J Seismol
https://doi.org/10.1007/s10950-018-9810-5

ORIGINAL ARTICLE

Automatic discrimination of earthquakes and quarry blasts using wavelet filter bank and support vector machine

Omar M. Saad · Ahmed Shalaby · Mohammed S. Sayed

Received: 16 January 2018 / Accepted: 4 December 2018

Abstract False discrimination between earthquakes and quarry blasts may lead to an unrealistic characterization of the natural seismicity of a region. The similarity in seismograms between earthquakes and quarry blasts is the primary reason for incorrect discrimination. Therefore, in this paper, we propose a discriminative algorithm utilizing wavelet filter bank to extract unique features between earthquakes and quarry blasts. The discriminative features are found to be in the first five seconds after the onset time. The proposed algorithm is divided into two stages: first, wavelet filter bank extracts the features of the seismic signals; then, support vector machine classifies the event based on these extracted features. The proposed algorithm achieves a discrimination accuracy of 98.5% when applied to 900 earthquakes and quarry blast waveforms.

O. M. Saad (✉)
National Research Institute of Astronomy and Geophysics (NRIAG), Helwan 11731, Egypt
e-mail: omar.saad@ejust.edu.eg

A. Shalaby
CS, Faculty of Computers and Informatics, Benha University, Banha 13511, Egypt

A. Shalaby · M. S. Sayed
ECE, E-JUST University, Alexandria 21934, Egypt

M. S. Sayed
ECE, Zagazig University, Zagazig 44519, Egypt

Keywords Seismic data classification · Wavelet filter bank · Particle swarm optimization · Support vector machine

1 Introduction

Earthquake catalogs usually include a complex mixture of natural and artifacts events. In such catalogs, artifacts can cause variations in seismicity patterns due to misinterpretation as precursors to past or upcoming main shocks. Hence, before any statistical analysis of the seismicity, artificial contaminations must be explored and discriminated to extract meaningful information (Mousavi 2017). The main challenge of the discrimination between earthquakes and quarry blasts is the similarity in waveforms. Manual discrimination is a time-consuming process, and in some cases, it becomes unreliable. Therefore, an automatic discrimination algorithm with high accuracy is mandatory.

Many techniques are proposed to discriminate between earthquakes and quarry blasts automatically. Taylor et al. (1989) adopted the ratio between the maximum amplitudes of the secondary and the primary waves $\left(R_{(S/P)}\right)$, while Kuyuk et al. (2014) utilized Linear Discriminant Function (LDF) and Quadratic Discriminant Function (QDF) as a basis for classification. K-mean and Gaussian mixture are proposed as a classifier instead of LDF and QDF by Kuyuk et al. (2012). Moreover, Gendron et al. (2000) proposed discrimination based on the time-frequency contents of P-wave and S-wave extracted by discrete wavelet transform (DWT). Gendron et al. (2000) adopted the posterior probability of the time difference between P- and S-waves. The discrimination was accomplished by computing the posterior probability of these features for every class, then choosing the maximum posterior.

Moreover, time-frequency content analysis of the seismic events was adopted to extract distinctive discrimination features. In Horasan et al. (2009), the complexity (CX) versus the spectral ratio (SR) of the seismogram methodology was estimated as basic features for classification. A new figure of merit defined as the product of the ($R_{(S/P)}$), the ratio of complexity and the spectral ratio was proposed in Kekovalı et al. (2012), where it was named the power of the event. While short-time Fourier transform (STFT) was utilized to support discrimination based on the complexity and spectral ratio (Yılmaz et al. 2013). Furthermore, in Farahani and Zaré (2014), the proportion between the spectrums of the horizontal and the vertical components was applied as an indicator of the discrimination, given the same station.

On the other hand, artificial networks were proposed to discriminate between earthquakes and quarry blasts. In Orlic and Loncaric (2010), a genetic algorithm (GA) was used as an optimization tool for a classifier utilizing the ($R_{(S/P)}$), complexity, and spectral ratio parameters. In Akhouayri et al. (2015), a fuzzy-based algorithm was proposed to extract the features of the seismic data. Meanwhile, Yıldırım E et al. (2011) took the advantage of the feed-forward neural network and adaptive neural fuzzy. In Farahani (2015), the neuro-fuzzy inference system (ANFIS) was utilized. Statistical time series classifier based on hidden Markov model tool was introduced in Quang et al. (2015) and Beyreuther et al. (2012). In Lyubushin et al. (2013), multi-fractal singularity spectral was used to extract some features which can characterize earthquakes and quarry blasts, while the authors in Kortström et al. (2016) adopted support vector machine (SVM) for discrimination. They filtered the seismic wave via many narrow band pass filters and divided them into four phase windows: P, Pcoda, S, and Scoda, then computed a short-term average (STA) to use them for training the SVM.

Most of the previous algorithms utilize certain features such as ($R_{(S/P)}$), the spectral ratio, complexity, S-wave time, and STFT. On the contrary, this paper proposes to extract unique features employing wavelet filter bank. The proposed algorithm depends on the energy distribution of the seismic signal on the time-frequency representation. The wavelet filter bank is utilized to obtain the frequency contents in a specific high-resolution time window; then, unique features are extracted and employed for the discrimination process. Particle swarm optimization (PSO) is utilized to determine the optimum detail/approximation and the optimum time window, in which unique discrimination features are found, after the onset time. Next, SVM is used as a classifier between earthquakes and quarry blasts. The rest of the paper is organized as follows. In Section 2, the proposed algorithm is illustrated. Section 3 shows the results and analysis. The discussion is presented in Section 4. Finally, Section 5 concludes the paper.

2 The proposed algorithm

In the proposed algorithm, we employ the wavelet filter bank to reach a high resolution in the time-frequency representation. Wavelet filter bank is a powerful tool to get the frequency contents for a particular time window. The wavelet transform performs constant relative bandwidth frequency analysis because it uses a short-time window for high frequencies and a long-time window for low frequencies. This feature is one of the benefits of using wavelet transform over STFT because the signal is analyzed under different resolutions at different frequencies so wavelet can obtain better representation for the signals with low-frequency (approximation) and high-frequency (detail) contents (Mousavi and Langston 2016, 2017). Therefore, a wavelet filter bank is used to obtain the jth detail and approximation for the recorded seismic signals. Afterwards, the PSO is applied to choose the best detail or approximation which contains unique features that can distinguish between earthquakes and quarry blasts. Finally, SVM is utilized as a classifier to classify the waveforms according to these extracted features.

Wavelet filter bank divides the input signal into several frequency bands (j) (Strang and Nguyen 1996). The input signal, x, passes through the wavelet filter bank and is scaled to reach the jth detail (high frequency) and jth approximation (low frequency). The frequency content is halved at each stage. For instance, in our dataset, the sampling rate is 100 samples per second. Therefore; the general equation of the frequency band of the jth stage can be defined as follows (Percival and Walden 2006) :

$$\frac{F}{2^{j+1}} \leq |f| \leq \frac{F}{2^j} \qquad j = 1, 2, 3, ... \tag{1}$$

In Eq. 1, F is 50 Hz and j is the stage number. Scaling and wavelet coefficients for the jth stage are shown in Eqs. 2 and 3, respectively (Percival and Walden 2006).

$$v_{j,t} = \sum_{l=0}^{L-1} g_{j,l}\, x_{2j(t+1)-l mod L} \tag{2}$$

$$w_{j,t} = \sum_{l=0}^{L-1} h_{j,l}\, x_{2j(t+1)-l mod L} \tag{3}$$

In both Eqs. 2 and 3, $t = 0, ..., L/2 - 1$ meanwhile g and h are quadrature mirror filters, and L is the filter width while l varies from 0 to $L - 1$. The scaling filter g is a quadrature mirror filter that is related to the wavelet filter h through the equation, $g_l = (-1)^{l+1} h_{L-1-l}$ and the inverse relationship is $h_l = (-1)^l g_{L-1-l}$. Meanwhile, x represents the input signal. Also, the approximation and detail for the jth stage are defined in Eqs. 4 and 5, respectively (Strang and Nguyen 1996).

$$A_j = \sum_{l=0}^{L-1} g^o_{j,l}\, v_{j,(t+1) mod L} \tag{4}$$

$$D_j = \sum_{l=0}^{L-1} h^o_{j,l}\, w_{j,(t+1) mod L} \tag{5}$$

In Eqs. 4 and 5, g^o and h^o are g and h periodized with length L, respectively (Percival and Walden 2006). Periodized means creating an infinite length signal from a finite length signal. In other words, periodized is replicating a finite length signal over and over to create an infinite length periodic version. A detailed definition of the periodized filters is presented by Percival in Percival and Walden (2006).

