

RESEARCH

Open Access



Enhancing analyst decisions for seismic source discrimination with an optimized learning model

Mohamed S. Abdalzaher^{1*}, Sayed S. R. Moustafa¹, W. Farid¹ and Mahmoud M. Salim^{2,3}

Abstract

Sustainable development in urban areas requires a wide variety of current and theme-based information for efficient management and planning. In addition, researching the spatial distribution of earthquake (EQ) clusters is an important step in reducing seismic risks and EQ losses through better assessment of seismic hazards, therefore it is desirable to acquire an uncontaminated database of seismic activity. Quarry blasts (QBs) conducted over the mapped area have tainted the seismicity inventory in the northwestern region of Egypt, which is the focus of this paper. Separating these QBs from the EQs is hence preferable for accurate seismicity and risk assessments. Consequently, we present a highly effective ML model for cleaning up the seismicity database, allowing for the accurate delineation of EQ clusters using data from a single seismic station, "AY7", which is part of the Egyptian National Seismic Network. The magnitudes ≤ 3 that are very uncertain as EQs or QBs and need a significant amount of time to analyze are the primary focus of the model. In order to find the best way to classify EQs and QBs, the method looks at a number of ML models before settling on the best one using eight features. The results show that the suggested method, which uses the quadratic discrimination analysis model for discriminating, successfully separates EQs and QBs with a 99.4% success rate.

Keywords Earthquakes, Quarry-blasts discrimination, ML, Seismicity contamination, Seismic hazard

Introduction

Egypt is thought to be a region of low-to-moderate earthquakes (EQs), with activity dispersed among multiple source regions. The primary influence on seismic activity comes from the relative motions of the plates in Africa, Arabia, and Eurasia. Since the majority of the country's residents and historical locations are

centered in the Nile Valley and Delta, determined EQs in Egypt are a main source of loss of life and property destruction (Badawy 1999). The latest destructive EQs have been reclassified as great magnitude 6.1 and 7.2, respectively, as the Alexandria 1955 and Aqaba 1995 EQs, from average magnitude 5.8 as the Aswan 1981 and Cairo 1992 EQs (Hussein et al. 2013).

The recent decades in Egypt have been characterized by increasing urbanization and the construction of specialized infrastructure. This development necessitates an enhanced awareness of the risks posed by natural phenomena, notably EQs. As part of Egypt's strategy to expand its land use, several new urban centers are slated for establishment (Hegazy and Kaloop 2015).

The mapped region (Fig. 1) is noteworthy due to its high seismic activity and the ongoing urban expansion both northward and southward. Furthermore, the

*Correspondence:

Mohamed S. Abdalzaher
msabdalzaher@nriag.sci.eg

¹ Seismology Department, National Research Institute of Astronomy and Geophysics, Helwan, Cairo 11421, Egypt

² Department of Electronics and Electrical Communications, October 6 University, 6 of October City, Giza 12585, Egypt

³ Interdisciplinary Research Center for Communication Systems and Sensing (IRC-CSS), King Fahd University of Petroleum and Minerals (KFUPM), Dhahran 31261, Saudi Arabia



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

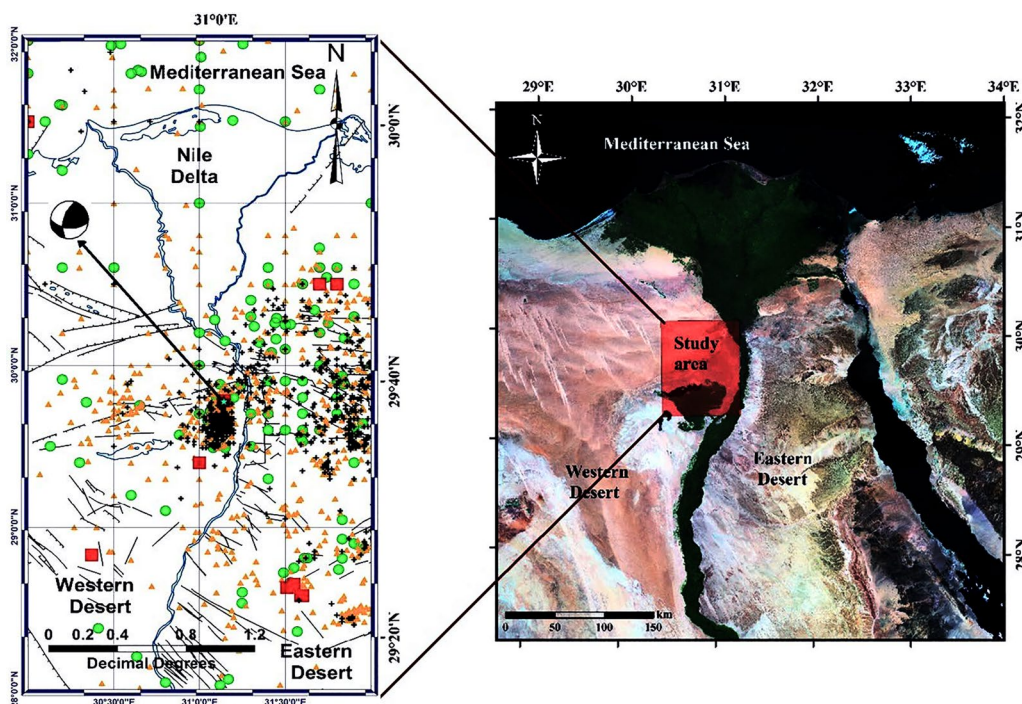


Fig. 1 Map depicting the location and recorded seismic activity in the northern part of Egypt (Hussein et al. 1996)

area encompasses the epicenter of the infamous Cairo EQ on October 12, 1992, situated in Dahshour, approximately 35 kms south of Cairo. This seismic event caused considerable destruction in Cairo, Giza, and Fayoum cities (Arafa-Hamed et al. 2023; Moustafa and Takenaka 2009).

The mapped area under investigation, situated within the Western Desert of Egypt, is not only prone to natural seismic events but also faces additional threats from anthropogenic activities, particularly quarry blasts from nearby cement companies. These industrial operations, while essential for economic development, introduce a significant challenge in distinguishing between earthquake tremors and blast-induced vibrations. This distinction is crucial for the accurate assessment and mitigation of seismic hazards (Abdalzaher et al. 2022).

Given the region's susceptibility to seismic activities, as evidenced by the notable Cairo earthquake of 1992, and the burgeoning urban expansion, understanding and differentiating these seismic sources becomes paramount. This study aims to provide a comprehensive analysis of the seismic events in the area, with a focus on discriminating between natural EQs and QBs. By doing so, it endeavors to enhance the region's seismic hazard mitigation strategies, thereby contributing to the safety and sustainability of the ongoing and future urban developments (Elhadidy et al. 2021; Abdalzaher and Elsayed 2019). Such efforts are vital in ensuring that

the region's economic growth does not compromise its resilience to seismic risks.

Egypt is experiencing an urbanization dilemma due to the country's 2% yearly population growth, which is causing many of its cities to get increasingly crowded. The major cities and the majority of economic activity are located in northern Egypt, which is typically wealthier than the southern region. A completely new metropolis called the recently developed administrative capital is being constructed in order to relieve traffic in Cairo. With a total area of 714 km², it is located 35 km east of Cairo in the middle of the Cairo-Suez road, the Cairo-regional ring road, and the Cairo-Ain El-Sokhna road (Elmouelhi 2019).

The Egyptian National Seismological Network (ENSN) was created by the National Research Institute of Astronomy and Geophysics (NRIAG) in order to reduce the seismic hazard and obtain more information about Egypt's seismic activity. The ENSN program began as a small-scale pilot project in the epicentral sector of Cairo EQ in 1997 and was eventually expanded to monitor the entire nation. As a result, there have been notable changes to Egypt's seismic activity in the present era. Besides, according to our knowledge, the relationship between tectonic and geological parameters and the most recent reported seismic activity is presented in Abdalzaher et al. (2020), Moustafa et al. (2022). Additionally, Egypt is improving

its ports, roads, highways, subways, and residential properties across a large portion of the nation. As a result, there are more blasting and underground mining incidents reported, which causes environmental degradation to worsen with time (Yan et al. 2020; Abdalzaher et al. 2022).

In order to accurately map an active fault, estimate stress values correctly, and accurately observe micro-EQ activity, it is required to differentiate between the different vibration types through the categorization of low-intensity explosion recordings as natural EQs (Moustafa et al. 2021). According to this viewpoint, the initial and crucial stage in any subsequent seismic probability and alleviation studies is the recognition of QBs. Remarkably, there is a resemblance between the frequency components of tectonic EQs, landslide events, volcanic-tectonic EQs, and QBs (Moustafa et al. 2022).

Carefully examining waveform recordings can yield valuable insights into distinguishing different event types with a reasonable level of confidence (Kim et al. 1994). However, conducting a comprehensive analysis of data like this takes time and expertise, and cannot be immediate. Automated systems, including deep learning (DL) and machine learning (ML) techniques, may offer a useful substitute since they produce objective, precise findings quickly, enabling the processing of more data with less effort (Abdalzaher et al. 2022; Puente-Sotomayor et al. 2021; Moustafa et al. 2021; Mousavi and Beroza 2022; Abdalzaher et al. 2023; Mfondoum et al. 2023; Krichen et al. 2023; Abdalzaher et al. 2023). For various monitoring purposes, these approaches can be advantageous since they enable the identification and differentiation of determined events relying on their signature at a single station (Chin et al. 2020; Nguyen and Bui 2019; Asim et al. 2020).

Because of this, the most recent research efforts (Renouard et al. 2021; Kim et al. 2020; Pu et al. 2020; Zhu et al. 2022; Dong et al. 2014) concentrate on developing an automated ML classification system that takes into account the challenge of separating small-magnitude EQs from QBs using the least amount of training data possible while also making it easier to discover highly variable event patterns. More specifically, a dataset collected in northwestern Egypt will be used to analyze the methodology. In the meantime, the features of the model that are obtained from seismogram records include the complexity (C), event power (P) and its logarithmic form ($\log P$), maximum peak amplitude of S-wave A_s and its logarithmic form ($\log A_s$), maximum peak amplitude of P-wave A_p , ratio between the maximum peak amplitudes (A_s/A_p), and spectral amplitude ratio (S_r).

