Math Geosci (2023) 55:59–80
https://doi.org/10.1007/s11004-022-10026-w

Advanced KNN Approaches for Explainable Seismic-Volcanic Signal Classification

Manuele Bicego[1] · Alberto Rossetto[1] · Matteo Olivieri[1] · John Makario Londoño-Bonilla[2,3] · Mauricio Orozco-Alzate[4]

Received: 4 May 2021 / Accepted: 5 September 2022 / Published online: 3 October 2022

Abstract Acquisition, classification, and analysis of seismic data are crucial tasks in volcano monitoring. The large number of seismic signals that are continuously acquired during the first monitoring stage poses a huge challenge for the human experts that must classify and analyze them. Several automatic classification systems have been proposed in the literature to alleviate such an overwhelming workload, each one characterized by different levels of accuracy, computational complexity, and interpretability. Considering this last perspective, which represents one of the recent key issues in geoscience, it is possible to find many accurate methods (in terms of classification accuracy) which however represent black boxes, not permitting a clear interpretation. On the other hand, there are other approaches, such as those based on support vector machines (SVM), random forests (RF), and K-nearest neighbor (KNN), which permit the interpretation of results, rules, and models at different levels. Among these last techniques, KNN approaches for volcanic signal classification typically do

✉ Manuele Bicego
manuele.bicego@univr.it

John Makario Londoño-Bonilla
jmakario@sgc.gov.co

Mauricio Orozco-Alzate
morozcoa@unal.edu.co

1 Dipartimento di Informatica, Università degli Studi di Verona, Strada le Grazie 15, 37134 Verona, Italy

2 Servicio Geológico Colombiano, Observatorio Vulcanológico y Sismológico de Manizales, Av. 12 de Octubre No. 15 - 47, Manizales 170001, Colombia

3 Facultad de Arquitectura e Ingeniería, Especialización en Prevención, reducción, y atención de desastres, Universidad Católica de Manizales, Cra. 23 No. 60-63, Manizales 170002, Colombia

4 Universidad Nacional de Colombia–Sede Manizales, Departamento de Informática y Computación, km 7 vía al Magdalena, Manizales 170003, Colombia

not achieve the satisfactory classification results obtained with RF and SVM. One possible reason is that in this context, the KNN rule has usually been applied in its basic version, not exploiting the different advanced KNN variants that have been introduced in recent years. This paper takes one step along this direction, investigating the suitability of a number of advanced versions of the KNN rule for the problem of classifying seismic-volcanic signals. The usefulness of these rules, in comparison with the original KNN rule as well as other interpretable classifiers, is evaluated within a real-world scenario involving a five-class dataset of seismic signals acquired at the Nevado del Ruiz volcano, Colombia. The results show that the classification accuracy of basic KNN is largely improved by these advanced variants, even surpassing that obtained with other classifiers like RF and SVM.

Keywords Advanced KNN rules · Automatic classification · Interpretability · Pattern recognition · Seismic-volcanic signals

1 Introduction

Over the past few decades, the geosciences have experienced a remarkable transition from a scarcity of data to a richness of data availability (Karpatne et al. 2018). Such a change has mainly been produced by the advances in acquisition tools—particularly in both sensor technology and telecommunication networks—which, in turn, have either taken advantage of or pushed forward the developments in data storage and computation. As a result, geoscientists are increasingly required to incorporate computer-based data analysis methods into their practices, which facilitate the processing of large amounts of data while also revealing patterns that would not likely be discovered by other means. Among the data analysis methods, those involving pattern recognition and machine learning (Bishop 2006) have been particularly useful for at least partially automating demanding tasks such as the classification of geophysical events that are typically registered as finely sampled data streams.

Volcano monitoring includes a par excellence case of the abovementioned demanding tasks, namely the classification of seismic-volcanic signals into a number of predefined categories. This task exhibits a streaming nature because data from seismometers continuously arrive via telemetry to the volcano observatories. It is also finely sampled because signals are generally acquired at a rate of hundreds of samples per second and using several strategically located triaxial sensors. The amount of data to be processed is, therefore, huge and always growing. Moreover, distinguishing among the different classes of seismic signals is not easy even for the most experienced analysts.

Several classification approaches have been used for seismic-volcanic signal discrimination, based on different representations and/or different classifiers—see the reviews in Orozco-Alzate et al. (2012), Malfante et al. (2018), and Carniel and Guzmán (2020). However, classification accuracy is not the only interesting criterion that can be used to judge a seismic classification system: actually, according to Karpatne et al. (2018), interpretability is a fundamental goal in geosciences. Interpretable methods are preferable, since they increase the confidence of the experts in automatic sys-

tems, allowing them to understand the reasons behind classifier outcomes and, in general, provide additional knowledge which may be of crucial importance for understanding seismic activity (Bicego et al. 2015). In Kostorz (2021), the importance of interpretable results in the geosciences is highlighted by saying that, when lacking explainable results—particularly regarding the errors—the classification models are expected to be accurate, rather than trusted. Again, Karpatne et al. 2018 suggest that transparent methods are preferred in most geoscience applications. Further, according to Talebi et al. (2020), interpretability belongs to the five minimum criteria which are required for machine learning methods applied to geosciences (the others are transparency, accuracy, credibility, and physical realism). All these arguments are well aligned with a recent trend which is growing in importance inside the artificial intelligence community and related fields, variously referred to as explainable artificial intelligence, interpretable or understandable machine learning, and so on (Adadi and Berrada 2018). Disregarding the particular adjective that is used, the common motivation in this counter-trend is giving importance not just to the assigned class label (the "what") but also to the reason for the assignment (the "why"). The former is sought to build accurate classifiers; the second enables an enhanced understanding of the phenomena as well as the algorithms themselves.

From this perspective, approaches to seismic-volcanic signal classification can be divided into two classes: approaches which do not have clear mechanisms of interpretability, such as those based on hidden Markov models (Bicego et al. 2013; Trujillo-Castrillón et al. 2018), Gaussian mixtures (Venegas et al. 2019), adaptive ensembles (Castro-Cabrera et al. 2021), and in particular artificial neural networks (Titos et al. 2018, 2019; Canário et al. 2020; Grijalva et al. 2021), and systems which permit interpretability. This latter class contains all those approaches based on support vector machines (SVM) (Lara-Cueva et al. 2016, 2017; Curilem et al. 2019; Peréz et al. 2020), decision trees/random forests (RF) (Lara-Cueva et al. 2016; Peréz et al. 2020), and K-nearest neighbor (Cárdenas-Peña et al. 2013; Castro-Cabrera et al. 2014; Orozco-Alzate et al. 2019b; Peréz et al. 2020), which permit, at different levels, the interpretation of results, rules, and models.

This paper is focused on the last method, namely, the nearest neighbor rule (NN) and its generalized version: K-nearest neighbor (KNN) (Cover and Hart 1967). This method implements an easy and human-understandable rule: in the NN rule, a test signal is assigned to the class of the signal in the training set which is most similar to it. More generally, the KNN assigns an object to the most frequent class among the K objects of the training set that are nearest to the test object (that is, $K = 1$ in the NN rule). This classifier, applied in many contexts due to its simplicity and flexibility, has also been widely studied from a theoretical perspective, for example, in asymptotic behaviors (Fukanaga 1990); further, it has been extended in different ways, such as by modifying the set of training objects (Triguero et al. 2012), dealing with high-dimensional spaces (Pal et al. 2016), weighting the different objects of the training set (Bicego and Loog 2016), or modifying the dissimilarities (Wang et al. 2007; Lopes and Ribeiro 2015; Orozco-Alzate et al. 2019a; Bicego and Orozco-Alzate 2020).

