IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received October 22, 2021, accepted November 9, 2021, date of publication November 15, 2021,
date of current version November 29, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3128284

A Quantitative Site-Specific Classification Approach Based on Affinity Propagation Clustering

SAYED S. R. MOUSTAFA[1], MOHAMED S. ABDALZAHER[1], (Member, IEEE), FARHAN KHAN[2], MOHAMED METWALY[3], ESLAM A. ELAWADI[4], AND NASSIR S. AL-ARIFI[2]

[1]Seismology Department, National Research Institute of Astronomy and Geophysics (NRIAG), Helwan, Cairo 11421, Egypt
[2]Geology and Geophysics Department, Faculty of Science, King Saud University, Riyadh 11451, Saudi Arabia
[3]Archaeology Department, College of Archaeology and Tourism, King Saud University, Riyadh 11451, Saudi Arabia
[4]Airborne Exploration Department, Exploration Division, Nuclear Materials Authority, Cairo 11936, Egypt

Corresponding author: Mohamed Metwaly (mmetwaly@ksu.edu.sa)

This work was supported in part by the Research Supporting Project, King Saud University, Riyadh, Saudi Arabia, under Grant RSP-2021/89; and in part by the Egyptian National Seismic Network (ENSN) Lab, National Research Institute of Astronomy and Geophysics (NRIAG), Cairo, Egypt.

ABSTRACT Investigations made to evaluate the site-effect characteristics and to develop a reliable site classification scheme have received paramount importance for the urban areas planning and reliable site-specific seismic hazard assessment. This paper presents a novel non-objective and data-driven approach for preliminary seismic site-specific classification maps using machine learning (ML) based on affinity propagation (AP) along with a selected set of representative horizontal to vertical spectral ratio (HVSR) curves inside King Saud University (KSU) campus, which is among the main areas in Saudi Arabia. Besides, the proposed model aims to overcome the clustering error due to the dependency of the interpreter's experience. Measurements of the ambient vibrations were performed to cover the entire campus area by about 307 stations. Recording at each station lasted for 20 minute length and a sample rate of 128 Hz for each station to satisfy the criteria for reliable and unambiguous HVSR results. Frequency and amplification values were used for subsequent site classification by passing messages between data points. The obtained results illustrate that the microtremor spectral ratio can be a remarkably robust tool in determining site effects. Accordingly, the proposed methodology can assist the decision-makers to set the priorities of managing land uses, estimating the earthquake losses, conducting programs for reducing the vulnerability of existing structures, enforcing building codes, planning for emergency response and long-term recovery, and designing and implementing phases of new constructions.

INDEX TERMS Machine learning, seismic site classification, HVSR, clustering, affinity propagation.

I. INTRODUCTION

Effects of near-surface soil conditions on ground motion and the consequent seismic response of the structures are common phenomena and produce huge effects on the characteristics of ground shaking during earthquakes [1], [2]. The site-effect analysis is an indispensable part of the present and any urban area planning. It is associated with the surface geology and geotechnical characteristics of soil deposits and has paramount importance on seismic ground motion [3]. Generally, site-effects include the modification of the characteristics (in terms of amplification, frequency content, and duration) are controlled by anomalies in the mechanical properties of the shallowest layers of subsoil, when it consists of soft sediments, or by the shape of surfaces layer discontinuity close to or coincident with the topographic surface [1], [4], [5].

Local site response can be estimated by theoretical and empirical approaches. Explanations of various procedures to estimate the local site effects based on the geology and topography properties of the studied area are given in many research efforts [3], [6], [7]. The evaluation of the site-effects

The associate editor coordinating the review of this manuscript and approving it for publication was Wentao Fan.

by using microtremor records as a tool has gained widespread popularity in the recent researches [8]. Microtremor studies were initially proposed by [9], which is strongly emphasized by [10], for inferring the dynamic site characteristics and associated subsurface soil structure at an observation point. Analysis of microtremor measurements can give useful information on dynamic properties of the site such as the predominant frequency and amplification. Microtremor observations are easy to perform, inexpensive, and can be applied to regions with low seismicity as well; hence, predominant frequency and amplification measurements can be used conveniently for performing a preliminary seismic micro-zonation as studied in [1].

It is worth pointing out that, after [1], numerous authors around the world tested the validity of the technique experimentally and theoretically [11]–[16]. They proved that it is successful for estimating the site response of surface deposits using ambient noise as a source. Applying the horizontal to the vertical spectral ratio (HVSR), for instance, [17], [18], obtained thicknesses map of soft sediments and estimate the frequency of the fundamental resonant mode and correctly predict the amplification level [10]. In [19], the authors obtained classification based on the HVSR method and validate the application for site-dominant frequency estimation and site classification. In spite of its limitations in determining true site amplification values, many researchers have proved that HVSR method is a robust technique and a preliminary step toward site characteristics estimation [20]–[22] and micro-zonation in the areas of interest [17], [18], [23]–[29].

Seismic site classification is the most widely accepted practical method in the design of seismic resistant infrastructure [30], [31]. The most elementary technique for site classification is the availability of borehole data. These soil classes are based on the average shear wave velocity at the upper 30 meters of the subsurface successive materials, and the dominant period. Both parameters also affect the normalized elastic response spectra [30]–[32]. In the recent studies on site characterization, the HVSR technique is one of the successfully used methods as it gives an accurate reading of the site's predominant frequency [1], [5], [31].

The conventional method employed in identifying seismic site-effect is intuitive and simple, but its corresponding interpretation and classification are very subjective due to personal experience bias. Given the complexity of the problem, one of the most promising approaches is to develop alternative techniques for the automatic identification of seismic site classification schemes. Therefore, it is necessary to seek a clustering method [33], [34], with rapid convergence, good global search capability, simple, and convenient implementation for engineering applications.

Nowadays, the utilization of machine learning (ML) techniques has become increasingly widespread in seismology, with applications ranging from identifying unseen signals and patterns to extracting features that might improve our geological understanding. A good survey for the ML methods is given in [35].

Finding groups in data is an important step in many fields of computer science analysis, and there exist algorithms to solve such problems [36]. Clustering, as an unsupervised data mining technique, deals with the problem of dividing a given set of entities into meaningful subsets [37], [38]. A survey for the clustering methods is presented in [39]. Usually, clustering methods satisfy the constraint that a user has to specify the initial number of clusters, and is very sensitive to that parameter [38], [40]. One solution to this problem is the popular AP clustering algorithm proposed by [41], [42]. The AP algorithm has been successfully employed in applications including face recognition, gene discovery, text mining, and image segmentation [43]–[46]. It uses a message-passing model between data points to form a collection of exemplars and respective clusters. It tries to solve the problem without needing to know the number of clusters beforehand, by only supplying a similarity criterion. It has the advantage of identifying clusters faster and with lower errors than other methods [47]. Besides, AP is a powerful clustering technique that sends affinity signals between paired points in a factor graph. Unlike standard techniques, the AP methodology may employ nonmetric similarities as input data, making data analysis exploration appropriate for atypical similarity metrics. Compared with conventional clustering methods such as K-means, this method is insensitive to initial cluster centers and is able to achieve a global optimum. Fortunately, ML and deep learning approaches offer great promise not only in seismological research as means for integrating numerous data into personalized indices of diagnostic and prognostic value [35], [48]–[51]. Table 1 lists a comparison of the complexity and corresponding tradeoff of the involved clustering algorithms in this study [52], [53].

The motivation beyond the proposed study is that the King Saud University (KSU) is considered one of the leading universities founded in 1957 in Saudi Arabia and has massive in future, which aims to serve the well-being of citizens, through an educational and new engineering construction projects such as residential suites, hotels and research centers in the study area. This available development site of interest included approximately 2,224 acres (9km^2) area. The rapid growth of these projects has heightened the importance of evaluation of site response study for detailed site investigation. The present study would be the first indispensable step for designing and implementing phases of new constructions inside the study area.

In this study, an alternative approach is proposed to characterize sites inside KSU campus by integrating unsupervised ML for clustering relying on adaptive affinity propagation (AAP) technique and the HVSR technique for analyzing ambient noise data. The proposed approach could be applied to quantify site-effects in the estimation of seismic site classes associated with seismic hazards [29], [54]. More particularly, for better site-specific characterization of the KSU region

TABLE 1. Comparison of clustering algorithms complexity and corresponding trade-off.

	DBSCAN	HDBSAN	K-means	Agglomerative	AP
Category	Density and distance-based clustering	Density and distance-based clustering	Kernel	Hierarchy	Graph-based clustering
Time complexity	$O(NlogN)$	$O(N^2)$	$O(tcN)$	$O(N^3)$	$O(N^2logN)$
Advantages	Suitable for data with arbitrary shape	-Suitable for data with arbitrary shape -Suitable for high dimensional feature space	-Suitable for data with arbitrary shape -Suitable for high dimensional feature space -Able to analyze noise and detect the clustering overlapping	-Suitable for data with arbitrary shape -Easy to detect the hierarchical relationship between clusters	-Simple and clear -Insensitive to the outliers -The number of clusters is not desired beforehand
Disadvantages	-Low quality with inequality of data space density -Clustering is highly sensitive to the parameters	-Low quality with inequality of data space density -Clustering is highly sensitive to the parameters -Not robust to noise -Influenced by chaining effects	-Not suitable for large-scale data -Clustering is highly sensitive to the parameters -High complexity at the beginning	- High time complexity -Number of clusters is needed beforehand	-High time complexity -Clustering is highly sensitive to the parameters
Key notes	N is the number of objects, c is the number of clusters, and t is the number of iterations.				

an AAP-based unsupervised ML model for clustering was implemented based on the obtained frequencies. AP clustering attempts to divide datasets into many clusters, with related data points remaining in the same group and dissimilar data points remaining in distinct groups. Due to the outperforming features of AP, we rely on an AAP data clustering for identifying important underlying patterns in data, with statistical distributions followed by different classes into which the data can be classified. To adopt a reasonable exhibiting of the proposed approach, K-mean is utilized as a bench mark due to the high uncertainty in determining the optimal number of clusters. Besides, the K-mean is a very popular bench mark used in the literature for clustering.

The contributions of this research work are five folds:

- We developed an AAP-based clustering is based on real measurements of more than 300 measured points in KSU, which is one of the main important areas in Saudi Arabia. Then, we have developed AAP-based ML site-specific classification map as an entirely data-driven map, which is independent of the interpreter's experience. The proposed AAP-clustering approach has been utilized as a proof-of-concept about the effectiveness of AP for such vulnerable application. To the best of our knowledge, this study would be the first indispensable step, which will provide insights into the seismic site response of the KSU campus for designing and implementing phases of new and development expansions.
- We integrate HVSR predominant frequency with an AAP algorithm [41] to develop a quantitative site-specific classification scheme. Moreover, we develop a more efficient vectorized version of the classical AP algorithm called AAP. The performance of this algorithm is compared to that of the K-means clustering [55], and discuss some advantages and disadvantages.
- The study performs a pilot site effects study on the KSU campus area by using the HVSR in conjunction with available geological information to estimate site-specific predominant frequency and amplification.
- We present a vital paradigm to deeply understand the seismic properties of the site. More particularly, the

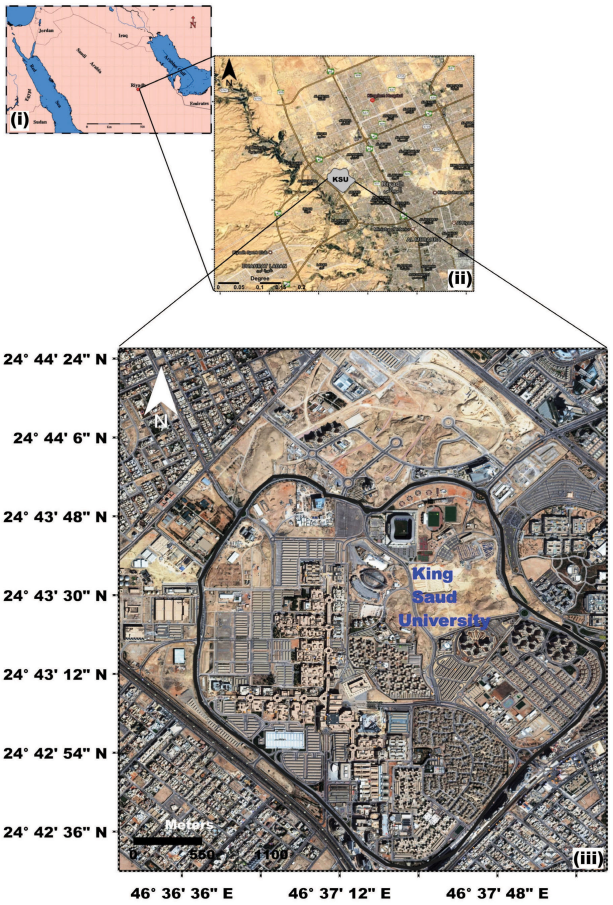

FIGURE 1. Location map of the study area. (i) Location of Riyadh, Capital of Saudi Arabia; (ii) General location of King Saud University (KSU) boundaries inside Riyadh City; (iii) General outline of the KSU campus.

obtained results of the quantitative site-specific classification are used to prepare a preliminary seismic site classification map of the KSU campus.
- The present study can be highly significant to estimate the fundamental frequency and a measure of the site amplification, and hence, can estimate the earthquake losses and related scenarios such as for designing and implementing phases of new constructions inside the study area.

