UNIVERSIDAD NACIONAL DE COLOMBIA

Pattern representations for classifying (non-)metric (non-)vectorial data

with applications in Structural Health Monitoring and geotechnical/natural-hazard engineering

Yesid Mauricio Ospina Dávila

Universidad Nacional de Colombia
Faculty of Engineering and Architecture
Department of Electrical, Electronic, and Computing Engineering
Manizales, Colombia
2022

Pattern representations for classifying (non-)metric (non-)vectorial data
with applications in Structural Health Monitoring and geotechnical/natural-hazard engineering

Yesid Mauricio Ospina Dávila

A dissertation submitted in partial fulfillment of the requirements for the degree of:
Doctor in Engineering – Automatics

Advisor:
Dr.Eng. Mauricio Orozco-Alzate

Research Line:
Statistical Pattern Recognition and Signal Processing

Universidad Nacional de Colombia
Faculty of Engineering and Architecture
Department of Electrical, Electronic, and Computing Engineering
Manizales, Colombia
2022

Representaciones de patrones para la clasificación de datos (no-)vectoriales (no-)métricos

con aplicaciones en el Monitoreo de Salud Estructural y la ingeniería de amenazas geotécnicas/naturales

Yesid Mauricio Ospina Dávila

Tesis presentada como requisito parcial para optar al título de:
Doctor en Ingeniería – Automática

Advisor:
Dr.Ing. Mauricio Orozco-Alzate

Research Line:
Reconocimiento de Patrones Estadístico y Procesamiento de Señales

Universidad Nacional de Colombia
Facultad de Ingeniería y Arquitectura
Departamento de Ingeniería Eléctrica, Electrónica y Computación
Manizales, Colombia
2022

Abstract

Nowadays, data-driven modelling in structural and geo-engineering problems using Statistical Pattern Recognition and Machine Learning provides powerful and more versatile tools within a predictive framework. In contrast to the mainstream orientations of the state-of-art in data-driven structural and geo-engineering surrogates, which are based on advanced and (hyper-)parametrized classifiers, this thesis is focused on data representation issues. Firstly, for vectorial slope/landslide data, feature-based vector spaces are enriched and enhanced according to the Occam's razor principle, which is achieved through three simple but powerful existing variants of a transparent classifier as the nearest neighbor rule. Secondly, for non-vectorial SHM data, powerful and highly discriminant dissimilarity-vector spaces are built-up using spectral/time-frequency information from structural states, adopting a proximity-based learning scheme. In both cases, the results show the importance of a proper data representation and its key role in a bottom-up design for surrogate modelling.

Keywords: Classifier system design, Data-driven surrogates, Dissimilarity pattern recognition, Landslides, Pattern representation, Slope stability, Structural health monitoring.

Resumen

Actualmente, el Reconocimiento de Patrones Estadístico y el Aprendizaje de Máquinas proveen herramientas poderosas y versátiles para el modelamiento predictivo de problemas de estructuras civiles, mecánicas y de la geo-ingeniería. A diferencia de las principales tendencias en el estado del arte en los sustitutos basados en datos en problemas de estructuras y de geo-ingeniería, esta tesis se enfoca en la representación de los datos. Primero, para datos vectoriales de taludes/deslizamientos, los espacios vectoriales basados en características son enriquecidos y mejorados de acuerdo al principio de la navaja de Occam o de parsimonia, el cual se logra mediante tres simples pero poderosos variantes ya existentes del clasificador de vecinos más cercanos. Segundo, para datos no-vectoriales pertenecientes al Monitoreo de Salud Estructural, son construidos, poderosos y altamente discriminantes, espacios de disimilitudes usando información espectral/tiempo-frecuencia, tomando un esquema de aprendizaje basado en proximidades. En ambos casos, los resultados demuestran la importancia de una apropiada representación de datos y su influencia en el diseño incremental de modelos sustitutos.

Palabras clave: Deslizamientos, Diseño de sistemas de clasificación, Estabilidad de taludes, Monitoreo de salud estructural, Reconocimiento de patrones basado en disimilitudes, Representación de patrones, Sustitutos basados en datos.

Table of contents

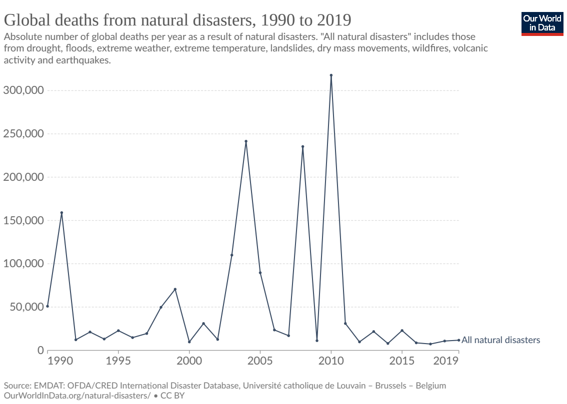
List of figures

List of tables

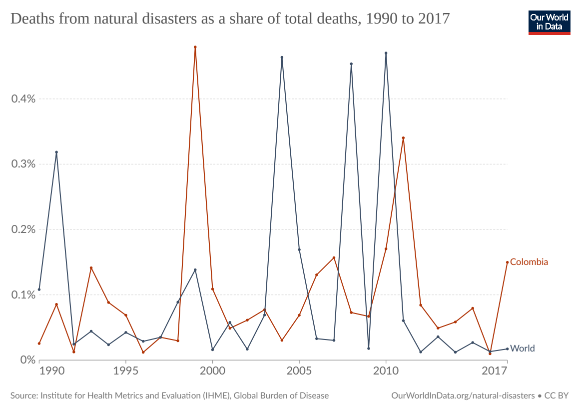
1 Introduction

1.1. Motivation

Some facts could give us a better understanding about the importance, in our society, of the disaster risk management and its effects on natural hazard engineering [134]: firstly, on average $60,000$ people are killed by natural disasters per year around the world (see Figure **1-1**a), where this death toll represents, approximately, a range from 0,1 % to 0,4 % of the global deaths (see Figure **1-1**b).



(a) Global deaths caused by all natural disasters, from 1990 to 2019.

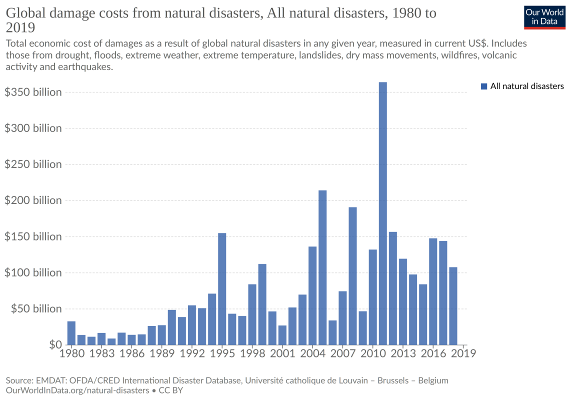


(b) Deaths from all natural disasters as a share of total deaths, from 1990 to 2017: World and Colombia tolls.

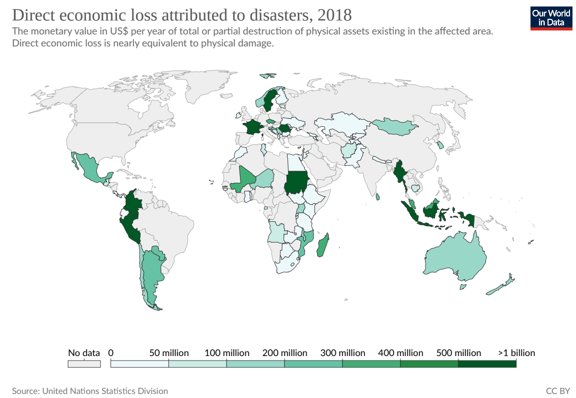
Figure 1-1: Natural disasters statistics: human deaths. Source [134]

Secondly, the global damage costs from all natural disasters have been increasing since the 90's, reaching global values of 350 billion USD (see Figure **1-2**a), and where, for instance, Colombia suffered a direct economic loss of 56,68 billion USD in 2018; in fact, Colombia is one of the countries with the highest economic loss level for that year (see Figure **1-2**b).

Catastrophic events, such as earthquakes, hurricanes, landslides or wind turbulence, can introduce unexpected degradation and damage in civil, aerospace or mechanical infrastructures; therefore, the application of damage assessments and structural performance strategies could be of paramount importance for preventing damage issues. In this sense, Structural Health Monitoring (SHM), —from a Pattern Recognition perspective—, assesses the *health state* of an infrastructure through data processing and statistical analysis [49, 172] i.e., SHM



(a) Total economic cost of damages in USD, from 1980 to 2019.



(b) Economic loss in USD per country in 2018.