The proposed algorithm is based on the hypothesis that earthquake and quarry blasts have different spectral characteristics. This phenomenon can be due to different source mechanisms of earthquakes and quarry blasts (Mousavi et al. 2016) . Our objective is to find unique features in earthquakes not shared with quarry blasts or vice versa. Our intuition that these features are in a certain period, this period starts at the onset time (t_0) and ends at a certain time which we call (t_2). Most of the previously published approaches depend on extracting different features from low and high frequencies (Horasan et al. 2009; Kekovalı et al. 2012; Yılmaz et al. 2013; Farahani and Zaré 2014). In this paper, for better representation of the seismic signals, we define a new parameter referred to as wavelet discriminator (WD), Eq. 6, to obtain the frequency contents in a specific time window with high resolution. Six stages of the wavelet filter bank are used to extract more features from allowable frequency range. In Eq. 6, $Q_{j,k}$ represents the jth detail or approximation according to the value of k, detail in case of k equals one or approximation when k equals 0, and $Q_{i,n}$ represents the ith detail or approximation according to the value of n, detail in case of n equals one or approximation when n equals 0.

Meanwhile, WD represents the ratio between (1) the integral of the square Q for jth stage in the time window [the onset time (t_0) $- t_1$] and (2) the integral of the square Q for ith stage in the time window [$t_1 - t_2$]. Also, Q belongs to $\{A_1, A_2, A_3, A_4, A_5, A_6, D_1, D_2, D_3, D_4, D_5, D_6\}$; A and D represent the approximation and detail, respectively. We anticipate that some of the discriminating features are located in low frequency while the others are located in high frequency. Therefore, we expect the stage number j represents the high-frequency band, while the stage number i represents the low-frequency band or vice versa. However, we do not know from which stage we should pick the Q. Moreover, we do not know the optimum limits of the time windows t_1 and t_2. Therefore, an optimization process is required.

$$WD = \frac{\int_{t_1}^{t_2} Q^2_{j,k}}{\int_{t_0}^{t_1} Q^2_{i,n}}$$
$$Q \in \{A_1, A_2, A_3, A_4, A_5, A_6, D_1, D_2, D_3, D_4, D_5, D_6\} \tag{6}$$

Six parameters have to be obtained; these parameters are shown in Table 1. The parameters are the two-time limits t_1 and t_2, k for the numerator, n for the denominator, the stage level of the numerator (j) and denominator (i). Optimum values for these parameters are needed to reach a robust discrimination technique. Therefore, we utilize particle swarm optimization (PSO) to obtain the optimum parameters for

Table 1 Input parameters of PSO

Parameters	Description	Constraints
t_1	Time limit 1	N/A
t_2	Time limit 2	N/A
j	Stage number for WD numerator	From 1 to 6
i	Stage number for WD denominator	From 1 to 6
k	jth Det.or Approx	If k=1, $Q_{j,k}$ is Det. If k=0, then $Q_{j,k}$ is Approx
n	ith Det.or Approx	If n=1, $Q_{i,n}$ is Det. If n=0, then $Q_{i,n}$ is Approx

the proposed algorithm. The objective function for the PSO of training set is to minimize the number of misclassified waveforms for earthquakes and quarry blasts.

PSO (Shi and Eberhart 1999) is widely used technique in many applications due to its advantages including simplicity, easy implementation and fast searching speed (Kennedy 2011; Hassan et al. 2005; Selvi and Umarani 2010; Bai 2010). Therefore, we propose to use PSO to find the optimum values for these parameters. The basic PSO algorithm consists of three steps generating positions of particles and velocities, velocity update, and position update (Selvi and Umarani 2010). Each particle represents a possible solution to the problem that changes its position from one iteration to another based on velocity updates. First, the positions, xp_i, and velocities, v_i, of the initial swarm of particles are randomly generated. The PSO consists of many particles which form a swarm, design space. At each step, each particle updates its velocity and distance according to Eqs. 7 and 8, respectively (Kennedy 2011).

$$v_i = v_i + c1 * rand * (P_i - xp_i) + c2 * rand * (P_{global} - xp_i) \quad (7)$$

$$xp_i = xp_i + v_i \quad (8)$$

In Eq. 8, the Pi represents the best previous position and the global best position is stored in P_{global}. The velocity update formula includes some random parameters ($rand$) to ensure excellent coverage of the design space. These random parameters are uniformly distributed. According to Shi and Eberhart (1999), the original PSO algorithm used the value of two for both constants $c1$ and $c2$. Once the PSO chooses the optimum parameter $t1$, $t2$, j, i, n, and k, these parameters are set to be the input parameters for the setup of the wavelet filter bank.

Finally, we propose to use support vector machine (SVM) as a classification tool to discriminate between earthquakes and quarry blasts. SVM is one of the standard tools for machine learning and data mining and a popular tool for a wide range of supervised pattern recognition problems (Cristianini and Shawe-Taylor 2000; Muller et al. 2001; Catanzaro et al. 2008; Schölkopf et al. 1999). SVM gives an unambiguous result for an ambiguous problem, which can easily be implemented into an automatic process. The SVM algorithm chooses the support vectors, points in the dataset, which can be used to classify the objects. SVM searches for the best support vector which reduces the misclassified points. In the proposed algorithm, we train the SVM with a linear kernel. The linear kernel is chosen because it is easier in implementation and faster than the nonlinear kernel (Tong and Chang 2001), given that the performance of the classifier will maintain the same result for the nonlinear kernel.

To evaluate SVM, we test the proposed algorithm using the dataset in Lyubushin et al. (2013). This dataset contains 143 waveforms: 75 for quarry blasts and 68 for earthquakes that should be discriminated with high accuracy. For the training dataset, we randomly select 30 waveforms: 15 waveforms for earthquakes and 15 waveforms for quarry blasts. These 30 waveforms represent 20% of the total dataset in Lyubushin et al. (2013). The two attributes of SVM are $log_{10}(WD)$ and $log_{10}(numerator\ of\ WD)$. These two attributes are used for training since they give the best discrimination accuracy between earthquakes and quarry blasts. Linear discriminant function for SVM is shown in Eq. 9 (Muller et al. 2001; Chi et al. 2008).

$$f(U) = w^T U + b \quad (9)$$

In Eq. 9, w, b, and U are the weight vector, bias, and input, respectively. The parameters w and b control the classification decision to specify the hyperplane of SVM. In the proposed algorithm, SVM aims to distinguish between the earthquakes and quarry blasts by minimizing the number of misclassified events. The sign of the weight is positive for support vectors belonging to the first group and negative for the second group (i.e., earthquakes and quarry blasts groups). SVM finds its optimum parameters, w, and b, by minimizing the objective function in Eq. 10 (Chi et al. 2008).

$$lossfunction = 0.5 * ||w||^2 + C * \sum_{i=1}^{n} H(y_i f(U_i)), \tag{10}$$

where C is regularization parameter (complexity constant) aies, and lower values create harder boundaries. SVM searches for a hyperplane with the largest minimum margin which can accurately separate many instances; this approach can be controlled by parameter C. The number of misclassified events at different C values is shown in Eq. 11. and n is the number of points in the training dataset. Meanwhile, $H(y_i f(U))$ is the loss for the training patterns x_i, defined by $H(index) = max(0, 1 - index)$ (Chi et al. 2008), where y_i is the desired output. In our case, $H(y_i f(U_i)) = max(0, 1 - y_i(W^T U_i + b))$. For SVM, we use soft-margin SVM which the C parameter control the boundaries of this margin, where higher C values allow for softer boundary.

$$Misclassified\ waveforms = \begin{cases} 3, & C = 0.1, 0.2, 0.3, 0.4 \\ 2, & C = 0.5, 0.6, .., 15 \end{cases} \tag{11}$$

In Eq. 11, C varies with fixed step size of 0.1. According to Eq. 11, the optimal value C is any value above 0.5. Therefore, we set the C parameter to its default value 1.