The main contributions of this paper are as follows:

- Eight features derived from data collected by a single-component seismometer at a single seismic station are used by the model realizing an accuracy rate of 99.4%.
- The proposed method evaluates multiple ML classifiers and identifies the most effective one for EQ and QB classification. This model can aid in early EQ detection and mitigation, thereby reducing EQ risk and accurately distinguishing tectonic events from non-tectonic ones.
- The effectiveness of the ML technique in separating tainted EQ data is demonstrated. The proposed approach's performance is validated through comprehensive comparisons using various metrics, including accuracy, F1-score, Kappa, MCC, confusion matrix, ROC curves, training vs. validation, and Precision-Recall.

The following is the outline of the paper. The relevant research is detailed in Section "[Related work](#)". The experimental setting of the data that was used is then thoroughly examined in Section "[Experiment setup](#)". In Section "[Method and analysis](#)", the suggested system model and the approaches used are subsequently examined and analyzed. Section "[Performance evaluation metrics](#)" describes the metrics used for evaluating ML models, and Section "[Results and discussions](#)" displays the outcomes of these evaluations. In Section "[Conclusion and future work](#)", the paper is finally concluded with future perspectives.

Related work

The abrupt release of energy that causes ground vibration on Earth's surface can be categorized as either natural or man-made, depending on the source. Numerous research has shown that distinguishing between natural and artificial seismic occurrences is a significant issue (Renouard et al. 2021; Kim et al. 2020; Pu et al. 2020; Zhu et al. 2022; Dong et al. 2014; Qi et al. 2020).

In the literature context, the SeisComP3 operational monitoring system has incorporated two classifiers. The first classifier eliminates false events from the identified events using a low short-term average/long-term average threshold. The second classifier then categorizes the remaining events as either EQs or QBs (Renouard et al. 2021). Other research efforts have integrated the SVM with the heterodyne laser interferometer to enhance the detection of seismic wave and contribute to the EQs and QBs discrimination (Kim et al. 2020). More investigations were conducted using ten common ML models to evaluate the performance of recognizing microseismic and QBs (Pu et al. 2020). Indeed, studying and discriminating the QBs is also beneficial for the

pipe-soil subject. Accordingly, ML has also been used to study the effect of QBs excavation and their interaction with urban soil-rock stratum (Zhu et al. 2022). It is worth meeting that precise discrimination can offer decision-makers dependable methods for managing disasters by defining the geographic distribution of a likely epidemic and its effects on the populace (Hamdy et al. 2022; Abdalzaher et al. 2021).

The efforts have been extended also to build an automatic classifying of seismic events using supervised learning (Malfante et al. 2018). The framework constructs a predictive model from labeled data through three steps: (i) signal preprocessing, (ii) feature space mapping, and (iii) automatic classifier training. The model achieves an accuracy of about 93.5%. ML has also proved beneficial in assessing the effectiveness of predicting and evaluating landslide susceptibility as presented in a case in Japan (Hokkaido Eastern Iburi earthquake, 2018), South Korea, and China (Nam and Wang 2019; Lee and Lee 2024; Zhou and Fang 2015). Moreover, the SVM has been employed to differentiate between earthquake-triggered landslides and rainfall-triggered, which is considered a more crucial step for landslide susceptibility mapping (Nam and Wang 2020).

Previous work (Ghamry et al. 2021; Zhu et al. 2024) suggested several methods relying on a purely mathematical basis to identify and distinguish blasting from seismicity records in order to discriminate reported seismic activities. These algorithms have been applied in a number of studies to remove reported QBs and clean up regional catalogs. By simulating QBs' spectra, Allmann et al. (2008) investigated the geographical discrimination of QBs and were able to anticipate spectral details indicative of QBs rather than EQs. Other research has examined the observed spectra of EQs and QBs and compared spectral slopes, spectral ratios, path-independent modulations, and time-independent modulations (Dong et al. 2014). A variety of methods, including spectrograms, and spectral analysis as well as quadratic and linear discrimination functions, have been regularly employed by those researchers and others. Although they have looked at several frequency windows, their overall findings remain the same. However, the majority of the previously suggested techniques only use one feature, are regarded as linear discriminant techniques, and are unable to identify the high complexity, discontinuities, and non-linearities in the recorded waveforms.

Following this, a variety of credit scoring models have been created using modern technology and ML (Liu and Zhong 2020), including affinity propagation, logistic regression (LR), and linear discriminate analysis (LDA) (Dong et al. 2016). But occasionally, the efficacy

of conventional statistical analysis approaches is insufficient. Prediction accuracy is therefore impacted because some of the assumptions that happened in these models may not be verified. The advancements in ML have allowed for improved performance over statistical models. Examples of these include the support vector machine (SVM) (Lara-Cueva et al. 2016; Chin et al. 2019) and Naïve Bayes (NB) (Dong et al. 2016) as well as random forest (RF) (Zhang et al. 2020).

While a lot of work has been done in the literature to distinguish between EQs and QBs, a dependable and flexible solution is still needed. As far as we are aware, no prior scheme has been put forth that achieves 99.4% classification accuracy between EQs and QBs while solely depending on the records of a single station.

Experiment setup

The discrimination technique is used in this study to determine whether EQs and QBs are written down in detail by the closest seismic station to the mining area in the studied region, the *AYT*. We utilized 100 km as the maximum epicentral distance from the *AYT* station and events with local magnitudes ≤ 3 . The *AYT* short-period station is equipped with only one component seismometer, 1 dB gain at 1 Hz, a frequency range of 1 Hz to 10 Hz, and a sampling rate of 100 samples per second. It is necessary to configure the environment before we start any ML approach. Feature preparation is thought to be the primary responsibility of ML algorithms. To put it another way, it is the process of enhancing prediction performance by generating suitable characteristics from the employed features.

Waveform data collection

Several tectonic events and QBs have been recorded at one of the Egyptian National Seismic Network (ENSN) seismic stations, named (*AYT*) in the mapped area as depicted in Figure 1. A conventional short-time-average via the long-time-average trigger (STA/LTA) can detect seismic transients automatically if at least four stations record the occurrence. The signal is then manually categorized later on the basis of the seismogram waveforms' visual aspect. The dataset utilized in this study is a part of the database provided by ENSN. More particularly, the data preprocessing is executed based on the calculations mentioned in the coming subsections.

The study area is characterized by a wadi plain that exhibits a subtle relief, with a gradient of approximately 1% in the north–south direction. The topography of the region varies significantly; in the southwest, the elevation exceeds 70 ms, while the northern part of the site sees elevations below 40 ms. The geological composition of this area is a complex amalgamation of sedimentary

rocks, spanning a broad temporal range from the Eocene to the Quaternary periods (Hemdan 1992).

Previous geological and geophysical research in this region has revealed the presence of multiple fault systems, which have a pervasive impact across the entire area (EI-HADIDY 1995). These findings underscore the geological diversity and complexity of the region, which is crucial for understanding its seismic behavior and potential hazards.

Seismic activity in a given region is predominantly influenced by the geological structures present, including faults, trenches, subduction zones, and the region's positioning within the tectonic plate. A notable example of this was observed on October 12, 1992, when a significant earthquake with a magnitude of 5.8 (Mb) struck southwest of Cairo, near the Dahshour region, as illustrated in Figure 1. This event occurred approximately 25 km southwest of Cairo and was notably north of the proposed site for El-Fayoum New City, at a distance of about 56 km. The earthquake had a focal depth of 23 km, making it the largest instrumentally recorded earthquake in the Dahshour region to date.

The October 12, 1992, earthquake marked the first catastrophic seismic event in this area since the earthquake of 1847, occurring after a quiescent period of 145 years (Hussein et al. 1996). The region has a rich history of seismic events, both in historical and contemporary times. Figure 1 delineates the recent seismic activity in and around the studied area from 1997 to the present, as recorded by the ENSN. This data underscores the ongoing seismic relevance of the region and the necessity for continued monitoring and analysis.

In order to differentiate surface QBs from tectonic EQs impacting the studied region, a total of 870 occurrences from 1997 to 2013 were chosen. Because of the substantial lateral heterogeneities in the crust (Nergizci et al. 2024), EQs in various regions show a surprising range of signal characteristics. Therefore, only seismograms, that took place near the *AYT* seismic station, were taken into consideration. The chosen seismograms have the deepest possible focus of 42 km and are captured with moment magnitudes ≤ 3.0 measured by high-gain in addition to short-period seismometers. Figure 2 shows the spatial distribution of these incidents, the cement quarries, and the used *AYT* station. The ENSN bulletins [53] contained the hypocentral parameters of the events that were gathered, and a sampling rate of 100 Hz was employed. In order to determine the distance between the epicenters of EQs and volcanoes, the ENSN reports provide both the position of the *AYT* seismic station and the epicenters of earthquakes (EQs). The highest distance computed for collected EQs was 166 km, and for QBs, it was around 114 km.

Table 1 provides basic descriptive statistics of the gathered waveform dataset's spatiotemporal distribution. Besides, it provides descriptive statistics, which serve as the foundation for synthesizing the gathered waveform. It is worth mentioning that the table is separated into dispersion measures and measurements of central tendency. The first and third quartiles (Q1 and Q3) are used to evaluate variability, whereas the mean and median are used to measure the center trend.