It is important to observe that in the specific context of seismic-volcanic signal classification, KNN approaches typically do not reach the satisfactory classification results obtained with RF and SVM. One possible reason is that in this field, the KNN

rule has usually been applied in its basic version (Castro-Cabrera et al. 2014; Peréz et al. 2020; Orozco-Alzate et al. 2019b); however, it is reasonable to expect improvements using more advanced versions, which have never been investigated in this field. This respresents the main goal of this manuscript. In particular, this paper addresses the question of whether some advanced versions of the KNN rule are suitable for the classification of seismic signals, permitting one to reach or exceed the accuracy of other interpretable techniques such as SVM or RF.

Five variants are investigated: two well-established rules, the adaptive nearest neighbor (ANN) (Wang et al. 2007) and the hypersphere classifier (HC) (Lopes and Ribeiro 2015); their two very recent variants, the ANN-HC interpolation (Orozco-Alzate et al. 2019a) and the PowerHC rule (Bicego and Orozco-Alzate 2020); and a new variant which is introduced here in this paper. The empirical evaluation is based on a dataset containing more than 1,000 volcanic earthquakes, collected at Nevado del Ruiz volcano in Colombia and encompassing five classes of volcanic activity: long-period (LP) events, screw-like (TO) earthquakes, volcano tectonic (VT) events, volcanic tremors (TR), and hybrid (HB) seismic events (Chouet and Matoza 2013). We will show that the classification accuracies of KNN are largely improved by its advanced variants, becoming better than those obtained with other classifiers such as RF and SVM.

The rest of the paper is organized as follows. The steps for a seismic signal classification system are explained in Sect. 2. Afterwards, the investigated variants of KNN are presented in Sect. 3. In Sect. 4, the experiments are shown. Finally, concluding remarks are given in Sect. 5.

2 The Basic Pipeline for Seismic Signal Classification

This section introduces the steps of the classification system used in the study, which is based on the standard KNN approach. The advanced variants investigated to improve the performances of the basic version are provided in the next section.

2.1 Representation and Dissimilarities

Continuous recordings acquired by the volcano observatories follow the nature of multichannel time series. Each channel may correspond to either a recording station or an acquisition axis from a single sensor. From the raw seismic recordings, digital signals are extracted by detecting the beginning and end of individual seismic events—for the precise implementation details, see the experimental section. As commonly done in this field (Castro-Cabrera et al. 2014; Orozco-Alzate et al. 2015), signals were represented using a windowed fast Fourier transform (FFT); that is, an FFT-based spectrogram. From the spectrograms, two distances have been derived: (i) the first is the Euclidean distance computed between averaged spectrograms (that is, the average of all FFT frames along time); (ii) the second works directly on the spectrograms and is based on the dynamic time warping (DTW) distance (Wang et al. 2013). The DTW distance is an elastic dissimilarity measure, suitable for comparing seismic signals since it is able to deal with different lengths; moreover, among the collection of available elastic measures, the DTW distance is still considered the par excellence

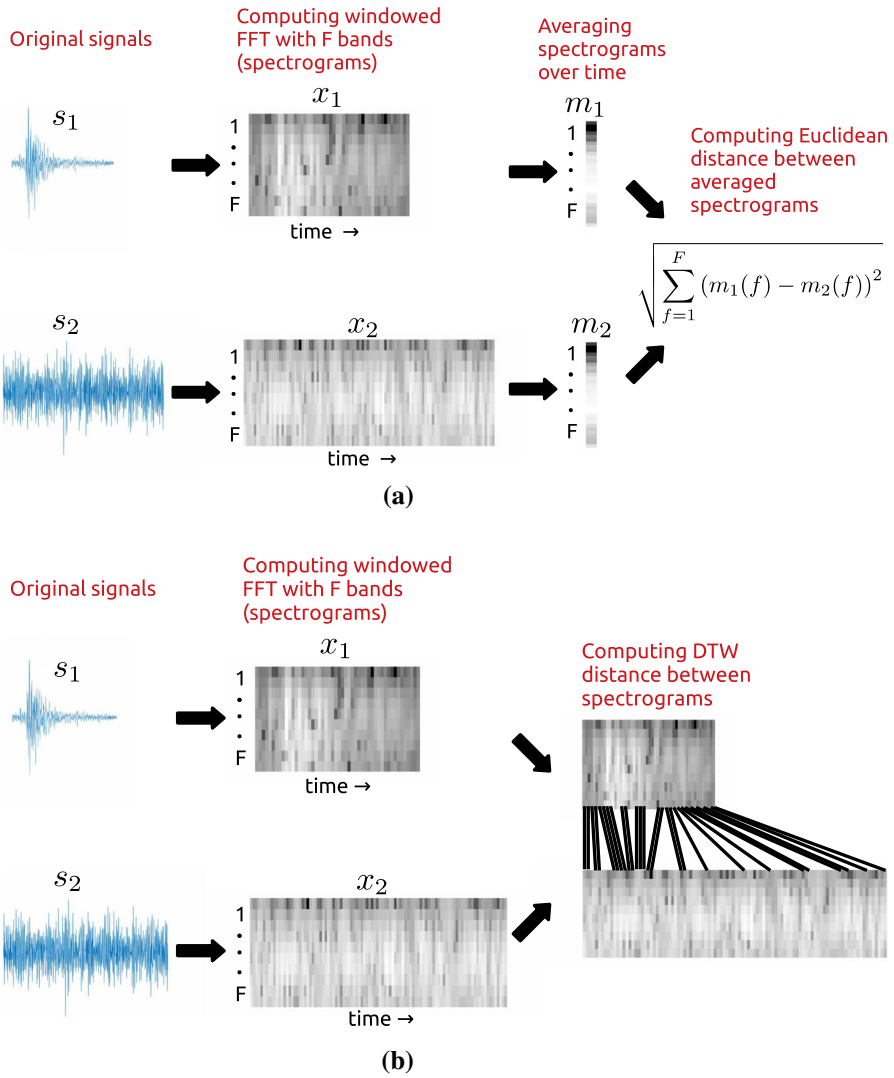

Fig. 1 Visual explanation of the proposed scheme for computing distances between volcanic signals: (top) with Euclidean distance, (bottom) with DTW distance

option to compare time series (Wang et al. 2013). The following subsection briefly summarizes this distance and its extension to spectrograms.

A visual explanation of the scheme used to compute the distance is reported in Fig. 1 for the Euclidean (part (a)) and for the DTW (part (b)) distances. In the former case, the signal s_1 is encoded with a windowed FFT (with F bands), then the t_1 frames are averaged to get the averaged spectrogram m_1; in the same way the averaged spectrogram m_2 of the second signal s_2 is obtained by averaging the t_2 frames of the windowed FFT of s_2; finally, the distance is computed using the Euclidean formula

between m_1 and m_2. In the DTW case, there is no average, and the distance is directly computed on the spectrograms, x_1 and x_2. Notice that even though the signals may have different durations, their impact on the computation of the averaged spectrogram reduces to the average values estimated with moderately different sample sizes; therefore, this effect can be safely neglected if no extremely short-duration events are included in the dataset. Similarly, even though an accurate segmentation is not always guaranteed, the coda signals are low-energy portions that do not contribute much to modify the average unless they are extremely long.

2.1.1 The DTW Measure

Consider two seismic signals $A = \{a_1, \ldots, a_n\}$ and $B = \{b_1, \ldots, b_m\}$ to be compared. The DTW measure computes the minimum cost to align A and B. In the first step, a matrix $\mathbf{D} \in \mathbb{R}^{n \times m}$ is built, where $D_{i,j} = d(a_i, b_j)$. Typically, $d(\cdot, \cdot)$ is chosen as $(a_i - b_j)^2$. The matrix $\mathbf{D}$ is then explored from the upper-left corner to the bottom-right one, by looking for a path that minimizes its cost: the sum of the traversed entries. At the end, the DTW distance between A and B corresponds to the square root of the path with minimum cost. This optimization problem is typically solved by dynamic programming[1] and only exploring a corridor around the pseudo-diagonal of the matrix; see Algorithm 1 for further implementation details, which is based on the DTW pseudo-code presented by Lin et al. (2012). The basic version of the DTW distance can also be extended for the comparison of spectrograms by just computing the entries of the matrix as the norm between spectrogram slices, that is: $||\mathbf{a}_i - \mathbf{b}_j||$. In this case, both $\mathbf{a}_i$ and $\mathbf{b}_j$ have as many entries as frequency bands used for computing the Fourier transform.