Figure 1-2: Natural disasters statistics: economical loss. Source [134]

is a damage detection strategy that seeks to extract information from samples of dynamic response measurements, which are recorded by a distributed array of sensors located on the structure, in order to predict its structural performance and safety [155].

Simultaneously, in very recent years, there has been a growing interest in the development of the so-called *data-driven* (also known as *data-based* or *model-free*) algorithms, to obtain useful information from large or small data sets with the aim of predicting future trends and to make suitable, and perhaps economical, decisions in science, engineering, industry and finance [18]. In this sense, the Statistical Pattern Recognition (SPR) and Machine Learning (ML) theory offers powerful and efficient tools to address this type of problems, especially when they imply uncertain and complex models, with a high degree of engineering knowledge and heavy hardware and software requirements [153]. It means that SPR/ML systems become *data-driven surrogates* when the original problem does not need a detailed engineering explanation regarding how the prediction was obtained [129]; moreover, they also allow to build a boosted physics-informed predictive system [80]. A graphical explanation about a data-driven approach using tools from the SPR/ML is shown in Fig. 1-3. Up to this point, the main components are highlighted: input data, algorithm for learning and prediction.

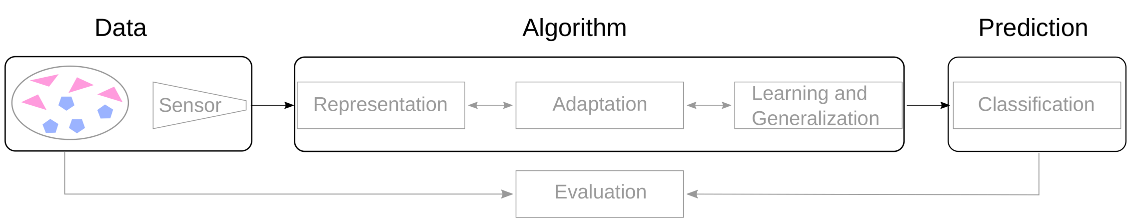


Figure 1-3: A generic data-driven approach using tools from the SPR/ML.

The predictive modelling on geotechnical/natural-hazard engineering and on the assessment of structural safety and performance illustrates the importance of this approach [177, 101]. Some particular examples from these areas can be found in flash flood prediction [165], stream-flow forecasting [180], landslide susceptibility mapping [128], slope stability evaluation [94], reliability analysis of geotechnical structures [90], earth science informatics [87], site-specific characterization of soils [28], pattern recognition for vibration measurements [172], damage prediction [159], building information modelling considering natural hazards [167], among others [7, 181]. In addition, several efforts to set up specialized datasets [1] [2] for prediction tasks in geotechnical/natural-hazard engineering have been carried out.

Two data-driven problems are of interest in this thesis. On the one hand, the slope stability analysis and the landslide susceptibility assessment which are well-established areas within geotechnical/natural-hazard engineering [74], due to their key role on the reliability of critical infrastructures such as networks of highways, roads, tunnels or bridges. On the other, SHM which is a comparatively recent discipline within the assessment of structural safety and performance that, originally, has involved scientific areas such as signal processing, SPR/ML methods, decision theory or probabilistic risk assessment [154, 52].

Although the aforementioned problems are different, from an engineering perspective, both require —given the high level of economical and life losses— data-driven approaches that will ensure its reproducibility and facilitate a fair comparison across diverse SPR/ML systems. Typically, these SPR/ML systems are composed of several steps (see Figure 1-4), including two key ones [43]: (1) representation and (2) learning and generalization.

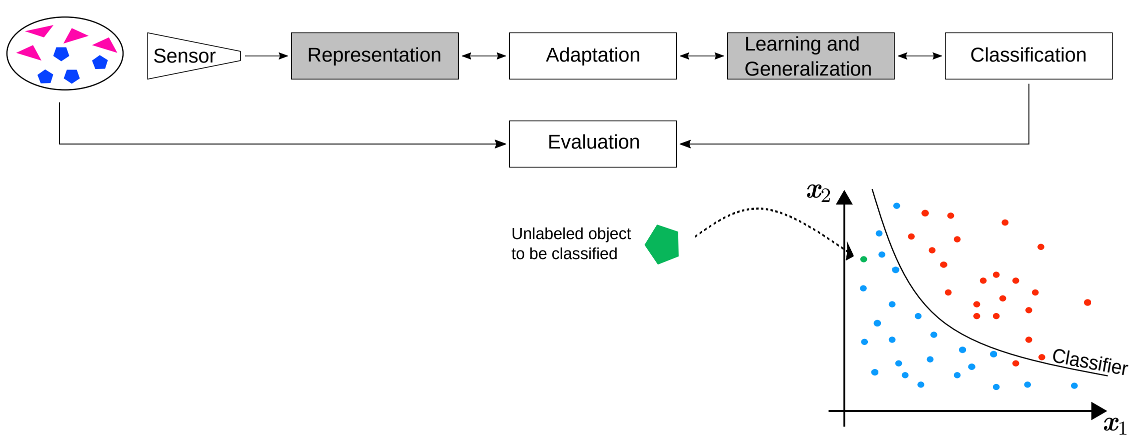


Figure 1-4: A schematic diagram of a SPR/ML system using a 2-D vector space representation. Adapted from [42]

The first one concerns how an individual real-world object or phenomenon is numerically described by feature vectors, probability models, (dis-)similarity representations, among others.

[1] https://hazmapper.org/
[2] https://www.nextgenerationliquefaction.org/

The last one is related to obtain decision boundaries between pattern classes through statistical decision theory or geometric tools. The adaptation is an intermediate step in which representations, learning methods or prior knowledge are simplified, enriched or 'adapted' for obtaining a satisfactory trade-off between recognition accuracy and required computational resources. In the evaluation stage, the performance of the SPR/ML system is estimated via some kind of loss function.

This thesis is devoted entirely to the issue of representation in slope/landslide and SHM predictions, considering them as classification tasks. Indeed, it should be noted that this stage is outside the mainstream orientations of nowadays data-driven slope/landslide and SHM assessments based on SPR/ML surrogates, which are mostly restricted to learning and generalization. The emphasis of this thesis, within a SPR/ML scheme, is depicted in the Fig. **1-5**.

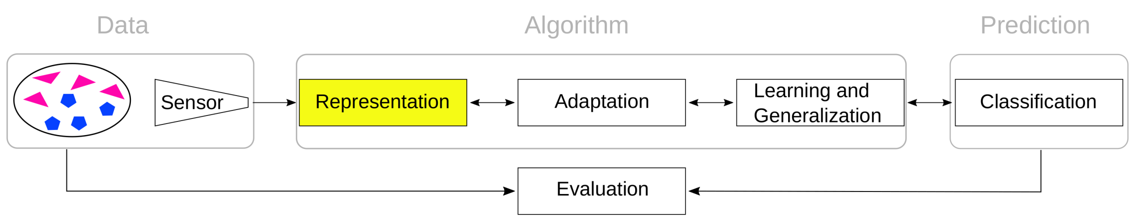


Figure 1-5: A generic data-driven approach using tools from the SPR/ML.

1.2. Problem statement

In slope stability analysis and landslide susceptibility, the data are usually represented by features such as cohesion (c), friction angle (φ), unit weight (γ) and geometric properties (slope angle or height) for the first case, or geological and topographic factors such as land coverage, type of slide material, rainfall conditions, etc., for the second one. The limited availability of these data, strictly from real-world cases, have led to the design of heavily (hyper-)parameterized SPR/ML systems for dealing with the uncertainty associated to this scarcity of data (see Section 3.1). In addition, these datasets are sparse, noisy, uncertain and spatially variable which, in turn, entails the need to understand the structure and distribution of the data [126].

In contrast with the feature-based approach, most data obtained from any real-world SHM implementation, such as digitally-acquired signals, contours or images of different sizes, are often considered as non-vectorial or syntactic data [52, 159, 9] i.e., they do not correspond —in their raw form– to a feature representation and, hence, no feature space is originally defined. Although the SHM community has paid a lot of attention to statistical model development for feature discrimination, obtaining a discriminative set of highly damage-sensitive features is a hard task for these non-vectorial or data [111, 51, 151]. To tackle this issue, once again, (hyper-)parameterized SPR/ML systems have also been designed, such as (deep-)neural networks, support vector machines (SVM), ensemble learning methods, etc. Indeed —in many cases— inappropriate setups are adopted i.e., SPR/ML systems which use, during training, data from both damaged and undamaged conditions providing, thus, additional information to any computational learning technique under unrealistic settings (see Section 3.2). A more fair or realistic scenario consists in only using undamaged examples for training, since data from damaged conditions are often unavailable or very costly to be acquired. Such a setting is known as a one-class classification approach.