Eliminating outliers from data shown in Figure 2 using quartiles is a robust statistical method that enhances

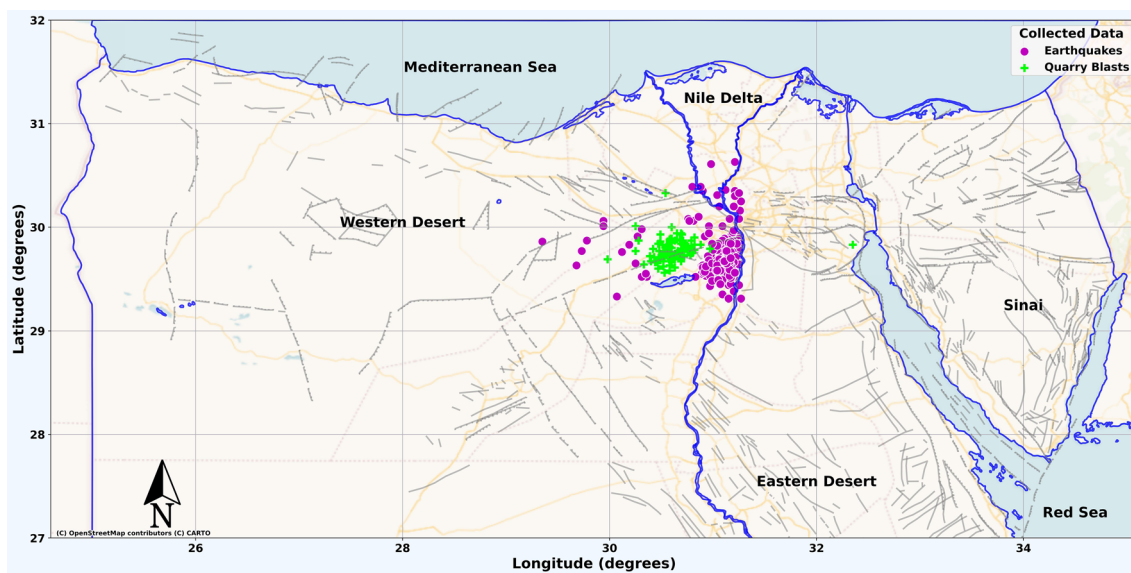


Fig. 2 Spatial location of utilized EQs and QBs in the mapped area recorded at *AYT* station

Table 1 Statistics of the employed features in the developed ML models

Metric	QB/EQ	A_p	A_s	A_s/A_p	C	S_r	P	$\log(P)$	$\log S$
Mean	QB	3786.824074	4597.37037	1.366985	3.845804	0.7974	6.144892	0.455352	3.519407
	EQ	1645.46861	7670.410314	4.518848	5.638497	2.309079	1486.85241	2.622628	3.5278
Std	QB	3085.000786	5190.224843	0.768135	1.443072	0.232093	8.219126	0.591598	0.318167
	EQ	2217.279822	11187.41449	2.07131	3.332058	0.676536	3325.712483	0.602524	0.579983
Min	QB	731	1080	0.402436	0.97463	0.3505	0.11717	-0.931183	3.033424
	EQ	45	148	1.824065	1.6523	0.93038	88.750583	1.948171	2.170262
25%	QB	1486.75	2034.25	0.793947	2.782075	0.642955	1.315614	0.119082	3.308186
	EQ	311.25	1163.75	3.318707	3.28415	1.84645	139.260292	2.143827	3.065823
50%	QB	2714	3037	1.16815	3.95825	0.77757	3.274657	0.515082	3.482323
	EQ	774	3442.5	3.7937	4.4816	2.1257	259.315451	2.413801	3.536802
75%	QB	5478.5	4593.75	1.66488	5.0143	0.950495	7.431209	0.87078	3.662167
	EQ	2117	8972.5	4.883396	7.253375	2.78795	779.473887	2.89167	3.952911
Max	QB	13837	29785	4.456908	5.992	1.5483	53.928891	1.731821	4.473998
	EQ	14386	77561	14.940639	22.8162	3.9983	25330.31375	4.403641	4.889643

the integrity and accuracy of the dataset. This process involves calculating the data’s first quartile (Q1), median, and third quartile (Q3), which partition the dataset into four equal parts. The interquartile range (IQR), defined as the difference between Q3 and Q1, measures statistical dispersion. To identify outliers, any data points lying below $(Q1 - 1.5 \times IQR)$ or above $(Q3 + 1.5 \times IQR)$ are considered outliers. This method effectively highlights anomalies without being influenced by extreme values. Removing these outliers makes the dataset more representative of the underlying population, leading to more reliable and valid analyses.

One-hot encoding

A typical encoding used in ML techniques is one-hot encoding. This methodology modifies the event type’s category data, which is implemented using a sophisticated algorithm. On the other hand, it allows us to collect the category data while retaining the essential information. With this encoding method, every event type is given a unique binary characteristic, with each EQ event being assigned a value of 1 and all other QB events receiving a value of 0. A database including 870 events was created for the study region between October 16, 1997, and November 21, 2013. 36.2069% (315 events) of the data are classified as EQs based on the obtained data, and the remaining 63.7931% are classified as QBs (555 events). As a result, as shown in Table 1, a number of significant values (maximum, minimum, mean, median, standard deviation, etc.) have been computed for the used input features (A_s , A_p , A_s/A_p , P , S_r , and C), which are utilized to discriminate between the EQs and QBs. In

the preliminary stage, manual labeling is utilized to facilitate the training procedure.

It is worth mentioning that, discriminating between EQs and QBs is crucial for accurate seismic monitoring. A_p and A_s help to distinguish these events, as EQs usually have a higher A_s to A_p ratio, while QBs often show a higher A_p amplitude. The parameter C differentiates the simpler waveforms of QBs from the more intricate waveforms of EQs. S_r reveals that EQs have a broader and more varied frequency spectrum, whereas QBs have concentrated spectral content with specific frequency peaks. P indicates the nature of energy release, with EQs exhibiting more sustained energy over time and QBs showing a quicker rise and fall. These features collectively enhance the ML model’s ability to accurately identify and classify seismic events.

Exploratory data analysis

Examining datasets is a key competency for a thorough knowledge of data. Consequently, regardless of the type of data, data analysis exploration is an important phase needed to complete algorithms. Additionally, it offers a thorough understanding of the goal and caliber of the task.

The plots of events (QB and EQ) frequency versus period of existence are shown in Figure 3. This plot is used to look at the distributions of the two variables and evaluate the relationship between them. Besides, the magnitude frequency of the EQs and QBs versus the existence period influencing the mapped area is depicted in Figure 4. Significant restrictions on the EQ operation posing over a wide-ranging length, from micro-EQs to big ones, are expected to be supported by

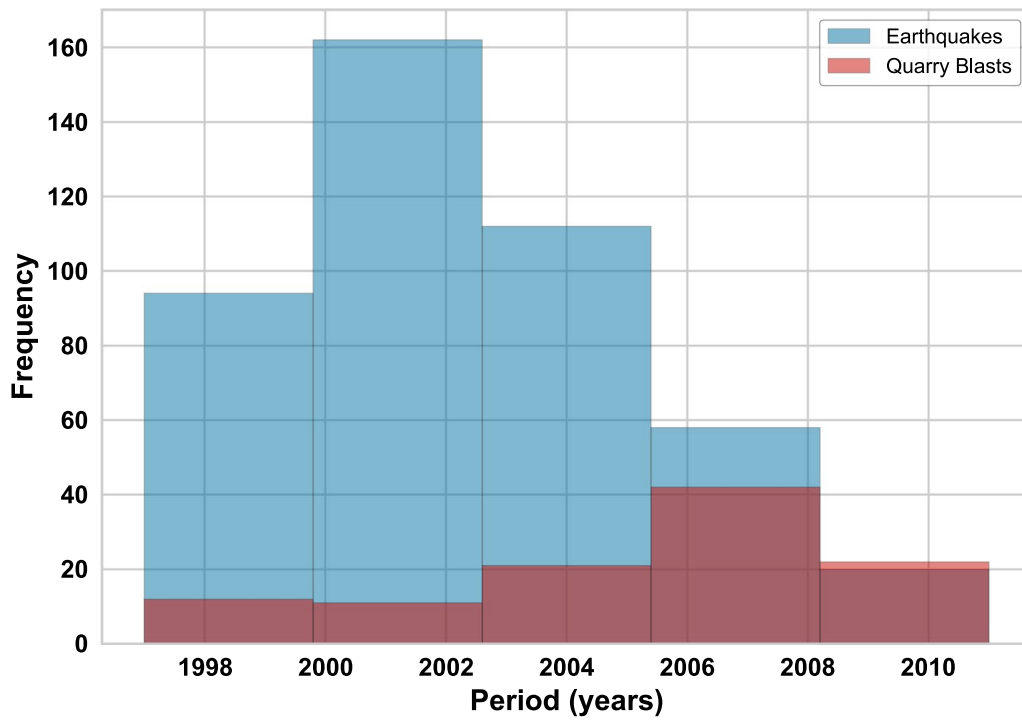


Fig. 3 Histogram of the observed EQs and QBs by AYT station with their frequency

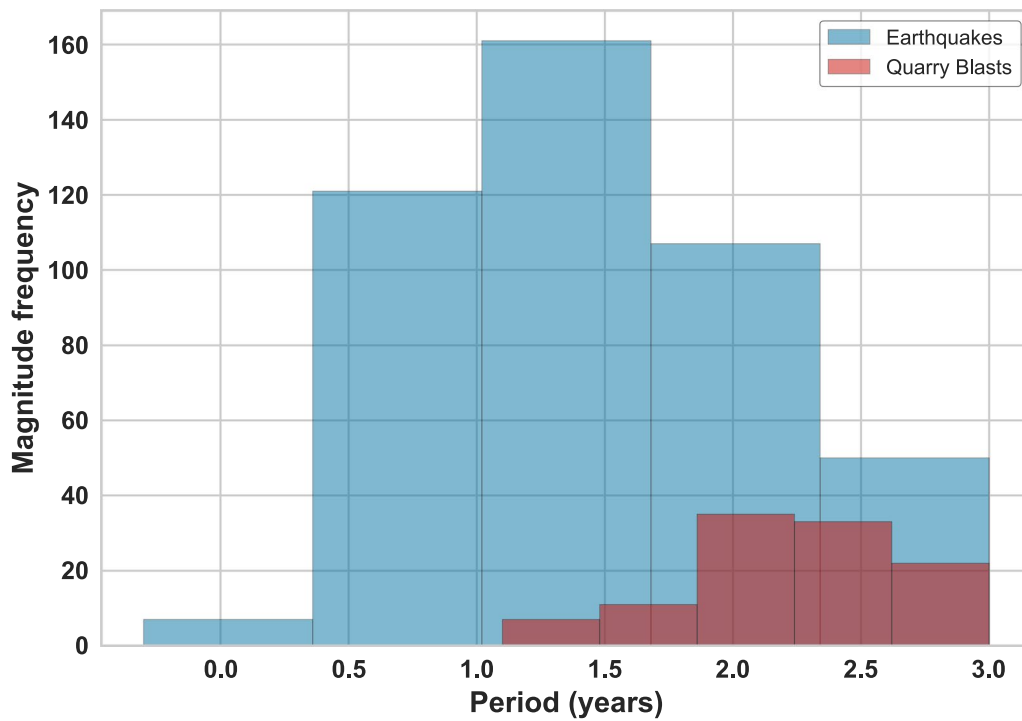


Fig. 4 Histogram of the observed EQs and QBs by AYT station with their magnitude frequency

the comprehensive analysis of the seismic activity of the delineated area at multiple spatiotemporal levels.