Algorithm 1 Computation by dynamic programming of the DTW distance between two signals $A = \{a_1, \ldots, a_n\}$ and $B = \{b_1, \ldots, b_m\}$

$D_{i,j} \leftarrow 0, \quad \forall i, j$ ▷ Initialize $\mathbf{D} \in \mathbb{R}^{n \times m}$
for $i = 1, \ldots, n$ **do**
  for $j = 1, \ldots, m$ **do**
    $D_{i,j} \leftarrow d(a_i, b_j) + \min\left(D_{i-1,j}, D_{i,j-1}, D_{i-1,j-1}\right)$
  end for
end for
return $D_{n,m}$ ▷ Dynamic programming leaves the minimum cost of the alignment in the bottom-right corner of $\mathbf{D}$

2.2 Classification

Consider the following notation given a distance measure $d(\cdot, \cdot)$ to be used by the KNN classifier: (i) $\mathbf{x}$ denotes the test object; (ii) $\{\mathbf{x}_i\}$, (with $1 \leq i \leq N$), is the

[1] The code is available at https://www.mathworks.com/matlabcentral/fileexchange/43156-dynamic-time-warping-dtw.

training set which contains N labelled patterns; (iii) let y_i denote the label of the i-th object ($1 \leq i \leq N$) such that, given a problem with C classes, y_i takes values in the set $\{1, \ldots, C\}$; and (iv) let $neigh_K(\mathbf{x}) = \{\mathbf{n}_1, \ldots, \mathbf{n}_K\}$ be a set that contains the K training objects which, according to $d(\cdot, \cdot)$, are at the minimum distance from $\mathbf{x}$ and whose corresponding labels are $\{y_{\mathbf{n}_1}, \ldots, y_{\mathbf{n}_K}\}$. According to the above-described notation, the original KNN approach classifies $\mathbf{x}$ as belonging to the class $\hat{c}$ which is more frequent in the set of $neigh_K(\mathbf{x})$; that is,

$$\mathbf{x} \longleftarrow \arg\max_c \left| \{\mathbf{n}_i : y_{\mathbf{n}_i} = c\} \right|, \tag{1}$$

where $|\cdot|$ indicates the cardinality of a set. When $K = 1$, this turns out to be the NN rule, which assigns the test object $\mathbf{x}$ to the class of its nearest neighbor

$$\mathbf{x} \longleftarrow y_{\mathbf{n}_1}, \tag{2}$$

where $\mathbf{n}_1$ is the object j, such that $j = \arg\min_{1 \leq i \leq N} d(\mathbf{x}, \mathbf{x}_i)$.

3 Advanced KNN Classification

Five different variants of the classic KNN rule are investigated, all based on the exploitation of the concept of the radius of a given training point $\mathbf{x}_i$, which represents the radius of the largest hypersphere having as its center $\mathbf{x}_i$ and not containing any training object coming from a different class. The value of r_i, which measures the dissimilarity from the nearest training object of $\mathbf{x}_i$ coming from a different class, is defined as

$$r_i = \min_{\mathbf{x}_j \in OT(\mathbf{x}_i)} d(\mathbf{x}_i, \mathbf{x}_j), \tag{3}$$

with

$$\mathrm{OT}(\mathbf{x}_i) = \{\mathbf{x}_k \text{ such that } y_k \neq y_i\}. \tag{4}$$

The idea of all the investigated variants is to use the radii to correct the distance between the test point $\mathbf{x}$ and all training points $\mathbf{x}_i$, such that points with large radii (that is, points well inside their class) become closer to the test object. Once given the correction, the final classification is performed using the standard KNN technique. The idea is that points well inside their class—which reasonably represent good exemplars of that class—should be more important for the classification of $\mathbf{x}$.

The correction can be implemented in different ways, as explained in the following:

1. ANN: the adaptive nearest neighbor rule (Wang et al. 2007). This was the first introduced rule of this family of variants, which performs the correction by dividing the distance by the radius

$$d_{\mathrm{ANN}}(\mathbf{x}, \mathbf{x}_i) = \frac{d(\mathbf{x}, \mathbf{x}_i)}{r_i}. \tag{5}$$

2. HC: the hypersphere classifier rule (Lopes and Ribeiro 2015). In this rule, the idea is also to reduce the distance from points with large radius; however, here the reduction is performed with a subtraction

$$d_{\mathrm{HC}}(\mathbf{x}, \mathbf{x}_i) = d(\mathbf{x}, \mathbf{x}_i) - r_i. \tag{6}$$

Note that Eq. (6) may return negative distances if the training object $\mathbf{x}_i$ is nearer to the test object $\mathbf{x}$ than to a training object of another class. However, this does not represent a problem when using the KNN rule, since this technique is based on ranking.

3. Interp: This variant starts from the observation that the behavior of the corrections in Eqs. (5) and (6) is rather different, since the penalization of the distance is much stronger in the first rule (ratio versus difference). Moreover, the correction in Eq. (5), in contrast to the one in Eq. (6), does not generate negative values but may diverge if $r_i \to 0$, leading to numerical inaccuracies. To keep the best variant, very recently a combination of the two was proposed in Orozco-Alzate et al. (2019a). Different rules were proposed to induce a smooth transition between them; here, the following one is used, which turned out to be the best in the above-cited paper by Orozco-Alzate et al. (2019a)

$$d_{\mathrm{Int}}(\mathbf{x}, \mathbf{x}_i) = \frac{(1-\lambda)d(\mathbf{x}, \mathbf{x}_i)}{(r_i + \sqrt{\lambda})} + \lambda(d(\mathbf{x}, \mathbf{x}_i) - r_i), \tag{7}$$

where the parameter λ measures the importance of each variant (the HC rule corresponds to $\lambda = 1$, whereas the ANN rule is obtained with $\lambda = 0$).

4. NormInterp: This represents a variant of the Interp rule introduced in this paper, in which each component of the convex combination is normalized before combining them.

$$d_{\mathrm{NInt}}(\mathbf{x}, \mathbf{x}_i) = \frac{(1-\lambda)d(\mathbf{x}, \mathbf{x}_i)}{z_1 r_i} + \frac{\lambda(d(\mathbf{x}, \mathbf{x}_i) - r_i)}{z_2}, \tag{8}$$

where

$$z_1 = \max_{ij} \frac{d(\mathbf{x}_j, \mathbf{x}_i)}{r_i} \quad z_2 = \max_{ij}(d(\mathbf{x}_j, \mathbf{x}_i) - r_i).$$

The main idea of this approach is that the HC- and ANN-corrected distances may be very different in terms of magnitude so that one of the two can dominate the integration; performing a normalization before interpolating may reduce this effect.

5. PowerHC: This represents a very recent variant of the HC rule introduced in Bicego and Orozco-Alzate (2020). The starting point, as shown in Orozco-Alzate et al. (2019a), is that the ANN rule can be seen as equivalent to the application of the HC rule to the logarithm of the original distances; that is, to the distances which are nonlinearly scaled. Nonlinear scaling of feature spaces and distances was recently shown to represent a powerful way to enhance discriminability (Duin et al. 2014; Bicego and Baldo 2016; Orozco-Alzate et al. 2016). In more detail, in the PowerHC

rule the distances are normalized with a power transformation before applying the HC rule

$$d_{\text{Pow}}(\mathbf{x}, \mathbf{x}_i) = d(\mathbf{x}, \mathbf{x}_i)^{\rho} - r_i^{\rho}, \tag{9}$$

ρ is a parameter that must be set in advance, which drives the strength of the nonlinear correction.