In parameter-laden SPR/ML systems, to determine suitable values for (hyper-)parameters by cross-validation is not a straightforward task, which, in most situations, lacks the interpretability and, besides, imposes pre-fixed concepts from a particular domain expertise [81, 82] to ensure a proper performance. Therefore, these parameter-laden algorithms may fail in finding the optimal values for their (hyper-)parameters, potentially causing overtraining or hindering the reproducibility of the results. Accordingly, in [56, p. 3134] it is claimed that "A researcher may not be able to use classifiers arising from areas in which he/she is not an expert (for example, to develop parameter tuning), being often limited to use the methods within his/her domain of expertise". In [158, p. 27] a long-pursued parameter-free approach, for SPR/ML, is referred as: "Ideally, a learning method is automatic, i.e., no parameters need to be set by the user". In this sense, a practical SPR/ML system should have as few parameters to tune as possible while, at the same time, reaching predictive models that are able to adequately explain the real-world data. This parameter-free or parameter-light approach implies simpler SPR/ML systems (see, for example, [30, pp. 3-7] or [5, p. 86]). Moreover, this reasoning is consistent with the so-called Occam's razor, which "is the principle of parsimony [which claims that] a model should be simple enough for efficient computation and complex enough to be able to capture data specifics" [14, pp. 104].

According to the above-listed issues, the starting point for this thesis is the observation that the ability to properly extract knowledge from a data set, to make decisions, is the cornerstone of the data-driven approach and, therefore, it is well-known that the issue of data representation is as important as the statistical model development [76]. In other words, the better data representation, potentially the clearer or sharper separation between classes is achieved and, hopefully, parameter-free or parameter-light SPR/ML systems will be sufficient for a successful classification. In turn, this parameter-free or parameter-light approach may be a very important step towards the design of *interpretable* or *explainable* SPR/ML systems [109, 12], focusing on the reasoning of "letting the data speak for themselves". Currently, an interpretable concept-learning is not only an active new research direction in the SPR/ML areas, but also in slope stability evaluation [135, 100] and SHM [121].

Consequently, this thesis seeks to respond the following questions:

– By default, data-driven slope stability evaluation is defined by features. Accordingly, is it possible to obtain competitive or better performances than state-of-art models, from a bottom-up design of SPR/ML systems, such that parameter-free/light feature-based models are accomplished? Moreover, could this feature representation be enhanced without a significant increase in the number of model parameters?

– If compelling evidence exists of alternative approaches to deal directly with non-vectorial data i.e., dissimilarity representations, is it possible to obtain competitive or better performances than state-of-art feature-based models in data-driven SHM, from alternative data representations and adopting a bottom-up design of SPR/ML systems, such that parameter-free/light dissimilarity-based models are accomplished?

1.3. Objectives

The main goals of this thesis are summarized as follows:

– Explore mathematical properties and practical possibilities of enriched dissimilarity representations, based on spectral/time-frequency information and one-class classifiers, for structural health monitoring.

– Evaluate advantages or disadvantages of employing (non-)metric information contained on dissimilarity spaces with applications to structural health monitoring.

– Examine the importance about the reproducibility and the interpretability issues on methods and/or research based on statistical pattern recognition, in particular, for the case of structural health monitoring and geotechnical/natural-hazard engineering, through the use of classifiers that have few, or none, (hyper-)parameter(s) to tune.

– Validate the proposed methodology using computational simulations.

1.4. Structure of the thesis

After this introduction, Chapter 2 starts with a basic background for this thesis, which includes a brief description of the learning mechanism in SPR/ML, followed by an explanation about nearest feature classifiers; in particular, the Nearest Feature Line and the Rectified Nearest Feature Line Segment classifiers are presented. Also, hypersphere-based scalings techniques and one-class classifiers within a parameter-free/light approach are described. Then, the fundamental concepts of proximity learning are provided, focusing on dissimilarity embeddings as an alternative to representation issues when we deal with non-vectorial or syntactic data.

Chapter 3 presents the related work for slope/landslide susceptibility analysis and SHM using SPR/ML techniques. Chapter 4 presents the main contributions of the thesis through three sections which correspond to the three associated papers. Paper 1 is based on the design of SPR/ML systems according to the Occam's razor principle for slope vectorial data. On the other hand, Paper 2 is dedicated to slope/landslide vectorial data with emphasis on an enrichment of the representational capacity of the data set or an enhancement of the distance learning. Lastly, Paper 3 proposes a dissimilarity-based pattern recognition approach for SHM non-vectorial data. As a whole, all three papers present an alternative and a bottom-up framework for the design of highly competitive SPR/ML systems for two particular problems in engineering. Finally, conclusions and future work are discussed in Chapter 5.

2 Learning classifier systems using (non-)metric and (non-)vectorial data

This chapter is devoted to basic concepts in Statistical Pattern Recognition (SPR). In this direction, it gives a special importance to representation i.e., how the object or data is *represented* and the choice of discriminant (invariant-)features, following the original, and main, aim of the field of SPR [1]: "... the core research topic of [statistical] pattern recognition, the topic that makes it different from the related domains [machine learning, neural networks and statistics], is to study object representations in which the relevant pattern classes naturally emerge". For years, a number of authors have already highlighted, —and supported—, this difference [76, 42]. The object or data representations typically used in SPR/ML are shown in Figure **2-1**, and the respective contributions (papers) of this thesis upon this roadmap are also indicated.

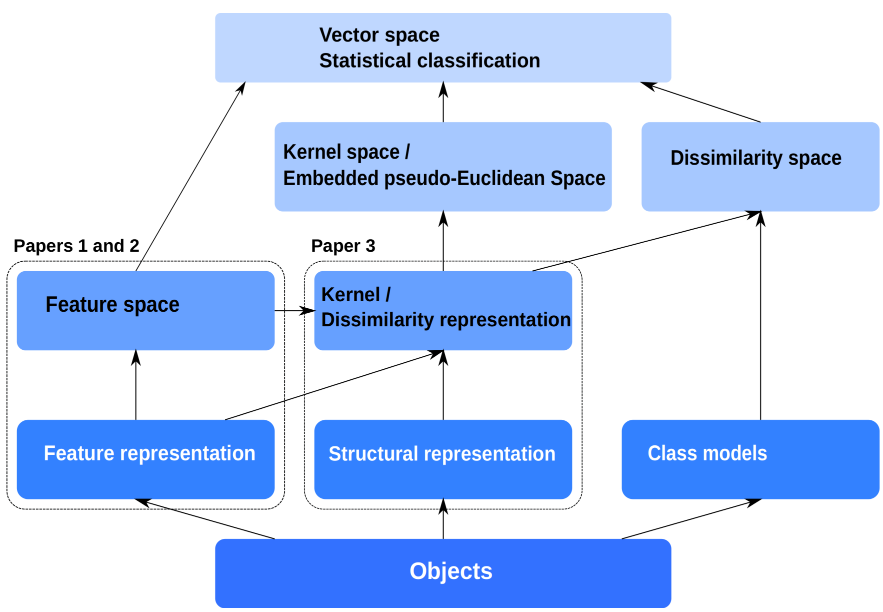


Figure 2-1: Typical object or data representations used in SPR/ML. Adapted from [40].

Within this framework, there has been a particular focus on classifier systems with few, or none, (hyper-)parameter(s) to tune, in order to study the learning and generalization

[1] https://37steps.com/108/pr-core-business/

in relation to the issue of representation providing interpretability of the results of applied SPR/ML systems.

2.1. Statistical pattern recognition

Pattern recognition (PR) is a scientific discipline of computer science that, according to Bishop [15, p. vii] originated in engineering. From a theoretical and practical perspective, PR studies methods for designing *intelligent machines* which learn to discriminate *patterns* for a specific recognition task under noisy or uncertain conditions, simulating the human and biological ability of perception and intelligence [5, 153] (see Fig. **2-2**). In this context, a pattern can be seen as "any relation present in the data, whether it be exact, approximate or statistical" [138, p. 8]. Practical applications of PR include machine vision, computer-aided diagnosis, remote sensing, speech recognition, financial forecasting, data mining and knowledge discovery, among others [163].

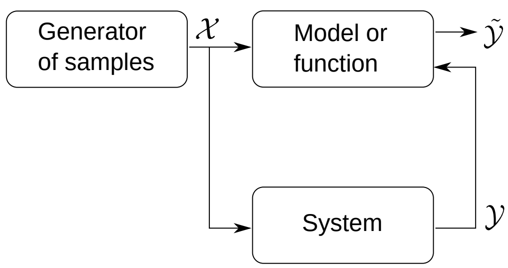


Figure 2-2: Model of a learning machine. Adapted from [26].

In SPR, each pattern is represented by a *feature vector* i.e., a d-dimensional vector of real numbers that characterize a certain object which is composed by d features or attributes, each one corresponding to a dimension in a proper vector space, called *feature-based vector space*. This feature-based vector space is a well-equipped vectorial space due to their powerful geometrical and analytical tools that, in many cases, are not available in other representations [163]. However, choosing the right set of features for a specific recognition problem is often a difficult and challenging task; indeed, this task usually ends up being a subjective and knowledge-dependent process [38].