The rest of this paper is organized as follows. The adopted materials and methods in the proposed study are discussed in Section II. The framework of the experimental setup and obtained results are illustrated in Section III. Then, the discussion is presented in Section IV. Finally, Section V concludes the paper.

II. MATERIALS AND METHODS

A. GEOLOGICAL SETTING

The present area of study is the campus of KSU, which is located at the northwest of Riyadh city, Saudi Arabia. Riyadh city is the most important political, economic, and densely populated region in the Kingdom of Saudi Arabia with a population of 7.6 million people (Figure 1). The Riyadh region is positioned at an elevation around 600 meters above mean sea level, and it covers almost 1,913 km^2 area in the central part of the Najd Plateau (Figure 1). More particularly, the campus area to witness accelerated construction expansion and new engineering construction projects. These activities intensified the need for site response evaluation. Many researchers utilized the microtremor measurements in many parts of Saudi Arabia for estimating the site response, in terms of the predominant frequency and concluded that the technique is very promising particularly for the densely populated areas like the new urban planning and reducing the vulnerability on the existing civil constructions [25], [26]. Hence, this study would be the first indispensable step, which will provide insights into the seismic site response of the KSU campus for designing and implementing phases of new and development expansions. The utilized parameters and notations are summarized in Table 2.

Geographically, the plateau extends from the Awanid scarp on the northern edge, to the Kharj rise on its southern edge, and from the Dahna sand belt on its eastern edge, to the Tuwaiq Mountains on its western edge [56]. It is largely Jurassic to Quaternary sediments, mostly composed of sandy limestone, siltstones, and shales [57]. The Najd Plateau has a great thickness of continental and shallow marine limestone deposits. In [58], the authors studied the geological setting of Riyadh and stated that the sedimentary section of the region can be characterized into a surface geological system composed of a mixture of Aeolian clay, silt, sand, and gravel deposits whereas subsurface geology is composed of the great thickness of shallow marine limestone with shale and clay intercalations. The local geologic section of the study site (Figure 2) has a main stratigraphical succession of Upper Jurassic Arab Formation comprises two main members: Arab-D member to the western side and a mixture of Arab (C and D members) to the eastern side [59].

TABLE 2. Notations of parameters and variables.

Symbol	Description
$HVSR$	Horizontal to vertical spectrum ratio
$S_{NS}(f)$	Fourier amplification spectra in the North-South direction
$S_{EW}(f)$	Fourier amplification spectra in the East-West direction
$S_Z(f)$	Fourier amplification spectra in the Vertical direction
$s(i,k)$	The similarity of two distinct points (i,k) in the similarity matrix
$\mathcal{X}$	A data set
E	The number of exemplars
$r(i,k)$	The responsibility matrix
$a(i,k)$	The availability matrix
$c(i,k)$	The criterion matrix
$s(k,k)$	An exemplar
a	The mean intra-cluster distance
b	The distance between a point and the nearest cluster
λ	The damping factor
p	the peak value denoted by *Silhouette* score
C and C^*	The number and optimal number of clusters
f_0	The lowest fundamental frequency
f_1 and f_2	The second and third peak frequencies

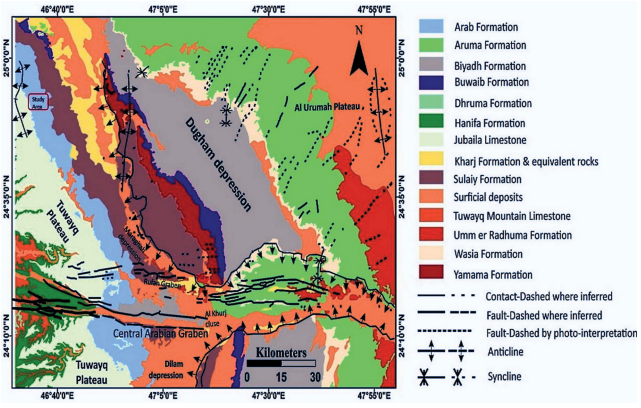

FIGURE 2. General geological setting of the Riyadh region [58]. Small red square indicates the location of KSU campus area. Adopted and modified from [60].

B. MICROTREMOR HVSR METHOD

Microtremor is introduced by [9] and later enhanced by [1], [12]. It is defined as a low amplification ambient vibration of the ground caused by man-made or atmospheric disturbances, like the wind, sea, or ocean waves, and vehicle vibrations that can describe the geological conditions of an area. In [1], the authors utilized a simple HVSR measurement in three orthogonal directions (two horizontal and one vertical). It is based on the assumption that the ratio of the horizontal spectrum and vertical surface vibration is a function displacement [12], [15]. According to the method proposed by [1], the dominant vibration frequency of a site (or engineering structure) can be determined from microtremor record. The record is composed of a triple component: a time domain

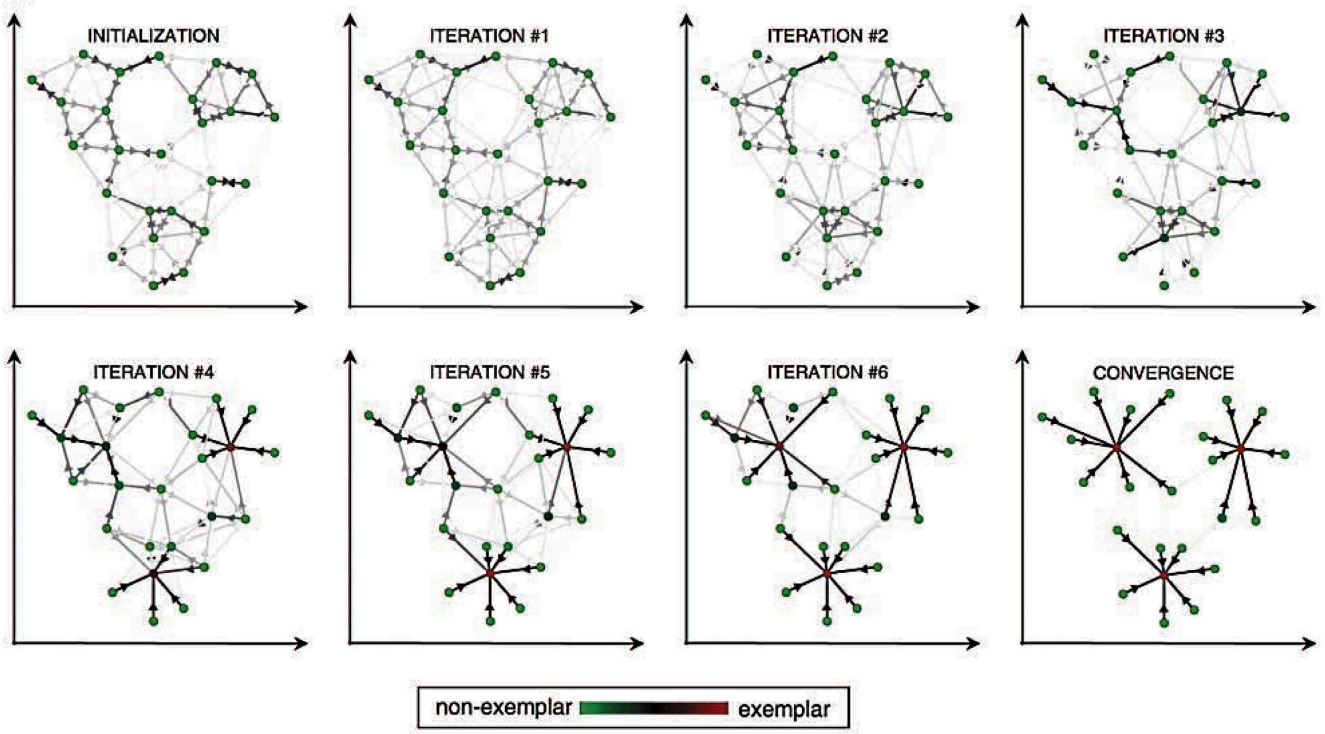

FIGURE 3. Visualized message-passing procedure showing the likelihood of a point being an exemplar, and its exemplar-relation to other points. A directed edge represents the likelihood for a point to choose another point as its exemplar. Adopted and modified from [41].

component, the transformed to spectral-domain component, and the ratio of horizontal and vertical components utilizing the following definition:

$$HVSR(f) = \frac{\sqrt{S_{NS}(f) \times S_{EW}(f)}}{S_Z(f)} \tag{1}$$

where $HVSR(f)$ is the horizontal to vertical spectrum ratio, $S_{NS}(f)$, $S_{EW}(f)$ and $S_Z(f)$ are the Fourier amplification spectra in the North-South, East-West and Vertical directions, respectively [12].

1) AFFINITY PROPAGATION CLUSTERING APPROACH

AP creates clusters by sending messages between pairs of samples until convergence. A dataset is then described using a small number of exemplars, which are identified as those most representative of other samples. The messages sent between pairs represent the suitability for one sample to be the exemplar of the other, which is updated in response to the values from other pairs (Figure 3). This updating happens iteratively until convergence, at which point the final exemplars are chosen, and hence the final clustering is given [41], [47]. **Algorithm 1** denotes the adopted AP algorithm [41].

AP can be interesting as it chooses the number of clusters based on the data provided. For this purpose, the two important parameters are the preference, which controls how many exemplars are used, the damping factor, which damps the responsibility and availability messages to avoid numerical oscillations when updating these messages [47].

It is worth mentioning that speed, wide application, and suitability for a large number of clusters are some of the benefits of the AP. There are some challenges to AP: it is difficult to determine what value of the parameter ''preferred'' would give optimal clustering solutions, and oscillations cannot be automatically removed if they occur. Accordingly, we develop an adaptive AP (AAP) model, which is illustrated in **Algorithm 2**, to enhance the traditional AP in the following areas: adaptive damping factor modification to minimize oscillations (also known as adaptive damping), adaptive escape oscillations, and adaptive exploring the space of preference parameter to identify the best clustering solution for a data set (called adaptive preference scanning). The employed AAP can beat the AP method in terms of clustering quality and oscillation removal. Besides, this developed model uses Silhouette indices to identify optimal clustering solutions. In other words, the model is employed to tackle two intrinsic challenges. First, it finds an optimal ''preference'' parameter for AP. Second, it eliminates the oscillations in the convergence behavior in the classical AP.