3.1 Properties

All these variants are based on the idea that it may be beneficial to correct the distance of the test object from a given training object by considering the so-called quality of such training object. The quality is measured by its radius, that is, by its distance from the nearest point of another class.

Even if these corrections have been shown to be beneficial in many cases (Wang et al. 2007; Lopes and Ribeiro 2015; Orozco-Alzate et al. 2019a; Bicego and Orozco-Alzate 2020), there are still some situations in which they may not lead to an improvement (Orozco-Alzate et al. 2019a). A possible explanation is that these approaches exploit in a deeper way the information contained in the dissimilarity measure; when this measure is not appropriate for the context at hand, the result can be wrong. In this sense, when using these KNN variants, it is even more fundamental than in the basic case to choose a reasonable distance measure: this will be evident also in the experimental session, where the improvements obtained with the DTW distance are definitely larger than those obtained with the Euclidean one.

Concerning the different variants, it is important to observe that for the first two methods (ANN and HC), there are no parameters to set; this represents an important feature of ANN and HC. The remaining ones have one tunable parameter: λ for Interp and NormInterp, and ρ for PowerHC. In general, setting these parameters may be critical, and may lead to different results. However, some suggestions on how to set them can be found in Orozco-Alzate et al. (2019a) and Bicego and Orozco-Alzate (2020); in particular, according to Orozco-Alzate et al. (2019a), good values for λ can be found in the range of 0.1 to 0.3; concerning ρ, in Bicego and Orozco-Alzate (2020), the authors propose to set it via cross validation, showing that the obtained value is very close to the optimal one.

A final comment on the computational overhead introduced by the advanced variants: it is important to observe that the variants introduce a correction of the distances; after the correction, the classifier works as the classic KNN. All the corrections are based on fast mathematical operations, and the only overhead is the computation of the radius of each training point, which requires the computation of the nearest point of a different class. However, this computation can be done only once (offline) and stored for all the objects in the training set.

4 Experiments and Discussion

In this section, the experimental evaluation is presented. First, the dataset and the details of the experiments are introduced; subsequently, a visual intuition of the effects of the

correction of the distance is provided, followed by results and discussion concerning the different variants. Finally, a comparison of the proposed scheme with alternative state-of-the-art interpretable methods is reported.

4.1 Experimental Details

The methods were tested on a set of signals collected by the Volcanological and Seismological Observatory of Manizales (OVSM by its acronym in Spanish) from Servicio Geológico Colombiano (SGC). In particular, the signals used in this study are related to Nevado del Ruiz volcano, the most active volcano in the northern volcanic segment of Colombia; in the last 30 years, this volcano has erupted several times. The signals used in this paper were gathered during the period from January 2010 to September 2013 from the BIS station. They belong to five classes, representing the most common seismic events related to volcanoes: volcano tectonic (VT) events, long-period (LP) events, volcanic tremors (TR), hybrid (HB) events and screw-like (TO) earthquakes. In particular, there are 153 VT events, 333 LP events, 224 TR events, 251 HB events, and 104 TO events, for a total of 1,065 signals. According to Trombley (2006), VT events have an abrupt start, exhibit a relatively high dominant frequency (e.g., 5 Hz), and are associated with rock fracture; LP events—in contrast to VT events—are lower in frequency (e.g., 3 Hz) and are associated with magma or gas flow; TR events can be understood as long and sustained LP quakes, and HB events correspond to a mixture or superposition of VT and LP quakes. The abovementioned frequency ranges for VT and LP events are typical but may vary from volcano to volcano, since they depend on particular properties such as the type of rock and the length or shape of the volcanic conduits. TO events also exhibit low frequencies but are mainly characterized by a quasi-monochromatic spectrum and low exponential amplitude decay, resembling thereby a screw shape. Representative signals per class, along with their corresponding spectrograms, are shown in Fig. 2.

Seismic sensors deliver three components (vertical, north-south, east-west): as commonly done in the seismic community, only the vertical signal is considered.[2] The signals were acquired at 100 Hz; before being segmented, they were quantized with a 16-bit analog-to-digital converter; therefore, their amplitudes may cover the range $\pm 32{,}768$. The segmentation was performed either automatically, with the so-called STA/LTA algorithm, or manually, depending on the events. As done in (Castro-Cabrera et al. 2014; Orozco-Alzate et al. 2015), signals are represented with spectrograms with a 128-point FFT (overlap 50%), smoothing each frame with a 64-point Hamming window.

In order to also have a vectorial representation, for each signal, the averaged spectrogram was computed. Two different distances were considered: the Euclidean distance between the averaged spectrograms, and the DTW distance between the spectrograms. The distances were normalized so that the maximum distance is 1: normalization has no effect on the NN rule (since this rule is based on the ranking, which remains

[2] However, it was very recently shown that the integration of the three components can be very useful (Orozco-Alzate et al. 2019b).

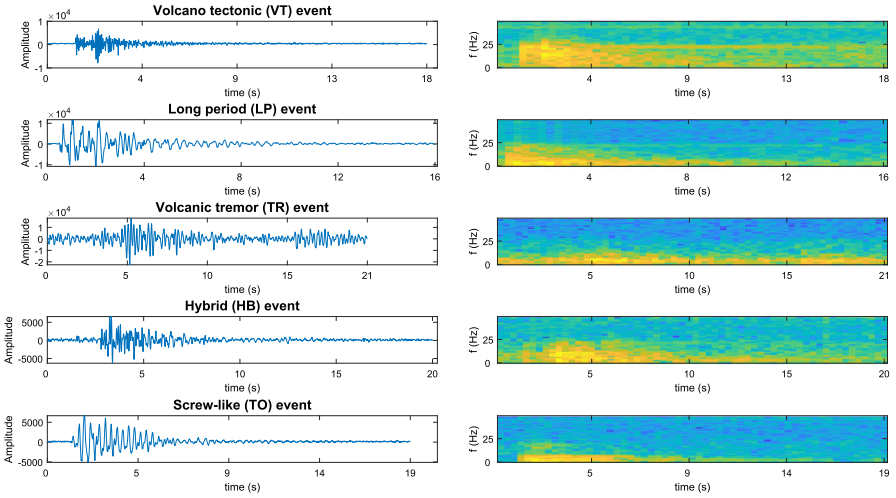

Fig. 2 Sample signals, one per class of seismic event, from the set of signals collected at Nevado del Ruiz volcano, Colombia. Their associated spectrograms are shown on the right side

unaltered when scaling the distance), but can have important consequences on its variants, as discussed in Bicego and Orozco-Alzate (2020). The normalized distances represent the input for the KNN technique and all its variants. Concerning Interp and NormInterp, the parameter λ was varied between 0 and 1 in steps of 0.05, choosing the value that minimized the classification error. In the PowerHC case, as done in Bicego and Orozco-Alzate (2020), the parameter ρ was varied between 0.2 and 10 with step 0.2, choosing again the value leading to the minimum error. Classification errors were computed using the classic Leave-One-Out (LOO) cross-validation procedure (Bramer 2016): within this protocol, the system is trained with all the signals except one, which is then used for testing. The procedure is iteratively repeated by leaving out at every iteration a different object, until all signals have been tested.

4.2 A Visual Explanation of the Methods

This section provides some evidence of the mechanisms behind the NN variants. The starting point is observing that all the considered variants are based on the concept of correcting the distance between the test object and a given training object using the radius: the goal is to push points that are not well inside their class (that is, those with a low radius) far away from the test point.