In our case, we use a linear kernel SVM as a classifier which can be modelled as shown in Eq. 12.

$$Out = Slope * Inp + b \tag{12}$$

In Eq. 12, Inp and Out are the input and the output of the linear classifier, respectively. According to SVM linear classifier, the value of $Slope$ and b are estimated. The pseudo-code for the proposed linear SVM is shown in Table 2. We multiply the first attribute, $log_{10}(numerator\ of\ WD)$, by the $Slope$ and subtract the second attribute, $log_{10}(WD)$, from the result of this product. The final decision of the system is one in case of an earthquake and a zero otherwise.

Figure 1 shows a flowchart summarizes the procedure of the proposed algorithm. According to Fig. 1, in the first iteration of PSO, the six parameters are randomly initialized, and then the WD values are obtained. Two attributes are fed to the SVM classifier: $log_{10}(WD)$ and $log_{10}(numerator\ of\ WD)$. Afterwards, the cost function is determined, the number of misclassified waveforms. For each iteration, the six parameters are updated according to PSO scheme; then, the newly updated parameters are used to obtain the cost function. If the cost function is less than the global minimum, the current parameters will be considered as the optimum parameters. Meanwhile, the cost function will reach 0 if the number of misclassified waveforms is 0. If this condition happens, the PSO will be stopped, but if this condition did not meet, the process would continue till the number of iterations reaches 1000, and the program will be halted.

Table 2 Pseudo-code for proposed linear support vector machine as a classifier

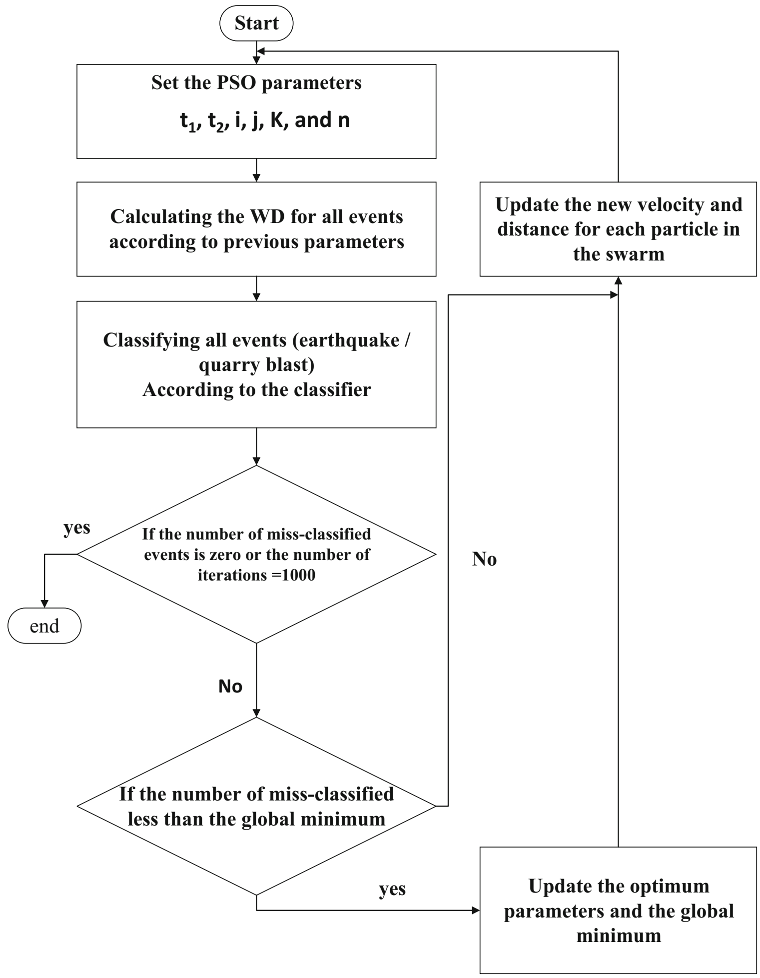
$attribute1 = log_{10}(numeratorofWD)$;
$attribute2 = log_{10}(WD)$;
//Perform linear SVM
$Yout = A * attribute1 + B$;
$PreDecision = Yout - attribute2$;
$IF\ \ PreDecision <= 0$
$Decision = 1$;
$Else$
$Decision = 0$;
$End\ IF$

3 Results and analysis

3.1 Description of dataset

In Egypt, Aswan is a high-activity region regarding the quarry blasts. The Egyptian National Seismic Network (ENSN) takes responsibility for monitoring, recording, and analyzing seismic data. In 2016, ENSN operated 69 seismic stations distributed all over Egypt to detect the occurrence of seismic ground motion. These seismic stations are gathered into groups according to their locations. For instance, Aswan region is a vital area because of the high dam and Nasser Lake. Therefore, Aswan sub-network consists of eleven seismic stations to detect and recognize quarry blasts for the safety of the region. We train and test the proposed algorithm using the dataset in Lyubushin et al. (2013). This dataset contains 68 earthquake waveforms and 75 quarry blasts wave-

Fig. 1 Flowchart of the PSO procedure



forms recorded by Aswan sub-network from 2004 to 2007. At that time, seismic stations were equipped with sensors which have vertical component only. Figure 2 shows the locations of the 11 seismic stations for Aswan sub-network and its corresponding longitude and latitude. In ENSN, each waveform of the events passes through several expert seismologists to decide either it is an earthquake or it is a quarry blast. This process takes hours, days, or maybe weeks. Therefore, we need a reliable discrimination algorithm to automatically act like the expert seismologists in just a few seconds.

3.2 First dataset discrimination results

The proposed algorithm is applied to the above-described dataset to find the optimum parameters for high discrimination accuracy. As a result, the optimum values for the six parameters mentioned in Table 1 are shown in Table 3. According to Table 3, the features that are required for an efficient discrimination exist in the first 5 s after the onset time. These features are found in details numbers 1 and 5 which represent the high-frequency band and the low-frequency band, respectively. According to Eq. 1, detail number 1 contains the frequency band of 12.5–25 Hz, while detail number 5 represents the frequency band of 0.7813–1.5625 Hz. These results comply with our hypothesis that there are some features in the high frequency and others in the low frequency in a certain time window that can discriminate between earthquakes and quarry blasts.