Clustering starts by estimating the Euclidean distance as a measure of the similarity matrix. In the similarity matrix, the similarity $s(i, k)$ (for two distinct points, indexed as i and k) indicates how well the data points with index k is suited to be the exemplar (i.e., point that serves as a cluster centre) for data point i. It is calculated using negative squared Euclidean

Algorithm 1: The Affinity Propagation Algorithm

Input $\mathcal{X}$, s ;
while $r(i,k)$ *and* $a(i,k)$ *have no updates* **do**
  if $(i,k) \in \mathcal{X}^2$ **then**
    Compute $r(i,k)$ by Eq. 2 ;
  else
    Break;
  end
  if $(i,k) \in \mathcal{X}^2$ **then**
    Compute $a(i,k)$ by Eq. 3 ;
  else
    Break;
  end
  Compute $c(i,k)$
end
Output Optimal number of clusters (C^*), $\forall\ (i,k) \in \mathcal{X}$;

distance in our implementation: $s(i,k) = -||\mathcal{X}[i] - \mathcal{X}[k]||^2$ for a data set $\mathcal{X}$. For the similarity of a point to itself, i.e., the ''self-preference'' for being an exemplar, we provide two options in our algorithm: setting all equal to the minimum $s(i,k)$ or to the median $s(i,k) \forall i,k$ with $i \neq k$. This is followed by evaluating the responsibility matrix $r(i,k)$ to quantify how well-suited k is to serve as the exemplar for i, relative to other candidates. To begin with, the availabilities are initialized to be zero. In later iterations, as a data point is assigned to an exemplar, the availability drops below zero and reduce the effect of similarity [41].

$$r(i,k) \leftarrow s(i,k) - \max_{k'\ S.t.\ k' \neq k} \left\{a(i,k') + s(i,k')\right\} \quad (2)$$

For points on the diagonal, $r(k,k)$ is calculated as the input preference that point k is chosen as an exemplar $s(k,k)$, minus the largest similarity between the point and all other candidate exemplars. The **availability matrix** $a(i,k)$ is used to represents how appropriate it is for i to pick k as its exemplar, taking into account other points' preference for k as an exemplar.

$$a(i,k) \leftarrow \min\left\{0, r(k,k) + \sum_{i' \neq \{i,k\}} \max\left\{0, r(i',k)\right\}\right\} \quad (3)$$

For points on the diagonal, the following equation is used.

$$a(k,k) \leftarrow \sum_{i'\ S.t.\ i' \neq k} \max\left\{0, r(i,k')\right\} \quad (4)$$

Quality of the estimated clusters is determined using the sum of the responsibility matrix and the availability matrix $c(i,k) \leftarrow r(i,k) + a(i,k)$ which is known as the criterion matrix $c(i,k)$.

2) FEATURES ENGINEERING

In order to generate outputs, all ML algorithms require some input data. The features in this input data are normally in the form of hierarchical columns. To function properly, algorithms need features with a special characteristic. The need for feature engineering emerges in this situation. When using ML or mathematical modeling to build a predictive algorithm, feature engineering refers to the method of choosing and transforming variables. The procedure entails a mixture of data interpretation, rule-of-thumb application, and judgment. For such big amount of data, the data optimization, feature enhancement, random weights can affect the model performance [61]–[64]. In this section, all the performed features engineering steps are outlined for creating a suitable input dataset that meets the specifications of the AP-based ML algorithm. For clustering both amplification and frequency, data is considered and normalized to have the same scale. HVSR measurements from ambient noise recordings imply both reliability of the results and rapidity of data collection. In order to initiate the HVSR technique, microtremor measurements were carried in a total of 307 points covering the KSU campus as represented by the red circles in Figure 4.

Algorithm 2: Adaptive Damping and Escape Exploration for AAP

Input $\lambda = 0.5$, $w = 40$, $MI = 0$, $ps = 0.05$;
while $MI < 50000$ **do**
  $MI = MI + 1$;
  $E_{set}(MI) = E$;
  if $MI > E$ **then**
    Calculate
    $mean(E(MI)) = mean(E_{set}(MI - w/8)\ \forall\ MI)$;
  else
    break;
  end
  if $mean(E(MI))_{new} - mean(E(MI))_{old} < 0$ **then**
    $E_d = 1$, E is decreased;
  else
    E is not decreased;
  end
  if $E_{set}(MI)_{new} - E_{set}(MI)_{old} == 0$ **then**
    $E_c = 1$, E is changed;
  else
    E is not changed;
  end
  if $E_d == 0 || E_c == 0$ **then**
    $E_b = 1$, No Oscillation occurred;
  else
    if $\sum E_b < \frac{2}{3w}$ **then**
      $\lambda = \lambda + ps$;
      if $\lambda > 0.85$ **then**
        $p = p + ps$;
      end
    end
  end
end
Output Exemplars (E) are obtained $\forall s(i,k)$;

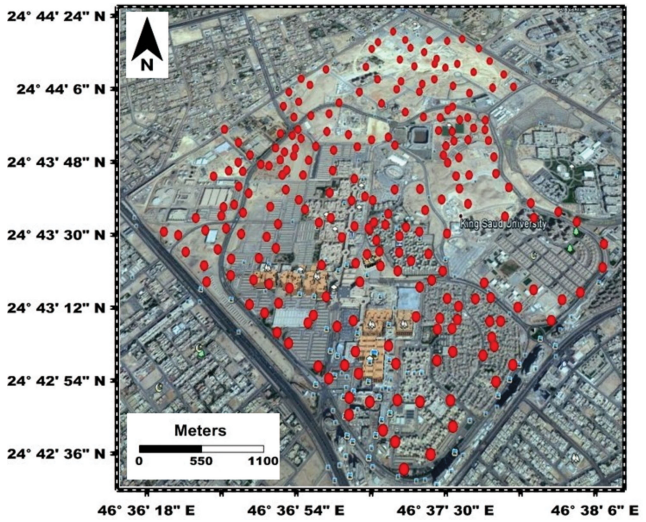

FIGURE 4. KSU campus map showing the location of microtremor sites surveyed.

HVSR measurements have been performed, quite uniformly distributed with a mean spacing of about 200 to 500 m. The utilized equipment for the free-field single-station microtremor data collection is a highly portable three-component seismic station called Tromino 3G ENGY [65], equipped with three velocity transducers. For each conducted measurement point, twenty minutes of ambient noise were recorded at the sampling rate of 128 Hz. The non-stationary portion of the recorded noise was excluded, thus considering only the low-amplification part of the signal, for the computation of the average HVSR function. Seismic station localities were carefully chosen to evade the impact of trees, sources of monochromatic noise, and strong topographic landscapes.

To prevent data from industrial sources, the measurements were taken from late night to early morning hours along the study area. Examples of the collected data are given in Figure 5. The whole measurements were achieved in accordance with the internationally accepted Site Effects Assessment using Ambient Excitation (SESAME) Project guidelines and precautions [66], [67]. For experimental aspects, all the site conditions and parameters were recorded at each station (e.g., recording parameters, recording duration, measurement spacing, in-situ soil-sensor coupling, artificial soil-sensor coupling, sensor setting, nearby structures, weather conditions, and the available geological information). The collected Tromino data were checked for abnormal noise levels and then processed and interpreted. The quality of the obtained microtremor records in our study is categorized into four main categories, high, intermediate, low, and worse quality raw signals of microtremor recording. In the current study, the SESAME guidelines and precautions [66], [67] were fulfilled in 273 sites. Thirty-four sites did not fulfill the standard criteria and were rejected. Rejections were mainly due to the presence of artificial noise or non-clear HVSR peak as the amplification of the peak is too small. An example of each category is given in Figure 5.

The data processing to get the HVSR curves at each individual site was implemented to determine the peak frequency of the soft sedimentary layer above a harder layer, providing a strong impedance contrast, and from that, the peak amplification of that signal was estimated [1]. The origin of the identified peaks of the predominant frequency has been tested first to check whether it is industrial or natural, and then, only natural peaks were considered for the unsupervised AP-based ML clustering. Prior to conducting the HVSR analysis, the GEOPSY damping toolbox [68] is adopted to detect the presence of any data originating from an industrial source utilizing a random decrement technique [69], [70]. An industrial origin is concluded if the damping is much lower than 3% and the frequency is sustained. This detection is important to justify the validity of the recorded ambient noise data used in the HVSR analysis and later in the clustering process. For each microtremor waveform in the database, HVSR is calculated using the geometric mean of the 5% damped velocity waveform of the two horizontal components divided by the corresponding spectral ordinates of the vertical component. Until dividing the horizontal spectra by the vertical one, in [71], logarithmic window (w) smoothing function was used to approximate and smooth the Fourier amplification spectra. The frequency corresponding to the largest peak of the HVSR curve represents the site predominant frequency. Complete HVSR analysis was performed using GEOPSY software developed within the framework of the SESAME project.

III. EXPERIMENTAL RESULTS

Microtremor data collected for the present study suggests that [1] method of using H/V spectral ratios can be remarkably robust tool in determining site effects. Nevertheless, it is suggested that the microtremor technique be used in conjunction to other geotechnical and geophysical studies. To the best of our knowledge, no site effect analyses have been carried out in the study area. Based on the proposed study, there is a great confidence that the obtained outcome can be highly significant to be used for estimating the earthquake losses and designing and implementing phases of new constructions.

HVSR data were processed to extract the frequency and amplification of peaks. The result of this analysis is a natural frequency and an amplification of the local site. Figure 6 shows 12 examples of the estimated HVSR values at different sites. Graphs show sites amplifications as a function of frequencies. More particularly, the effect of surface geology and topography were noticed on microtremor measurements. Based on the obtained frequencies, the area was classified into three classes named as standard grounds, soft rocks and stiff rocks.

In the current study, a visual inspection was used to pick both the frequency and the corresponding amplification, as we noticed that the utilized software in most of the cases could not pick the correct peak. So we did not rely on the plotted (gray) line, which indicates the peak selected by the

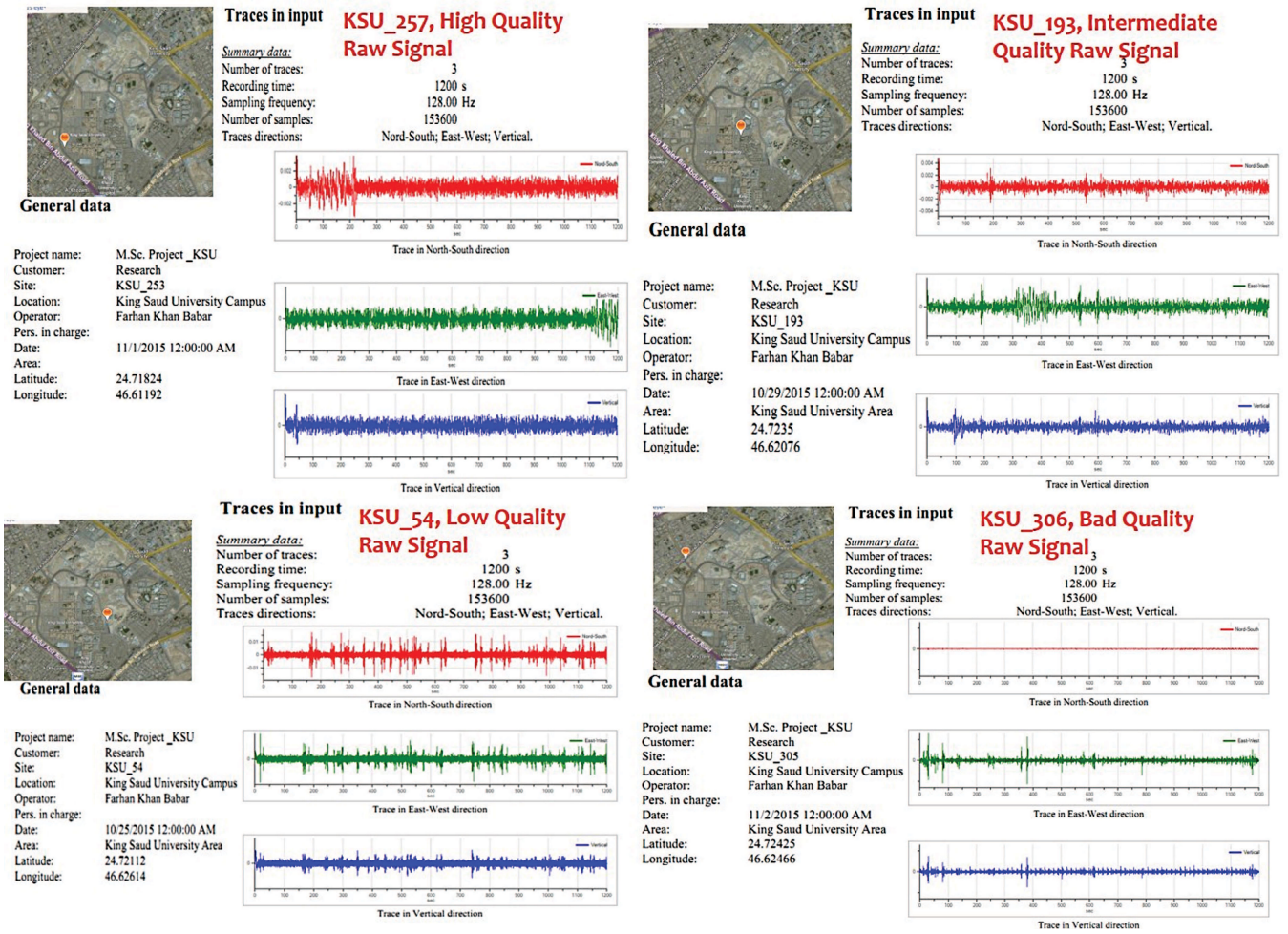

FIGURE 5. Example of high and intermediate quality (top panel) raw signal of microtremor recording at site KSU257 and KSU193. Another example of low and bad quality (bottom panel) raw signal of microtremor recording at site KSU054 and KSU306.