Discrimination attributes extraction

In this case, characteristics taken from a collection of seismograms are used to automatically classify the data. Furthermore, one of the finest ways to improve the discriminating model and raise the learning algorithm's predictive capacity is the creation of new features from the raw data. Using gathered waveform records, we assume three different seismic source discrimination features in this step.

Seismic waves amplitudes

The amplitude of the P- and S-waves in the waveform analysis is represented by their maximum values, which are indicated by the symbols (A_p and A_s), respectively. Indeed, the vertical velocity component of the EQs and QBs seismograms determines the values of A_p and A_s . The source data are adjusted for linear trend and instrument reaction for every seismogram. It is possible to see differences in waveform features between the tectonic and human activities. Natural event waveforms are full of high-frequency signals. While the anthropogenic events show emergent Sg signals and are relatively low in high-frequency energy, they have an impulsive Sg signal with a rapid amplitude drop-off. Using the frequency-dependent Q-model (Moustafa et al. 2023; Morozov 2010), we accounted for geometrical dispersion and perceptual attenuation when we adjusted the computed amplitude values.

Two additional features were computed from A_p and A_s to improve discrimination. The maximum S-wave amplitude's logarithm is the first feature, and the maximum S-wave maximum to maximum P-wave amplitude ratio is the second. These amplitude properties have been demonstrated in numerous studies to be useful markers for distinguishing between EQs and QBs at distances of 50 to 200 km (Dong et al. 2016; Kim et al. 2020). EQs frequently show that $A_s \geq A_p$. As with QBs, they demonstrate that $A_p \geq A_s$.

Complexity

The parameter of complexity (C) can be declared as the ratio of merged powers of the velocity seismogram $s^2(t)$ in the chosen windows length (t_0 represents the P-wave onset time; t_1 and t_2 represent the first and second window lengths). One possible representation of the parameter C is Horasan et al. (2009).

$$C = \frac{\int_{t_1}^{t_2} s^2(t) dt}{\int_{t_0}^{t_1} s^2(t) dt} \quad (1)$$

It is observed that the ideal C values for EQs and QBs with equal magnitudes were determined empirically by determining the limits of the integrals (t_0 , t_1 , and t_2). given example, given the windows of $t_0 = 0s$, $t_1 = 2s$, and $t_2 = 4s$, the C results of the employed events (QB and EQS) reveal 0.59 and 11, respectively. For $t_0 = 0s$, $t_1 = 2s$, and $t_2 = 4s$, the C results of QBs grow as the chosen window length increases ($C = 0.59$; $C = 0.62$ for $t_0 = 0s$, $t_1 = 2s$, and $t_2 = 5s$). Consequently, the computed C values serve as the basis for choosing the greatest window for differentiating between EQs and QBs. Table 1 provides an overview of the signal properties used in the process of discrimination.

Spectral ratio

This parameter called Sr was employed. The FFT is used to calculate the Sr , a potential criterion of classification between the shallow EQs and the QBs. To put it more specifically, Allmann et al. (2008) may represent Sr .

$$Sr = \frac{\int_{f_{h1}}^{f_{h2}} a df}{\int_{f_{l1}}^{f_{l2}} a df} \quad (2)$$

where f_{h1} represents a band of high frequency ranges between f_{h1} and f_{h2} , while the band of low frequency ranges between f_{l1} and f_{l2} is the low-frequency band, and a stands for the seismogram integrated ratio of Sr .

The bounds of Sr integrals are chosen via high power and lower power in the spectra of the employed events. Indeed, the low and high frequencies are 1.10 Hz and 10.20 Hz, respectively. According to our observations and other research (Allmann et al. 2008), QBs have a limited band and EQs have a wide band. As a result, we choose to employ the spectrum's maximum frequency values as a variation parameter. One of the primary characteristics used for QB discrimination is the event power (P). This parameter is dependent on the amplitude schemes and spectral qualities (Kekoval et al. 2012). It takes into account A_s/A_p , C , and Sr in more detail. As so, P can be written as.

$$P = (A_s/A_p)^2 \times C \times Sr^2. \quad (3)$$

Finding information in a dataset with high discriminative power as features is known as feature engineering (Domingos 2012). Such information would be a set of traits that might distinguish between two classes of data points in a binary classification task. Coping, removing the features with high association, and separating the class from the remaining data is the first step. Then, for our ML estimators to operate with certain distinguished

results in the proposed categorical variables, we need to cast them into a numerical format.

According to recent studies on seismogram interpretation and EQ data analysis (Kulhánek 2012; Abdalzaher 2024a; Abdalzaher et al. 2024b), local EQ catalogs involve events recorded at epicentral distances closer to 150 km by a network of stations with station spacing of tens of kilometers. The initial seismic waves to reach the recording station at these short distances are represented as Pg and Sg. In the crust’s granitic layer, they stand in for shear or compressional waves.

Method and analysis

This section suggests a classification scheme that divides EQs from QBs based on ML algorithms. The suggested method uses very little training data and describes how to recognize extremely varied event patterns. The suggested method is then demonstrated by depending on such models.

Developed ML model and benchmarks

Quadratic discriminant analysis (QDA)

The development of QDA aims to address the limitations of LR in various domains, including parameter estimation. If we have a sample with a small size and the distribution of the predictors is normal. QDA is a better fit than the linear discrimination analysis model when the data deviates solely in terms of being more or less than the linear assumption (James et al. 2013)..

Support vector machine (SVM)

To determine the ideal hyperplane for K-features, the SVM is employed. The features are mitigated to produce the best hyperplane in a K-dimensional space. The hyperplane is reached if the data marks are deviated by the greatest possible distance. SVMs are used to affect the behavior of the hyperplane and to show the data points that are closest to it. Additionally, a linear support vector machine that delivers values between “-1” and “1” to an output linear function resembling LR is another improvement made to the SVM. It is important to note that the SVM uses a regularization parameter of an objective function to balance the loss and boundary maximization. The goal of this procedure is to maximize the hyperplane’s border between data points (Chang and Lin 2008).

Naive bayes (NB)

In general, the NB classifier is nonlinear. In contrast, if the probability factors depend on exponential relations, the NB classifier is handled as a linear classifier. For continuous-value features, the NB Gaussian technique is employed (Perez et al. 2006; Abdalzaher et al. 2021). As a result, it is regarded as a Gaussian NB (GNB), a probabilistic method

that uses likelihood to forecast values in a deterministic manner. The following formula is used to compute the probability:

$$p(d = F|C) = \frac{1}{\sqrt{2\pi\sigma^2}} \times e^{-\frac{(F-\mu)^2}{2\sigma^2}} \tag{4}$$

where *d* represents continuous input data, *C* signifies the class, σ denotes the variance, *F* corresponds to the probability density, and μ stands for the mean.

K-nearest neighbors (KNN)

This model relies on a boundary to ascertain an input value for each set of its closest neighbors during the decision-making process (Tan 2006; Abdalzaher et al. 2023). The distance parameter, which is used to choose the nearby neighbors, and the neighbors’ set, which is used to regress the input, are the hyper-parameters of this technique. The lowest KNN is expected for one neighbor since the approach can become overly dependent on outliers. According to their cosine similarity to their input, KNN models can also assess the trainset votes (Abdalzaher et al. 2021).

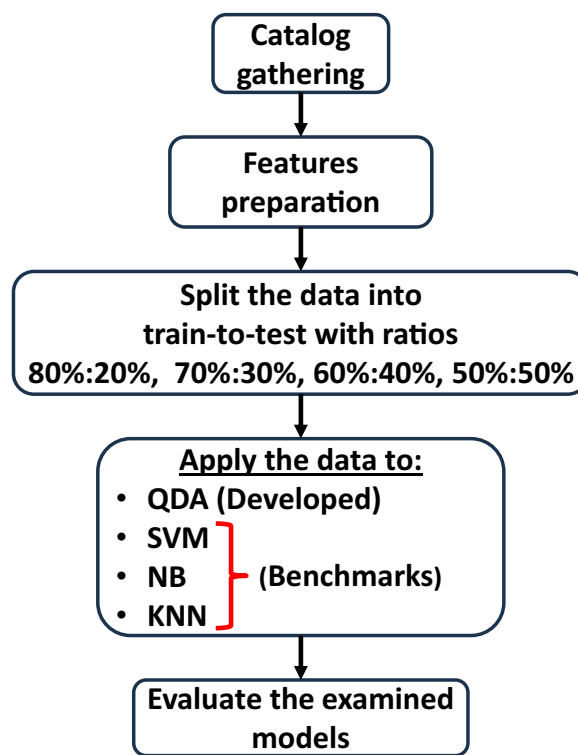


Fig. 5 Proposed ML System Model for EQs and QBs Discrimination

Proposed system model

Figure 5 illustrates the proposed method, which combines the QDA, SVM, NB, and KNN ML techniques. We performed a thorough analysis using these ML models. First, we have extracted the features employed in the suggested supervised method. Eight characteristics (As , $\log(As)$, Ap , As/Ap , P , $\log(P)$, Sr , and C) are first calculated. This is where the process of maximizing the quantity of features used to categorize EQs and QBs starts. Large-scale experiments are then created with these models. Moreover, the number of events (EQs and QBs) utilized in our study is 870. To ensure the model's performance stability and accurate discrimination between EQs and QBs, we examined the model based on four training dataset ratios (50% to 50%, 60% to 40%, 70% to 30%, and 80% to 20%). Once more, the four split ratios are considered to employ the utilized ML models until the last evaluation step, where the best model with the best performance is considered. The evaluation measures utilized in the suggested approach are provided in detail in the section that follows.

Performance evaluation metrics

Improvements in computational power and data storage capacity enable far more thorough analysis than were before possible, leading to a better knowledge of the ways in which QB and EQs differ from one another. Hence, we use a number of criteria to evaluate the classifiers' performance as follows.