The WD equation can be written in the final formula as shown in Eq. 13, where $Q_{1,1}$ represents the first detail of wavelet filter bank, while $Q_{1,5}$ repre-

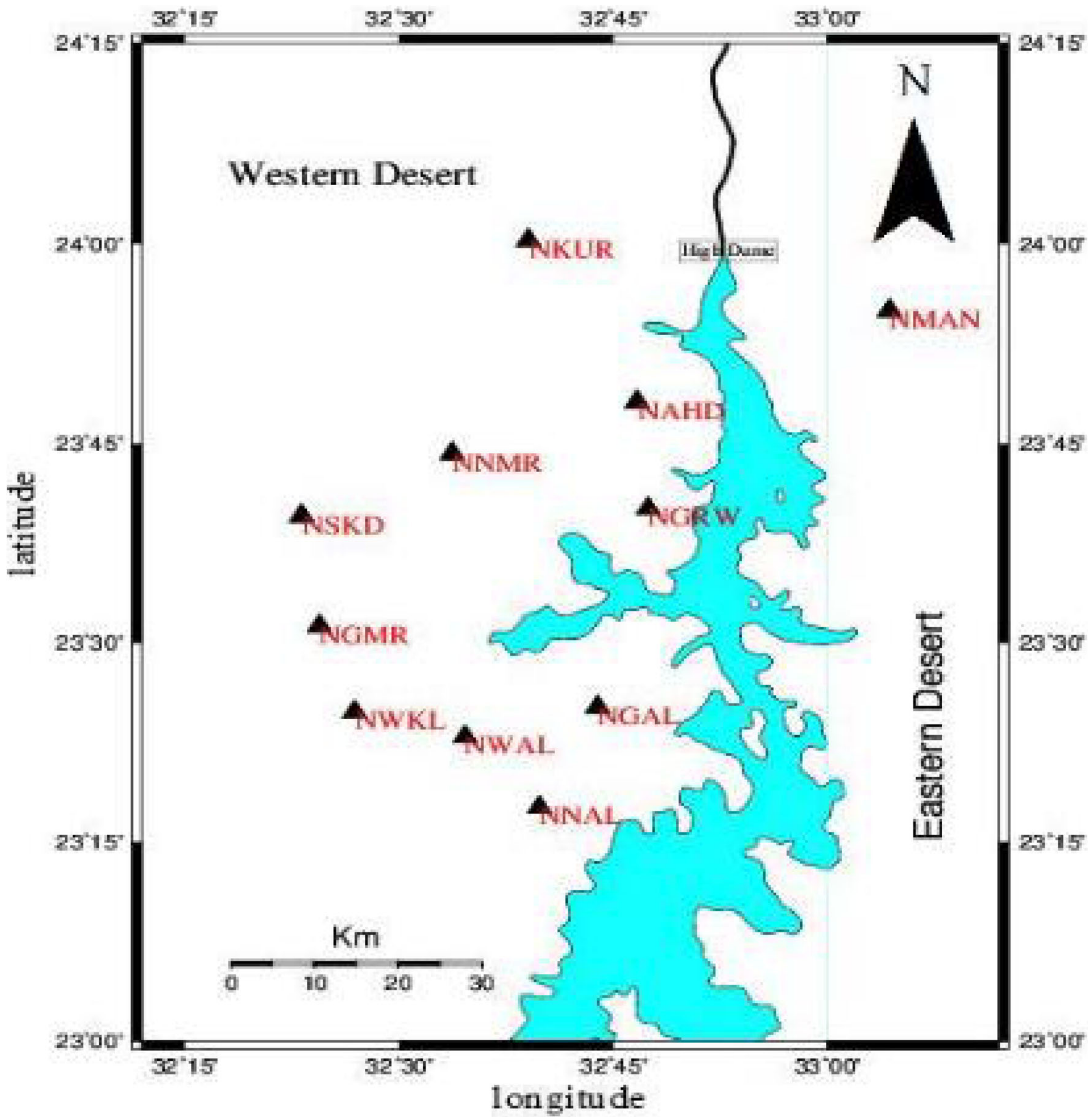


Fig. 2 Location of Aswan sub-network and its seismic stations

Table 3 The optimum input parameter values obtained by PSO.

Parameters	Values
t_1	3.5 s
t_2	5 s
j	1
i	5
k	1 (detail)
n	1 (detail)

sents the fifth detail of wavelet filter bank and t_0 is the onset time of the event.

$$WD = \frac{\int_{3.5s}^{5s} Q_{1,1}^2}{\int_{t_0}^{3.5s} Q_{5,1}^2} \tag{13}$$

Figures 3 and 4 show an example of the quarry blast and earthquake, respectively, with a similar shape. Figures 3a and 4a show a quarry blast and earthquake events (i.e., waveforms) with a similar shape in the time domain, respectively. The squared detail num-

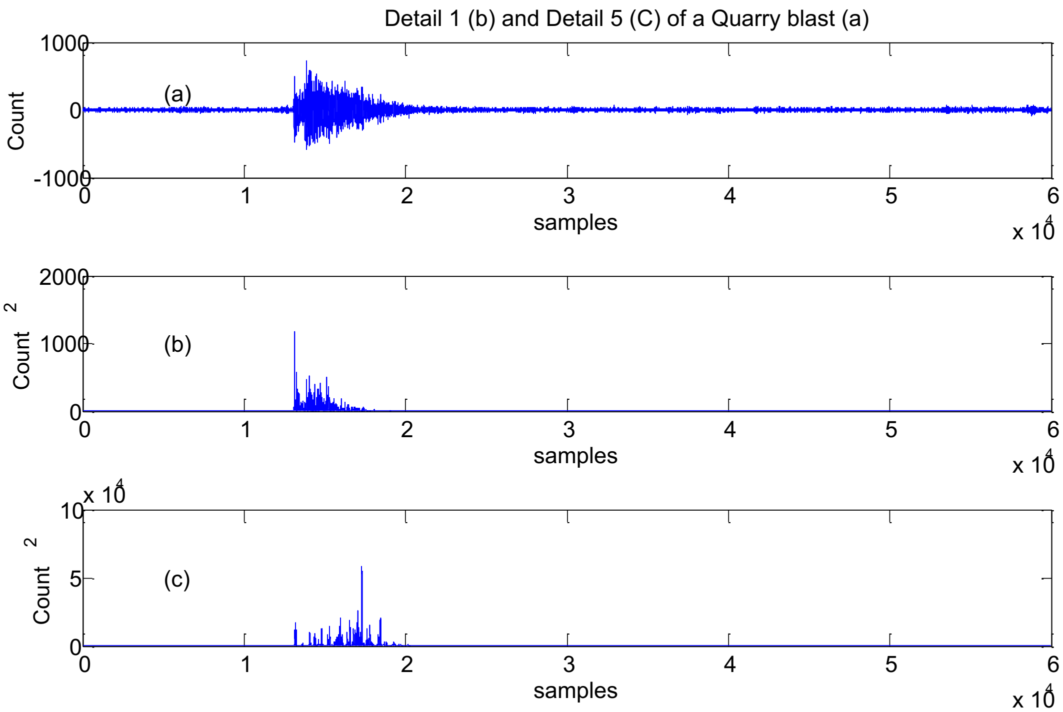


Fig. 3 **a** Quarry blast. **b** The detail number 1 of this quarry blast. **c** The detail number 5 of this quarry blast

ber 1 and squared detail number 5 of the quarry blast are shown in Fig. 3b and c, respectively. Meanwhile, the squared detail number 1 and squared detail number 5 of the earthquake are shown in Fig. 4b and c, respectively. For the same earthquake and quarry blast, Figs. 5 and 6 show detail number 1 and detail number 5 in the first 5 s after the onset time for quarry blast and earthquake waveforms, respectively.