TABLE 3. The estimated HVSR results and presented using median and 25^{th} to 75^{th} percentiles.

	min	max	mean	95%
Frequency	1.37	23.0	8.5	(3.75-11.0)
Amplification	0.10	4.5	1.43	(0.90-1.65)

software. Table 3 indicates the general descriptive statistics of the estimated HVSR data.

A bi-variate distribution of frequency and amplification variables is shown in Figure 7. This multi-panel figure shows both the bi-variate (or joint) relationship between the frequency and amplification variables along with the univariate (or marginal) distribution of each on separate axes. The frequency variable shows larger fluctuations than the amplification variable, which fits in well with the geological settings of the region.

After the identification of the peaks of the HVSR curves attributable to resonance effects, it is essential to delineate regions of similar features characteristics. Non-objective and data-driven decisions based on the approximate frequency and amplification values should be used to define those areas. As it is possible to identify on the HVSR curves, different peaks associable with different resonance frequencies of the investigated site, data-driven procedures are used [72], [73] based on the results of the AP clustering approach. Hence, for the KSU campus, and to group peaks to be attributed to the same origin (e.g., stratigraphic, topographic, anthropogenic, or other sources), a multi-parametric clustering procedure utilizing the AP clustering method [41] has been adopted for better quantitative data-driven site-specific classification. In [74], the error of the presented clustering algorithm was much smaller than that of other algorithms such as K-means as it does not require the number of clusters to be determined or estimated before running the algorithm.

During the implementation of the AP module from the scikit-learn library, it was noted that the likelihood of any HVSR data sample (an exemplar) is strongly influenced by an input parameter known as preference or p [75]. During the iterative procedure utilized to choose the optimal value of p, information is spread among the frequency and amplification points. These data points are handled as a network in which

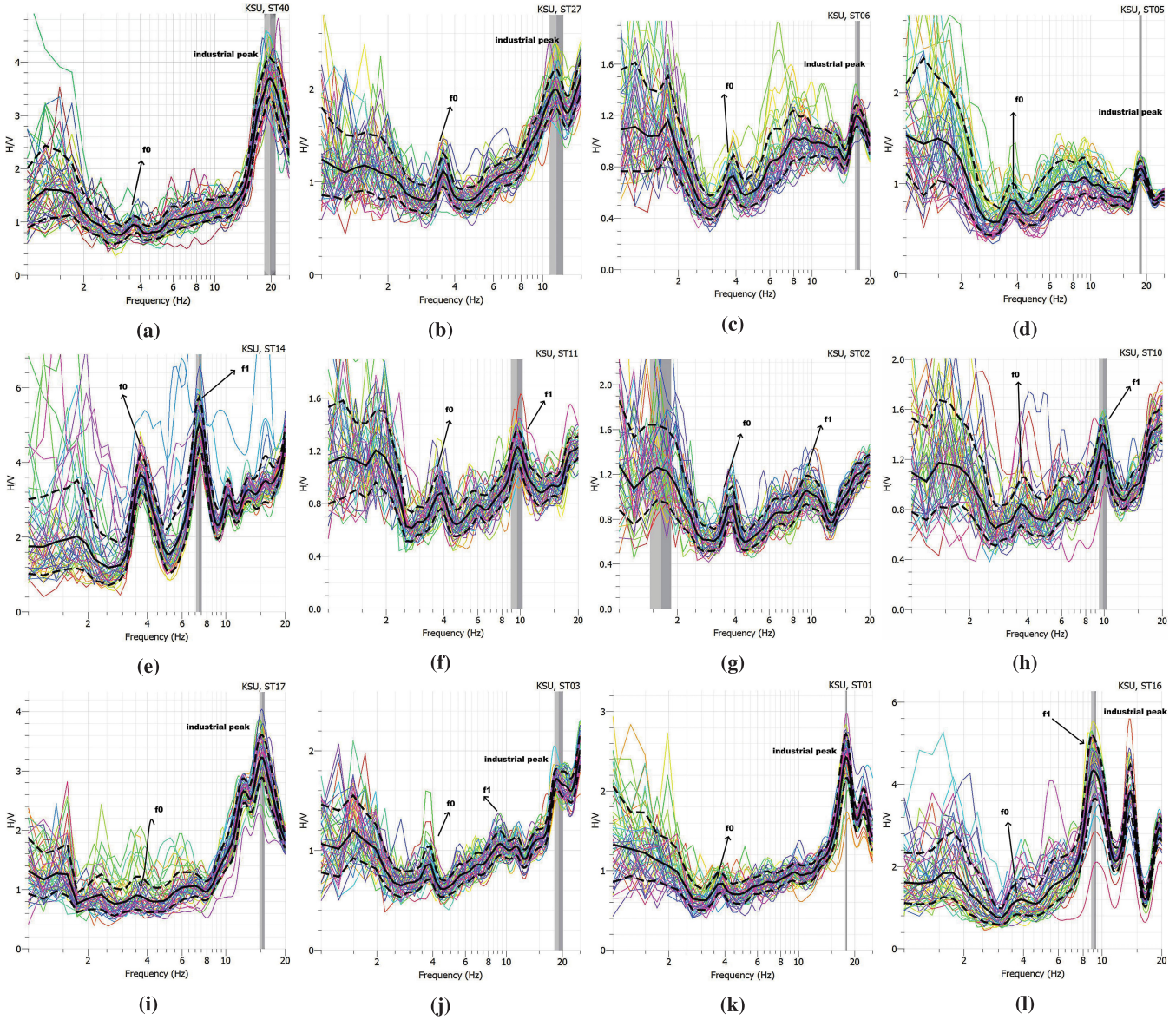

FIGURE 6. Examples of the estimated HVSR frequency and amplification values at different sites in the study area. Graphs show sites amplifications as a function of frequencies. Several peaks some of which are related to local geology of the KSU campus and others are artificial peaks from engineering constructions inside the site. Peaks are labeled based on the results of clustering.

messages are sent back and forth between pairs of frequency and amplification samples (see Figure 3). The algorithm was very sensitive to the input parameters and did not produce a unique number of exemplars. In this case, the AP algorithm is unable to cluster the data points into optimal clusters, and a systematic way to determine the correct optimal number of clusters is needed.

To address the above problem, a modified version of the AP module was coded in this paper along with the global *Silhouette* coefficient [76] as a validity index to overcome the above problem. It is calculated as an average of all samples in clustering and is given by:

$$Silhouetee = \frac{b-a}{\max(a,b)}, \tag{5}$$

where a is the mean intra-cluster distance, and b is the distance between a point and the nearest cluster.

The variation of the *Silhouette* score with the AP various parameters is depicted in Figure 8. Two hundred iterations are performed to calculate the *Silhouette* score utilizing the Squared Euclidean distance between two HVSR data points. Each iteration of AP is consisted of updating all responsibilities given the availabilities, and finally, combining availabilities and responsibilities to monitor the exemplar decisions and terminate the algorithm when these decisions do not change for 15 iterations. When the AP algorithm fails to converge, the damping factor λ is increased to avoid numerical oscillations. The λ close to one has a greater capability to guarantee computational stability but larger time is

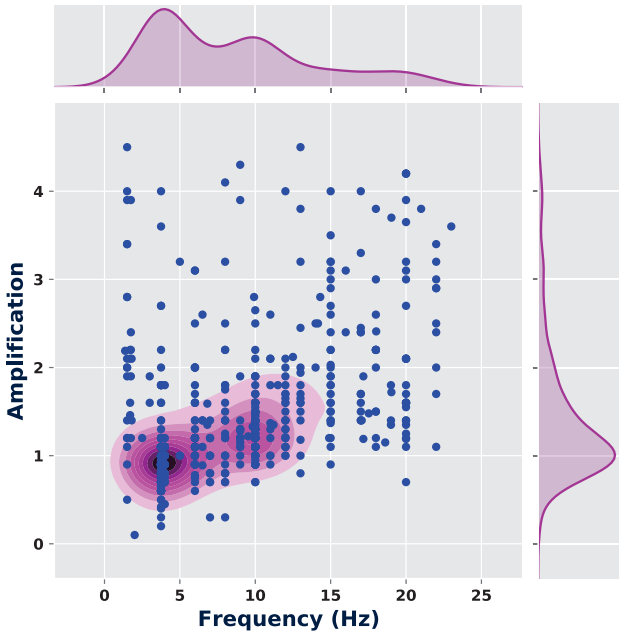

FIGURE 7. Bivariate distribution of HVSR frequency and amplification.

taken [75]. In our experiment, $\lambda = 0.95$ was used to guarantee the stability of the AP algorithm. The peak or maximum value of the *Silhouette* score for the entire clustering near the center of the distribution indicates the optimal parameter value as shown in Figure 8. Using the optimal λ and p as suggested by the *Silhouette* score peak values, the AP clustering procedure has created three clusters from the total estimated observations. More concretely, Figure 8 is performed to show the preliminary step to select the optimum parameters (preferences and damping) for successful AP clustering and compare it to the K-means as a bench mark utilizing maximum value *Selhotti* coefficient as a stopping criteria of the optimization process.

Before we start interpreting the results of the AP clustering algorithm, we benchmark and calibrate the algorithm deduced the number of clusters results by comparing the AP clustering with the optimal number of clusters for HVSR data obtained from the K-means clustering algorithm results. This step is a critical step that will help us in discriminating between observed peaks clusters, which may be caused by source effects and those due to site effects [73], [77], [78].

The K-means clustering algorithm creates clusters by separating data points into the number of clusters or groups k. The value of k is needed to be inputted into the algorithm. The clusters are determined by minimizing the inertia, or the within-cluster sum-of-squares. The inertia is a measure of how coherent the clusters are. By minimizing the inertia, the algorithm tries to minimize the difference between the mean value of a cluster and the values of points in the cluster. The inertia is not normalized, but lower values are better and zero is the optimum value.

For KSU data, to find the optimal number of clusters C^*, the elbow method is firstly utilized. In Figure 8 we can see that the total within-cluster sum of squares has been plotted against the number of clusters C. The bend (elbow) in the graph is detected where the value of C is three. Therefore, for KSU data, the elbow method suggests three-cluster solutions. Secondly, the *Silhouette* score plot as a method of interpretation and validation of consistency within clusters of data also suggests that three clusters are an optimal number of clusters for HVSR data as can be seen in Figure 8. The final AP suggested clusters which represent groups of observations with similar frequency and amplification characteristics are shown in Figure 9.

IV. DISCUSSION

It should be noted that the study area has an explicit increase of urban expansion and new construction projects such as research centers, residential suites, and hotels. Accordingly, the evaluation of site-specific classification strictly needs to a more reliable and intellectual solution for detailed site investigation. To achieve these objectives, a pilot free-field site-response study of the King Saud University Area (KSUA) is undertaken by using the HVSR in conjunction with available geological information to estimate dynamic soil properties and soil amplification ratio (Figure 4). However, the task becomes particularly difficult to identify different peaks of the HVSR curves associable with different resonance frequencies (Figures 5 and 6). To overcome the aforementioned problems, an automatic procedure based on cluster analysis was implemented to create a site-specific classification map for the area.