Under this rationale, the purpose of a SPR/ML system can be defined as follows [69]: on the basis of n samples, denoted as $\mathcal{Z} := \{(\mathbf{x}_1, y_1), ..., (\mathbf{x}_n, y_n)\} \in \mathcal{X} \times \mathcal{Y}$, to build a model or function, $f(\mathbf{x}, \gamma) : \mathcal{X} \to \mathcal{Y}$, that predicts unseen data, $\mathcal{T} = \{\mathbf{t}_1, ...\mathbf{t}_k\}$. Conventionally, $\mathcal{X} \in \mathbb{R}^d$ is the set of feature vectors or inputs and $\mathcal{Y} \in \mathbb{R}$ is the set of associated measured output or labels, where $\mathcal{Z}$ set is called *training set*, the set of unseen data is called *test set* and the function or model $f(\mathbf{x}, \gamma)$ is the *learning machine*. Furthermore, the samples in $\mathcal{Z}$

are assumed independently identically distributed from an unknown probability measure, $dP(\mathbf{x}, y)$, and $\gamma \in \Gamma$ is the set of (hyper-)parameters of a particular function or model.

From a statistical learning perspective, the quality of a learning machine is determined by the risk function:

$$R(\gamma) = \int_{\mathcal{X}} L(y, f(\mathbf{x}, \gamma))\, dP(\mathbf{x}, y) \tag{2-1}$$

which is a Stieltjes integral defined as the expected value of the loss function $L(y, f(\mathbf{x}, \gamma))$ that, in essence, estimates the difference between the learning machine and the actual output. If $P(\mathbf{x}, y)$ is derivable, then Eq. 2-1 becomes:

$$R(\gamma) = \int_{\mathcal{X}} L(y, f(\mathbf{x}, \gamma))\, p(\mathbf{x}, y)\, d\mathbf{x}dy \tag{2-2}$$

Thus, the aim of a parsimonious SPR/ML system is to minimize the value of the risk function with the simplest learning machine e.g., to obtain the learning machine with the lowest number of (hyper-)parameters to adjust in Eq. 2-2, according to the so-called Occam's razor (see Section 1.2). Since the expected risk can be minimized only by a finite set of samples, then it is only possible to compute an approximation of the expected risk, known as Empirical Risk Minimization principle (ERM) and defined as:

$$R_e(\gamma, n) = \frac{1}{n} \sum_{i=1}^{n} L(f(\mathbf{x}_i, \gamma)) \tag{2-3}$$

The main learning modes considered in SPR/ML are [138]: (un)supervised learning, semi-supervised learning and reinforcement learning. This thesis pays special attention to supervised learning and, in particular, to the classification task. In this case and particularly for two classes, the data consists in n pairs of feature vectors $\mathbf{x}_i \in \mathbb{R}^d$ and labels $y_i \in \{+1, -1\}$, so that $\mathbf{z} := \{(\mathbf{x}_i, y_i), \forall i \in \{1 : n\}\}$. For classifications involving multiples $\mathcal{C}$ clases, $y_i \in \{1, 2, ..., \mathcal{C}\}$. The loss function assumes the form:

$$L(y, f(\mathbf{x}_i, \gamma)) = \begin{cases} 0, & \text{if } y = f(\mathbf{x}, \gamma) \\ 1, & \text{otherwise} \end{cases} \tag{2-4}$$

and the ERM is defined as:

$$R_e(\gamma, n) = \frac{1}{n} \sum_{i=1}^{n} \mathcal{I}(y_i = f(\mathbf{x}_i, \gamma)) \tag{2-5}$$

being $\mathcal{I}$ the indicator function.

2.2. Learning parameter-free/light classifier systems

The design of parameter-free or parameter-light SPR/ML systems is a suitable step towards the concept-learning of interpretable Artificial Intelligence [109] and, within this framework, the nearest neighbor (NN) classifier is recognized as a *transparent classifier* [10]. That means that, it is —by itself— highly understandable for a human due to its clear geometrical interpretation [18, p. 192].

The NN classifier builds a local decision boundary using the nearest prototype feature point, $\mathbf{x}_i^c$, to the unlabeled point, $\mathbf{x}$. In its most basic form this method chooses the label of the $\mathbf{x}_i^c$ and the nearness is conventionally quantified using the Euclidean distance [13]. A simple extension considers k nearest prototype feature points to the unlabeled point, $\mathbf{x}$, thus, this rule classifies $\mathbf{x}$ by assigning it the class label most frequently represented of the k nearest prototype feature points. The empirical risk, for a binary problem, is defined as [26]:

$$R_e(\gamma, n) = \frac{1}{k}\sum_{i=1}^{n}\left(y_i - f\left(\mathbf{x}_i,\gamma\right)\right)\mathcal{I}_k\left(\mathbf{x}_i,\mathbf{x}\right) \tag{2-6}$$

where k is the number of nearest prototype feature points and $\mathcal{I}_k\left(\mathbf{x}_i,\mathbf{x}\right) = 1$ if $\mathbf{x}_i$ is one of the k feature point nearest to $\mathbf{x}$ and zero in other case. Here, $f\left(\mathbf{x}_i,\gamma\right) \in \{+1,-1\}$. The value $f\left(\mathbf{x}_i,\gamma\right)$ for which the empirical risk is minimized is [26]:

$$f\left(\mathbf{x}_i,\gamma\right) = \begin{cases} 1 & \frac{1}{k}\sum_{i=1}^{n} y_i\mathcal{I}_k\left(\mathbf{x}_i,\mathbf{x}\right) > 0{,}5 \\ 0 & \text{otherwise} \end{cases} \tag{2-7}$$

NN is one of the simplest non-parametric classifiers [31] with a very competitive performance[2] in complex SPR/ML problems subject to (non-)vectorial data [174, 12] and its accuracy, in Euclidean spaces, achieves at most twice the Bayes error rate in an asymptotic sense [118].

However, its performance is closely related to: (1) the representational capacity of the data set i.e., the smaller the training set size, the higher the loss in its performance, as well as, (2) on the structure of the data e.g., unbalanced feature scales or meaningless features. Similarly, its performance depends on the choosen distance [39]. To cope these limitations, extensive enhanced methods based on geometrical strategies [21] or adaptive (non-)metric distance learning [61] have been proposed.

This thesis pays special attention, in the first case, to one of the most well-known enrichment rules of the NN method —originated from the Machine perception and Computer vision community— called the nearest feature rules, which employ a linear interpolation and extrapolation between sample points or prototype feature points of the same class [92, 93, 27]. For

[2] An example is the technological strategy that won the first Netflix progress prize, in 2007, which was based on the NN rationale [11].

the later case, a family of (non-)metric learning distances, based on the fundamental concept of territorial hyperspheres, are studied for complex and noisy datasets. Both approaches are detailed in the following sections. In addition to that, data descriptors or one-class classifiers are briefly mentioned.

2.2.1. Nearest feature classifiers

In general, classification methods based on nearest feature rules are designed under simple but powerful geometrical concepts whose major advantage is supported by the enrichment of the feature space or the enhancement of the distance learning. This enrichment is achieved through some kind of approximation between the prototype feature points which, similarly to NN, does not imply any tunning parameter.

The Nearest Feature Line classifier

The Nearest Feature Line (NFL) classifier was originally proposed in [92]. It is a template matching procedure which generalizes each pair of prototype feature points, $\{\mathbf{x}_i^c, \mathbf{x}_j^c\}$, in the same class by a linear approximation, L_{ij}^c, called *feature line* (see Figure 2-3 left side). This feature line is defined by the span $L_{ij}^c = sp(\mathbf{x}_i^c, \mathbf{x}_j^c)$. A new unlabeled point $\mathbf{x}$ (also known as *query*) must be projected onto L_{ij}^c as follows:

$$\tilde{\mathbf{x}}_{ij}^c = \mathbf{x}_i^c + \tau(\mathbf{x}_j^c - \mathbf{x}_i^c) \tag{2-8}$$

where τ is the position parameter given by $\tau = (\mathbf{x} - \mathbf{x}_i^c) \cdot (\mathbf{x}_j^c - \mathbf{x}_i^c)/\|\mathbf{x}_j^c - \mathbf{x}_i^c\| \in \mathbb{R}$ and $\tilde{\mathbf{x}}_{ij}^c$ is the projection of the query point.