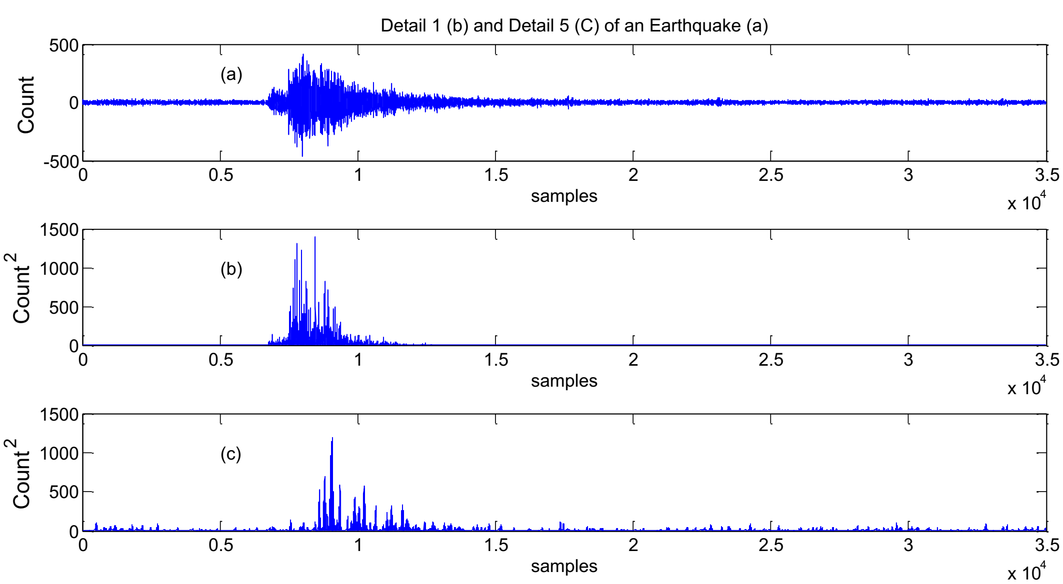


Fig. 4 **a** Earthquake. **b** The detail number 1 of this earthquake. **c** The detail number 5 of this earthquake

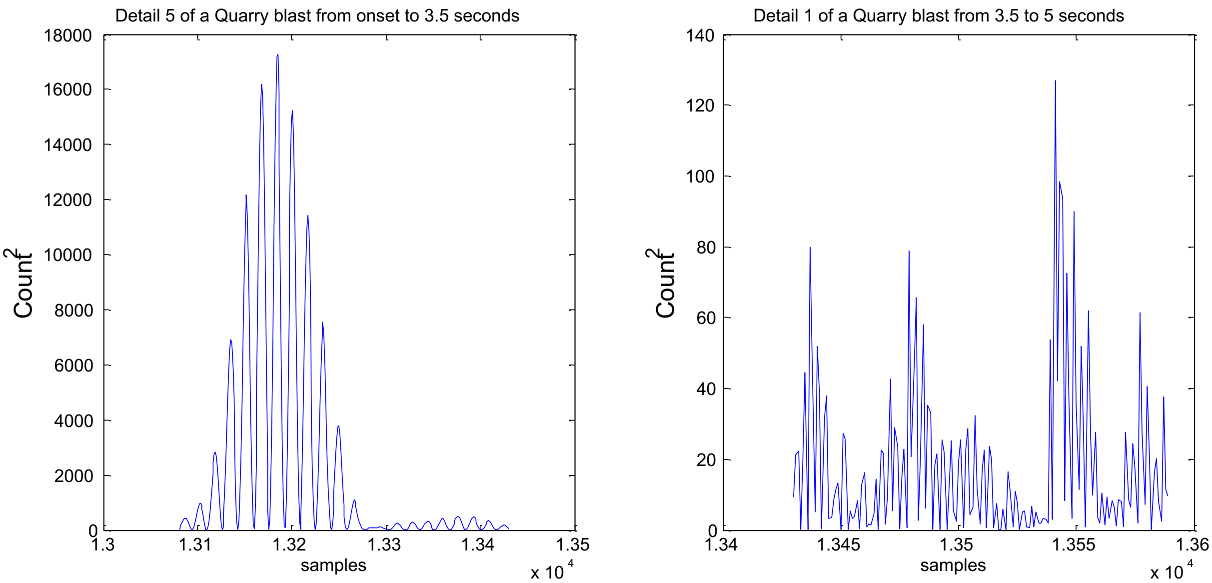


Fig. 5 Left hand side, the detail number 5 of a quarry blast from the onset time to 3.5 s. Right hand side, the detail number 1 of a quarry blast, in Fig. 3, from 3.5 to 5 s after the onset time

The left-hand side of Fig. 5 shows the squared detail number 5 of the quarry blast in a time window of 3.5 s starting from the onset time, while the right-hand side of Fig. 5 shows the squared detail number 1 for the same quarry blast from 3.5 to 5 s from the onset time. On the other hand, the time window of 3.5 s starting from the onset time for the detail number 5 of the earthquake is shown on the left-hand side of Fig. 6,

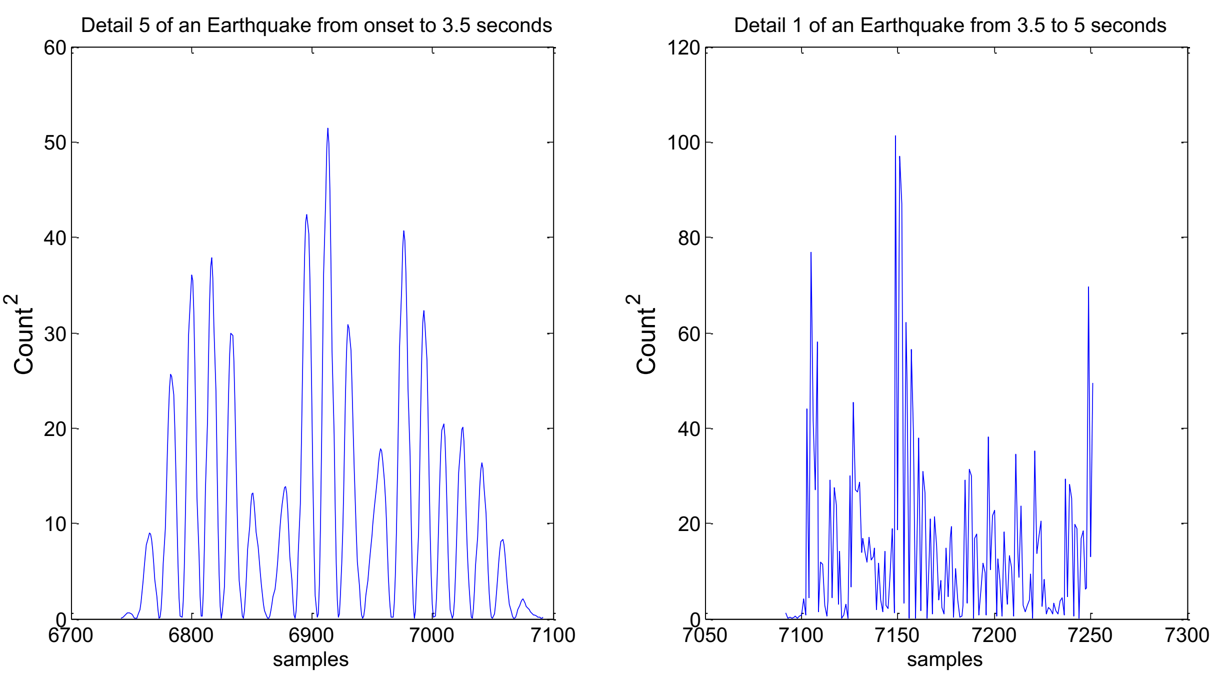


Fig. 6 Left hand side, the detail number 5 of an earthquake from the onset time to 3.5 s. Right hand side, the detail number 1 of an earthquake, in Fig. 4, from 3.5 s after onset to 5 s after the onset time

while the right-hand side of Fig. 6 represents the detail number 1 of the same earthquake from 3.5 to 5 s from the onset time. Furthermore, the integral value of the wavelet detail on the right-hand side for both Figs. 5 and 6 represent the numerator of WD, while the integral value of the wavelet detail on the left-hand side for both Figs. 5 and 6 represent the denominator of WD. Then, the WD values for quarry blast and earthquake can be obtained according to Eq. 13 from Figs. 5 and 6, respectively.