To show this mechanism, a test object is selected together with another 20 points: the first 10 are the 10 nearest neighbors according to the original distance, whereas the second 10 are the 10 nearest neighbors according to the corrected distance (in this example the ANN version was used with the Euclidean distance computed on the averaged spectrograms). These 21 points (20 neighbors plus the test point) were then projected on a two-dimensional space using multidimensional scaling (Cox and Cox 1994), visualizing them in the resulting vectorial space. The obtained plots are shown

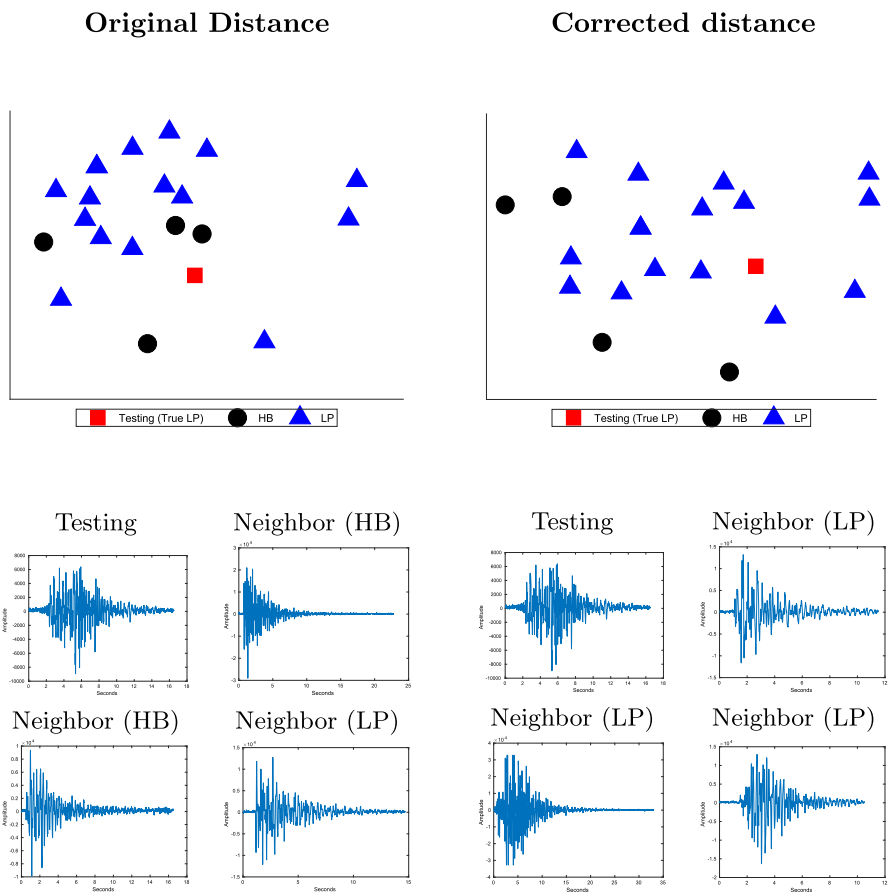

Fig. 3 Visual interpretation of the effect of the distance correction (ANN variant). Top plots: MDS spaces, Bottom plots: test signal plus three neighbors

in Fig. 3 (top part), for the original distance (left) and for the corrected one (right). The bottom part also displays the test signal together with its three nearest neighbors, according to the original distance (left) and to the corrected one (right). The top part of the image shows that the test object (square), which is from the LP class, is incorrectly assigned to the HB class by the KNN rule for $K = 1, 2$ or 3. Actually, two of the nearest training signals (displayed in the bottom plot) belong to the HB class (circles in the plot). However, these signals are not well inside the HB class, thus they have a low radius. When correcting the distance with the ANN rule, such signals are pushed away from the test objects (together with the other points with low radius), whereas points well inside their class (triangles in this case) are made closer. After this correction, all the closest neighbors (displayed in the bottom-right plot) are from the LP class.

Table 1 LOO errors for the original and advanced NN rules, using both the Euclidean and the DTW distances

Method	Euclidean	DTW
Orig	0.2385	0.2582
ANN	0.1399	0.1484
HC	0.1408	0.1362
Interp	0.1399	0.1286
NormInterp	0.1390	0.1362
PowerHC	0.1371	0.1211

4.3 Analysis of the Different Variants

In this section, the classification accuracies obtained with the different variants are compared to the original version of the KNN classifier. Results and comments are reported for both the NN rule ($K = 1$) and the general KNN.

4.3.1 Nearest Neighbor Results

As a first analysis, Table 1 shows the LOO errors obtained using NN with both the Euclidean and the DTW distances. For a deeper insight, the corresponding confusion matrices are also reported in Table 2. For every confusion matrix, the true labels are on the rows, whereas the assigned labels are on the columns. The different classes are represented in this order: 1 HB, 2 LP, 3 TO, 4 TR, and 5 VT. NN was used with the original distance (Orig) and with the five variants (ANN, HC, Interp, NormInterp, PowerHC).

Different observations can be derived from Table 1. First, notice that all the variants drastically improve the original NN, in most of the cases with an improvement which is larger than 0.1. This confirms the intuition behind this paper: variants of NN can drastically improve the performance of NN, driving this interpretable tool to the levels of alternative high complex interpretable classifiers such as SVM and RF (see the next section for a deeper analysis). It is important to note that the improvements obtained in this scenario are drastically larger than those obtained in other classification scenarios, such as those presented in Wang et al. (2007), Lopes and Ribeiro (2015), and Orozco-Alzate et al. (2019a), suggesting that their exploitation can be very beneficial in the specific field of classification of volcanic seismic events.

Second, the behavior with the two distances is different: with the Euclidean distance, the error with the original NN is lower than with the DTW distance. However, when using the variants, the improvement is larger with DTW, reaching in the best case (the PowerHC variant) a remarkable difference of 0.137 (from 0.2582 to 0.1211). This confirms the intuitions provided at the end of Sect. 3: the variants exploit in a deeper way the information contained in the distance, and more sophisticated measures (such as the DTW) may permit one to reach better results.

Third, concerning the variants, there is not a great difference between them when using the Euclidean distance: ANN, HC, and their combinations/improvements all remain on the same level, with a slight decrease in the error when going to more

Table 2 Confusion matrices of the different methods with the Nearest Neighbor rule

Method	Euclidean					DTW				
Orig	154	91	5	8	9	142	68	15	34	13
	69	220	13	3	0	102	252	9	6	3
	6	7	82	0	0	2	12	78	2	0
	14	15	4	212	1	0	0	0	181	0
	8	0	0	1	143	5	1	2	1	137
ANN	183	46	7	7	2	164	26	9	0	2
	46	276	6	0	0	59	280	2	1	0
	6	2	90	1	0	6	13	90	0	1
	8	8	1	216	0	12	11	3	223	0
	8	1	0	0	151	10	3	0	0	150
HC	183	46	7	7	2	169	28	9	0	3
	47	276	7	0	0	60	290	3	2	0
	5	2	89	1	0	7	6	90	1	0
	8	8	1	216	0	8	7	2	221	0
	8	1	0	0	151	7	2	0	0	150
Interp	183	46	7	7	2	177	36	8	0	2
	46	276	6	0	0	55	288	4	1	1
	6	2	90	1	0	8	6	92	1	1
	8	8	1	216	0	0	3	0	222	0
	8	1	0	0	151	11	0	0	0	149
NormInterp	184	46	7	7	2	168	28	9	0	3
	45	276	6	0	0	61	290	3	2	0
	6	2	90	1	0	6	6	90	0	0
	8	8	1	216	0	8	7	2	222	0
	8	1	0	0	151	8	2	0	0	150
PowerHC	185	44	7	7	1	180	27	9	3	3
	46	278	7	0	0	54	297	3	1	0
	5	2	89	1	1	6	5	91	2	0
	8	8	1	216	0	4	4	1	218	0
	7	1	0	0	151	7	0	0	0	150

For each confusion matrix: true labels are on rows, whereas assigned labels are in the columns, for the different classes. The classes are: 1: HB, 2: LP, 3: TO, 4: TR, 5: VT

complex variants (such as NormInterp and PowerHC). When using DTW, on the contrary, there is a clearer trend: the decrease in the error when using complex variants is more relevant.