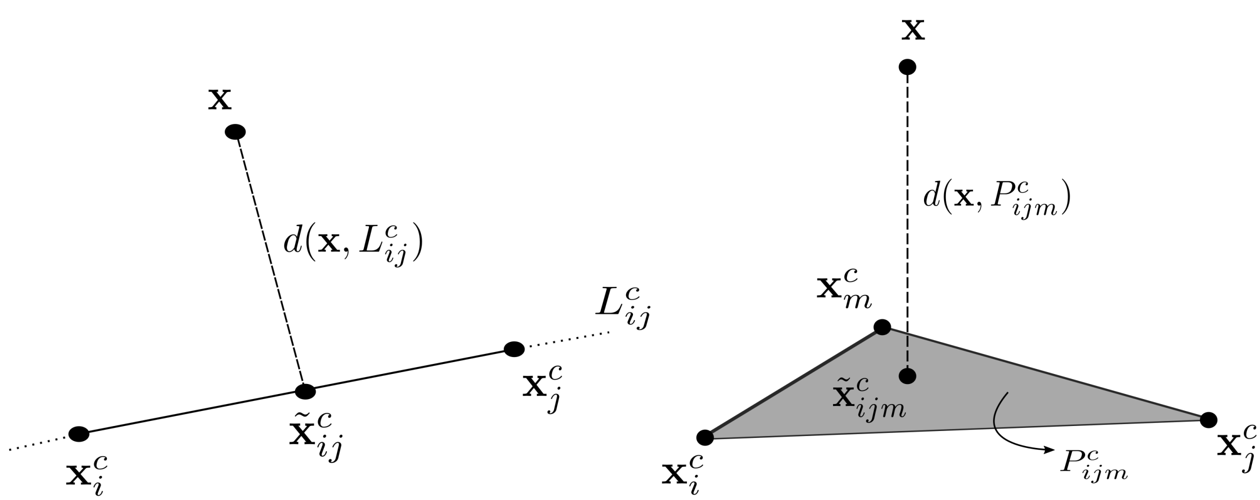


Figure 2-3: Query point and its distance to feature line (left) or to feature plane (right). Adapted from [114].

The classification of $\mathbf{x}$ is done by assigning the class label $\hat{c}$ to it according to the most frequent label among the k nearest feature lines:

$$d(\mathbf{x}, L_{\hat{i}\hat{j}}^{\hat{c}}) = \min_{1 \leqslant c \leq C,\, 1 \leq i,j \leq n_c,\; i \neq j} d(\mathbf{x}, L_{ij}^{c}) \tag{2-9}$$

where

$$d(\mathbf{x}, L_{ij}^{c}) = \|\mathbf{x} - \tilde{\mathbf{x}}_{ij}^{c}\| \tag{2-10}$$

A theoretical justification of NFL was shown in [189]. The authors proved its effectiveness on problems with small datasets, due to additional information provided by the feature lines passing through each pair of samples; thus, the representativeness of the prototype feature points set is generalized.

Similarly, the Nearest Feature Plane (NFP) classifier belongs to this family of algorithms. The formulation of NFP is a geometrical extension of NFL but, in this case, three prototype points of the same class $\{\mathbf{x}_i^c, \mathbf{x}_j^c, \mathbf{x}_m^c\}$ are used during the classification, obtaining a generalization through a triangular subspace or *feature plane*, $\mathbf{P}_{ijm}^c$, in a Euclidean space [27]. This feature plane is defined as $P_{ijm}^c = sp(\mathbf{x}_i^c, \mathbf{x}_j^c, \mathbf{x}_m^c)$. A new unlabeled point $\mathbf{x}$ is projected onto $\mathbf{P}_{ijm}^c$, as follows:

$$\tilde{\mathbf{x}}_{ijm}^{c} = \mathbf{X}_{ijm}^{c} \left(\mathbf{X}_{ijm}^{c}{}^{\top} \mathbf{X}_{ijm}^{c} \right)^{-1} \mathbf{X}_{ijm}^{c}{}^{\top} \mathbf{x} \tag{2-11}$$

with $\mathbf{X}_{ijm}^c = \left[\mathbf{x}_i^c \, \mathbf{x}_j^c \, \mathbf{x}_m^c\right]$. Considering $k = 1$, the classification of the query point, $\mathbf{x}$, is done similarly to Eq. 2-9:

$$d\left(\mathbf{x}, P_{\hat{i}\hat{j}\hat{m}}^{\hat{c}}\right) = \min_{1 \leqslant c \leq C,\, 1 \leq i,j,m \leq n_c,\; i \neq j \neq m} d(\mathbf{x}, P_{ijm}^{c}) \tag{2-12}$$

where

$$d(\mathbf{x}, P_{ijm}^{c}) = \|\mathbf{x} - \tilde{\mathbf{x}}_{ijm}^{c}\| \tag{2-13}$$

Regarding NFL, the interpolating or extrapolating parts of some feature lines could cause two trespass mistakes: extrapolation inaccuracy pointed out in [185] and interpolation inaccuracy considered in [37]; a graphical explanation is given in Figure **2-4**. In fact, there exists a third drawback: its large computation cost [190].

Nevertheless, several modified or refined NFL approaches have been reported in the literature in order to handle these issues. In [185], the Nearest Neighbor Line (NNL) was presented; it is a fast algorithm that deals with the extrapolation inaccuracy. For the study reported in [188], the authors extended the NFL method with a new distance metric. In [190] a pattern classification method, the Nearest Feature Midpoints (NFM), was developed. In NFM, the classification is based on the minimum Euclidean distance between a query sample and the

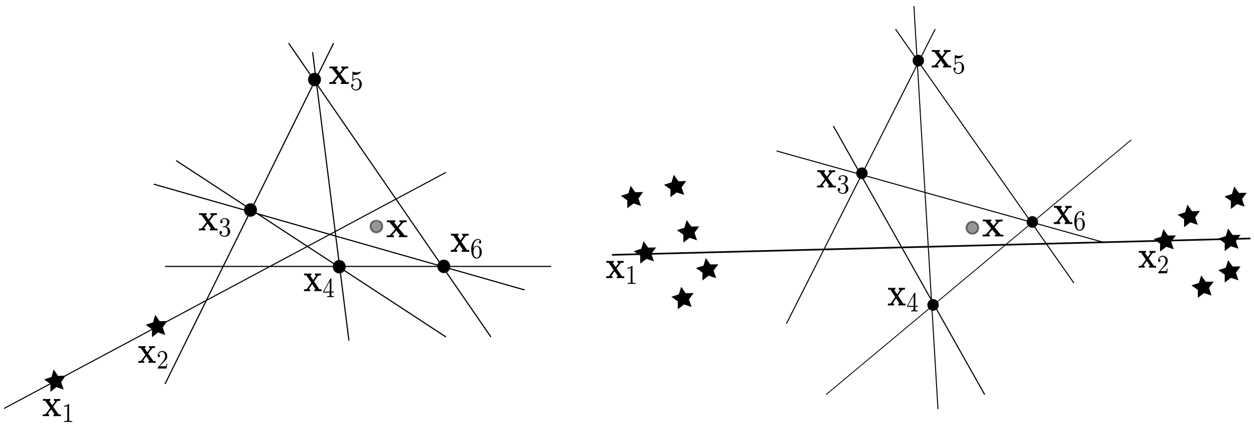


Figure 2-4: Inaccuracies caused by trespassing: extrapolation (left) and interpolation (right). Adapted from [37].

middle point on a feature line. NFM deals with the extrapolation problem under a simple and fast computing framework.

A Center-based Nearest Neighbor (CNN) classifier was reported in [60] based on the Euclidean norm between the query point and the feature line defined by the center of the class and a training sample. CNN can significantly reduce the computational cost. The Restricted Nearest Feature Line with Angle (RNFLA) was suggested in [54], which solves the problem of far away prototypes in NFL i.e., the RNFLA classifier computes the angle between the query point and each of the feature lines to overcome the extrapolation inaccuracy. In all cases, the refined NFL approaches borrow the geometrical NFL concept to redefine a new metric. Hence, refined NFL rules inherit somehow the drawbacks of the original NFL, specially concerning the trespassing issue.

The Rectified Nearest Feature Line Segment classifier

Among these methods, a refined approach which overcomes both interpolation and extrapolation inaccuracies linked to NFL, called the Rectified Nearest Feature Line Segment (RNFLS), was developed in [37]. RNFLS tackles both issues in a two-stage procedure: first, when the query lies in the extrapolation part, just a feature line segment is used, denoted by $\widetilde{L_{ij}^c}$. That is, the distance to the line is assumed between the query, $\mathbf{x}$, and the closest point of the feature line segment, $\boldsymbol{z} \in \widetilde{L_{ij}^c}$. Thus, Eq. 2-10 becomes:

$$d(\mathbf{x}, \widetilde{L_{ij}^c}) = \min_{\mathbf{z} \in \widetilde{L_{ij}^c}} \|\mathbf{x} - \mathbf{z}\| \tag{2-14}$$

Secondly, if the query lies in the interpolation part, an examination of the territories of each class is carried out. Besides if the feature line trespasses a sample territory which belongs to other class, this feature line would be removed. This sample territory, $T_{\mathbf{x}_i^c} \subseteq \mathbb{R}^n$, is defined by:

$$T_{\mathbf{x}_i^c} = \left\{ \mathbf{h} \in \mathbb{R}^n : \|\mathbf{h} - \mathbf{x}_i^c\| \leq \rho_{\mathbf{x}_i^c} \right\} \tag{2-15}$$

where $\rho_{\mathbf{x}_i^c}$ is expressed as:

$$\rho_{\boldsymbol{x}_i^c} = \min_{1 \leqslant c,r \leq C,\ c \neq r} \left\| \boldsymbol{x}_i^c - \boldsymbol{x}_j^r \right\| \quad \forall j = 1, \ldots, n \tag{2-16}$$

and where the union of the sample territories, beloging to the same class, leads up to the class territory $\mathcal{T}_c = \bigcup_c T_{\mathbf{x}_i^c}$.

