confidence measure) and sets all the others to zero, as shown in Fig. 5. If all the classifiers are set to zero, this measure retrieves the least confident negative output (lowest confidence measure) and sets it to one. As a consequence, the confidence measure always retrieves the activation of only one output: the most reliable label.

3.2. Station-independent classifying structures

The general structure proposed in Curilem et al. (2014a) is presented in Fig. 6. This is a station-independent classifying structure, as it classifies events coming from any station. The classifying structure was developed based on a database that stored events coming from any station. Thus a general learning process was performed.

This procedure was applied in the present work. Events randomly selected from the database of each station were used to form a single database, like in previous works. This database was used to design a station-independent classifying structure.

3.3. Station-dependent classifying structures

The automatic classification method presented above (Curilem et al., 2014a) was applied in this work considering separately the information of the different stations. The same events recorded by three different stations were organised in separate databases that were used to implement station-dependent classifying structures. It should be noted that, as the events included in the databases had to be recorded in at least two of the three stations considered, the databases of this work did not contain the same events as those considered in our previous work.

As in our previous work, we propose here a classifying structure consisting of four SVM classifiers, one for each kind of event, and a confidence step to guarantee a unique output. Here, the ranges of the grid search took values from -5 to 10 for c and from -15 to 9 for σ, with a step of 1. Unlike the previous work, the five features were used in all the classifiers to simplify comparison of their performances, and a cross-validation strategy was used to obtain the decision hyperplane of each SVM and to validate the values of c and σ. The data were partitioned into 5 folds or subsets. As the databases stored 768 events, each fold contained approximately 153 events. The amount of data in each validation fold ensured an adequate variance of the classification performance, making more complex strategies unnecessary, for example, the leave-one-out or the 10-folds cross-validation (Kalayeh and Landgrebe, 1983; Mitchell, 1997). Five classifying structures were then trained and validated: the first classifier was trained using four folds and its performance indices were calculated using the remaining fifth fold. The next classifier was trained using other four folds and validated with the remaining fold. This process was repeated until all the folds were used for training and validation (5 iterations). The classifier that achieved the best performance was selected and the other four classifiers were discarded; however, the final performance indices were calculated as the mean of the performance of the five classifiers, that is, considering all the folds (the whole data set). This process is repeated for each class. The main advantage of the cross-validation process is that all the examples in the dataset are used for both training and testing, avoiding over-fitting.

3.3.1. Strategy 1: Combination of the features

In this approach, the database stores the records from each station separately and the pre-processing step is separated for each station, as presented in Fig. 7. This proposal enabled us to assess whether the fact of having input information from different stations improves the performance of the classification structure. The procedure is as follows: each event from the database is recorded by the three stations under study. The five features are extracted from the three records, thus obtaining a dimension features vector: n features $\times$ m stations per event. Here, $n = 5$ and $m = 3$, thus 15 features per event. In this proposal, the combination of the features is very simple: the feature vector is formed by all the extracted features. This vector is the input to the classifying structure that performs the identification. As shown in Fig. 7, the classification step is unique for all the stations and only the features are station-dependent.

Fig. 7 presents the feature extraction process performed separately from the stations of the volcano and the combination of features before the classifying structure is applied. The training and validation of the classifiers were performed with this new input vector. It may be observed that the general method for the design of the pattern recognition system remains the same as presented in Fig. 4.

3.3.2. Strategy 2: Combination of the classifiers

In the second approach, each seismic station of Llaima volcano had a complete pattern recognition structure, independent from the others, as shown in Fig. 8. This means that the general pattern recognition structure of Fig. 4 was replicated for each station. The station-dependent classifying structures were trained with signals coming exclusively from their own station. This proposal enabled us to assess two situations: whether the system performance is improved by having specialised classifiers for each station and whether this performance is improved by merging the classifiers to generate a single decision. As Fig. 8 shows, in this strategy, the classifiers are station-dependent. A final stage is added to merge the decisions of the classifiers and deliver a single system output.

The combination of the classifiers merges the output of each station-dependent classifying structure. Three merging criteria were developed in this work: one used a simple majority vote and the other two used the confidence criterion described in Section 3.1.2. In the first case, the final decision corresponds to the class which received the most votes from the set of classifiers. In the second case, a confidence measure for each classifier decision is used (Huenupan et al., 2008). BBCM is used as a confidence measure and the two schemes implemented for merging the classifiers' outputs were maximum and mean BBCM. In the merging with maximum BBCM, the output of each local classifier is the class selected with its corresponding confidence, and the final decision is one of the classes previously selected by local classifiers with the highest reliability. In the merging with mean BBCM, the output of each classifier is the reliability for each class (there is no local decision). The final decision is the class that achieves the highest mean of the all output BBCM values given by the local classifiers. For more details see Appendix 1.

Table 3
Contingency table of the station-independent classifying structure.