AP was implemented for a non-objective and data-driven site-specific characterization scheme. HVSR frequency and corresponding amplification values (Figure 7) were used as input data for clustering. The optimum number of clusters was found by estimating the *silhouette* score. In this method, a graphical validation (Figure 8) was used for evaluating the number of clusters and comparing different scenarios [76]. In Figure 8, *Silhouette* score plot shows that HVSR data can be optimally divided into three clusters. This was validated by comparing *Silhouette* score plots of both AP and a well know K-means clustering algorithm (Figure 8). Distributions of each cluster are shown in Figures 9 and 10, while the histogram analysis of the outlined three clusters is shown in Figure 11.

The frequencies of the observed HVSR peaks are distributed in the wide range of 1.37 to 23 Hz (Figures 9 and 11), but 50% of them is in the range 1.37 to 7 Hz, 32.32% are below 13 Hz and 17.68% above 13 Hz. The majority of HVSR spectral ratios calculated has the first peak with the lowest fundamental frequency f_0 corresponding to the overall limestone thick deposits covering the area. The amplification of the HVRS peaks is distributed in a range of 0.5 to 4.5 for the delineated clusters. The observed amplification peaks are related to the impedance contrast between the surface layer and the underlying bedrock, to the lateral heterogeneities,

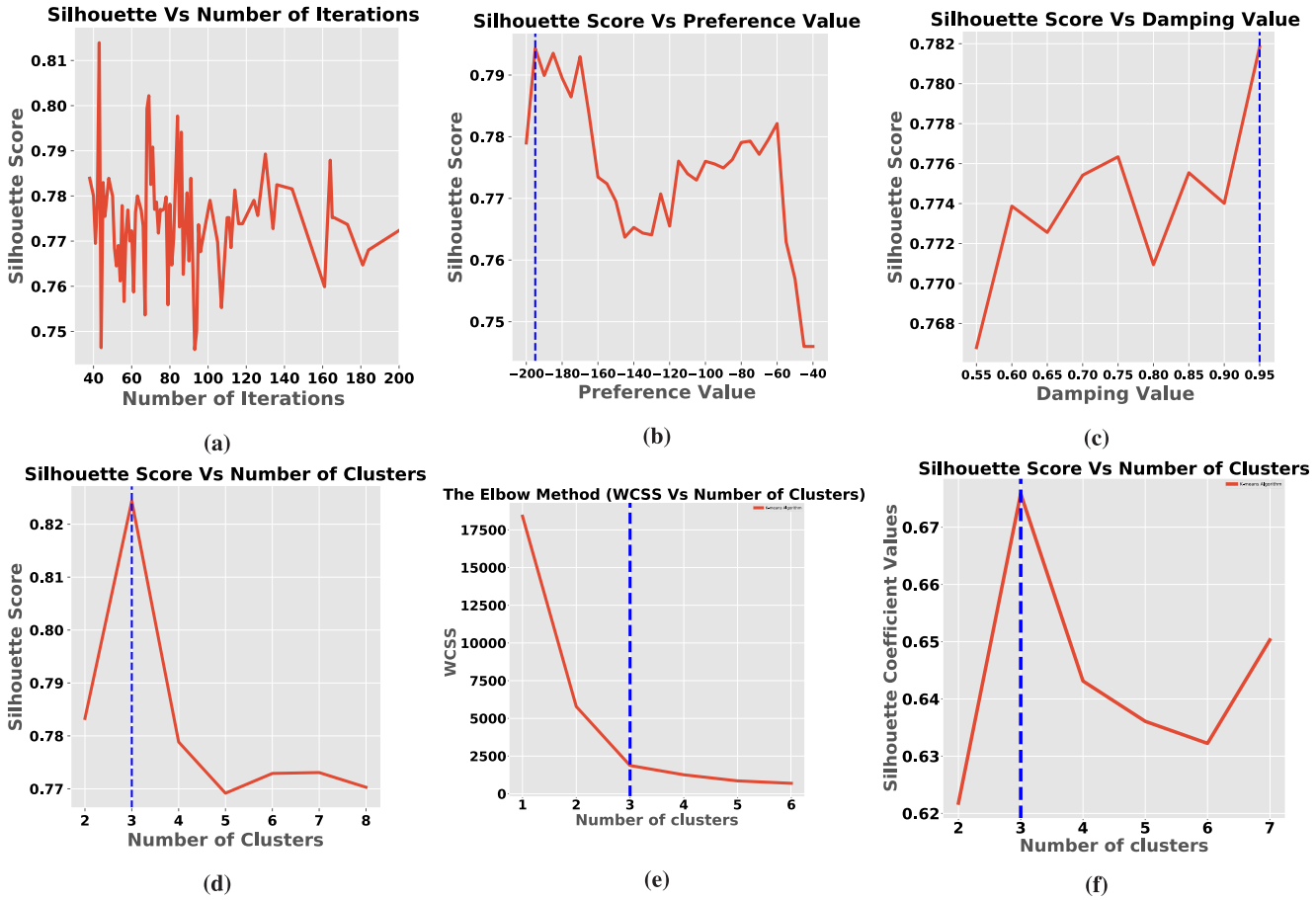

FIGURE 8. Variation of *Silhouette* score with various AP hyperparameters.

to the material damping of sediments, and to the characteristics of the incident wave-field [79]. The spatial distribution of the observed frequency and the corresponding amplification revealed that the subsurface column below the recording stations consist of three layers. The first layer (shallow layer) is overburden and rock fragments. The second layer is considered compact alluvium and/or fractured limestone rocks. The third is the hard and massive limestone bedrock (deep layer), which is mapped below this layer.

Based on the results of the clustering analysis and the collected geological information, it is quite clear that the first and second classes are mainly due to site-specific effects while the third may be due to the very thin soil layer brought to the area for agricultural purposes. The first peak (first cluster) with the lowest frequency was interpreted to be the deepest compacted limestone layer while the other peaks correspond to the highly weathered limestone and the superficial layers (see Figures 6 and 10). The second frequency peak (second cluster) in the HVSR plot was observed at many sites. Identifying a second frequency peak is significant because the amplification of ground motion may also occur at frequencies higher than the fundamental mode even when thick sediments are present [80]. There are two possible explanations for the second frequency peak. The first explanation is that it could be the first higher harmonics of the fundamental frequency of the site. This higher-mode frequency would be expected to be at about three times the fundamental frequency [29], [80], which is what was observed (see Figures6 and 10). The second explanation is that the second frequency peak could be a resonance of a soft soil layer over a shallow stiffer layer [81].

To judge how well the implemented ML models performs, several evaluation metrics were calculated [82] as given in Table 4. Here, clusters are evaluated based on some similarity or dissimilarity measure such as the distance between cluster points cluster analysis for developing site-specific frequency and amplification three and four clusters to demonstrate the efficacy of the proposed approach.

Internal cluster validation indices, which represent the compactness, connectivity, and separability of the clusters, are used to evaluate clustering effectiveness. The connectivity index was estimated to measure to what extent frequency and amplification values are placed in the same cluster as their nearest neighbors in the observations space. For the current clustering experiment, the lower connectivity index which indicates better clustering was estimated for three clusters produced from implementing the AP algorithm. Table 4 summarizes the results of the validation indexes used, with the

TABLE 4. Effectiveness evaluation of clustering methods for the KSU site-specific classification.

Algorithm	C	Connectivity	Dunn	Caliński-Harabasz	Davies-Bouldin	*Silhouette*
DBSCAN	3	10.981	0.5354	84.2252	0.95	0.7517
	4	15.354	0.5123	202.3392	1.0451	0.4041
HDBSCAN	3	11.254	**0.6751**	256.922	0.6422	0.7539
	4	11.892	0.3982	478.7319	0.7513	0.4605
K-means	3	20.5603	0.1576	483.2728	0.8165	0.6787
	4	19.9687	**0.5423**	227.7127	0.8044	0.4857
Agglomerative	3	18.943	0.3259	491.0284	1.0132	0.6439
	4	17.947	0.4700	577.9471	0.8025	0.4789
AAP	3	**11.4929**	0.2123	**644.4392**	**0.8625**	**0.824**
	4	12.8206	0.3389	348.4452	0.8415	0.4737

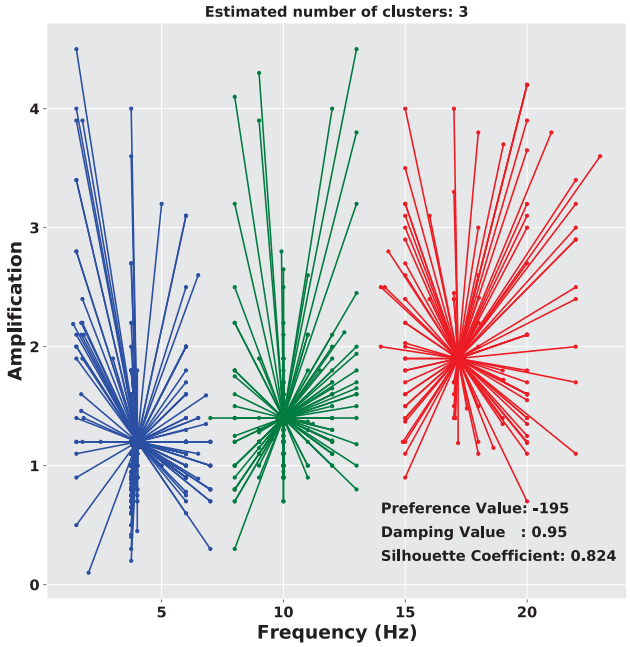

FIGURE 9. Three color-coded exemplars identified by AP clustering algorithm of HVSR data. Optimal clustering parameters are given.

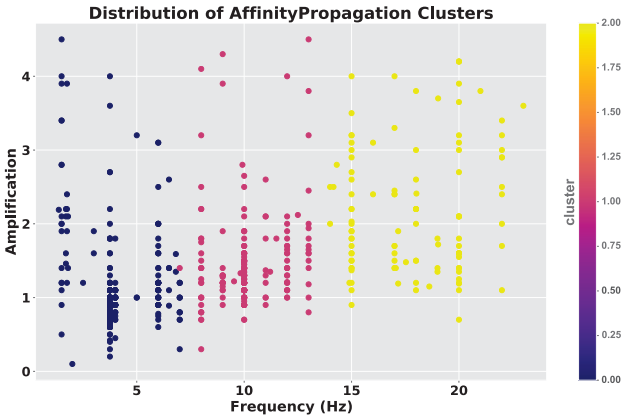

FIGURE 10. Three color-coded spatial distribution of clusters as identified by the AP clustering algorithm of HVSR frequency and amplification data.

best overall results in bold. On three indices (Connectivity, Dunn, Caliński-Harabasz, Davies-Bouldin, and *Silhouette*) AP has the best results, indicating that it is the best clustering algorithm for the site-specific classification of the KSU campus.

The second index used is called the Dunn index. This index aims to identify sets of clusters that are compact and well-separated. That is, when the diameter of the clusters is expected to be small and the distance between clusters to be large. For the current benchmarking, the higher Dunn index which indicates better clustering was obtained from the K-means clustering that has four clusters as given in Table 4. The third Caliński-Harabasz index which considered a variance-ratio criterion (the variance of all cluster centroids from the centroid of the observations) to evaluate the cluster validity was estimated. Higher Caliński-Harabasz index values estimated when implementing the AP algorithm with only three clusters indicate better clustering, for the current clustering configuration. The final cluster validation index implemented is known as the Davies-Bouldin index. This is an internal evaluation scheme, where the validation of how well the clustering has been done is made using quantities and features inherent to the dataset. Higher Davies-Bouldin index values estimated when implementing the AP algorithm with only three clusters indicate better clustering, for the current clustering configuration as given in Table 4.

It is commonly recognized from different studies that the frequency of the HVSR peak replicates the predominant frequency of overburden sedimentary rocks. Its spectral amplification mainly depends on the impedance contrast with deep-seated bedrock and cannot be used as a quantify of amplification of the mapped site. However, the comparison with the standard reference site spectral ratio procedures results has revealed that the maximum amplification of HVSR underestimates the actual site amplification [66], [67]. Hence, we depend mainly on fundamental frequency rather than amplification in our characterization of the KSU campus.