Accuracy

It is worth noting that the accuracy gives us the product of the number of targets properly guessed and the target number, which is represented by.

$$\text{Accuracy} = \frac{\sum_{j=1}^S \gamma(\hat{y}_j = y_j)}{S} \quad (5)$$

where y_j is our classifier's output, γ shows the indicator function, S indicates the number of samples, and \hat{y}_j is the estimated target. In addition, we tested the classifiers ten times over using the cross-validation function. It is important to note that if the accuracy score is 1, the predicted values exactly match the target values. The true positive (TP) in a two-class classification problem denotes the correct predictions (> 0), whereas the false positive (FP) denotes the mistakenly classified samples as positive. True negatives (TN) and false negatives (FN) are the two categories of negative findings. Thus, the precision can be written as follows:

$$\text{Accuracy} = \frac{TP + TN}{S} \quad (6)$$

MCC

Matthews (1975) introduced the MCC, a popular index in ML for binary classification. It could evaluate simultaneously the accuracy of the classification of explosions and micro-seismic events. A coefficient called MCC is used to compare predicted and observed binary classifications. The value range of MCC is $[-1, 1]$. "0" indicates that no further prediction beats a random prediction, "-1" shows complete falseness between observation and estimation, and "1" indicates absolute prediction accuracy. The MCC can be written as

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP) \cdot (TP + FN) \cdot (TN + FP) \cdot (TN + FN)}} \quad (7)$$

F1-score

A popular statistic to show the degree of ML model accuracy dependent on the input dataset is the F1 score. In specific, it uses a binary categorization technique that divides the choice into "positive" and "negative" categories. It combines the model's harmonic mean recall and precision. It is provided by:

$$F1 = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \quad (8)$$

Cohen Kappa score (Kappa)

The accuracy that would have resulted from using random estimates is taken into account by the Kappa score. The ideal accuracy score is obtained if Kappa = 1 (McHugh 2012). It is provided by:

$$\kappa = \frac{2 \times (TP \cdot TN - FP \cdot FN)}{(TP + FP) \cdot (TP + FN) \cdot (TN + FP) \cdot (TN + FN)} \quad (9)$$

Confusion matrix

To find and show the category where a classifier fails the most, confusion matrices are used. A confusion matrix is generated by categorizing the test cases according to their expected and ground truth labels. According to the standard, the confusion matrix for an N-class model is an $N \times N$ matrix where the true class is indexed in the row dimension and the predicted class is indexed in the column dimension. Finally, the rates of true and false positives (TPR and FPR) are given by

$$TPR = \frac{TP}{TP + FN}, \quad (10)$$

$$FPR = \frac{FP}{FP + TN} \tag{11}$$

Results and discussions

Initially, we did not eliminate any aspect that we deemed to be irrelevant during our training. Furthermore, we have normalized the features due to significant disparities in the ranges of our continuous features. Data imputation was unnecessary as our dataset is devoid of any missing data. The data frame is partitioned into two distinct datasets: the training dataset and the test dataset. The training dataset comprises 50%, 50%, 60%, 70%, or 80% of the main data, whereas the test dataset consists of the remaining ratios of 40%, 30%, or 20% of the main data, respectively. Subsequently, during the training phase, we fine-tune the hyperparameters. Next, we compute the error rate for each ratio using all the ML classification models that were investigated.

To be more specific, the suggested analytical framework relies on three main components: feature calculation, development of an ML system, and evaluation models against the top-performing categories. Considering the magnitudes of the recorded EQs/QBs presented in Table 1, it is anticipated that aftershocks will occur. It is important to note that the aftershocks may

occur after the mainshock, ranging from weeks to years. Additionally, the quantity of aftershocks is contingent upon the magnitude of the mainshock.

In the proposed approach, we have utilized four ML classification models. The classification system used focuses on two distinct classes: class “0” for EQs and class “1” for QBs. We have created Python code that depends on the Scikit-Learn module. In addition, the configuration settings of the top-performing model that result in the optimal performance.

Here, we present the findings acquired from the proposed criterion. Put simply, we present the key findings that lead to the most effective methodology for accurately classifying EQs and QBs. In order to facilitate the comparison of the remaining top characteristics selected from the initial optimization step, we have incorporated F1-score, MCC, accuracy, as well as Kappa score into Fig. 6. The optimal model is obtained using the “QDA” algorithm, which is compared against the SVM, NB, and KNN models. Specifically, this model attains optimal performance with 99.4% accuracy and 98.2% F1 score, as well as 97.8% Kappa and MCC. Furthermore, while employing the “SVM” model, its results yield an accuracy of 93.4%, an F1-Score of 75.6%, a Kappa score of 72.0%, and an MCC score of 75.0%. When using the NB model, it

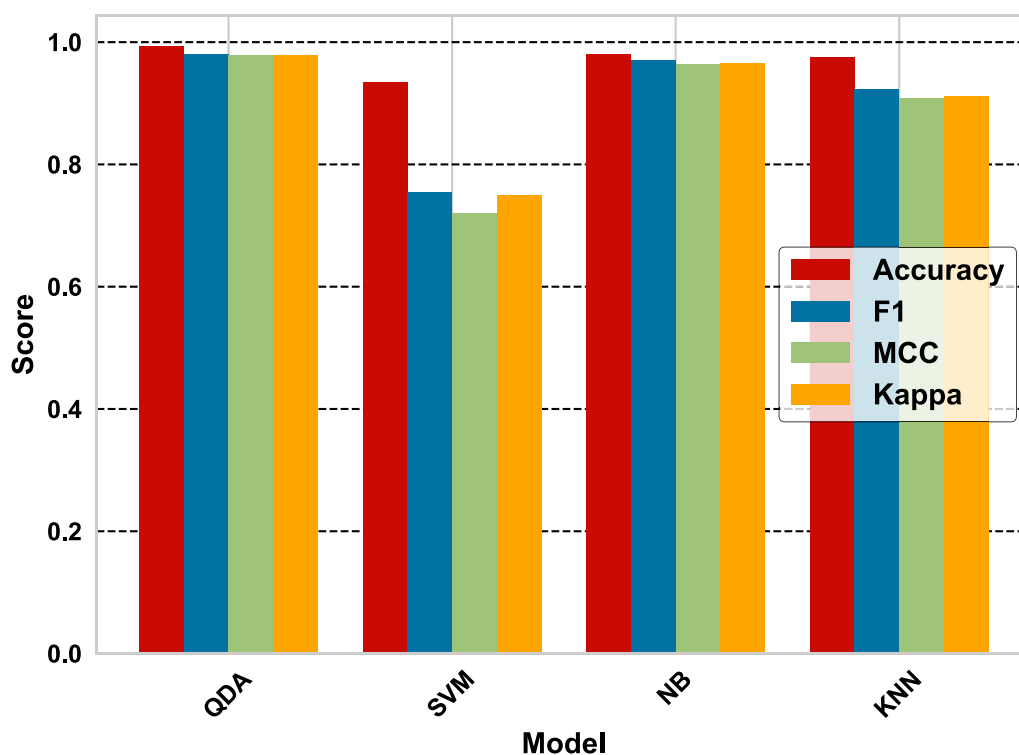


Fig. 6 Scores of the accuracy metrics of the developed ML models for EQs and QBs discrimination

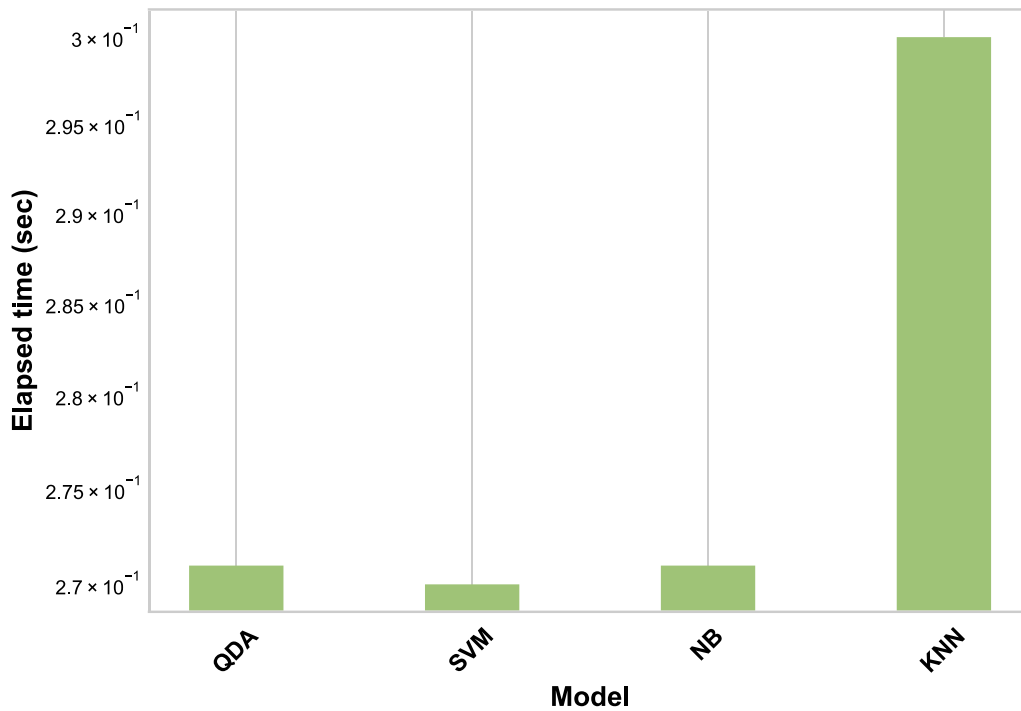


Fig. 7 Comparison of the Elapsed time of the developed ML models for EQs and QBs discrimination

gives an accuracy of 99%, an F1-Score of 97.1%, and a Kappa and MCC scores of 96.6%. Regarding the KNN model, its accuracy is 98.0%, its F1-Score is 92.3% with a Kappa score of 90.9% as well as 91.3% for the MCC score.