Expert \ Classifier	OT	LP	VT	TR	**Total**
OT	116	20	26	4	**166**
LP	26	222	41	7	**296**
VT	9	43	81	1	**134**
TR	13	10	10	140	**173**
Total	**164**	**295**	**158**	**152**	**769**

Table 4
Performance indices of the individual classifiers of the station-independent classifying structure.

Class	OT	LP	VT	TR
Exactitude	**87.26**	**80.88**	**83.09**	**94.15**
Error	12.74	19.12	16.91	5.85
Sensitivity	69.88	75.00	60.45	80.92
Specificity	92.04	84.57	87.87	97.99

3.4. Classifier performance indices

The classification results allow constructing the contingency tables of the classifiers. The contingency table is a tool which allows us to evaluate the performance of a classifier, compared to an expert. It consists of a square matrix containing all the classes considered: the rows contain the events of each class as classified by the expert; the columns contain the events of each class as classified by the classification structure. The diagonal therefore shows the agreements while the rest of the cells show how many events of a class were erroneously classified in another class.

The kappa coefficient (Landis and Koch, 1977; Witten et al., 2011) is a measure of agreement between the classification of the expert and the classifying structure, for this multiclass approach. Kappa values near to one show excellent levels of agreement. However, to simplify the comparison of the results, the performance will also be evaluated in a binary approach, that is considering separately the performance of each classifier (LP, TR, VT, and OT) in the one versus all configuration. To perform the binary evaluation, four statistical indices were used to evaluate the performance of each classifier: *sensitivity* (Se) shows how good the classifier is at recognizing the positive class; *specificity* (Sp) measures the ability of recognizing the events that do not belong to the positive class; *exactitude* (Ex) and *error* (Er) show the total successes and errors, respectively, among the total classified events.

To calculate the performance in the multiclass approach, for n samples of C classes, the kappa coefficient (K) is computed from the contingency table using Eq. (1), that is, the agreement minus the

Table 5
Contingency table of the station-dependent classifying structure implemented with the multi-station feature classifiers.

Expert \ Classifier	OT	LP	VT	TR	**Total**
OT	157	4	5	0	**166**
LP	9	281	5	1	**296**
VT	8	4	122	0	**134**
TR	0	0	0	173	**173**
Total	**174**	**289**	**132**	**174**	**769**

Table 6
Performance indices (%) of the individual classifiers of the station-dependent classifying structure implemented with the multi-station feature classifiers.

Class	OT	LP	VT	TR
Exactitude	**96.62**	**97.01**	**97.14**	**99.87**
Error	3.38	2.99	2.86	0.13
Sensitivity	94.58	94.93	91.04	100.00
Specificity	97.18	98.31	98.43	99.83

probability that the agreement is due to chance (Landis and Koch, 1977).

$$K = \frac{P_o - P_e}{1 - P_e} \tag{1}$$

where P_o is the observed agreement between the classifier and the expert (diagonal), and P_e is the probability that the agreement is due to chance (see Eqs. (2) and (3)).

$$P_o = \sum_{i=1}^{C} p_{ii} \tag{2}$$

$$P_e = \sum_{i=1}^{c} p_{i\cdot} p_{\cdot i} \tag{3}$$

where p_{ii} is the joint proportion of the agreement and $p_{i\cdot}$ and $p_{\cdot i}$ are the sum of the joint proportions of the classifier (rows) and the expert (column), respectively, for each class.

To calculate the performance in the binary approach, the following values can be extracted from the contingency table: true positives (TP) is the number of events correctly classified as belonging to a specific class; true negatives (TN) is the number of events correctly classified as not belonging to a specific class; false positives (FP) and false negatives (FN) are the number of events classified erroneously, and the statistical indices can be obtained in percentages using Eqs. (4) to (7), where n is the total number of events (Witten et al., 2011).

$$Ex = \frac{TP + TN}{n} \times 100 \tag{4}$$

$$Er = \frac{FP + FN}{n} \times 100 \tag{5}$$

$$Se = \frac{TP}{TP + FN} \times 100 \tag{6}$$

$$Sp = \frac{TN}{TN + FP} \times 100 \tag{7}$$

It is important to remind that the contingency tables presented in the next sections show the results of the whole dataset, as all the folds of the cross-validation strategy were used as validation set.

4. Results

In this section, we make a comparative analysis of the results obtained in the previous work with the present proposal of the two combination strategies. As stated above, the classifying structures that are being evaluated consist of four classifiers, one for each class. The individual performances of each classifier of the classifying structure are presented for each strategy.

4.1. Station-independent classifiers

As in our previous work (Curilem et al., 2014a), we implemented the station-independent classifying structure shown in Fig. 6, but here using the present database. This database was implemented selecting random events from the three stations, avoiding to read the same event in more than one station (that is, just one record of each event was considered). Table 3 shows the contingency table obtained from this classifying structure. As mentioned before, the diagonal in Table 3 contains the agreement between the reference class (set by the expert) and the output of the automatic classification. The other cells show the misclassified events: for example, in the OT line, 20 OT were incorrectly classified as LP, 26 as VT, and 4 as TR. The kappa value of this classifying system was $\kappa = 0.63$. Table 4 shows the individual performance indices of the classifiers per class. This table shows that the exactitude (Ex) is superior to 80% for all the classifiers. It can be seen from Table 4 that the sensitivity is low for VT and OT (<70%) and specificity is low for LP.