A final comment can be derived from the confusion matrices reported in Table 2: notice that most of the errors are between the classes LP (long period) and HB (hybrid events), this being in line with findings in the literature (Trombley 2006). These errors are less evident when using the advanced KNN variants. On the contrary, the easiest pair of classes to be discriminated are the last two, namely, TR (volcanic tremors) and

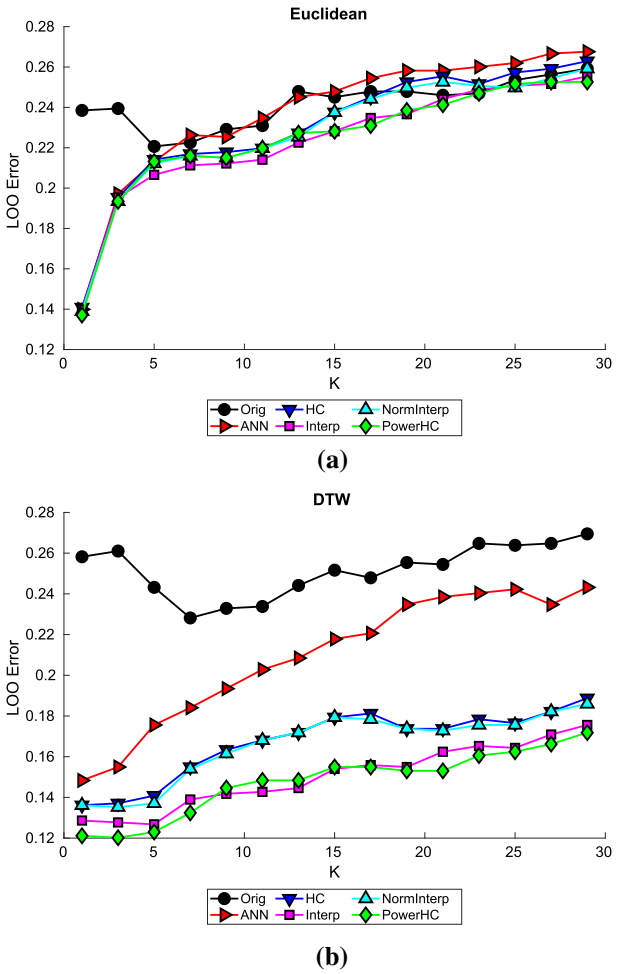

Fig. 4 LOO errors when varying the parameter K of the KNN for: **a** Euclidean distance and **b** DTW distance

VT (volcano tectonic). In this case, very few wrong assignments are found even for the basic NN rule.

4.3.2 KNN Results

To gain a deeper understanding of the analyzed methods, the previous analysis was repeated while using KNN, varying K from 1 to 30 (step 2). LOO errors are shown in Fig. 4a, b for the Euclidean distance and the DTW distance, respectively.

In the Euclidean case, notice that the errors of the variants drastically increase when increasing K, becoming approximately on the same level of the error with the original distance. The behavior with the DTW distance is more interesting; also, in this case, there is a worsening of the performances with larger values of K, but the difference

Table 3 Results with the KNN rule, for **a** the Euclidean distance and **b** the DTW distance

Method	Euclidean	
	Average	Best
(a)		
Orig	0.2421 (3.60e−04)	0.2207 (3)
ANN	0.2371 (1.04e−03)	0.1399 (1)
HC	0.2302 (9.84e−04)	0.1408 (1)
Interp	0.2235 (9.11e−04)	0.1399 (1)
NormInterp	0.2279 (9.64e−04)	0.1390 (1)
PowerHC	0.2243 (9.08e−04)	0.1371 (1)
Method	DTW	
	Average	Best
(b)		
Orig	0.2516 (3.95e−04)	0.2282 (4)
ANN	0.2093 (9.79e−04)	0.1484 (1)
HC	0.1671 (5.23e−04)	0.1362 (1)
Interp	0.1503 (4.86e−04)	0.1268 (3)
NormInterp	0.1658 (5.28e−04)	0.1352 (2)
PowerHC	0.1477 (5.04e−04)	0.1202 (2)

Reported results are the average among all K (first column, between brackets the standard error of the mean) and with the best K (second column, the best K is between brackets)

with the original version remains significant. In this second case, advanced rules such those introduced in Orozco-Alzate et al. (2019a) and Bicego and Orozco-Alzate (2020) represent the best choice, significantly outperforming the original rule as well as the less recent variants such as ANN and HC. To confirm numerically, the best and the average errors among the different values of K of the different techniques are reported in Table 3. It can be seen that, on average, variants outperform the original rule, with a remarkable improvement for the DTW distance. Note that in this case, due to a lack of space, the confusion matrices are not reported; however, those matrices showed trends similar to those shown in Table 2.

The best number of neighbors is a trade-off between the local and global considerations for the decision. A too large number of neighbors turns KNN into the nearest mean classifier (thereby, a global decision is made) while, in contrast, the one-nearest neighbor rule (that is, KNN with $K = 1$) assigns the label by using a local criterion. Even though the best number of neighbors is highly dependent on the nature of the dataset, we may say that in complex real-world problems, it is generally expected that taking decisions based on local information is preferred over doing so by taking into account global behaviors. Moreover, the fact of confirming that the best value for K is a small one reveals that classes in the representation space tend to be non-compact or sparse, either due to the intrinsic complexity of the problem itself or because of a small sample size.

To enable a deeper comparison between the standard KNN approach and the advanced variants, a further test is performed. In particular, the test is aimed at comparing the "averaged confidence" of the decisions in the different cases. The confidence

represents how confident the classifier is in assigning a label; for example, in a binary classification problem, with $K = 10$, if all the first 10 neighbors are from the same class, then the classifier is very sure about the decision, whereas if only six of them are from the same class and the remaining four are from the other, the classifier is less confident in the classification. To quantify this, the "confidence" $c(\mathbf{x})$ of KNN on the decision for an object $\mathbf{x}$ is defined here as the number of neighbors, among the K nearest ones, with the label equal to the assigned one

$$c(\mathbf{x}) = \frac{|\{\mathbf{n}_i : y_{\mathbf{n}_i} = \text{KNN}(\mathbf{x})\}|}{K}, \tag{10}$$

where $\text{KNN}(\mathbf{x})$ is the label assigned by KNN to $\mathbf{x}$ using the rule defined in Eq. (1). In the above example, the decision on the first object has a confidence of 1 (10/10), whereas the one on the second has a confidence of 0.6 (6/10). Given this definition, it is possible to compute the averaged confidence $\tilde{c}(\{\mathbf{x}_i\})$ of a set of objects $\{\mathbf{x}_i\}$ as the average of the confidence $c(\mathbf{x}_i)$ on the decision of every single object in the set $\{\mathbf{x}_i\}$. Figure 5 reports the averaged confidences obtained in the experiments, varying K, for the basic KNN and for the variants. In particular, plot (a) reports the behavior for the Euclidean distance, whereas part (b) is devoted to the DTW distance. The plots do not report the value for $K = 1$ since, in that case, the confidence is 1 for all methods by definition. From the plots it is evident that the advanced variants have larger confidence values than the original KNN, this being valid for both distances. By correcting the distance, the classifier is more sure about the classification. Further, it can be observed that all variants perform similarly with the Euclidean distance, whereas for the DTW, Interp and PowerHC perform better. This confirms the observations made for the accuracy plots in Fig. 4.