Figure 7 illustrates the comparison of elapsed time among the four presented models. The “QDA”, “SVM” and the “NB” models have almost the same elapsed time of 0.271, 0.27, and 0.271 s, respectively. Whereas, the KNN shows a slightly longer elapsed time of 0.3 s. However,

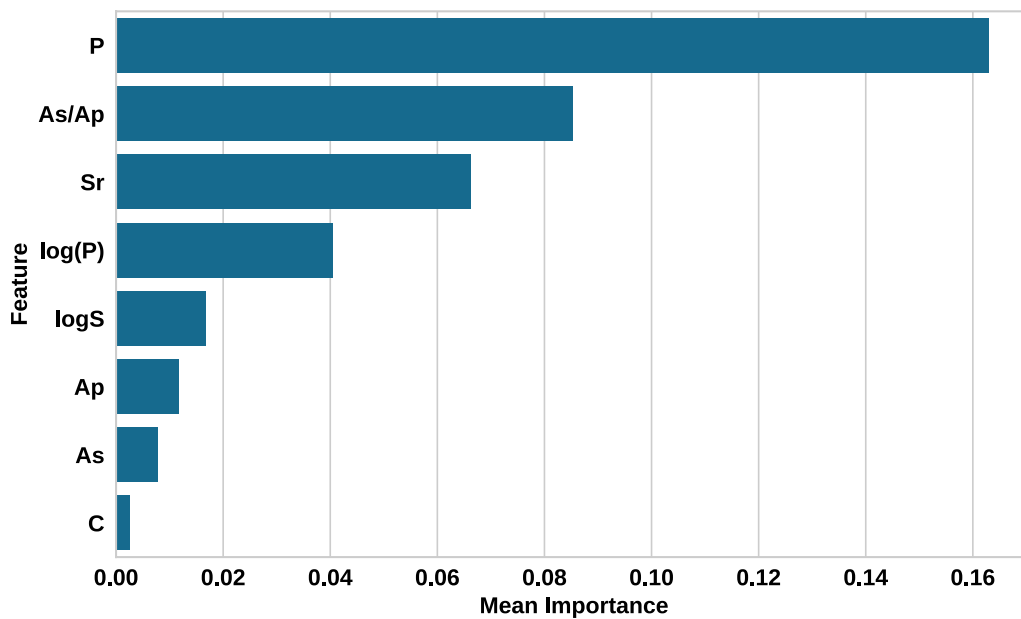


Fig. 8 Scores of the feature importance of the best-developed ML model (QDA) for EQs and QBs discrimination

these models do not exhibit the optimal classification performance. Clearly, the most accurate classification model is the QDA, which has a decent elapsed time.

Figure 8 presents the mean importance of various features used in a model. The feature *P* stands out as the most significant, having the highest mean importance. Following *P*, the features *As/Ap* and *Sr* are also notably important, though their contributions are considerably less than *P*. The features *log(P)* and *logS* show moderate importance, while *Ap* and *As* have relatively low importance. The feature *C* is the least influential, with the lowest mean importance value among all the features evaluated.

In order to demonstrate the efficacy of our QDA best model, we show its learning curves in Fig. 9 for both the training and cross-validation. The figure states that the QDA model achieves the highest level of performance with 100% accuracy in training and about 99% in cross-validation. To be more specific, the no fluctuating curve means no more improvements can be achieved. Furthermore, the training and cross-validation curves exhibit a strong resemblance, indicating the model's efficacy.

The ROC curve represents the relationship between the true positive rate (TP) and the false positive rate (FP) of the projected class. Based on our extensive analysis, the QDA model consistently outperforms the others in terms of classification accuracy. Figure 10 displays the ROC curves obtained by classifying EQs and QBs using

the superior model QDA model. The classification of EQ class "0" and QB class "1" shows success rates of 99% and 96%, respectively.

The confusion matrix is a primary methodology employed in assessing a classification problem, providing an indication of the performance of the projected classes. Once again, after conducting a thorough investigation, we have found that the QDA model obtains the highest level of classification accuracy in distinguishing between the two events' types (EQs and QBs), as demonstrated in Fig. 11. The results indicate that all employed datasets accurately predict the labels of EQ "0" with a 100% accuracy rate. Whereas, the label prediction of QB "1" has an accuracy rate of 96.4%.

Figure 12 displays the relationship between the precision and recall for the most accurate approach in terms of the accuracy of classification. Specifically, we exclusively present the precision-recall outcome based on the QDA (best model) after exhaustive efforts made with the other accessible ML models. To clarify, the QDA model is able to accurately distinguish between the two event types (EQs and QBs) with a precision rate of 100%.

Conclusion and future work

Distinguishing between small-magnitude EQs and QBs is a crucial difficulty in resolving catalogs that have been affected by contamination. This differentiation plays a vital role in evaluating seismic risks, especially in relation

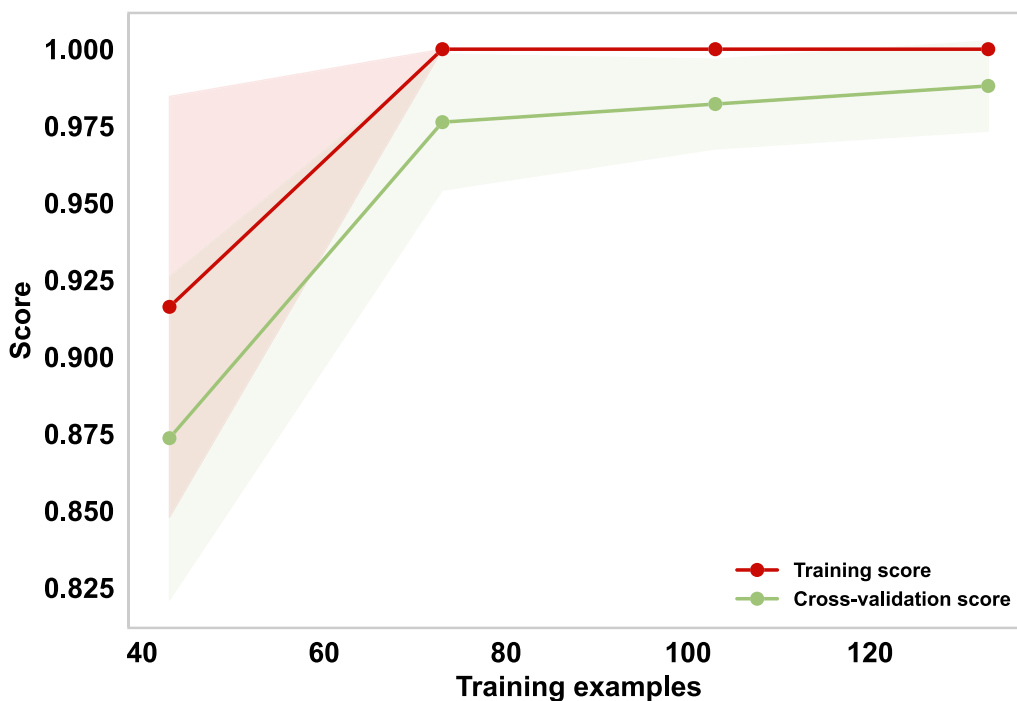


Fig. 9 Learning curves of the best-developed ML model (QDA) for EQs and QBs discrimination

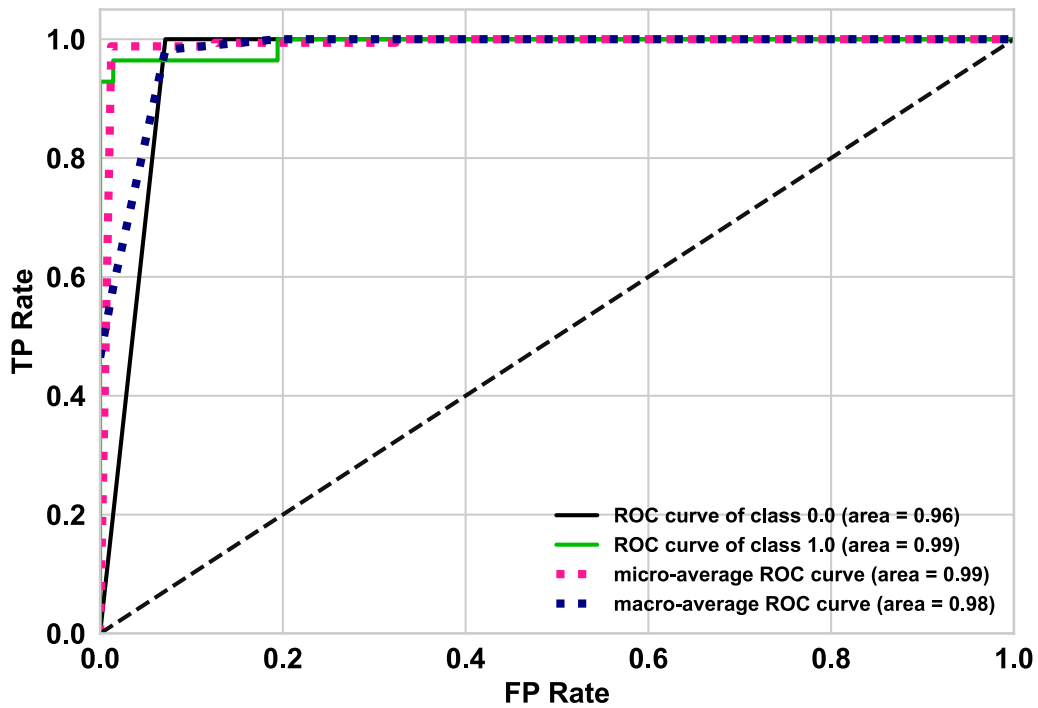


Fig. 10 ROC curves of the best-developed ML model (QDA) for EQs and QBs discrimination