The major strength of the proposed algorithm is its ability to determine the energy of the seismic signal in a specific time window in a particular frequency band. We calculate this power by calculating the square value of detail numbers 5 and 1 in a time window of 5 s starting from the onset time. This segment contains features that can discriminate between earthquakes and quarry blasts. It appears that the energy of the quarry blasts for low-frequency content between the intervals from the onset time to 3.5 s is higher than the energy for the earthquake in the same interval. For example, the numerator value of WD for the quarry blast in Fig. 3a is 1420 which can be determined from the right-hand side of Fig. 5, while the denominator value of WD for the same quarry blast is 153398 which can be calculated from the left-hand side of Fig. 5. Also, the numerator and denominator of WD for the earthquake in Fig. 4 can be calculated from the right- and left-hand sides of Fig. 6, respectively. The numerator and denominator of WD for this earthquake are 578 and 11440, respectively. Hence, the values of WD for quarry blast and earthquake are 0.0092 and 0.051, respectively. Therefore, the proposed algorithm can easily identify the earthquake from the quarry blast, according to the value of WD. To make it clear, we calculate the average values for numerator and denominator of WD for all waveforms. For all waveforms of quarry blasts, the average value for the numerator of WD is 1223.9, while the average value for the denominator of WD is 10681.7. For all waveforms of earthquakes, the average value for the numerator of WD is 1500.4, while the average value for the denominator of WD is 150624.

After obtaining the WD values for all the waveforms in the dataset (i.e., 143 waveforms), we calculate the logarithm of WD for all waveforms and plot it versus the logarithm of the numerator of WD. The reason for selecting these two parameters, the logarithm of WD and the logarithm of the numerator of WD, is that we find they give the best discrimination accuracy between earthquakes and quarry blasts. In Fig. 7, the red circles represent quarry blasts while blue squares are the earthquakes. On the other hand, the misclassified waveforms appear as black stars in Fig. 7. By

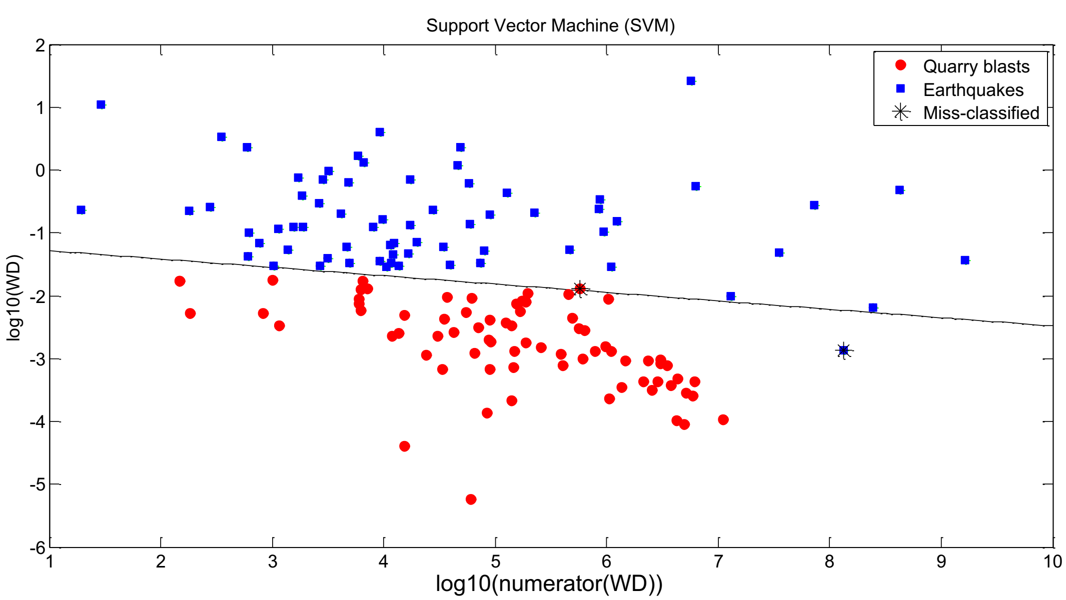


Fig. 7 Classification result according to the proposed algorithm

Table 4 The accuracy of different mother wavelet types and order

Mother wavelet type	Order	Accuracy
Daubechies	From 2 to 7	85 to 89%
Daubechies	8	98.5%
Daubechies	From 9 to 15	80 to 85%
Daubechies	From 16 to 45	< 80%
Haar	Haar	83%
Sym	From 1 to 45	80 to 89%
Bior1	.1, .3, and .5	75 to 80%
Bior2	All orders	< 80%
Bior3	All orders	< 80%
Bior4	.4	< 80%
Bior5	.5	< 80%
Bior6	.8	< 80%

using the SVM as a classifier, the discrimination accuracy is 98.5% and 98.6% for earthquakes and quarry blasts, respectively. In the proposed algorithm, we try several mother wavelets to reach maximum accuracy, and the best mother wavelet which gives the best discriminating features is Daubechies 8. Table 4 shows the accuracy of discrimination between earthquakes and quarry blasts using different mother Wavelets.

3.3 Testing dataset

We gathered most of the events which happened during 2014 in Aswan sub-network. Seven hundred fifty-seven waveforms from the 11 stations in Aswan sub-network are collected. This dataset consists of 574 waveforms for earthquakes and 183 waveforms are for quarry blasts. We choose Z-direction component because each station in the first dataset, from 2004 to 2007, was equipped by seismometer which has only one channel, vertical channel. In 2014, the seismometer used in the stations has three channels. However, to be more consistent and to keep the same features extracted from the training dataset, we use only the Z-component. From each station, we gathered

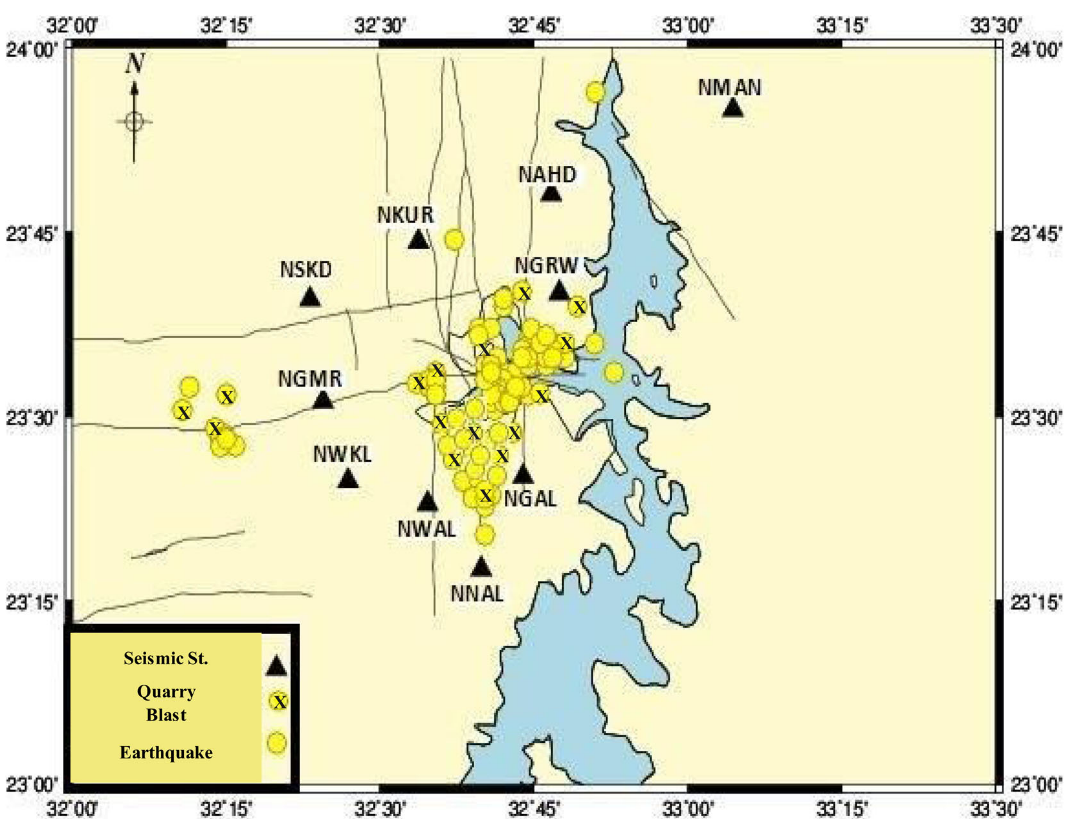


Fig. 8 Location of testing dataset waveforms and the stations of Aswan sub-network