As a final step to developing a non-objective and data-driven map for site classification, the results of the AP clustering procedure were used to draw a map of HVSR average curves classes related to different sites, and to identify areas characterized by site effects probably caused by the same buried structure. Considering the parameters of such peaks as a sampling of spatial trends that are continuous on the KSU campus, it is possible to estimate the expected peak amplification and frequency at each point of

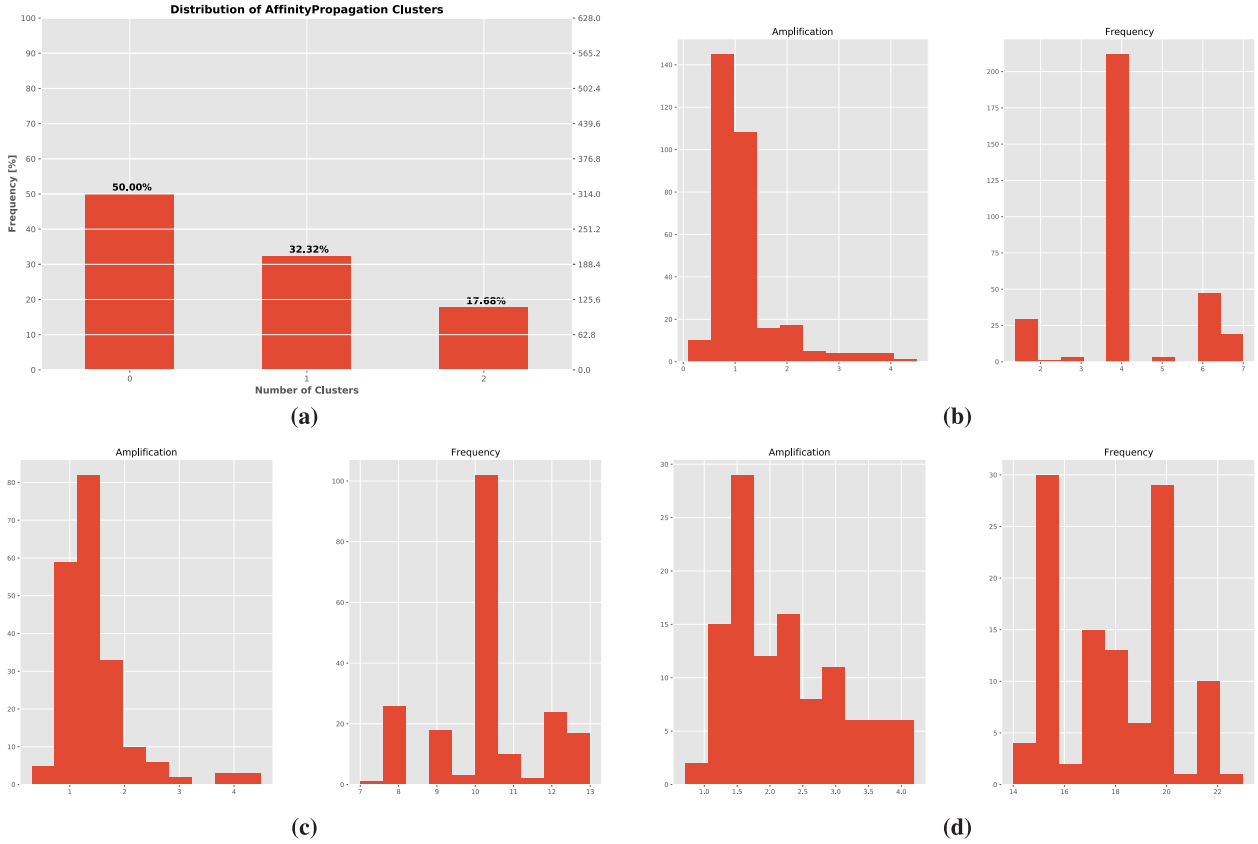

FIGURE 11. (a) Histogram distribution of observed frequency and corresponding amplification of HVSR curves; (b) Low class; (c) Medium class; and (d) High class of observed frequency and corresponding amplification inside KSU campus.

the area by two-dimensional interpolation techniques. The choice of the interpolation method depends on the analysis of the variability scale and the parameter to be studied. In our case, the context is relatively simple, where the geological variations are relatively mild and do not present a strong discontinuity. The natural neighbor interpolation method is used for spatial analysis of resonance frequencies distribution and to build the frequency map shown in Figure 12.

The newly developed fundamental soil frequency map for the KSU campus shown in Figure 12, provides valuable information for assessing soil-structure resonance using the relationship between fundamental building frequency and the height of RC (Reinforced Concrete) structures. To investigate the validity of seismic site response characteristics map estimated from microtremors with the help of unsupervised AAP-based ML shown in Figure 12a was compared with [79] classification, where more than 270 research works have relied on, shown in Figure 12b. Both maps show quite well correlation of classes. The area is clearly could be classified into three classes: standard grounds, soft rocks, and stiff rocks. From ground classifications based on observed frequencies, it can be noted that the KSU region has mainly soft and stiff limestone rocks and along with few sites, the standard ground is observed.

Another validation of our results is performed by comparing ML ground classifications with the nearby King Abdulaziz City for Science and Technology (KACST) and Almalqa area studies. There are almost in good agreement. On closer comparing of the KACST study, the first peak varies from 7.85 Hz to 8.72 Hz which clarifies the impedance contrast between the uppermost soil surface and the underlying completely weathered limestone, while the second peak ranges between 1.41 Hz and 1.46 Hz that corresponds to the impedance contrast between the completely weathered limestone and the underlying hard limestone rocks. Whereas, based on the frequency of the HVSR peak, the present study area has the first peak with the lowest fundamental frequency for ranging from 3.0 Hz to 8.25 Hz while the second peak f_1 and the third peak f_2 varies from 9.5 Hz to 13.0 Hz and from 13.0 Hz to 20 Hz, respectively. As within the study area, the majority of HVSR calculated has two peaks (f_0 and f_1) data which indicating agreement with the nearby KACST microtremor studies. The same comparison for the other nearby Almalqa area microtremor studies yielded results as: three zones of different frequencies, zone-1 up to 1.7 Hz, zone-2 from 1.7 Hz to 3.5 Hz, and the lastly zone-3 from 3.5 Hz to 10 Hz. These results suggest that the microtremor HVSR spectrum is a reliable tool to estimate the fundamental frequency.

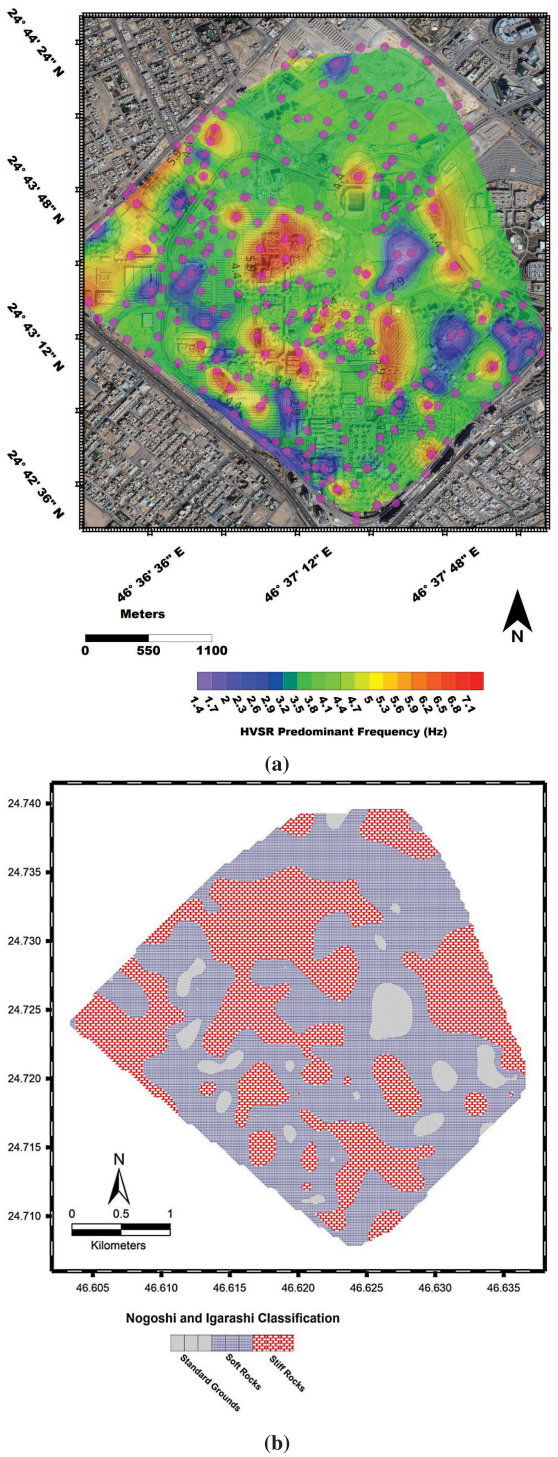

FIGURE 12. **(a) Map of the fundamental HVSR soil frequencies obtained for the KSU campus. Circles indicate microtremor measurements; (b) Ground classifications based on observed frequencies and the proposed method in [79].**

V. CONCLUSION

Although, the effect of surface geology and topography were noticed on the measured microtremor data sets, there is no previous site effect analyses have been carried out on the KSU campus. Besides, the majority of HVSR spectral ratios calculated in the study area have two peaks (f_0 and f_1). Based on the obtained frequencies, the area was classified into three classes named as standard grounds, soft rocks, and stiff rocks. To automatically determine the number of clusters and the corresponding cluster centers, we adopted AP clustering which is one of the unsupervised ML techniques. Moreover, a successive clustering procedure has been used to group the main HVSR clusters of peaks and categorized them with areas characterized by site effects reasonably caused by the same lithological features. Assuming that the three main identified clusters contain peaks produced by resonance effects of layers with varying thickness, the possible trend of the top of the seismic bedrock was reconstructed by inversion of the HVSR curves constrained with geological and lithological information and considering the minimum lateral variability of the physical and geometrical parameters. It is concluded that the cluster center generates the strongest interference compared to other cluster members.

Results of the current microtremor studies are consistent with previous work in the nearby KACST and Almalqa areas and suggest that the HVSR method is useful in evaluating seismic ground response. Besides, it is worth emphasizing that the site-effect analysis can assist the decision-makers to set the priorities in managing land uses, estimating the earthquake losses, conducting programs for reducing the vulnerability of existing structures, enforcing building codes, planning an emergency response and long-term recovery, and designing and implementing phases of new constructions. Furthermore, the site response variations are significant over very short distances. Accordingly, it is strongly recommended that the estimation of any future earthquake loss scenarios for the KSU or Riyadh city should be based on the site-response functions obtained over a relatively dense grid of measurement points. In the future work, we plan to extend the proposed AP-based clustering scheme to other urban areas using the microtremors data and site response functions.

ACKNOWLEDGMENT

The authors would like to thank the Research Supporting Project number (RSP-2021/89), King Saud University, Riyadh, Saudi Arabia for funding this work.

REFERENCES

[1] Y. Nakamura, "A method for dynamic characteristics estimation of subsurface using microtremor on the ground surface," *Railway Tech. Res. Inst., Quart. Rep.*, vol. 30, no. 1, pp. 25–33, 1989.
[2] M. K. Koçkar and H. Akgán, "Evaluation of the site effects of the Ankara basin, Turkey," *J. Appl. Geophys.*, vol. 83, pp. 120–134, Aug. 2012.
[3] S. Bonnefoy-Claudet, F. Cotton, and P.-Y. Brad, "The nature of noise wavefield and its applications for site effects studies: A literature review," *Earth-Sci. Rev.*, vol. 79, pp. 205–227, Dec. 2006.
[4] K. Aki, "Local site effects on weak and strong ground motion," *Tectonophysics*, vol. 218, nos. 1–3, pp. 93–111, Feb. 1993.
[5] P.-Y. Bard, "Microtremor measurements: A tool for site effect estimation," *Effects Surf. Geol. Seismic Motion*, vol. 3, pp. 1251–1279, Jan. 1999.
[6] F. J. Sánchez-Sesma, V. J. Palencia, and F. Luzón, "Estimation of local site effects during earthquakes: An overview," in *From Seismic Source to Structural Response: Contributions Professor Mihailo D. Trifunac*. India: Indian Soc. Earthquake Technol., 2004, pp. 44–70.
[7] K. Atakan, "A review of the type of data and the techniques used in empirical estimation of local site response," in *Proc. Int. Conf. Seismic Zonation*, 1996, pp. 1451–1460.