4.2. Station-dependent classifiers: Combination of features

Table 5 shows the contingency table obtained by the classifier designed with the combination of the features per station, as explained in Section 3.3.1. The kappa value of this classifying system was $\kappa = 0.94$. Table 6 shows the performance indices of the best classifiers per class. A significant improvement in the classification of the OT and TR events can be seen in Table 6. The corresponding classifiers have errors inferior to 4% and a sensitivity superior to 90% for all classes. The improvements for the TR group were the most significant.

Table 7
Contingency table of the station individual classifying structures.

STATION: MOT

		Classifier				
		OT	LP	VT	TR	**Total**
Expert	OT	129	23	4	10	**166**
	LP	24	266	2	4	**296**
	VT	14	37	83	0	**134**
	TR	3	0	0	170	**173**
Total		**170**	**326**	**89**	**184**	**769**

STATION: LAV

		Classifier				
		OT	LP	VT	TR	**Total**
Expert	OT	134	5	25	2	**166**
	LP	12	273	11	0	**296**
	VT	6	2	126	0	**134**
	TR	1	3	0	169	**173**
Total		**153**	**283**	**162**	**171**	**769**

STATION: LLA

		Classifier				
		OT	LP	VT	TR	**Total**
Expert	OT	141	1	23	1	**166**
	LP	16	264	16	0	**296**
	VT	12	11	110	1	**134**
	TR	9	0	0	164	**173**
Total		**178**	**276**	**149**	**166**	**769**

Table 8
Performance indices (%) of the individual classifiers of the classifying structures implemented for each station.

	STATION: MOT				STATION: LAV				STATION: LLA			
	Classifier				Classifier				Classifier			
	OT	LP	VT	TR	OT	LP	VT	TR	OT	LP	VT	TR
Ex	**89.86**	**88.30**	**92.59**	**97.79**	**93.37**	**95.71**	**94.28**	**99.22**	**91.94**	**94.28**	**91.81**	**98.57**
Er	10.14	11.70	7.41	2.21	6.63	4.29	5.72	0.78	8.06	5.72	8.19	1.43
Se	77.71	89.86	61.94	98.27	80.72	92.23	94.03	97.69	84.94	89.19	82.09	94.80
Sp	93.20	87.32	99.06	97.65	96.85	97.89	94.33	99.66	93.86	97.46	93.86	99.66

4.3. Station-dependent classifiers: Combination of classifiers

As mentioned in Section 3.3.2, the combination of classifiers required one classifying structure per station; for this reason, the individual classifying structures have to be evaluated separately for each station. The classifying structures were trained and validated with events coming exclusively from their station. Table 7 shows the contingency tables obtained separately by the classifiers designed for each station. The kappa values for MOT, LAV, and LLA, respectively, were $\kappa = 0.78$, $\kappa = 0.88$, and $\kappa = 0.84$. Table 8 shows their individual performance indices. When the results of our previous work, shown in Table 4, are compared with the results shown in Table 8, it is interesting to notice the improvement in the classifying performance of the OT and TR groups for the individual station classifiers. LP classification improved in some stations and VT recognition performance decreased in all the stations. It is also interesting to notice that some classes are better recognised in some stations than in others. The LAV station presented the best recognition performance, while MOT presented the lowest performance.

After the classifiers have been evaluated separately by station, their combination will retrieve a unique final label. Table 9 shows the contingency tables obtained by the merging of the classifying structures implemented with the voting, BBCMmax and BBCMmean strategies. The kappa values for the voting, BBCMmax, and BBCMmean merging strategies was the same: $\kappa = 0.0.91$. Table 10 shows the performance indices of the combination for each strategy. Table 10 shows that the results of the three combination strategies are not significantly different, although the voting strategy presented a slight improvement. Comparing the results shown in Table 6 with those shown in Table 10, the merging of features outperformed the merging of classifiers. However, when the results of Table 10 are compared to the results of the previous work (Table 4) it is interesting to notice the significant improvement of the merging of classifiers results for the OT, LP, and TR groups.

4.4. Final comparison of the results

Table 11 shows the kappa values for the best models of the different combination strategies. Table 12 shows the performance indices of the station-independent strategy and the two station-dependent strategies proposed here. This table shows the better results achieved by treating the information of each station separately, for all classes. In particular, the exactitude is significantly superior when compared with the classifiers of the station-independent strategy proposed in this work and in our previous work (Curilem et al., 2014a). The improvements in the sensitivity are outstanding for the combination of features strategy. Analysing the sensitivity indices, OT and VT are constantly the classifiers with the lowest performance; however, they improved for the combination of the features strategy.