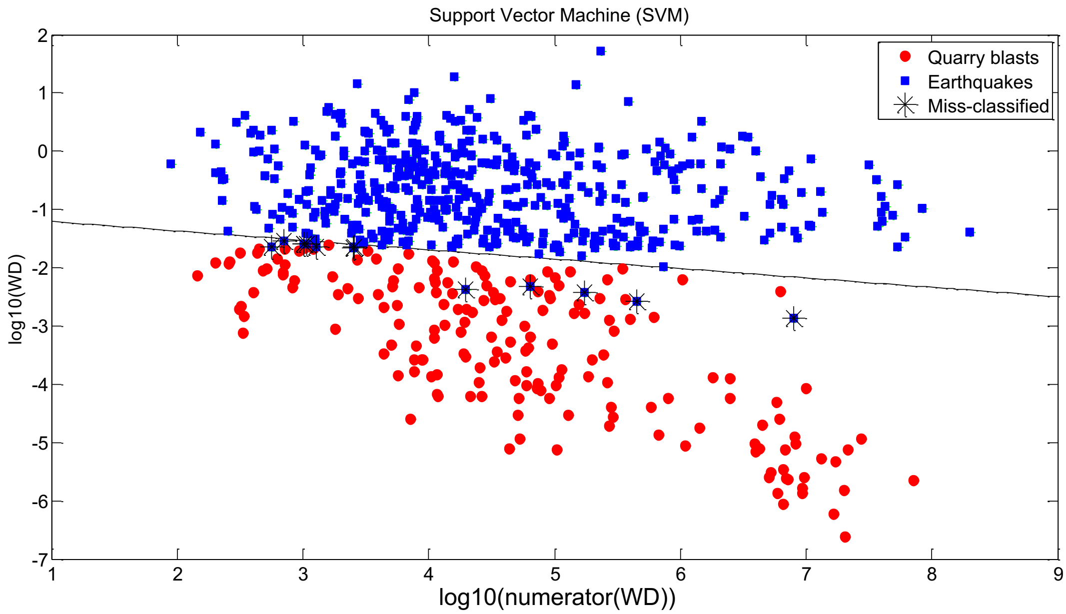


Fig. 9 Classification result of the test dataset using SVM

68 waveforms except the NGAL station; we gathered 77 waveforms. The locations of the events and stations are shown in Fig. 8. We employ the proposed algorithm to classify this dataset. For these events, the classification process cannot be reached using location and depth because both of location and depth of earthquakes and quarry blasts are similar. For this dataset, the depth range varies from 1 to 8 km while the magnitude, ML, varies from 0.8 to 4.2 ML. The WD values for all waveforms are determined and plotted versus the numerator of the WD. Then, the SVM classifier is applied to classify the earthquakes and quarry blasts. As a result, the discrimination accuracy of the proposed algorithm is 97.9% for classification of earthquakes (i.e., 12 waveforms for earthquakes are misclassified) and to 99.5% for classification of quarry blasts (i.e., 1 waveform for quarry blast is misclassified). Figure 9 shows the logarithm of WD versus the logarithm of WD numerator. In Fig. 9, quarry blasts are the red circles, and earthquakes are the blue squares while the black stars represent the misclassified waveforms. According to these results, there are several explosions that have occurred in the Aswan region. These explosions may be a quarry blast or any other mining activities and should be reported to the government. The proposed algorithm is an automatic discrimination tool which releases an alarm once a quarry blast is detected. The final proposed algorithm is shown in Fig. 10.

4 Discussion

In the proposed algorithm, for quarry blasts, the first 3.5 s after the onset time has larger energy than earthquakes, and this appears in the WD numerator values. However, once we get close to the S-wave, the energy of the earthquakes becomes larger than the quarry blasts. This phenomenon is reflected in the denominator of the WD which presents the energy stored in the interval from 3.5 to 5 s after the onset time. Furthermore, for quarry blasts, the low frequency is dominated for the first 3.5 s after the onset time. Meanwhile, for earthquakes, the low frequency is dominated in the interval from 3.5 to 5 s after the onset time. The results matched the fact that the energy of P-wave is greater than that of the S-wave for quarry blasts and, the energy of S-wave is greater than that of the P-wave for earthquakes.

The final proposed algorithm, shown in Fig. 10, utilizes wavelet filter bank and SVM classifier. The proposed algorithm achieves an accuracy of 98.3%. Table 5 compares between the proposed algorithm and previous algorithms (Horasan et al. 2009; Yılmaz

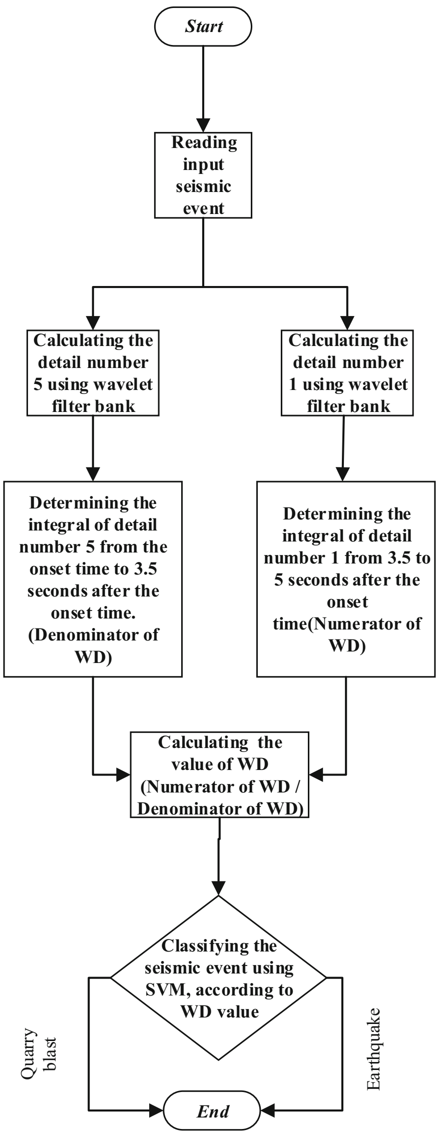


Fig. 10 Flowchart of the final proposed algorithm

et al. 2013; Lyubushin et al. 2013) when we apply them to our dataset. In Table 5, the accuracy and the number of waveforms are reported for each work. As observed, the proposed algorithm outperforms the other algorithms.

Moreover, the proposed algorithm is flexible and fully automated since the parameters are set automatically in the small initial training phase. For instance, the training dataset is 30 waveforms from a total of 143 waveforms. For Aswan region, to obtain the optimum parameters of the proposed algorithm, the training process including PSO process consumed time of 7708 s using Intel Processor Core i7-3612QM CPU @ 2.10 GHz, RAM of 8 Gb, and Windows 10, 64-bit operating system. To further evaluate the performance of the proposed algorithm, we obtain the confusion matrix for all the tested waveforms (i.e., 870 waveforms excluding the training dataset which contains 30 waveforms). The confusion matrix can be found in Eq. 14.

Table 5 The accuracy of different mother wavelet types and order

Method	Waveforms number	Accuracy
Horasan et al. (2009)	900	55%
Yılmaz et al. (2013)	900	77%
Lyubushin et al. (2013)	900	96%
This work	900	98.3%

$$\text{Confusion matrix} = \begin{bmatrix} 242 & 2 \\ 13 & 613 \end{bmatrix} \tag{14}$$

In the proposed algorithm, the true positive represents the correct classification of quarry blasts while the true negative represents the correct classification of earthquakes. Also, the false positive represents misclassification of earthquakes while the false negative represents misclassification of quarry blasts. From Eq. 14, the true positive (TP) is equal to 242 waveforms while the true negative (TN) is equal to 613 waveforms; the false positive (FP) is equal to 13 waveforms while the false negative (FN) is equal to 2 waveforms. Then, we obtain the sensitivity and specificity as shown in Eqs. 15 and 16, respectively.