[8] M. Mucciarelli and M. R. Gallipoli, "A critical review of 10 years of microtremor hvsr technique," *Boll. Geof. Teor. Appl*, vol. 42, nos. 3–4, pp. 255–266, 2001.
[9] M. Nagoshi and T. Igarashi, "On the amplitude characteristics of microtremors (Part 2)," *Zisin*, vol. 45, pp. 26–40, Dec. 1971.
[10] Y. Nakamura, E. D. Gurler, J. Saita, A. Rovelli, and S. Donati, "Vulnerability investigation of Roman Colosseum using microtremor," in *Proc. WCEE*, 2000, pp. 1–4.
[11] J. Lermo and F. J. Chávez-García, "Site effect evaluation using spectral ratios with only one station," *Bull. Seismol. Soc. Amer.*, vol. 83, no. 5, pp. 1574–1594, Oct. 1993.
[12] J. Lermo and F. J. Chávez-García, "Are microtremors useful in site response evaluation?" *Bull. seismol. Soc. Amer.*, vol. 84, no. 5, pp. 1350–1364, 1994.
[13] C. Lachetl and P.-Y. Bard, "Numerical and theoretical investigations on the possibilities and limitations of Nakamura's Technique," *J. Phys. Earth*, vol. 42, no. 5, pp. 377–397, 1994.
[14] E. Field and K. Jacob, "The theoretical response of sedimentary layers to ambient seismic noise," *Geophys. Res. Lett.*, vol. 20, no. 24, pp. 2925–2928, Dec. 1993.
[15] M. Mucciarelli, "Reliability and applicability of nakamura's technique using microtremors: An experimental approach," *J. Earthq. Eng.*, vol. 2, no. 4, pp. 625–638, Oct. 1998.
[16] A. M. Lontsi, F. J. Sánchez-Sesma, J. C. Molina-Villegas, M. Ohrnberger, and F. Kräger, "Full microtremor H/V(z, f) inversion for shallow subsurface characterization," *Geophys. J. Int.*, vol. 202, no. 1, pp. 298–312, Jul. 2015.
[17] M. Ibs-von Seht and J. Wohlenberg, "Microtremor measurements used to map thickness of soft sediments," *Bull. Seismol. Soc. Amer.*, vol. 89, no. 1, pp. 250–259, Feb. 1999.
[18] B. Tian, Y. Du, Z. You, and R. Zhang, "Measuring the sediment thickness in urban areas using revised H/V spectral ratio method," *Eng. Geol.*, vol. 260, Oct. 2019, Art. no. 105223.
[19] A. K. Panah, N. H. Moghaddas, M. Ghayamghamian, M. Motosaka, M. Jafari, and A. Uromieh, "Site effect classification in east-central of Iran," *J. Seismol. Earthq. Eng.*, vol. 4, no. 1, p. 37, 2002.
[20] M. Qaisar, K. Karam, I. Talat, M. Tariq, and M. Daud Shah, "Fateh Jang (Pakistan) Earthquake of February 17, 1993: Source mechanism and intensity distribution," *J. Himalayan Earthq. Sci.*, vol. 41, pp. 45–52, Dec. 2008.
[21] M. Y. Walling, W. K. Mohanty, S. K. Nath, S. Mitra, and A. John, "Microtremor survey in Talchir, India to ascertain its basin characteristics in terms of predominant frequency by Nakamura's ratio technique," *Eng. Geol.*, vol. 106, nos. 3–4, pp. 123–132, Jun. 2009.
[22] M. S. Fnais, K. Abdelrahman, and A. M. Al-Amri, "Microtremor measurements in Yanbu city of Western Saudi Arabia: A tool for seismic microzonation," *J. King Saud Univ.-Sci.*, vol. 22, no. 2, pp. 97–110, Apr. 2010.
[23] A. J. Choobbasti, S. Rezaei, and F. Farrokhzad, "Evaluation of site response characteristics using microtremors," *Gradevinar*, vol. 65, pp. 731–741, 2013.
[24] S. S. Moustafa, "Microtremor analysis of marsa matrouh industrial area using horizontal to vertical spectral ratio method," *EJGE*, vol. 20, pp. 1591–1602, Dec. 2015.
[25] M. Al-Malki, M. Fnais, A. Al-Amri, and K. Abdelrahman, "Estimation of fundamental frequency in Dammam city, eastern Saudi Arabia," *Arabian J. Geosci.*, vol. 8, no. 4, pp. 2283–2298, Apr. 2015.
[26] K. Alyousef, K. Aldamegh, K. Abdelrahman, O. Loni, R. Saud, A. Al-Amri, and M. Fnais, "Evaluation of site response characteristics of King Abdulaziz City for Science and Technology, Saudi Arabia using microtremors and geotechnical data," *Arabian J. Geosci.*, vol. 8, no. 7, pp. 5181–5188, Jul. 2015.
[27] M. Hellel, E. Oubaiche, J.-L. Chatelain, R. Bensalem, N. Amarni, M. Boukhrouf, and M. Wathelet, "Efficiency of ambient vibration HVSR investigations in soil engineering studies: Backfill study in the Algiers (Algeria) harbor container terminal," *Bull. Eng. Geol. Environ.*, vol. 78, n. 7, pp. 4989–5000, 2019.
[28] P. Clemente, G. Delmonaco, L. M. Puzzilli, and F. Saitta, "Stability and seismic vulnerability of the stylite tower at umm ar-rasas," *Ann. Geophys.*, vol. 61, p. 49, Jan. 2019.
[29] P. Anbazhagan, K. N. Srilakshmi, K. Bajaj, S. S. R. Moustafa, and N. S. N. Al-Arifi, "Determination of seismic site classification of seismic recording stations in the Himalayan region using HVSR method," *Soil Dyn. Earthq. Eng.*, vol. 116, pp. 304–316, Jan. 2019.
[30] Y. Fukushima, L. F. Bonilla, O. Scotti, and J. Douglas, "Site classification using horizontal-to-vertical response spectral ratios and its impact when deriving empirical ground-motion prediction equations," *J. Earthq. Eng.*, vol. 11, no. 5, pp. 712–724, Oct. 2007.
[31] C. Di Alessandro, L. F. Bonilla, D. M. Boore, A. Rovelli, and O. Scotti, "Predominant-period site classification for response spectra prediction equations in Italy," *Bull. Seismol. Soc. Amer.*, vol. 102, no. 2, pp. 680–695, Apr. 2012.
[32] S. Yaghmaei-Sabegh and R. Rupakhety, "A new method of seismic site classification using HVSR curves: A case study of the 12 November 2017 Mw 7.3 Ezgeleh earthquake in Iran," *Eng. Geol.*, vol. 270, Jun. 2020, Art. no. 105574.
[33] M. S. Abdalzaher, M. El-Hadidy, H. Gaber, and A. Badawy, "Seismic hazard maps of Egypt based on spatially smoothed seismicity model and recent seismotectonic models," *J. Afr. Earthq. Sci.*, vol. 170, Oct. 2020, Art. no. 103894.
[34] M. Elhadidy, M. S. Abdalzaher, and H. Gaber, "Up-to-date PSHA along the Gulf of Aqaba-Dead Sea transform fault," *Soil Dyn. Earthq. Eng.*, vol. 148, Sep. 2021, Art. no. 106835.
[35] Q. Kong, D. T. Trugman, Z. E. Ross, M. J. Bianco, B. J. Meade, and P. Gerstoft, "Machine learning in seismology: Turning data into insights," *Seismolog. Res. Lett.*, vol. 90, no. 1, pp. 3–14, 2018.
[36] G. Adelfio, M. Chiodi, A. D'Alessandro, D. Luzio, G. D'Anna, and G. Mangano, "Simultaneous seismic wave clustering and registration," *Comput. Geosci.*, vol. 44, pp. 60–69, Jul. 2012.
[37] J. A. Hartigan, *Clustering Algorithms*. Hoboken, NJ, USA: Wiley, 1975.
[38] R. Xu and D. Wunsch, *Clustering*. Hoboken, NJ, USA: Wiley, 2009.
[39] A. S. Sabau, "Survey of clustering based financial fraud detection research," *Inf. Economica*, vol. 16, no. 1, p. 110, 2012.
[40] G. Gan, C. Ma, and J. Wu, *Data Clustering—Theory, Algorithms, Application*. Philadelphia, PA, USA: SIAM, 2007.
[41] B. J. Frey and D. Dueck, "Clustering by passing messages between data points," *Science*, vol. 315, no. 5814, pp. 972–976, Feb. 2007.
[42] R. Refianti, A. B. Mutiara, and A. A. Syamsudduha, "Performance evaluation of affinity propagation approaches on data clustering," *Int. J. Adv. Comput. Sci. Appl.*, vol. 7, no. 3, pp. 420–429, 2016.
[43] Z. Tian, X. Tang, M. Zhou, and Z. Tan, "Fingerprint indoor positioning algorithm based on affinity propagation clustering," *EURASIP J. Wireless Commun. Netw.*, vol. 2013, no. 1, p. 272, 2013.
[44] K. Wang, J. Zhang, D. Li, X. Zhang, and T. Guo, "Adaptive affinity propagation clustering," 2008, *arXiv:0805.1096*.
[45] Y. Fujiwara, G. Irie, and T. Kitahara, "Fast algorithm for affinity propagation," in *Proc. 32nd Int. Joint Conf. Artif. Intell.*, 2011, pp. 2238–2243.
[46] A. F. El-Samak and W. Ashour, "Optimization of traveling salesman problem using affinity propagation clustering and genetic algorithm," *J. Artif. Intell. Soft Comput. Res.*, vol. 5, no. 4, pp. 239–245, Oct. 2015.
[47] D. Dueck, "Affinity propagation: Clustering data by passing messages," Ph.D. dissertation, Dept. Elect. Comput. Eng., Univ. Toronto, Toronto, ON, Canada, 2009.
[48] M. S. Abdalzaher, M. Elwekeil, T. Wang, and S. Zhang, "A deep autoencoder trust model for mitigating jamming attack in IoT assisted by cognitive radio," *IEEE Syst. J.*, early access, Aug. 10, 2021, doi: 10.1109/JSYST.2021.3099072.
[49] J. MacCarthy, O. Marcillo, and C. Trabant, "Seismology in the cloud: A new streaming workflow," *Seismol. Res. Lett.*, vol. 91, no. 3, pp. 1804–1812, May 2020.
[50] S. S. Moustafa, M. S. Abdalzaher, M. H. Yassien, T. Wang, M. Elwekeil, and E. A. H. Hafiez, "Development of an optimized regression model to predict blast-driven ground vibrations," *IEEE Access*, vol. 9, pp. 31826–31841, 2021.
[51] M. S. Abdalzaher, M. S. Soliman, S. M. El-Hady, A. Benslimane, and M. Elwekeil, "A deep learning model for earthquake parameters observation in IoT system-based earthquake early warning," *IEEE Internet Things J.*, early access, Sep. 22, 2021, doi: 10.1109/JIOT.2021.3114420.
[52] D. Xu and Y. Tian, "A comprehensive survey of clustering algorithms," *Ann. Data Sci.*, vol. 2, no. 2, pp. 165–193, 2015.
[53] L. McInnes and J. Healy, "Accelerated hierarchical density based clustering," in *Proc. Int. Conf. Data Mining Workshops (ICDMW)*, 2017, pp. 33–42.
[54] M. S. Abdalzaher, S. S. R. Moustafa, M. Abd-Elnaby, and M. Elwekeil, "Comparative performance assessments of machine-learning methods for artificial seismic sources discrimination," *IEEE Access*, vol. 9, pp. 65524–65535, 2021.
[55] J. Wu, "Cluster analysis and K-means clustering: An introduction," in *Advances in K-means Clustering*. Berlin, Germany: Springer, 2012, pp. 1–16.