Analysing the exactitude, sensitivity, and specificity of the classifiers presented in Table 12, it can be observed that strategy 1: Combination of the features is the strategy with the higher performances. Fig. 9 presents the impact of the combination strategies for the classification of each class, compared with the station-independent strategy. For all classes, the combination of features strategy gave better results than the combination of classifiers strategy. The considerable improvement produced by combination strategies on the performance of the VT, OT, and TR groups can be appreciated.

Fig. 10 presents the mean error of the station-independent classification structure proposed in the earlier work and of all the models proposed in the present work. The improvement produced by having individual classifiers by station and the even more significant improvement produced by the proposed combination strategies can

Table 9
Contingency table of the merging of the classifying structure implemented for each station.

MERGE: voting

		classifier				
		OT	LP	VT	TR	**Total**
Expert	OT	144	3	19	0	**166**
	LP	8	286	2	0	**296**
	VT	14	3	117	0	**134**
	TR	1	0	0	172	**173**
Tot al		**167**	**292**	**138**	**172**	**769**

MERGE: BBCM max

		classifier				
		OT	LP	VT	TR	**Total**
Expert	OT	140	4	19	3	**166**
	LP	6	288	2	0	**296**
	VT	6	9	119	0	**134**
	TR	1	0	0	172	**173**
Tot al		**153**	**301**	**140**	**175**	**769**

MERGE: BBCM mean

		classifier				
		OT	LP	VT	TR	**Total**
Expert	OT	141	3	19	3	**166**
	LP	7	287	2	0	**296**
	VT	7	8	119	0	**134**
	TR	1	0	0	172	**173**
T otal		**156**	**298**	**140**	**175**	**769**

Table 10
Performance indices (%) for each class of the combination of the classifying structure implemented for each station.

	MERGE: voting				MERGE: BBCM max				MERGE: BBCM mean			
	OT	LP	VT	TR	OT	LP	VT	TR	OT	LP	VT	TR
Ex	**94.15**	**97.92**	**95.06**	**99.87**	**94.93**	**97.27**	**95.32**	**99.48**	**94.80**	**97.40**	**95.32**	**99.48**
Er	5.85	2.08	4.94	0.13	5.07	2.73	4.68	0.52	5.20	2.60	4.68	0.52
Se	86.75	96.62	87.31	99.42	84.34	97.30	88.81	99.42	84.94	96.96	88.81	99.42
Sp	96.19	98.73	96.69	100	97.84	97.25	96.69	99.50	97.51	97.67	96.69	99.50

Table 11
Kappa indices of the classifying structures implemented with the different strategies.

	Station–independent classification.	Best individual station classifiers	Station–dependent classifiers: Combination of the features	Station–dependent classifiers: Combination of the classifiers
Kappa Index	0.63	0.88	0.94	0.91

be appreciated, as can the fact that the strategy of merging the features delivered the best results.

5. Discussion

Like in Duin et al. (2010), the results of this work show a significant improvement in the discrimination capacity of the four classes of events considered when station-dependent classification strategies are applied. This is shown by the indices of agreement presented in Table 11, and whether in the general performance indices by strategy (Table 12) or by class (Fig. 9). Fig. 10 shows how considering the information from individual stations results in a reduction in the global error of the classification structures.

If the results obtained in the station-independent strategy (Table 4), where the classifier was designed to classify signals from any station are compared with the results obtained by the individual classifiers for each station (Table 8), an improvement in the classification by station may be observed. This is particularly significant in certain stations such as LAV, where the mean of exactitude is 95.6%, compared to 86.3% obtained by the station-independent strategy; in other words, an increase of 9.3% is achieved for exactitude, representing a total reduction of 67.9% in the error (from 13.7% to 4.4%). So although there is no combination of the information, this result shows that the design of individual classifiers by station already represents an improvement over a general classifier for all the stations on the volcano. However, when the station-dependent classification strategies are analysed, these results improve still further. Fig. 10 shows that the mean error of the four classes falls from 4.4% to 3.3% for the combination of classifiers and from 4.4% to 2.3% for the combination of features; in other words, the two strategies for combining the information from the stations produced reductions of 25% and 48%, respectively, in the mean classification error. It is interesting to note that, unlike the work of Duin et al. (2010), the combination of classifiers improved significantly the accuracy of classification.

These results are consistent with the procedure used by the human analyst, who – among other strategies – reviews the signals from several stations to obtain a more certain classification. Reviewing several stations is particularly important for the analyst when he/she has to differentiate signals of volcanic origin from others. Volcanic tremors tend to be confused with signals which persist in time such as environmental noise caused by wind, rain, avalanches, etc., and which present similar patterns. In these cases, it is very important to have at least three observation stations located in different geographical quadrants with respect to the crater (taken as the origin of the coordinates).