Although its main drawback is the computational cost, this method enhances the classification ability and is applicable to very complex problems in pattern recognition [91, 113].

2.2.2. Hypersphere-based scalings

Two hypersphere-based scalings are considered in this thesis: the Hypersphere classifier (HC) [97] and the Adaptive Nearest Neighbor (ANN) classifier [169], which are powerful methods that adapt the metric used in such a manner that weigh distances to prototype feature points which are well inside their class [112]. In other words, HC and the ANN classifier involve a locally adaptive distance measure of a new unlabeled point $\mathbf{x}$ by means of an weighting process. Notice that, these techniques are particularly useful when $\mathbf{x}$ is near the class boundaries where there is a possibility of overlap between classes or noise level [169, 12].

For HC [97] the region of influence of a given prototype feature point $\mathbf{x}_i^c \in \mathbb{R}^d$ is defined as $\eta_i = \rho_i \,/\, 2$, where ρ_i is its radius computed by Eq. 2-16. So, the distance from $\mathbf{x}$ to $\mathbf{x}_i^c$, for HC, is given by:

$$d_{HC}\left(\mathbf{x}, \mathbf{x}_i^c\right) = \|\mathbf{x} - \mathbf{x}_i^c\| - g\eta_i \tag{2-17}$$

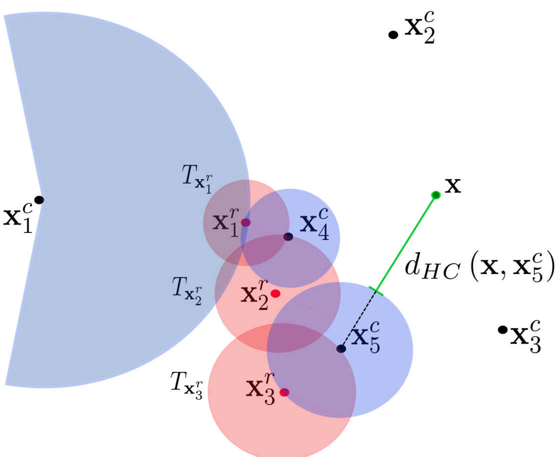
where g is the parameter that controls the overlapping between hyperspheres from different classes. Fig. **2-5**a shows a graphical explanation about HC which considers two classes: the c class and the r class. For example, the 'adapted' distance from $\mathbf{x}$ to $\mathbf{x}_5^c$ is depicted by the green line, $d_{HC}\left(\mathbf{x}, \mathbf{x}_5^c\right)$. In this case, it is assumed that the 'contrary' nearest prototype of $\mathbf{x}_5^c$ is $\mathbf{x}_3^r$.
The original version of the HC method proposes a value of $g = 2$, resulting in:

$$d_{HC}\left(\mathbf{x}, \mathbf{x}_i^c\right) = \|\mathbf{x} - \mathbf{x}_i^c\| - \rho_i \tag{2-18}$$

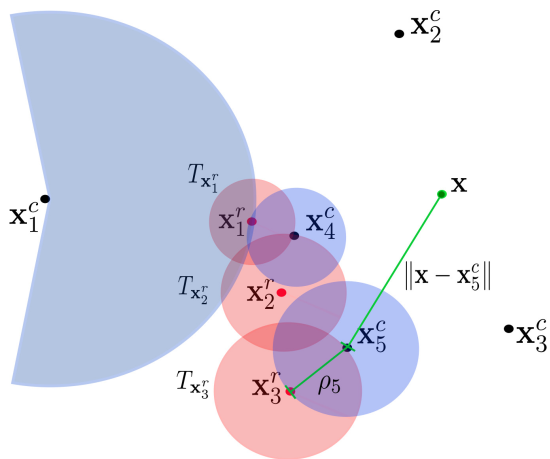
For the ANN classifier [169], the distance is scaled as follows (see Fig. **2-5**b):

$$d_{ANN}\left(\mathbf{x}, \mathbf{x}_i^c\right) = \frac{\|\mathbf{x} - \mathbf{x}_i^c\|}{\rho_i} \tag{2-19}$$

where the two components of the Eq. 2-19 are highlighted by two green segments.



(a) The Hypersphere classifier (HC).



(b) The Adaptive Nearest Neighbor (ANN) classifier.

Figure 2-5: Hypersphere-based scalings techniques considered in this thesis.

2.2.3. One-class classifiers

One-class classification, also called data description, is a SPR/ML approach, where the main goal is to describe and distinguish one class of objects, called *target class*, from all other possible objects, called *outlier class*. This classification problem is subject to classes with very poorly balanced data or where one of the classes is totally absent [161]. This approach is motivated by real-world problems such as anomaly/damage detection for machine condition monitoring or SHM [50, 139, 9], or imbalanced data sets for structural/geotechnical reliability [106], where most of the data corresponds to undamaged, safe or "healthy" data.
In general, all one-class classifiers have the following form [118]:

$$\begin{aligned}\mathcal{O}(\mathbf{x} \mid \mathcal{X}, \gamma) &= \mathcal{I}\left(d\left(\mathbf{x} \mid \mathcal{X}, \gamma\right) < \theta\right) \\ &= \begin{cases} 1, & \text{if } \mathbf{x} \text{ is classified as a target} \\ -1, & \text{otherwise} \end{cases}\end{aligned} \tag{2-20}$$

where $\mathcal{I}(\cdot)$ is the indicator function, which defines if a query $\mathbf{x}$ is accepted or rejected as target object given a (dis)similarity measure d between the query vector $\mathbf{x}$ and the training set $\mathcal{X}$. The model complexity parameter is referred as γ which is determined during the training phase e.g., proper tuning of penalty value, kernel function and its width for one-class support vector machines, and θ represents a rejection fraction of the target class such as 0.1, for instance. Similarly, the rejection fraction is defined as:

$$\min \theta \tag{2-21}$$

$$\text{s.t.} \quad \frac{1}{n}\sum_{i=1}^{n} \mathcal{I}\left(d\left(\mathbf{x}|\mathcal{X}, \gamma\right) \geq \theta\right) = \varepsilon \tag{2-22}$$

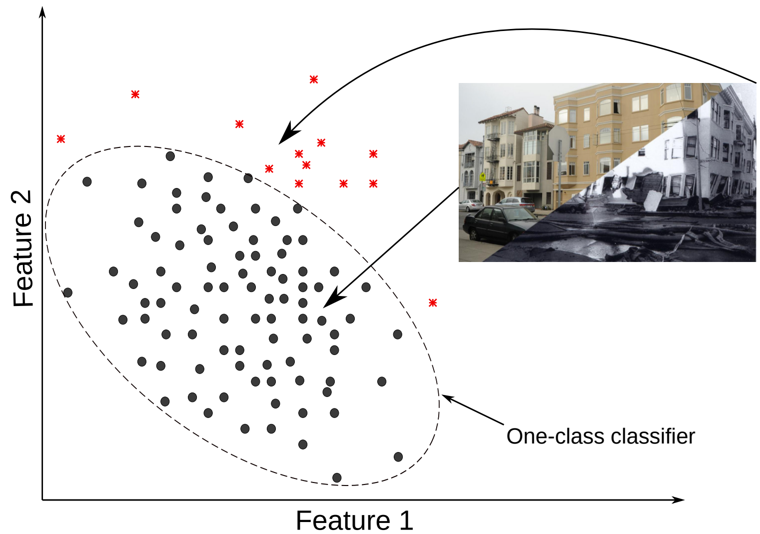


Figure 2-6: One-class classifier representation for (un)damaged structures. The dashed line describes the target class corresponding to undamaged data (black points), while the outlier class is outside the enclosed region, representing therefore the damaged data (red asterisk).