[56] M. Searle, "Arabia: Geography, history and exploration," in *Geology of the Oman Mountains, Eastern Arabia*. Berlin, Germany: Springer, 2019, pp. 3–25.
[57] T. Al-Refeai and D. Al-Ghamdy, "Geological and geotechnical aspects of Saudi Arabia," *Geotechn. Geol. Eng.*, vol. 12, no. 4, pp. 253–276, Dec. 1994.
[58] D. Vaslet, M. Al-Muallem, S. Maddeh, J. Brosse, J. Fourniquet, J. Breton, and Y. Le Nindre, "Explanatory notes to the geologic map of the Ar Riyad quadrangle, sheet 24 I, Kingdom of Saudi Arabia," Saudi Arabian Deputy Ministry Mineral Resour., Jeddah, Saudi Arabia, Tech. Rep. GM-121, 1991, vol. 54.
[59] B. Soleimani, S. Brumand, and F. Khoshbakht, "Petrophysical evaluation of Arab formation using multimin, petrography and petrography carbonate methods in one of Iranian oilfields, Persian Gulf," *Int. J. Sci. Technol.*, vol. 4, no. 2, p. 75, 2016.
[60] E. Ibrahim, "Aeromagnetic data interpretation to locate buried faults in Riyadh Region, Saudi Arabia," *Sci. Res. Essays*, vol. 7, no. 22, pp. 2022–2030, Jun. 2012.
[61] C. Wei, J. Zhang, T. Valiullin, W. Cao, Q. Wang, and H. Long, "Distributed and parallel ensemble classification for big data based on Kullback-Leibler random sample partition," in *Proc. Int. Conf. Algorithms Archit. Parallel Process.* Cham, Switzerland: Springer, 2020, pp. 448–464.
[62] Z. Xie, W. Cao, and Z. Ming, "A further study on biologically inspired feature enhancement in zero-shot learning," *Int. J. Mach. Learn. Cybern.*, vol. 12, no. 1, pp. 257–269, Jan. 2021.
[63] W. Cao, L. Hu, J. Gao, X. Wang, and Z. Ming, "A study on the relationship between the rank of input data and the performance of random weight neural network," *Neural Comput. Appl.*, vol. 2021, pp. 1–12, Jan. 2020.
[64] W. Cao, X. Wang, Z. Ming, and J. Gao, "A review on neural networks with random weights," *Neurocomputing*, vol. 275, pp. 278–287, Jan. 2018.
[65] M. Spa, *Manuale Tromino Eng Tr-Engy Plus*. Chennai, India: MSPA, 2018.
[66] P. Bard and S. Participants, "The sesame project: An overview and main results," in *Proc. 13rd World Conf. Earthq. Eng.*, Vancouver, BC, Canada, Aug. 2004, pp. 1–6.
[67] C. Acerra, "Guidelines for the implementation of the H/V spectral ratio technique on ambient vibrations measurements, processing and interpretation," Eur. Commission, Brussels, Belgium, Tech. Rep. EVG1-CT-2000-00026 SESAME, 2004.
[68] M. Wathelet, "Geopsy geophysical signal database for noise array processing," in *Software*. Grenoble, France: LGIT, 2005.
[69] F. Dunand, P. Bard, J. Chatelain, P. Guéguen, T. Vassail, and M. Farsi, "Damping and frequency from randomdec method applied to *in situ* measurements of ambient vibrations. evidence for effective soil structure interaction," in *Proc. 12nd Eur. Conf. Earthq. Eng.*, London, U.K., 2002, pp. 1–4.
[70] M. G. Koller, J.-L. Chatelain, B. Guillier, A.-M. Duval, K. Atakan, C. Lacave, and P. Bard, "Practical user guidelines and software for the implementation of the H/V ratio technique: Measuring conditions, processing method and results interpretation," in *Proc. 13rd world Conf. Earthq. Eng.*, Vancouver, BC, Canada, 2004, pp. 1–3.
[71] K. Konno and T. Ohmachi, "Ground-motion characteristics estimated from spectral ratio between horizontal and vertical components of microtremor," *Bull. Seismol. Soc. Amer.*, vol. 88, no. 1, pp. 228–241, 1998.
[72] P. L. Bragato, G. Laurenzano, and C. Barnaba, "Automatic zonation of urban areas based on the similarity of H/V spectral ratios," *Bull. Seismol. Soc. Amer.*, vol. 97, no. 5, pp. 1404–1412, Oct. 2007.
[73] R. Martorana, P. Capizzi, A. D'Alessandro, D. Luzio, P. Di Stefano, P. Renda, and G. Zarcone, "Contribution of HVSR measures for seismic microzonation studies," *Ann. Geophys.*, vol. 61, no. 2, pp. 1–17, Jun. 2018.
[74] Y. Zhu, J. Yu, and C. Jia, "Initializing K-means clustering using affinity propagation," in *Proc. 9th Int. Conf. Hybrid Intell. Syst.*, vol. 1, 2009, pp. 338–343.
[75] C. Sun, C. Wang, S. Song, and Y. Wang, "A local approach of adaptive affinity propagation clustering for large scale data," in *Proc. Int. Joint Conf. Neural Netw.*, Jun. 2009, pp. 2998–3002.
[76] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *J. Comput. Appl. Math.*, vol. 20, no. 1, pp. 53–65, 1987.
[77] A. D'Alessandro, P. Capizzi, D. Luzio, R. Martorana, and N. Messina, "Improvement of hvsr technique by cluster analysis," in *Proc. FIST GEOITALIA Forum Scienze della Terra*, 2013, p. 193.
[78] P. Capizzi, R. Martorana, G. Stassi, A. D'alessandro, and D. Luzio, "Centroid-based cluster analysis of hvsr data for seismic microzonation," in *Near Surface Geoscience*, vol. 1. Dubai, India: European Association of Geoscientists & Engineers, 2014, pp. 1–5.
[79] M. Nogoshi and T. Igarashi, "On the propagation characteristics of microtremors," *J. Seism. Soc. Jpn.*, vol. 23, pp. 264–280, Dec. 1970.
[80] S. Parolai, S. M. Richwalski, C. Milkereit, and P. Bormann, "Assessment of the stability of H/V spectral ratios from ambient noise and comparison with earthquake data in the cologne area (Germany)," *Tectonophysics*, vol. 390, nos. 1–4, pp. 57–73, Oct. 2004.
[81] P. Bodin, K. Smith, S. Horton, and H. Hwang, "Microtremor observations of deep sediment resonance in metropolitan Memphis, Tennessee," *Eng. Geol.*, vol. 62, nos. 1–3, pp. 159–168, Oct. 2001.
[82] A. Thalamuthu, I. Mukhopadhyay, X. Zheng, and G. C. Tseng, "Evaluation and comparison of gene clustering methods in microarray analysis," *Bioinformatics*, vol. 22, no. 19, pp. 2405–2412, 2006.

SAYED S. R. MOUSTAFA received the B.Sc. degree in geophysics from Cairo University, Egypt, in 1990, the M.Sc. and Ph.D. degrees in geophysics from Ain Shams University, Cairo, Egypt, in 1997 and 2002, respectively, and the Diploma degree in seismology and earthquake engineering from the International Institute of Seismology and Earthquake Engineering (IISEE), Japan, in 2001. Since December 1994, he has been with the Egyptian National Seismic Network Lab (ENSN), Department of Seismology, National Research Institute of Astronomy and Geophysics, Cairo, where he was an Assistant Professor, became an Associate Professor in 1998, and a Professor in 2002. From 2009 to 2019, he was a Professor with the Geology and Geophysics Department, College of Science, King Saud University, Saudi Arabia. His current research interests include earthquake rupture mechanics, numerical methods for wave propagation, spectral element method for ground motion simulation, site response and seismic hazard, characterization of sedimentary basins, and simulation of their seismic response. He is a member of the American Geophysical Union (AGU), the Society of Exploration Geophysicists (SEG), and the Egyptian Geophysical Society (EGS).

MOHAMED S. ABDALZAHER (Member, IEEE) received the B.Sc. degree (Hons.) in electronics and communications engineering and the M.Sc. degree in electronics and communications engineering from Ain Shams University, Cairo, Egypt, in 2008 and 2012, respectively, and the Ph.D. degree from the Department of Electronics and Communications Engineering, Egypt-Japan University of Science and Technology, Madinet Borg Al Arab, Egypt, in 2016.

He was a Special Research Student with Kyushu University, Fukuoka, Japan, from 2015 to 2016. From April 2019 to October 2019, he was with the Center for Japan-Egypt Cooperation in Science and Technology, Kyushu University, where he was a Postdoctoral Researcher. He is currently an Associate Professor with the National Research Institute of Astronomy and Geophysics, Cairo. His research interests include earthquake engineering, data communication networks, wireless communications, WSNs security, the IoT, and deep learning.

Dr. Abdalzaher is a TPC Member of the Vehicular Technology Conference and International Japan-Africa Conference on Electronics, Communications and Computers and a Reviewer of the IEEE Internet of Things Journal, IEEE Systems Journal, IEEE Access, *Transactions on Emerging Telecommunications Technologies*, *Applied Soft Computing*, *Journal of Ambient Intelligence and Humanized Computing*, and IET journals.

FARHAN KHAN received the B.Sc. degree in applied geophysics from Quaid-e-Azam University, Islamabad, in 2009, and the M.Sc. degree in engineering geophysics from King Saud University (KSU), Riyadh, Saudi Arabia, in 2017. From 2013 to 2017, he was a Researcher with the Geology and Geophysics Department, KSU. In 2017, he joined Saudi National Hydrographic Office at General Authority for Survey and Geospatial Information, in Marine Cartography and Data Processing Section, for chart production and QA/QC of the different Navigational Products in the Red Sea and Arabian Gulf as per IHO standards. In March 2021, he joined the Gulf Center for Geophysical and Water Consulting (GCGC) and started working on engineering geophysical projects for near surface geophysical investigation in Kingdom of Saudi Arabia. His current research interests include the seismic methods (MASW, Refraction), horizontal to vertical spectral ratio (HVSR), electrical methods (Resistivity, IP, and SP), electromagnetic methods (TEM, VLF), borehole geophysics, and ground penetrating radar (GPR). He was a member of the Society of Exploration Geophysicist (SEG) and Dhahran Geoscience Society (DGS) (Saudi Aramco's Organization) for the 3rd SEG/DGS Middle East Geoscience Young Professionals and Student Event in Manama, Bahrain.

MOHAMED METWALY received the master's degree in applied geophysics from Mansoura University, Egypt, in 1999, and the joint Ph.D. degree from ETH Zürich, Zürich, and Mansura University, in 2004. He attended to Tokyo University, from 2006 to 2008. His main field of research is the applied geophysics and its contributions to solve the environmental and engineering problems. He has been working with King Saud University, Saudi Arabia, since 2008, in the field of applied environmental, archaeological, and engineering geophysics. He has over 50 published articles, five published books, and book chapters as well as five awarded research prices.

ESLAM A. ELAWADI received the Ph.D. degree in engineering geophysics from the Faculty of Engineering, Kyushu University, Japan, in 2003. In 2007, he was a Postdoctoral Research Fellow with the Department of Earth and Environmental Sciences, University of Kentucky, USA. From 2008 to 2013, he worked with Geology and Geophysics Department, Faculty of Science, King Saud University, Saudi Arabia, as an Assistant Professor. In 1989, he joined the Exploration Division, Nuclear Materials Authority (NMA) of Egypt, as an Assistant Researcher. His research interests include potential field interpretation and applications, application of ground geophysical methods for mineral and groundwater exploration, and environmental and engineering applications.

NASSIR S. AL-ARIFI received the Ph.D. degree from The University of Manchester, U.K. He is currently the Director of the Visiting Professor Program and the Vice Rectorate for scientific research at King Saud University and a Full Professor specialized in earthquake seismology with the Geology and Geophysics Department, College of Science, King Saud University, Riyadh. He has published more than 45 articles in peer-reviewed journals and supervised more than 17 M.Sc. theses. His current research interest is in the field of geothermal exploration using multi method applied geophysics and has published more than 15 articles in this field. He translated two books in the field of hydrogeology and geophysics.