In Table 3, which shows the results of the station-independent strategy, it can be seen that the station-independent classification tends to confuse VT events with OT and LP, as the sensitivity of these classifiers

Table 12
Performance indices of the classifying structures of three combining strategies.

	Station–independent classification				Station–dependent classifier: combinationof the features				Station–dependent classifier: combinationof the classifiers			
	OT	LP	VT	TR	OT	LP	VT	TR	OT	LP	VT	TR
Ex	**87.26**	**80.88**	**83.09**	**94.15**	**96.62**	**97.01**	**97.14**	**99.87**	**94.15**	**97.92**	**95.06**	**99.87**
Er	12.74	19.12	16.91	5.85	3.38	2.99	2.86	0.13	5.85	2.08	4.94	0.13
Se	69.88	75.00	60.45	80.92	94.58	94.93	91.04	100.00	86.75	96.62	87.31	99.42
Es	92.04	84.57	87.87	97.99	97.18	98.31	98.43	99.83	96.19	98.73	96.69	100.00

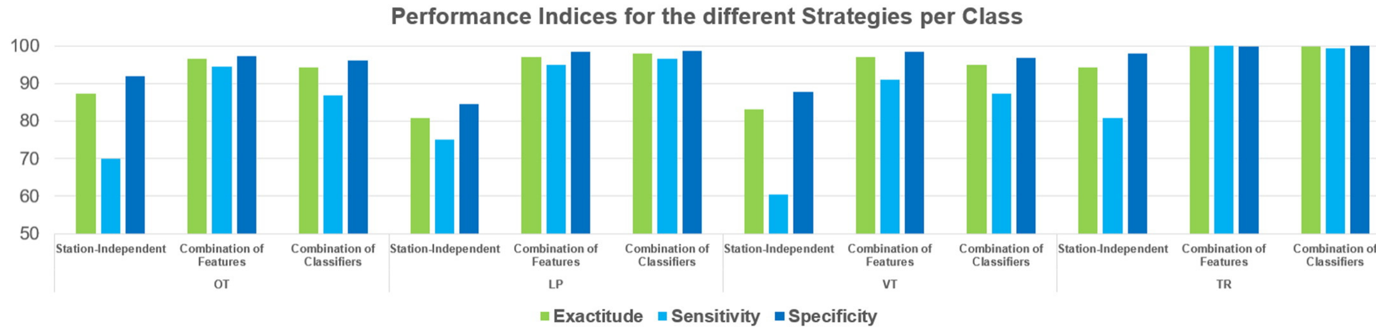

Fig. 9. Comparison of the exactitude, sensitivity, and specificity for each kind of event (OT, LP, VT, TR) for previous work, combination of features, and combination of classifiers.

is diminished. The results of the combination strategies presented in Tables 6 and 10 show that the capacity to discriminate between these events increases the sensitivity of the classifiers significantly: the first strategy presents net increases of 24.7% and 30.6% for OT and VT, respectively, while with the second strategy, the increases are 16.9% and 26.9%. This notable improvement, mainly in the combination of features strategy, may be explained as follows. In the case of a LP from a specific source, which is recorded in different stations around the volcano, some of the seismic features will be observed in more than one station preserving common features (e.g. frequency information of the source, arrival times, and magnitudes). The first strategy combines the features from several stations before going on to the classification stage, thus the classifier has information from each station in order to make a decision and classify the event correctly. The second strategy combines the decisions of the individual classifiers of each station. As these analyse the information of only one station each, the decision of the classifiers is based on less information than in the first strategy, resulting in lower performance.

In terms of sensitivity, the OT and VT classifiers present the lower performances (Fig. 9). This is due mainly to the fact that the OT group includes seismic events of different origins (tectonic, avalanches, environmental noise, etc.) which tend to be confused with VT due to their generation mechanism, form, and frequency content, especially in the LAV and LLA stations (Table 7).

Table 8 shows that MOT is the station which presents the lowest sensitivity results, especially for VT events. The results obtained with the new combination strategies show that the sensitivity increased from 60.45% (Table 4) to 91% (Table 6) with the first combination strategy and to 87.3% (Table 10) with the second. There are several reasons for this. Some of the VT seismic events present an epicentral distribution to the south of the volcano (Franco et al., 2014), enabling nearer stations to record the features of these events better. It commonly occurs that when the station is farther from the source, a loss occurs in the energy of high frequencies in the seismic records; thus in stations far from the source, VT signals may be confused with LP signals, as can be seen in the VT line of Table 7, specifically in the MOT station. Apart from this aspect related to source-reception distance, VT signals produced mainly at the south of the crater have to cross the volcano to be recorded by MOT, passing through low velocity zones (Bouvet de Maisonneuve et al., 2012) in which their spectral features are modified. This, added to the low magnitudes of VT, makes it more likely that this class will be confused with LP events, as the MOT station shows. It remains to be determined why VT events present a lower sensitivity (82%) in the LLA station, located at the south of the volcano, compared to the LAV station (94%) located farther away from the source. Some aspects directly linked to the generation mechanism, seismic wave radiation pattern,s and the directivity of the source may explain this fact. Thus, the poor performance of the MOT station classifier affects the performance of the second combination strategy. LP events are more restricted to sectors which are close to the main crater and more equidistant from the three stations, so their readings are consistent in the three stations considered in the study.