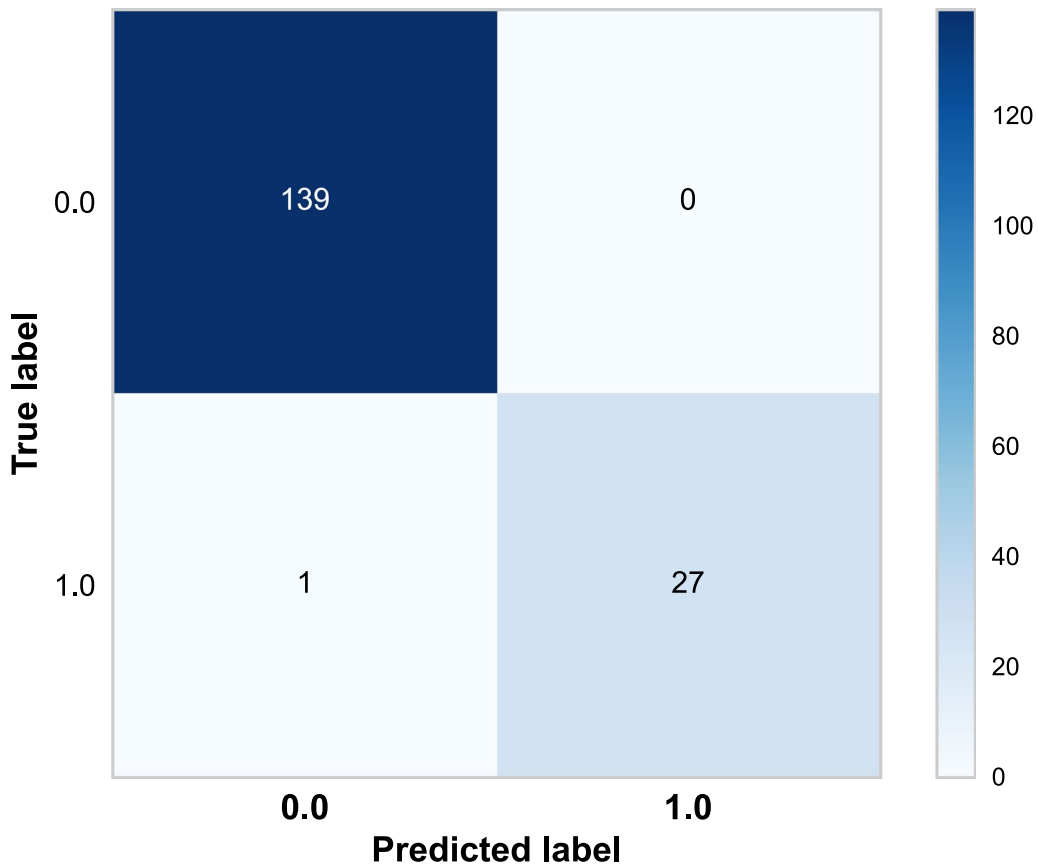


Fig. 11 Confusion matrix of the best-developed ML model (QDA) for EQs and QBs discrimination

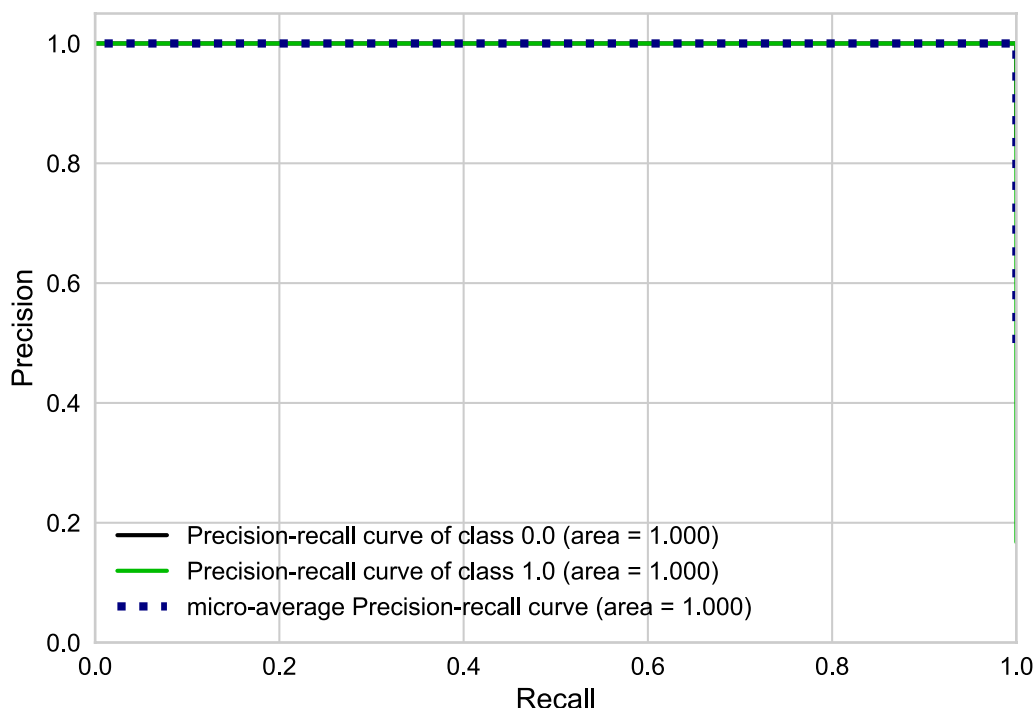


Fig. 12 Precision and Recall of the best-developed ML model (QDA) for EQs and QBs discrimination

to hazards. The paper employed the QDA ML classifier to differentiate between EQs and QBs, as compared to the SVM, NB, and KNN classifiers. Through extensive experimentation, we have determined that the QDA classifier is the optimal model for achieving accurate categorization between two classes: EQs and QBs. The QDA model achieved a promising classification accuracy of 99.4% in differentiating between the two classes. This analysis was conducted using a dataset consisting of 870 seismic events (EQs and QBs) that were monitored exclusively by a single seismic station within the designated study area. Therefore, we recommend that the stakeholders employ the proposed technique to precisely identify EQ clusters, particularly when working with seismicity catalogs that may be influenced by contamination. This methodology is also applicable for conducting comprehensive hazard assessments.

Our future work aims to refine the model by incorporating instances where natural transients display emergent onsets, which can further complicate the classification task. Additionally, we plan to extend the dataset by including seismic events from multiple stations to validate the model's performance across different geological settings and network configurations. We also intend to explore advanced feature extraction techniques and incorporate additional discriminative features that could improve the classifier's accuracy and generalization capability. Furthermore, integrating real-time processing

capabilities and optimizing the model for lower latency could make it more suitable for operational use in seismic monitoring networks.

Author contributions

Conceptualization, M.S.A. and S. S. M.; methodology, M.S.A., S. S. M., W. F., and M. M.S.; validation, M.S.A., S. S. M., and M. M.S.; formal analysis, M.S.A., and M. M.S.; investigation, M.S.A., S. S. M., W. F., and M. M.S.; resources, M.S.A., S. S. M., and W. F.; writing-original draft preparation, M.S.A., S. S. M., and M. M.S.; writing-review and editing, M.S.A., S. S. M., and M. M.S.; visualization, M.S.A. and M. M.S.; supervision, M.S.A., and S. S. M.; project administration, M.S.A., S. S. M., and M. M.S. All authors have reviewed the manuscript.

Funding

Open access funding provided by The Science, Technology & Innovation Funding Authority (STDF) in cooperation with The Egyptian Knowledge Bank (EKB). No funding was received for conducting this study.

Availability of data and materials

Available upon reasonable request.

Code availability

Applicable upon reasonable request.

Declarations

Ethical approval

All the author declare their ethics approval.

Consent to participate

All the author declare their consent to participate.

Consent for publication

All the author declare their consent for publication.

Competing interest

The authors have no Conflict of interest/Conflict of interest to declare that are relevant to the content of this article.