4.4 Comparison with Alternative Interpretable Approaches

In this section, the proposed schemes are compared with some alternative approaches, based on interpretable techniques appearing in the seismic literature. In particular, following the same evaluation protocol described in Sect. 4.1, LOO errors and the corresponding confusion matrices were computed for a set of recent classification approaches based on KNN (Cárdenas-Peña et al. 2013; Castro-Cabrera et al. 2014; Peréz et al. 2020; Orozco-Alzate et al. 2019b), SVM (Lara-Cueva et al. 2016, 2017; Curilem et al. 2019; Peréz et al. 2020), and RF (Lara-Cueva et al. 2016; Peréz et al. 2020). For every approach, the setting indicated in the most recent paper which provided an empirical evaluation on it was used. All the classifiers were applied to the averaged spectrograms; as is often done for kernel machines, a classic z-score standardization was performed before applying SVM classifiers. Specifically, for KNN, the version introduced in the very recent study by Peréz et al. (2020) was adopted, which uses the Euclidean distance and finds the optimal value of K in the range [1,20] via a 10-fold cross validation on the training set. Then, three versions of SVM were used: with linear (SVMlin), polynomial (SVMpoly), and radial basis function (SVM-rbf) kernels. Again following the suggestions in Peréz et al. (2020), the parameter

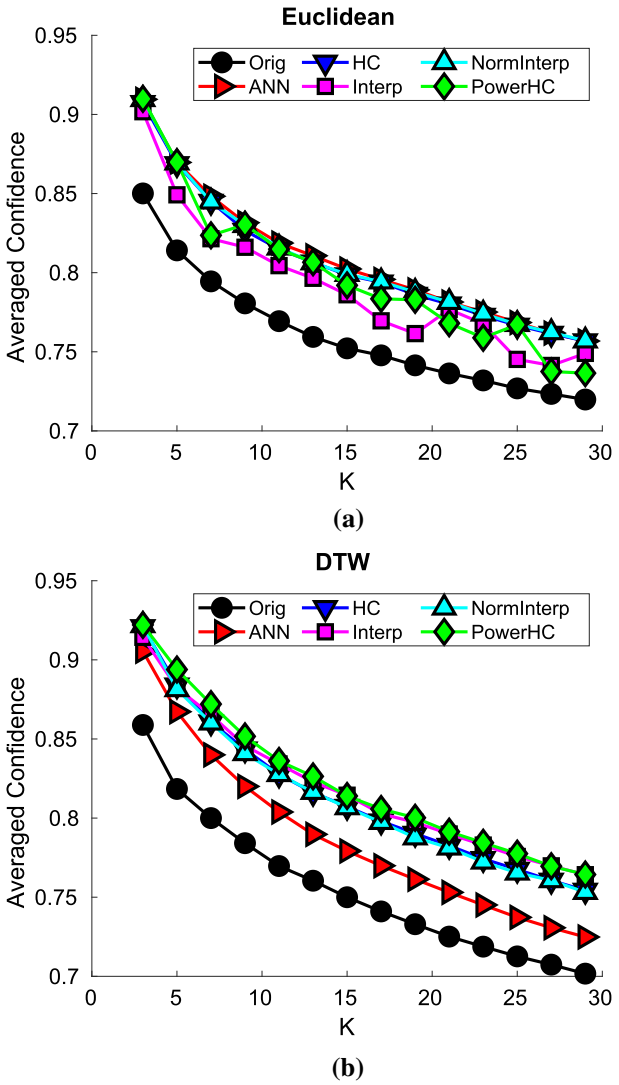

Fig. 5 Averaged confidence in the decision when varying the parameter K of the KNN for **a** Euclidean distance and **b** DTW distance

C (the cost) of the SVM was chosen via a 10-fold cross validation method on the training set, within a range from 10^{-4} to 10^{4} (increasing by a factor of 10). The other parameters of the kernels were set as the default values in WEKA software (again following Peréz et al. 2020). Finally, the RF classifier was also employed. In this case, again following Peréz et al. (2020), the optimal number of trees was found using a 10-fold cross validation on the training set, selecting the number in the range of 100 to 1,000 (with step 10).

The obtained LOO errors are shown in Table 4.

Table 4 Comparison with alternative interpretable methods used in the volcano-seismic community

Method	References	LOO error
KNN	Cárdenas-Peña et al. (2013), Castro-Cabrera et al. (2014) and Peréz et al. (2020)	0.2197
SVMlin	Lara-Cueva et al. (2016) and Peréz et al. (2020)	0.2573
SVMpoly	Peréz et al. (2020)	0.1831
SVMrbf	Lara-Cueva et al. (2016), Curilem et al. (2019) and Peréz et al. (2020)	0.1765
RF	Lara-Cueva et al. (2016) and Peréz et al. (2020)	0.1962
PowerHC with DTW	This paper, cf. Table 3(b)	0.1202 (for $K = 2$)

The overall best result obtained with the advanced KNN approaches is included in the last row for the sake of an easy comparison

To have a deeper analysis on the results, the corresponding confusion matrices are also reported in Table 5.

Different observations can be derived from such tables. Firstly, these results provide a further confirmation that RF and SVM (especially in the more flexible and sophisticated version with rbf and polynomial kernels) are typically better than standard KNN approaches in literature. Looking at results reported in Tables 1 and 3, it can also be observed that changing the distance (e.g., DTW) seems to not be sufficient to reach the accuracies of SVM and RF. However, when considering the advanced KNN versions, a clear improvement can be observed. Actually, all the proposed variants show a LOO error that is lower than those of SVMs or RF, with an improvement of 5% in the best case (PowerHC with the DTW distance). Remarkably, a significant improvement over the state-of-the-art is present even with the simple (and parameterless) advanced KNN variant; that is, the ANN (or HC) with the NN rule.

5 Conclusion

This paper addressed the exploitation of advanced variants of the KNN rule in the problem of seismic-volcanic signal classification. In particular, a number of state-of-the-art KNN variants, based on correction of the distances, were investigated. The KNN variants were evaluated in a challenging multi-class dataset of seismic signals from Nevado del Ruiz volcano. These signals were represented with spectrograms and compared with the DTW distance (computed on the spectrograms) and the Euclidean distance (computed on the averaged spectrograms). The experiments revealed that the KNN variants enable one to improve the original KNN rule, with a reasonably small increase in the computational load. Further, classification accuracies were also better than those obtained with alternative interpretable methods, such as RF or SVM. As future work, an enrichment of the signal representation by using not just dissimilar-

Table 5 Confusion matrices of the different literature methods

Method	Confusion matrix				
KNN	165	79	7	13	8
	62	231	10	3	0
	4	4	84	1	0
	11	19	3	207	1
	9	0	0	0	144
SVMlin	124	50	8	18	1
	90	244	12	5	3
	5	16	78	3	1
	17	13	5	197	0
	15	10	1	1	148
SVMpoly	161	54	7	7	2
	71	261	12	1	0
	5	11	81	0	0
	7	4	3	216	0
	7	3	1	0	151
SVMrbf	161	55	7	6	2
	70	266	8	4	0
	4	7	87	1	0
	8	2	1	213	1
	8	3	1	0	150
RF	163	54	7	17	6
	71	263	10	8	0
	3	7	87	3	0
	5	5	0	196	0
	9	4	0	0	147

ities between spectrograms but also traditional signal features—for instance, event duration—might be considered in either a stacked or combined classification pipeline.

Acknowledgements The authors acknowledge the Cooperation and Academic Exchange Agreement between Universidad Nacional de Colombia and Università degli Studi di Verona, which is available at https://tinyurl.com/ae4dt7d5.