The first strategy, using the merging of the features extracted from the three stations before the classification stage, presented better performance. This strategy is closest to that used by a human expert, since it includes information from several stations before taking a decision (Duin et al., 2010). However, there is an important consideration for the implementation of this strategy: as the number of stations

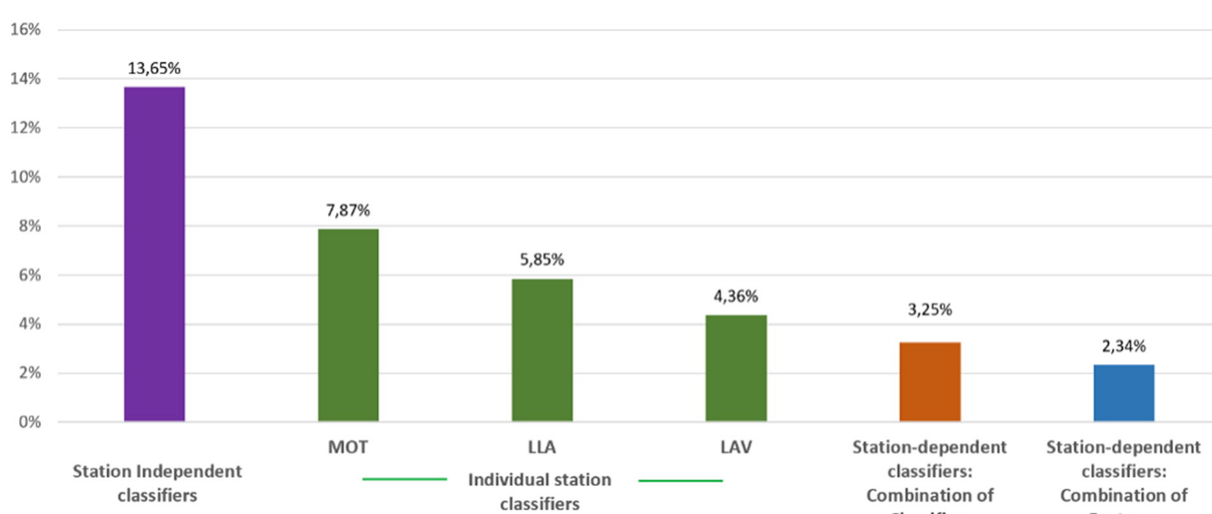

Fig. 10. Mean of the error index for the four classes of the classifying structures for previous work, individual classifiers for each station, combination of classifiers, and combination of features.

increases, the number of inputs of the SVM will increase proportionally, thus a bigger set of data will be necessary for training and validation, to handle the complexity of the resulting model. However, a system might be conceived which, by incorporating information on the epicentres of events, could determine which stations should participate in the strategy and which should be excluded, thus limiting the number of input features. Many configurations have to be trained separately in that case. Another scenario to analyse is when one or more stations are down. In this case, the classification is directly affected because one or more inputs will be missing. Different strategies have to be implemented to assess these scenarios. For example, training the classifiers to prevent missing inputs (SVM are robust to missing or incorrect input data) or selecting from the many trained configurations the ones that have the most reliable inputs. In the future, we plan to perform a study on the robustness of the system by conducting tests in different scenarios including station failures.

6. Conclusions

This paper presented a comparative study of the performance of automatic classifiers of seismic events in Llaima volcano. Two classification strategies which combine information from different seismic stations are compared with a station-independent classification strategy. The results show a significant improvement in classification performance when an approach of combi\ning information from the stations is used. In particular, a significant improvement is achieved in discriminating between VT and OT events.

The results show that the strategy which merges the features before classification proved to be superior to the strategy of merging the results of the individual classifiers by station. It is interesting to note that the first strategy is closer to what a human analyst does when classifying a seismic event, because it uses the information from several stations to make a decision. The second strategy depends on the performance of individual classifiers and it is interesting to see that some stations systematically produce a better performance than others.

The performance improvement of the station-dependent strategies is significant, but direct comparison with other methods is difficult because of the use of different volcanoes, features, type of events, and methods. The exactitude and error values obtained in this work are not significantly higher than those available in literature. What is very attractive in this work is that the high performances achieved for the Llaima volcano were obtained using very simple features, methods, and combination strategies. This is why, the study showed a very interesting approach for further improvements: integrating all the stations into a global automatic classifying system per volcano.