A wide set of one-class classifiers, belonging to proximity and density-based models, are considered in this thesis [162, 78]:

1. The nearest neighbor based data description (NNDD): This method uses the distance from a query object, $\mathbf{x}$, to the first nearest neighbor in the target-training set, $NN^{tr}(\mathbf{x})$, for its labelling. NNDD does not need any (hyper-)parameter to be optimized. In this case, the indicator function (Eq. 2-20) becomes:

$$\mathcal{I}\left(d\left(\mathbf{x} \mid \mathcal{X}, \gamma\right) < \theta\right) = \mathcal{I}\left(\frac{\left\|\mathbf{x} - NN^{tr}\left(\mathbf{x}\right)\right\|}{\left\|NN^{tr}\left(\mathbf{x}\right) - NN^{tr}\left(NN^{tr}\left(\mathbf{x}\right)\right)\right\|} \leqslant 1\right) \tag{2-23}$$

2. The k-nearest neighbor data description (k-NNDD): This method is similar to NNDD, however, it uses the distance to the k-nearest neighbor where k is found by a leave-one-out density estimation over the target-training set.

3. The Parzen density data description (parzenDD): This classifier is an extension of the Gaussian mixture classifier, where the density estimation is computed by a mixture of Gaussian kernels centered at the individual target-training objects following the expression:

$$\begin{aligned}\mathcal{I}\left(d\left(\mathbf{x} \mid \mathcal{X}, \gamma\right) < \theta\right) &= \mathcal{I}\left(p_p\left(\mathbf{x}\right) \leqslant \theta\right)\\ p_p\left(\mathbf{x}\right) &= \frac{1}{n}\sum_i p_n\left(\mathbf{x}, \mathbf{x}_i, h\mathcal{I}\right)\end{aligned} \tag{2-24}$$

where p_n is the kernel and $h > 0$ is the bandwidth. If a sample is out of the estimator, then it is assumed as outlier.

4. The support vector data description (SVDD): In this case, the description of the target-training set is carried out through a hypersphere with minimal volume around a center, $\mathbf{a}$. A flexible shape decision boundary can be obtained using a plethora of kernel settings. The SVDD is defined by:

$$\begin{aligned}\mathcal{I}\left(d\left(\mathbf{x} \mid \mathcal{X}, \gamma\right) < \theta\right) &= \mathcal{I}\left(\|\mathbf{x} - \mathbf{a}\|^2 \leqslant R^2\right) \\ &= \mathcal{I}\left((\mathbf{x} \cdot \mathbf{x}) - 2\sum_i \alpha_i \left(\mathbf{x} \cdot \mathbf{x}_i\right) + \sum_{i,j} \alpha_i \alpha_j \left(\mathbf{x}_i \cdot \mathbf{x}_j\right) \leqslant R^2\right)\end{aligned} \tag{2-25}$$

with $0 \leqslant \alpha_i \leqslant C$ being the Lagrange multipliers, and the parameter C is a tradeoff between the volume of the description and the errors. Now, R^2 is computed as the distance from the center of the sphere $\mathbf{a}$ to one of the support vectors on the decision boundary:

$$R^2 = \left(\mathbf{x}_k \cdot \mathbf{x}_k\right) - 2\sum_i \alpha_i \left(\mathbf{x}_i \cdot \mathbf{x}_k\right) + \sum_{i,j} \alpha_i \alpha_j \left(\mathbf{x}_i \cdot \mathbf{x}_j\right) \qquad \exists \mathbf{x}_k \in SV_{\text{bound}} \tag{2-26}$$

5. The minimum spanning tree data description (mstDD): This is a graph-based learning method which generates linear subspaces between target-training data, obeying the minimum spanning tree principle. The mstDD follows a form given by:

$$\begin{aligned}\mathcal{I}\left(d\left(\mathbf{x} \mid \mathcal{X}, \gamma\right) < \theta\right) &= \mathcal{I}\left(d_{mst}\left(\mathbf{x} | \mathcal{X} \leqslant \theta\right)\right) \\ d_{mst}\left(\mathbf{x} | \mathcal{X}\right) &= \min_{e_{ij} \in \text{mst}} d\left(\mathbf{x} | e_{ij}\right)\end{aligned} \tag{2-27}$$

If the value of τ, defined in Eq. 2-8, is $0 \leqslant \tau \leqslant 1$ then the distance d_{mst} can be obtained via Eq. 2-9 making appropriate changes [78]. Otherwise, the distance is computed under the same rationale from Eq. 2-14. Conventional selection of the rejection fraction is the minimum or maximum length of the linear subspace generated by each class.

2.3. Generalized kernels and statistical/machine learning with proximity data

Most real-world data are not *originally* defined by a set of isolated features. Examples of this include spectra, time and biological sequences, images, graphs, etc. In fact, there exists evidence that a human observer is primary led by *differences* between objects, later similarities come and finally a description is made by means of feature and models [45]. In other words, the basis of the human recognition and perception of patterns is mainly composed by the totality of the information contained in the object/data [44]. This process has two steps: (1) the objects are detected in their totality and (2) the isolated object is recognized. These two steps are interconnected and verified by each other.

At first, the pioneering work proposed in [65] emphasized the importance about the notion of proximity in statistical/machine learning method, where a proximity measure, $\mathscr{P} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, can be understood as a function from an arbitrary pair of data to a real value. However, the full potential of proximity-based learning starts with the formalization of the kernel methods that mainly learn from proximity data by means of similarity functions which are related to the inner product [69, 104].

For kernels it is asssumed that there exists a function $\Phi_{\mathcal{H}} : \mathcal{X} \to \mathcal{H}$ that maps the input data to a high-dimensional feature (Hilbert) space which is associated to the inner product $\langle \Phi_{\mathcal{H}}(\mathbf{x}_i), \Phi_{\mathcal{H}}(\mathbf{x}_j) \rangle_{\mathcal{H}}$. So, kernels are functions $\mathscr{K} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ associated to $\Phi_{\mathcal{H}}$ and $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ such that $\mathcal{K}(\mathbf{x}_i, \mathbf{x}_j) = \langle \Phi_{\mathcal{H}}(\mathbf{x}_i), \Phi_{\mathcal{H}}(\mathbf{x}_j) \rangle_{\mathcal{H}}$ [138]. This generalization from a inner product to a kernel function is known as the 'kernel trick'. The kernel matrix $\mathbf{K} \in \mathbb{R}^{n \times n}$ resulting from this kernel trick must be symmetric and positive definite i.e., $\forall \mathbf{q} \in \mathbb{R}^n : \mathbf{q}^\top \mathbf{K} \mathbf{q} \geqslant 0$.

As SPR/ML applications increased, the considered input data become more and more complex. Hence, nowadays SPR/ML methods must deal with non-vectorial, structural and, in general, with generic proximity data, e.g., text data, graphs, digital images, time series, genome-sequences, probabilistic models, etc. [108]. Conversely, kernel methods were originally designed for vectorial data. So, such methods are not directly applicable to generic proximity data and, therefore, kernels for structured data have been proposed in order to represent this non-vectorial information such as, graph kernels, gap-weighted subsequences kernels, P-kernels, among others [138]. Nevertheless, this kind of information is compositional and coding it in a set of fixed-length and highly discriminant vectors is still a very challenging task [41, 107].

An appealing alternative to learn from proximity data is offered by proximity-based representation spaces built with distances or dissimilarities [118]. This thesis is focused on these representation spaces, paying special attention to classification in a simple yet powerful proximity-based representation space so-called the dissimilarity-based vector space.

2.3.1. Proximity-based representation spaces

The representation spaces are data-dependent inner product spaces that encode the proximity information via domain-specific measures [122]. Suitable structure-aware measures have been devised by different learning frameworks; some examples are: Smith Waterman alignment in bioinformatics [107], compression distance for text analysis [82], non-metric measures in template matching [41], edit distances in time series analysis [170], Kullback–Leibler divergence measures in information-theoretic domains [117], or tangent distances [68], among others. One of the main reasons for using these domain-specific measures is their degree of accuracy and the intrinsic incorporation of prior knowledge about the problem [122, 108].

A special elastic measure, called Dynamic Time Warping (DTW) [132], which provides non-metric information between time series is used in this thesis. A time series is a sequence $\boldsymbol{s}$ of size $[t] \times [w]$. In order to compute the distance between two time series $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$ of length j and l, respectively, we need to define a warping or alignment path of order $j \times l$ as a sequence $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_q) \in \mathcal{A}$, being $\mathcal{A}$ an array of size $[j] \times [l]$. Two conditions must be met: the first one is $\alpha_1 = (1,1)$ and $\alpha_q = (j,l)$ for any warping path, and the second one is $\alpha_{r+1} - \alpha_r \in \{(1,0), (0,1), (1,1)\}$, $\forall 1 \leq r < q$. The cost function of a warping along α for $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$ is obtained by:

$$C_\alpha(\boldsymbol{s}_1, \boldsymbol{s}_2) = \sum_{(i,r)\in\alpha} \| \boldsymbol{s}_1 - \boldsymbol{s}_2 \|^2, \, \alpha \in \mathcal{A} \tag{2-28}$$

where $\| \cdot \|$ is the Euclidean norm and where the DTW is then defined as:

$$d_{dtw}(s_1, s_2) = \min_\alpha C_\alpha(\boldsymbol{s}_1, \boldsymbol{s}_2) \tag{2-29}$$

In general, in many theoretical or practical problems, domain-specific measures are often either non-metric or non-Euclidean, and thus violate mathematical assumptions of SPR/ML algorithms; that means, the positive definiteness and symmetry conditions are unfulfilled [122, 45]. Moreover, the proximity-based representations are related to the field of non-Mercer kernels in SPR/ML theory [119], where it is well recognized that non-metric or non-Euclidean information generates powerful vector spaces to classify non-vectorial or syntactic data [148, 149].