Declarations

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

References

Adadi A, Berrada M (2018) Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). IEEE Access 6:52138–52160

Bicego M, Baldo S (2016) Properties of the Box-Cox transformation for pattern classification. Neurocomputing 218:390–400
Bicego M, Loog M (2016) Weighted k-nearest neighbor revisited. In: Proceedings of international conference on pattern recognition, pp 1642–1647
Bicego M, Orozco-Alzate M (2020) PowerHC: non linear normalization of distances for advanced nearest neighbor classification. In: Proceedings of international conference on pattern recognition, pp 1205–1211
Bicego M, Acosta-Muñoz C, Orozco-Alzate M (2013) Classification of seismic volcanic signals using hidden-Markov-model-based generative embeddings. IEEE Trans Geosci Remote Sens 51(6):3400–3409
Bicego M, Londoño-Bonilla JM, Orozco-Alzate M (2015) Volcano-seismic events classification using document classification strategies. In: Proceedings of international conference on image analysis and processing, pp 119–129
Bishop CM (2006) Pattern recognition and machine learning. Springer, New York
Bramer M (2016) Principles of data mining, 3rd edn., Chap 7: estimating the predictive accuracy of a classifier, Springer, pp 79–92
Canário JP, Mello R, Curilem M, Huenupan F, Rios R (2020) In-depth comparison of deep artificial neural network architectures on seismic events classification. J Volcanol Geoth Res 401(106):881
Cárdenas-Peña D, Orozco-Alzate M, Castellanos-Domínguez G (2013) Selection of time-variant features for earthquake classification at the Nevado-del-Ruiz volcano. Comput Geosci 51:293–304
Carniel R, Guzmán SR (2020) Machine learning in volcanology: a review. In: Volcanoes: updates in volcanology, IntechOpen
Castro-Cabrera PA, Orozco-Alzate M, Adami A, Bicego M, Londoño-Bonilla JM, Castellanos-Domínguez G (2014) A comparison between time-frequency and cepstral feature representations for seismic-volcanic pattern classification. In: Proceedings of Iberoamerican congress on pattern recognition, pp 440–447
Castro-Cabrera P, Castellanos-Dominguez G, Mera-Banguero C, Franco-Marín L, Orozco-Alzate M (2021) Adaptive classification using incremental learning for seismic-volcanic signals with concept drift. J Volcanol Geoth Res 413(107):211
Chouet B, Matoza R (2013) A multi-decadal view of seismic methods for detecting precursors of magma movement and eruption. J Volcanol Geoth Res 252:108–175
Cover T, Hart P (1967) The nearest neighbor decision rule. IEEE Trans Inform Theory IT 13:21–27
Cox TF, Cox MAA (1994) Multidimensional scaling. Chapman & Hall, London
Curilem M, Soto R, Huenupan F, Cardona C, Franco L, San Marin C (2019) Hierachical classification structure based on SVM for volcano seismic events. In: Proceedings of IEEE Latin American conference on computational intelligence (LA-CCI), pp 1–6
Duin RPW, Bicego M, Orozco-Alzate M et al (2014) Metric learning in dissimilarity space for improved nearest neighbor performance. In: Proceedings of joint international workshop on structural, syntactic and statistical pattern recognition, pp 183–192
Fukanaga K (1990) Introduction to statistical pattern recognition, 2nd edn. Academic press, San Diego
Grijalva F, Ramos W, Peréz N, Benítez D, Lara-Cueva RA, Ruiz M (2021) ESeismic-GAN: a generative model for seismic events from Cotopaxi volcano. IEEE J Sel Top Appl Earth Observ Remote Sens 14:7111–7120
Karpatne A, Ebert-Uphoff I, Ravela S, Babaie HA, Kumar V (2018) Machine learning for the geosciences: challenges and opportunities. In: IEEE transactions on knowledge and data engineering, pp 1–12
Kostorz W (2021) A practical method for well log data classification. Comput Geosci 25(1):345–355
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):1–11
Lara-Cueva R, Benítez DS, Paillacho V, Villalva M, Rojo-Álvarez JL (2017) On the use of multi-class support vector machines for classification of seismic signals at Cotopaxi volcano. In: Proceedings of IEEE international autumn meeting on power, electronics and computing, pp 1–6
Lin J, Williamson S, Borne KD (2012) Pattern recognition in time series. In: Advances in machine learning and data mining for astronomy, chap 28, CRC Press, pp 617–646
Lopes N, Ribeiro B (2015) Incremental hypersphere classifier (IHC). In: Machine learning for adaptive many-core machines: a practical approach, vol 7, Springer, chap 6, pp 107–123

Malfante M, Dalla Mura M, Métaxian JP Mars JI, Macedo O, Inza A (2018) Machine learning for volcano-seismic signals: challenges and perspectives. IEEE Signal Process Mag 35(2):20–30
Orozco-Alzate M, Acosta-Muñoz C, Londoño-Bonilla JM (2012) The automated identification of volcanic earthquakes: concepts, applications and challenges. In: Earthquake research and analysis: seismology, seismotectonic and earthquake geology. InTech, chap 19, pp 345–370
Orozco-Alzate M, Castro-Cabrera PA, Bicego M, Londoñ-Bonilla JM (2015) The DTW-based representation space for seismic pattern classification. Comput Geosci 85:86–95
Orozco-Alzate M, Duin RPW, Bicego M (2016) Unsupervised parameter estimation of non linear scaling for improved classification in the dissimilarity space. In: Proceedings of joint international workshop on structural, syntactic, and statistical pattern recognition, pp 74–83
Orozco-Alzate M, Baldo S, Bicego M (2019a) Relation, transition and comparison between the adaptive nearest neighbor rule and the hypersphere classifier. In: Proceedings of international conference on image analysis and processing, pp 141–151
Orozco-Alzate M, Londoño-Bonilla JM, Nale V, Bicego M (2019) Towards better volcanic risk-assessment systems by applying ensemble classification methods to triaxial seismic-volcanic signals. Eco Inform 51:177–184
Pal AK, Mondal PK, Ghosh AK (2016) High dimensional nearest neighbor classification based on mean absolute differences of inter-point distances. Pattern Recogn Lett 74:1–8
Peréz N, Venegas P, Benítez D Lara-Cueva R, Ruiz M (2020) A new volcanic seismic signal descriptor and its application to a data set from the Cotopaxi volcano. IEEE Trans Geosci Remote Sens 58(9):6493–6503
Talebi H, Peeters LJM, Mueller U, Tolosana-Delgado R, van den Boogaart KG (2020) Towards geostatistical learning for the geosciences: a case study in improving the spatial awareness of spectral clustering. Math Geosci 52(8):1035–1048
Titos M, Bueno A, García L, Benitez C (2018) A deep neural networks approach to automatic recognition systems for volcano-seismic events. IEEE J Sel Top Appl Earth Observ Remote Sens 11(5):1533–1544
Titos M, Bueno Á, García L, Benítez MC, Ibañez J (2019) Detection and classification of continuous volcano-seismic signals with recurrent neural networks. IEEE Trans Geosci Remote Sens 57(4):1936–1948
Triguero I, Derrac J, García S, Herrera F (2012) A taxonomy and experimental study on prototype generation for nearest neighbor classification. IEEE Trans Syst Man Cybern Part C Appl Rev 42(1):86–100
Trombley RB (2006) The forecasting of volcanic eruptions. Ed. iUniverse
Trujillo-Castrillón N, Valdés-González CM, Arámbula-Mendoza R, Santacoloma-Salguero CC (2018) Initial processing of volcanic seismic signals using Hidden Markov Models: Nevado del Huila, Colombia. J Volcanol Geoth Res 364:107–120
Venegas P, Peréz N, Benítez D, Lara-Cueva R, Ruiz M (2019) Combining filter-based feature selection methods and Gaussian mixture model for the classification of seismic events from Cotopaxi volcano. IEEE J Sel Top Appl Earth Observ Remote Sens 12(6):1991–2003
Wang J, Neskovic P, Cooper LN (2007) Improving nearest neighbor rule with a simple adaptive distance measure. Pattern Recogn Lett 28(2):207–213
Wang X, Mueen A, Ding H, Trajcevski G, Scheuermann P, Keogh E (2013) Experimental comparison of representation methods and distance measures for time series data. Data Min Knowl Disc 26(2):275–309