Some future works are proposed based on these two strategies to improve the robustness of the whole system. For the first strategy, we propose to assess which stations deliver the best results in relation to the epicentres of events; for the second, we propose a study of the “quality” of the stations, in order to give some stations priority in the combination process, discarding stations which might have a negative influence on the final classification. Another important future work is to make the classification system function on-line. An automatic event segmentation system needs to be implemented, based on continuous recordings, in order to apply the automatic classification system proposed here to the selected segments.

Acknowledgments

The authors would like to thank DIUFRO10-0020 project supported by the Dirección de Investigación at the Universidad de La Frontera, project CONICYT-PIA ANILLO ACT 1120 and CONICYT-FONDEF IDeA CA13I10273for financing the present work. Also many thanks to the support of the Project STIC-AmSud 15STIC-06 and to OVDAS, who provided the data and geological knowledge to perform the simulations and analyse the results.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jvolgeores.2016.02.006.

References

Álvarez, I., García, L., Cortés, G., Benítez, C., De la Torre, Á., 2012. Discriminative feature selection for automatic classification of volcano-seismic signals. IEEE Geosci. Remote Sens. Lett. 9 (2), 151–155.

Beyreuther, M., Carniel, R., Wassermann, J., 2008. Continuous hidden Markov models: application to automatic earthquake detection and classification at Las Canãdas caldera, Tenerife. J. Volcanol. Geotherm. Res. 176 (4), 513–518.

Bicego, M., Londoño-Bonilla, J.M., Orozco-Alzate, M., 2015. Volcano-seismic events classification using document classification strategies. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 9279, pp. 119–129.

Bouvet de Maisonneuve, C., Dungan, M.A., Bachmann, O., Burgisser, A., 2012. Insights into shallow magma storage and crystallization at Llaima volcano (andean southern volcanic zone, Chile). J. Volcanol. Geotherm. Res. 211-212, 76–91.

Carniel, R., 2014. Characterization of volcanic regimes and identification of significant transitions using geophysical data: a review. Bull. Volcanol. 76 (8), 1–22.

Carniel, R., Jolly, A.D., Barbui, L., 2013. Analysis of phreatic events at Ruapehu volcano, New Zealand using a new SOM approach. J. Volcanol. Geotherm. Res. 254, 69–79.

Cortés, G., García, L., Álvarez, I., Benítez, C., de la Torre, T., Ibáñez, J., 2014. Parallel system architecture (PSA): An efficient approach for automatic recognition of volcano-seismic events. J. Volcanol. Geotherm. Res. 271, 1–10.

Cortés, G., Benitez, M.C., Garcia, L., Alvarez, I., Ibanez, J.M., 2015. A comparative study of dimensionality reduction algorithms applied to volcano-seismic signals. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. http://dx.doi.org/10.1109/JSTARS.2015.2479300.

Curilem, G., Vergara, J., Fuentealba, G., Acuña, G., Chacón, M., 2009. Classification of seismic signals at Villarrica volcano (Chile) using neural networks and genetic algorithms. J. Volcanol. Geotherm. Res. 180, 1–8.

Curilem, M., Vergara, J., San, Martin C., Fuentealba, G., Cardona, C., Huenupan, F., Chacón, M., Khan, S., Hussein, W., Becerra, N., 2014a. Pattern recognition applied to seismic signals of the Llaima Volcano (Chile): an analysis of the events' features. J. Volcanol. Geotherm. Res. 282, 134–177.

Curilem, M., Huenupan, F., San, Martin C., Fuentealba, G., Cardona, C., Franco, L., Acuña, G., Chacón, M., 2014b. Feature analysis for the classification of volcanic seismic events using support vector machines. In: Gelbukh, A., et al. (Eds.), MICAI 2014, Part II, Lecture Notes in Artificial Intelligence, LNAI 8857. © Springer International Publishing Switzerland, pp. 160–171.

Chouet, B.A., 1996. Long-period volcano seismicity: its source and use in eruption forecasting. Nature 380 (6572), 309–316.

Chouet, B.A., Matoza, R.S., 2013. A multi-decadal view of seismic methods for detecting precursors of magma movement and eruption. J. Volcanol. Geotherm. Res. 252, 108–175.

Duin, R.P.W., Orozco-Alzate, M., Londoño-Bonilla, J.M., 2010. Classification of volcano events observed by multiple seismic stations. Paper Presented at the Proceedings - International Conference on Pattern Recognition, pp. 1052–1055.

Esposito, A.M., D'Auria, L., Giudicepietro, F., Peluso, R., Martini, M., 2013. Automatic recognition of landslides based on neural network analysis of seismic signals: an application to the monitoring of stromboli volcano (Southern Italy). Pure Appl. Geophys. 170 (11), 1821–1832.

Franco, L.E., Palma, J.L., Gil-Cruz, F., San Martín, J.J., 2014. La actividad sísmica asociada a erupciones estrombolianas violentas en el volcán Llaima, Chile (2007–2010). In III Congreso Latinoamericano de Sismología CLAS. Abstrac in Earth Sciences Research Journal 18 (Special Issue july 23–25, 2014 [ISSN 1794-6190]).