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP+FN}} \tag{15}$$

Table 6 Confusion matrix, sensitivity, and specificity

Parameter	Value
True positive (TP)	242
True negative (TN)	613
False positive (FP)	13
True negative (FN)	2
Sensitivity $(\frac{\text{TP}}{\text{TP+FN}})$	99.1%
Specificity $(\frac{\text{TN}}{\text{TN+FP}})$	97.9%

$$\text{Specificity} = \frac{\text{TN}}{\text{TN+FP}} \quad (16)$$

According to Eqs. 14, 15, and 16, the sensitivity and specificity values are 99.1% and 97.9%, respectively as shown in Table 6.

5 Conclusion

Quarry blasts and earthquake signals are similar in their shape and are difficult to be distinguished. There are several drawbacks to misclassify quarry blast as an earthquake. For example, some analysis and models, such as crustal model, will be incorrect. This work proposes a discrimination algorithm utilizes a wavelet filter bank to extract unique features of quarry blasts and earthquakes signals. This is accomplished by calculating the squared detail number 5 and squared detail number 1 in a time window of 5 s starting from the onset time. Choosing detail numbers 1 and 5 as well as selecting the time window length is done by applying particle swarm optimization (PSO). Then, the SVM algorithm is used to classify between earthquakes and quarry blasts automatically based on the extracted features. The proposed algorithm is trained using the dataset in Lyubushin et al. (2013); 30 waveforms are used to train the proposed algorithm which represents 20% of the total dataset (i.e., 15 waveforms for earthquakes and 15 waveforms for quarry blasts). The accuracy of the proposed algorithm reaches 98.6%. Also, we gathered the most of the seismic events which happened during 2014 in Aswan sub-network. This dataset consists of 757 waveforms recorded by Egyptian National Seismic Network (ENSN). After applying the proposed algorithm to this dataset, the proposed algorithm reaches a discrimination accuracy of 98.5%.

Acknowledgements We would like to thank Egypt-Japan University of Science and Technology (E-JUST) for the continuous support. Also, we would like to thank the National Research Institute of Astronomy and Geophysics (NRIAG) for providing the seismic data used in this paper.

Funding information This work received funding from the Egyptian Ministry of Higher Education.

References

Akhouayri ES, Agliz D, Zonta D, Atmani A et al (2015) A fuzzy expert system for automatic seismic signal classification. Expert Syst Appl 42(3):1013–1027

Bai Q (2010) Analysis of particle swarm optimization algorithm. Comput Inf Sci 3(1):180

Beyreuther M, Hammer C, Wassermann J, Ohrnberger M, Megies T (2012) Constructing a hidden Markov model based earthquake detector: application to induced seismicity. Geophys J Int 189(1):602–610

Catanzaro B, Sundaram N, Keutzer K (2008) Fast support vector machine training and classification on graphics processors. In: Proceedings of the 25th international conference on machine learning, ACM, pp 104–111

Chi M, Feng R, Bruzzone L (2008) Classification of hyperspectral remote-sensing data with primal SVM for small-sized training dataset problem. Adv Space Res 41(11):1793–1799

Cristianini N, Shawe-Taylor J (2000) An introduction to support vector machines and other kernel-based learning methods. Cambridge University Press, Cambridge

Farahani JV (2015) Discrimination of quarry blasts and microearthquakes using adaptive neuro-fuzzy inference systems in the Tehran region. Episodes 38(3):162–168

Farahani JV, Zaré M (2014) Site characterizations for the Tehran network (tdmmo) in Tehran region using microearthquake, microtremor and quarry blast data. Soil Dyn Earthq Eng 63:235–247

Gendron P, Ebel J, Manolakis D (2000) Rapid joint detection and classification with wavelet bases via Bayes theorem. Bull Seismol Soc Am 90(3):764–774

Hassan R, Cohanim B, De Weck O, Venter G (2005) A comparison of particle swarm optimization and the genetic algorithm. In: Proceedings of the 1st AIAA multidisciplinary design optimization specialist conference, vol 18, p e21

Horasan G, Güney AB, Küsmezer A, Bekler F, Öğütçü Z, Musaoğlu N (2009) Contamination of seismicity catalogs by quarry blasts: an example from istanbul and its vicinity, northwestern Turkey. J Asian Earth Sci 34(1):90–99

Kekovalı K, Kalafat D, Deniz P (2012) Spectral discrimination between mining blasts and natural earthquakes: application to the vicinity of tunbilek mining area, western Turkey. Int J Phys Sci 7(35):5339–5352

Kennedy J (2011) Particle swarm optimization. In: Encyclopedia of machine learning, Springer, pp 760–766

Kortström J, Uski M, Tiira T (2016) Automatic classification of seismic events within a regional seismograph network. Computers & Geosciences 87:22–30

Kuyuk H, Yildirim E, Dogan E, Horasan G (2012) Application of k-means and gaussian mixture model for classification of seismic activities in istanbul. Nonlinear Process Geophys 19(4):411–419

Kuyuk HS, Yildirim E, Dogan E, Horasan G (2014) Clustering seismic activities using linear and nonlinear discriminant analysis. J Earth Sci 25(1):140

Lyubushin A, Kaláb Z, Lednická M, Haggag H (2013) Discrimination of earthquakes and explosions using multi-fractal singularity spectrums properties. J Seismol 17(3):975–983

Mousavi SM (2017) Comment on recent developments of the Middle East catalog by Zare et al. J Seismol 21(1):257–268

Mousavi SM, Langston CA (2016) Hybrid seismic denoising using higher-order statistics and improved wavelet block thresholding. Bull Seismol Soc Am 106(4):1380–1393

Mousavi SM, Langston CA (2017) Automatic noise-removal/signal-removal based on general cross-validation thresholding in synchrosqueezed domain and its application on earthquake data. Geophysics 82(4):V211–V227

Mousavi SM, Horton SP, Langston CA, Samei B (2016) Seismic features and automatic discrimination of deep and shallow induced-microearthquakes using neural network and logistic regression. Geophys J Int 207(1):29–46

Muller KR, Mika S, Ratsch G, Tsuda K, Scholkopf B (2001) An introduction to kernel-based learning algorithms. IEEE Trans Neural Netw 12(2):181–201

Orlic N, Loncaric S (2010) Earthquake—explosion discrimination using genetic algorithm-based boosting approach. Comput Geosci 36(2):179–185

Percival DB, Walden AT (2006) Wavelet methods for time series analysis, vol 4. Cambridge University Press, Cambridge

Quang PB, Gaillard P, Cano Y, Ulzibat M (2015) Detection and classification of seismic events with progressive multi-channel correlation and hidden Markov models. Comput Geosci 83:110–119

Schölkopf B, Burges CJ, Smola AJ (1999) Advances in kernel methods: support vector learning. MIT Press, Cambridge

Selvi V, Umarani DR (2010) Comparative analysis of ant colony and particle swarm optimization techniques. Int J Comput Appl 5(4):0975–8887

Shi Y, Eberhart RC (1999) Empirical study of particle swarm optimization. In: Proceedings of the 1999 congress on evolutionary computation, 1999. CEC 99. IEEE, vol 3, pp 1945-1950

Strang G, Nguyen T (1996) Wavelets and filter banks. SIAM, Philadelphia

Taylor SR, Denny MD, Vergino ES, Glaser RE (1989) Regional discrimination between nts explosions and western us earthquakes. Bull Seismol Soc Am 79(4):1142–1176

Tong S, Chang E (2001) Support vector machine active learning for image retrieval. In: Proceedings of the ninth ACM international conference on multimedia, ACM, pp 107–118

Yıldırım E, Gülbağ A, Horasan G, Doğan E (2011) Discrimination of quarry blasts and earthquakes in the vicinity of istanbul using soft computing techniques. Comput Geosci 37(9):1209–1217

Yılmaz Ş, Bayrak Y, Çınar H (2013) Discrimination of earthquakes and quarry blasts in the eastern Black Sea region of Turkey. J Seismol 17(2):721–734