Received: 24 February 2024 Accepted: 24 July 2024

Published online: 08 August 2024

References

- Badawy A (1999) Historical seismicity of egypt. *Acta Geodaetica et Geophys Hungarica* 34(1–2):119–135
- Hussein H, Elenean KA, Marzouk I, Korrat I, El-Nader IA, Ghazala H, ElGaby M (2013) Present-day tectonic stress regime in egypt and surrounding area based on inversion of earthquake focal mechanisms. *J Afr Earth Sc* 81:1–15
- Hegazy IR, Kalooop MR (2015) Monitoring urban growth and land use change detection with gis and remote sensing techniques in daqahlia governo-rate egypt. *Int J Sustain Built Environ* 4(1):117–124
- Arafa-Hamed T, Marzouk H, Elbarbary S, Abdel Zaher M (2024) A geophysical investigation of the urban-expanding area over the seismologically active Dahshour region. *Egypt Acta Geophys* 72(2):743–57
- Moustafa SS, Takenaka H (2009) Stochastic ground motion simulation of the 12 october 1992 dahshour earthquake. *Acta Geophys* 57:636–656
- Abdalzاهر MS, Moustafa SS, Hafez HA, Ahmed WF (2022) An optimized learning model augment analyst decisions for seismic source discrimination. *IEEE Trans Geosci Remote Sens* 60:1–12
- Elhadidy M, Abdalزاهر MS, Gaber H (2021) Up-to-date psha along the gulf of aqaba-dead sea transform fault. *Soil Dyn Earthq Eng* 148:106835
- Abdalزاهر MS, Elsayed HA (2019) Employing data communication networks for managing safer evacuation during earthquake disaster. *Simul Model Pract Theory* 94:379–394
- Elmouelhi H (2019) New administrative capital-cairo. power, urban development and social injustice-the official egyptian model of neoliberalism. *Neoliberalere Urbanisierung Stadtentwicklung* 215–254
- Hussein H, Korrat I, Abd Fattah A (1996) The october 12, 1992 cairo earthquake a complex multiple shock. *Bull Int Inst Seismol Earthq Eng* 30:9–21
- Abdalزاهر MS, El-Hadidy M, Gaber H, Badawy A (2020) Seismic hazard maps of egypt based on spatially smoothed seismicity model and recent seismotectonic models. *J Afr Earth Sc* 170:103894
- Moustafa SS, Abdalزاهر MS, Naeem M, Fouda MM (2022) Seismic hazard and site suitability evaluation based on multicriteria decision analysis. *IEEE Access* 28(10):69511–30
- Yan Y, Hou X, Fei H (2020) Review of predicting the blast-induced ground vibrations to reduce impacts on ambient urban communities. *J Clean Prod* 260:121135
- Abdalزاهر MS, Elsayed HA, Fouda MM (2022) Employing remote sensing, data communication networks, ai, and optimization methodologies in seismology. *IEEE J Selected Top Appl Earth Observ Remote Sens* 15:9417–9438
- Moustafa SS, Abdalزاهر MS, Yassien MH, Wang T, Elwekeil M, Hafez HEA (2021) Development of an optimized regression model to predict blast-driven ground vibrations. *IEEE Access* 9:31826–31841
- Moustafa SS, Abdalزاهر MS, Abdelhafez H (2022) Seismo-lineaments in egypt: Analysis and implications for active tectonic structures and earthquake magnitudes. *Remote Sens* 14(23):6151
- Kim W-Y, Simpson D, Richards PG (1994) High-frequency spectra of regional phases from earthquakes and chemical explosions. *Bull Seismol Soc Am* 84(5):1365–1386
- Abdalزاهر MS, Soliman MS, El-Hady SM, Benslimane A, Elwekeil M (2022) A deep learning model for earthquake parameters observation in iot system-based earthquake early warning. *IEEE Internet Things J* 9(11):8412–8424
- Puente-Sotomayor F, Mustafa A, Teller J (2021) Landslide susceptibility mapping of urban areas: Logistic regression and sensitivity analysis applied to quito, ecuador. *Geoenviron Disasters* 8(1):19
- Moustafa SSR, Abdalزاهر MS, Khan F, Metwaly M, Elawadi EA, Al-Arifi NS (2021) A quantitative site-specific classification approach based on affinity propagation clustering. *IEEE Access* 9:155297–155313
- Mousavi SM, Beroza GC (2022) Deep-learning seismology. *Science* 377(6607):4470
- Abdalزاهر MS, Elsayed HA, Fouda MM, Salim MM (2023) Employing machine learning and iot for earthquake early warning system in smart cities. *Energies* 16(1):495
- Mfondoum AHN, Nguet PW, Seuwei DT, Mfondoum JVM, Ngenyam HB, Diba I, Tchindjang M, Djiangoue B, Mihi A, Hakdaoui S et al (2023) Stepwise integration of analytical hierarchy process with machine learning algorithms for landslide, gully erosion and flash flood susceptibility mapping over the north-moungo perimeter, cameroon. *Geoenviron Disasters* 10(1):22
- Krichen M, Abdalزاهر MS, Elwekeil M, Fouda MM (2023) Managing natural disasters: An analysis of technological advancements, opportunities, and challenges. *Internet Things Cyber-Phys Syst* 4:99–109
- Abdalزاهر MS, Krichen M, Yiltas-Kaplan D, Ben Dhaou I, Adoni WYH (2023) Early detection of earthquakes using iot and cloud infrastructure: A survey. *Sustainability* 15(15):11713
- Chin T-L, Chen K-Y, Chen D-Y, Lin D-E (2020) Intelligent real-time earthquake detection by recurrent neural networks. *IEEE Trans Geosci Remote Sens* 58(8):5440–5449
- Nguyen H, Bui XN (2019) Predicting blast-induced air overpressure: a robust artificial intelligence system based on artificial neural networks and random forest. *Nat Resources Res* 28(3):893–907
- Asim KM, Moustafa SS, Niaz IA, Elawadi EA, Iqbal T, Martínez-Álvarez F (2020) Seismicity analysis and machine learning models for short-term low magnitude seismic activity predictions in cyprus. *Soil Dyn Earthq Eng* 130:105932
- Renouard A, Maggi A, Grunberg M, Doubré C, Hibert C (2021) Toward false event detection and quarry blast versus earthquake discrimination in an operational setting using semiautomated machine learning. *Seismol Soc Am* 92(6):3725–3742
- Kim S, Lee K, You K (2020) Seismic discrimination between earthquakes and explosions using support vector machine. *Sensors* 20(7):1879
- Pu Y, Apel DB, Hall R (2020) Using machine learning approach for microseismic events recognition in underground excavations: Comparison of ten frequently-used models. *Eng Geol* 268:105519
- Zhu B, Jiang N, Zhou C, Luo X, Li H, Chang X, Xia Y (2022) Dynamic interaction of the pipe-soil subject to underground blasting excavation vibration in an urban soil-rock stratum. *Tunn Undergr Space Technol* 129:104700
- Dong L, Li X, Xie G (2014) Nonlinear methodologies for identifying seismic event and nuclear explosion using random forest, support vector machine, and naive bayes classification. In: *Abstract and Applied Analysis*, 2014:1–8. Hindawi Limited
- Qi Y, Wu L, Mao W, Ding Y, He M (2020) Discriminating possible causes of microwave brightness temperature positive anomalies related with May 2008 wenchuan earthquake sequence. *IEEE Trans Geosci Remote Sens* 59(3):1903–1916
- Hamdy O, Gaber H, Abdalزاهر MS, Elhadidy M (2022) Identifying exposure of urban area to certain seismic hazard using machine learning and gis: A case study of greater cairo. *Sustainability* 14(17):10722
- Abdalزاهر MS, Moustafa SS, Abd-Elnaby M, Elwekeil M (2021) Comparative performance assessments of machine-learning methods for artificial seismic sources discrimination. *IEEE Access* 9:65524–65535
- Malfante M, Dalla Mura M, Mars JI, Métaux J-P, Macedo O, Inza A (2018) Automatic classification of volcano seismic signatures. *J Geophys Res: Solid Earth* 123(12):10–645
- Nam K, Wang F (2019) The performance of using an autoencoder for prediction and susceptibility assessment of landslides: A case study on landslides triggered by the 2018 hokkaido eastern iburi earthquake in japan. *Geoenviron Disasters* 6:1–14
- Lee S-M, Lee S-J (2024) Landslide susceptibility assessment of south korea using stacking ensemble machine learning. *Geoenviron Disasters* 11(1):1–17
- Zhou S, Fang L (2015) Support vector machine modeling of earthquake-induced landslides susceptibility in central part of sichuan province, china. *Geoenviron Disasters* 2(1):1–12
- Nam K, Wang F (2020) An extreme rainfall-induced landslide susceptibility assessment using autoencoder combined with random forest in shimane prefecture, japan. *Geoenviron Disasters* 7(1):1–16
- Ghamry E, Mohamed EK, Abdalزاهر MS, Elwekeil M, Marchetti D, De Santis A, Hegy M, Yoshikawa A, Fathy A (2021) Integrating pre-earthquake signatures from different precursor tools. *IEEE Access* 9:33268–33283

- Zhu J, Fang L, Miao F, Fan L, Zhang J, Li Z (2024) Deep learning and transfer learning of earthquake and quarry-blast discrimination: applications to southern california and eastern kentucky. *Geophys J Int* 236(2):979–993
- Allmann BP, Shearer PM, Hauksson E (2008) Spectral discrimination between quarry blasts and earthquakes in southern california. *Bull Seismol Soc Am* 98(4):2073–2079
- Liu Y, Zhong Y (2020) Machine learning-based seafloor seismic prestack inversion. *IEEE Trans Geosci Remote Sens* 59(5):4471–4480
- Dong L, Wesseloo J, Potvin Y, Li X (2016) Discrimination of mine seismic events and blasts using the fisher classifier, naive bayesian classifier and logistic regression. *Rock Mech Rock Eng* 49(1):183–211
- Lara-Cueva RA, Benítez DS, Carrera EV, Ruiz M, Rojo-Álvarez JL (2016) Automatic recognition of long period events from volcano tectonic earthquakes at cotopaxi volcano. *IEEE Trans Geosci Remote Sens* 54(9):5247–5257
- Chin T-L, Huang C-Y, Shen S-H, Tsai Y-C, Hu YH, Wu Y-M (2019) Learn to detect: Improving the accuracy of earthquake detection. *IEEE Trans Geosci Remote Sens* 57(11):8867–8878
- Zhang H, Zhou J, Armaghani DJ, Tahir M, Pham BT, Huynh VV (2020) A combination of feature selection and random forest techniques to solve a problem related to blast-induced ground vibration. *Appl Sci* 10(3):869
- Hemdan M (1992) Pliocene and Quaternary deposits in Bani Suef-East Fayoum area and their relationship to the geological evolution of River Nile. Cairo University Cairo Egypt
- El-HADIDY S (1995) Crustal structure and its related causative tectonics in northen egypt using geophysical data. Ph. D. thesis
- Nergizci M, Abbak RA, Arisoy MO (2024) The effect of crustal density heterogeneity on determining gravimetric geoid: Example in Central Anatolia, Türkiye. *J Asian Earth Sci* 1(264):106037
- Egyptian National Seismic Network. <https://doi.org/10.7914/SN/EY>
- Moustafa SS, Mohamed G-EA, Elhadidy MS, Abdalzاهر MS (2023) Machine learning regression implementation for high-frequency seismic wave attenuation estimation in the aswan reservoir area, egypt. *Environ Earth Sci* 82(12):307
- Morozov IB (2010) On the causes of frequency-dependent apparent seismological q. *Pure Appl Geophys* 167(10):1131–1146
- 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
- Kekoval 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
- Domingos P (2012) A few useful things to know about machine learning. *Commun ACM* 55(10):78–87
- Kulhánek O (2012) Anatomy of Seismograms: For the IASPEI/Unesco Working Group on Manual of Seismogram Interpretation. Elsevier, ???
- Abdalzاهر MS, Moustafa SS, Yassien M (2024) Development of smoothed seismicity models for seismic hazard assessment in the red sea region. *Natural Hazards* 1–30
- Abdalzاهر MS, Soliman MS, Krichen M, Alamro MA, Fouda MM (2024) Employing Machine Learning for Seismic Intensity Estimation Using a Single Station for Earthquake Early Warning. *Remote Sens* 16(12):2159
- James G, Witten D, Hastie T, Tibshirani R (2013) *An Introduction to Statistical Learning*, vol 112. Springer, ???
- Chang Y-W, Lin C-J (2008) Feature ranking using linear svm. In: *Causation and Prediction Challenge*, pp. 53–64. PMLR
- Perez A, Larranaga P, Inza I (2006) Supervised classification with conditional gaussian networks: Increasing the structure complexity from naive bayes. *Int J Approx Reason* 43(1):1–25
- Tan S (2006) An effective refinement strategy for knn text classifier. *Expert Syst Appl* 30(2):290–298
- Abdalzاهر MS, Sami Soliman M, El-Hady SM (2023) Seismic intensity estimation for earthquake early warning using optimized machine learning model. *IEEE Transactions on Geoscience and Remote Sensing* 1–11
- Matthews BW (1975) Comparison of the predicted and observed secondary structure of t4 phage lysozyme. *Biochim et Biophys Acta (BBA)-Protein Struct* 405(2):442–451
- McHugh ML (2012) Interrater reliability: the kappa statistic. *Biochem Med* 22(3):276–282

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.