Giacco, F., Esposito, A.M., Scarpetta, S., Giudicepietro, F., Matinaro, M., 2009. Support vector machines and MLP for automatic classification of seismic signals at Stromboli Volcano. Proceedings of the 2009 Conference on Neural Nets WIRN09, pp. 116–123.

Hamel, L., 2009. Knowledge Discovery with Support Vector Machines. John Wiley & Sons.

Hsu, C.-W., Chang, C.-C., Lin, C.-J., 2003. A Practical Guide to Support Vector Classification. Department of Computer Science, National Taiwan University.

Huenupan, F., Yoma, N.B., Molina, C., Garreton, C., 2008. Confidence based multiple classifier fusion in speaker verification. Pattern Recogn. Lett. 29 (7), 957–966.

Ibáñez, J.M., Benítez, C., Gutiérrez, L.A., Cortés, G., García-Yeguas, A., Alguacil, G., 2009. The classification of seismo-volcanic signals using hidden Markov models as applied to the Stromboli and Etna volcanoes. J. Volcanol. Geotherm. Res. 187, 218–226.

Kalayeh, H.M., Landgrebe, D.A., 1983. Predicting the required number of training samples. IEEE Trans. Pattern Anal. Mach. Intell. PAMI-5 (6), 664–667.

Lahr, J.C., Chouet, B.A., Stephens, C.D., Power, J.A., Page, R.A., 1994. Earthquake classification, location and error analysis in a volcanic environment: implications for the magmatic system of the 1989-1990 eruptions at Redoubt Volcano, Alaska. J. Volcanol. Geotherm. Res. 62 (1–4), 137–151.

Landis, J.R., Koch, G.G., 1977. The measurement of observer agreement for categorical data. Biometrics 33, 159–174.

Langer, H., Falsaperla, S., Messina, A., Spampinato, S., Behncke, B., 2011. Detecting imminent eruptive activity at Mt Etna, Italy, in 2007–2008 through pattern classification of volcanic tremor data. J. Volcanol. Geotherm. Res. 200 (1–2), 1–17.

Masotti, M., Falsaperla, S., Langer, H., Spampinato, S., Campanini, R., 2006. Application of support vector machine to the classification of volcanic tremor at etna, Italy. Geophys. Res. Lett. 33 (20).

Mitchell, T., 1997. Machine Learning. McGraw Hill.

Mora-Stock, C., Thorwart, M., Wunderlich, T., Bredemeyer, S., Hansteen, T.H., Rabbel, W., 2012. Comparison of seismic activity for Llaima and Villarrica volcanoes prior to and after the Maule 2010 earthquake. International Journal of Earth Sciences 1–14.

Naranjo, J.A., Moreno, H. 2005. Geología del volcán Llaima, Región de la Araucanía. *Servicio Nacional de Geología y Minería*, Carta Geológica de Chile, Serie Geología Básica, No. 88, 33p., 1 mapa escala 1:50.000, Santiago, Chile.

Newhall, C.G., Self, S., 1982. The volcanic explosivity index (VEI): an estimate of explosive magnitude for historical volcanism. J. Geophys. Res. 87, 1231–1238.

Oppenheim, A.V., Schafer, R.W., R. B.J., 1999. Discrete-Time Signal Processing. Prentice Hall, Upper Saddle River, New Jersey.

Rouland, D., Legrand, D., Zhizhin, M., Vergniolle, S., 2009. Automatic detection and discrimination of volcanic tremors and tectonic earthquakes: an application to Ambrym Volcano, Vanuatu. J. Volcanol. Geotherm. Res. 181 (3–4), 196–206.

Scarpetta, S., Giudicepietro, F., Ezin, E.C., Petrosino, S., Del Pezzo, E., Martín, M., M. M., 2005. Automatic classification of seismic signals at Mt Vesuvius Volcano, Italy, using neural networks. Bull. Seismol. Soc. Am. 95 (1), 185–196.

Stern, 2004. Active Andean volcanism: its geologic and tectonic setting. Rev. Geol. Chile 31 (2), 161–206.

SWARM, 2011. Seismic wave analysis and real-time Monitor: user manual and reference guide, February 6. USGS, USA (http://volcanoes.usgs.gov/software/swarm/index.php).

Vapnik, V., 1995. The Nature of Statistical Learning Theory. Springer Verlag.

Witten, I.H., Frank, E., Hall, M.A., 2011. Data Mining: Practical Machine Learning Tools and Techniques. third ed. (Morgan Kaufmann Publishers. Elsevier, Chapter 5).

Yoma, N.B., Carrasco, J., Molina, C., 2005. Bayes-based confidence measure in speech recognition. IEEE Signal Process Lett. 12 (11), 745–748.

Zobin, V.M., 2012. Seismic Monitoring of Volcanic Activity and Forecasting of Volcanic Eruptions, Introduction to Volcanic Seismology. 407-431. Elsevier.