The dissimilarity-based vector space

The dissimilarity representation for SPR/ML consists of a finite numerical representation based on relative differences between objects, which might span a space i.e., consider the collection of these dissimilarity values as entries of a vector in a so-called *dissimilarity-based vector space*, with suitable properties for the use of statistical/machine learning tools [120, 118] (see Figure **2-7**). Thus, this proximity-based learning has a direct application in the following SPR/ML problems:

- When the data are non-vectorial type such as shapes, digital images, spectra or time series. In SHM non-vectorial data correspond, for instance, to accelerations or displacements measured or (un)damaged images (see Section 4.3).

- When the data are structural type, i.e., strings, graphs. In SHM this one corresponds, for instance, to recent applications of transfer learning [62].

- When the vector representation lies in a high-dimensional space;

- When a feature representation is composed by mixed types;

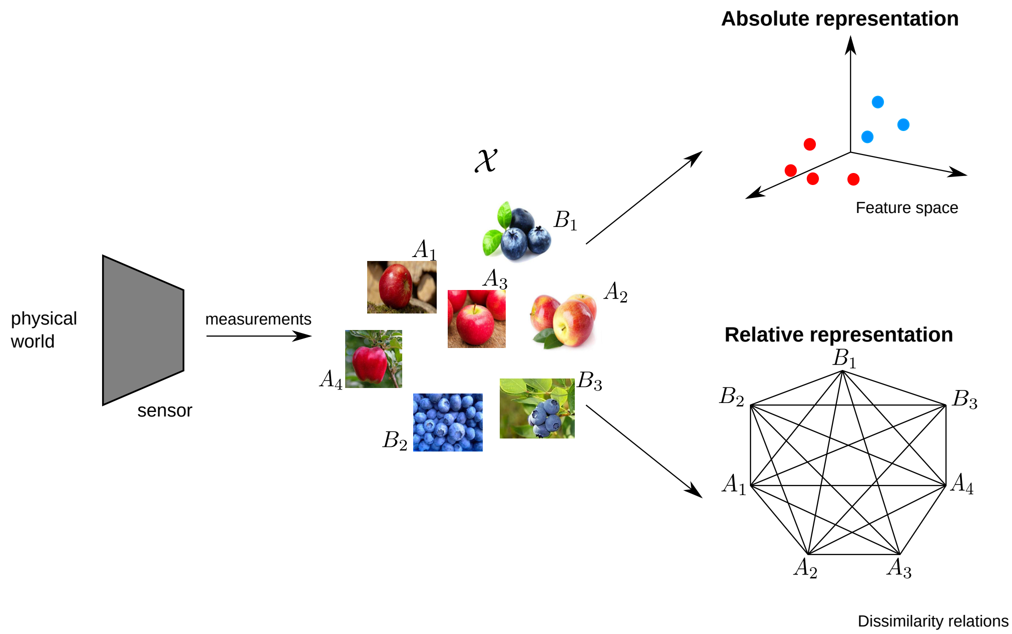


Figure 2-7: Dissimilarity-based representation in comparison to the feature-based representation. Adapted from [116].

- To design nonlinear classifiers from a given feature space.

The formalization of that representation is as follows. Let $\mathcal{X} = \{x_1, x_2, \ldots, x_N\}$ be a training set of N objects and $\mathcal{P} = \{p_1, p_2, \ldots, p_n\} \subseteq \mathcal{X}$ be a set of n prototype objects, which are representative instances of the classes. If d_{ds} is a dissimilarity measure between objects, then an object $x \in \mathcal{X}$ is represented by a row vector of dissimilarities given by:

$$\mathbf{d}\left(x, \mathcal{P}\right) = \left[d_{ds}\left(x, p_1\right)\ d_{ds}\left(x, p_2\right) \ldots d_{ds}\left(x, p_n\right)\right] \tag{2-30}$$

Therefore, the dissimilarity representation for the training set $\mathcal{X}$ is then obtained when the dissimilarity matrix

$$\mathbf{D}\left(\mathcal{X}, \mathcal{P}\right)\ \forall x_i \in \mathcal{X}, i \in \{1 : N\} \tag{2-31}$$

is computed. Similarly, the dissimilarity matrix $\mathbf{D}\left(\mathcal{T}, \mathcal{P}\right)$ must be also obtained for the test set $\mathcal{T} = \{t_1, t_2, \ldots, t_k\}$ with k objects.

In general, a dissimilarity matrix $\mathbf{D}_{N \times n} = (d_{ij})\ \forall i, j$ has the following properties [118]:

1. Non-negativity: $d_{ij} \geq 0$

2. Identity of indiscernibles: $d_{ij} = 0$ iff objects i and j are identical

3. Symmetry: $d_{ij} = d_{ji}$

4. Triangle inequality: $d_{ij} < d_{ik} + d_{kj}$

5. Euclidean property: A $N \times n$ matrix $\mathbf{D} = (d_{ij})$ is Euclidean if there exists an isometric Euclidean embedding into a Euclidean space

6. Compactness property: If objects i and j are very similar, then $d_{ij} < \delta$

7. True representation: If $d_{ij} < \delta$ then objects i and j are very similar

8. Continuity property: Given $(\mathcal{X}, d_{ds})$ a generalized metric space, the dissimilarity measure $d_{ds} : \mathcal{X} \times \mathcal{X} \longrightarrow \mathbb{R}_0^+$ is continuous at x, y, if x_n, $y_n \in \mathcal{X}$ with $\lim_{n\longrightarrow\infty} x_n = x$ and $\lim_{n\longrightarrow\infty} y_n = y \Rightarrow \lim_{n\longrightarrow\infty} d(x_n, y_n) = d(x, y)$

Fulfillment of the first two properties produces positive definite dissimilarity matrices and fulfillment of the first four properties defines a metric space. This approach is adopted in this thesis, where the dissimilarity matrix is adressed as a data-dependent mapping $\mathbf{D}(\cdot, \mathcal{P}) : \mathcal{X} \rightarrow \mathbb{R}^n$, from the representation set to the so-called dissimilarity-based vector space.

In this regard, a linear classifier trained in a dissimilarity space $\mathbf{D}(\mathcal{X}, \mathcal{P})$ is given by:

$$f\left(\mathbf{D}\left(x, \mathcal{P}\right)\right) = \sum_{j=1}^{n} \varpi_j d\left(x, p_j\right) + \varpi_0 = \mathbf{w}^\top \mathbf{D}\left(x, \mathcal{P}\right) + \varpi_0 \tag{2-32}$$

where $f(\mathbf{D}(x, \mathcal{P})) = 0$ defines the classifier. According to this one, for example, a linear Bayesian classifier for a binary problem is defined as:

$$f\left(\mathbf{D}\left(x, \mathcal{P}\right)\right) = \mathbf{w}^\top \mathbf{D}\left(x, \mathcal{P}\right)^\top + \varpi_0 \tag{2-33}$$

where $\mathbf{w} = C_\varpi^{-1}(\mathbf{m}_1 - \mathbf{m}_2)$ and $\varpi_0 = -\frac{1}{2}(\mathbf{m}_1 - \mathbf{m}_2)^\top C_\varpi^{-1}(\mathbf{m}_1 - \mathbf{m}_2) + \log\left(\frac{p_{w_1}}{p_{w_2}}\right)$. Here, p_{w_1} and p_{w_2} corresponds to class prior probabilities and C_ϖ is the sample covariance matrix; $\mathbf{m}_1$ and $\mathbf{m}_2$ are the class mean in the dissimilarity space. For more details and classifiers in dissimilarity spaces, see [118].

Finally, the dissimilarity based vector space is an Euclidean one, which preserves all the non-Euclidean or non-metric information [122, 41], and is equipped with its own norm and inner product. Each dimension, given by Eq. 2-30, corresponds to the dissimilarity $\mathbf{d}(\cdot, p_i)$ to a prototype p_i. Notice that, $\mathbf{d}(\cdot, p_i)$ can be seen as a dissimilarity-based feature vector.

Pseudo-Euclidean linear embedding space

In this approach, the dissimilarity matrix $\mathbf{D}(\mathcal{X}, \mathcal{X})$ is embedded by a proper isometric (distance-preserving) mapping in a pseudo-Euclidean spaces $\mathcal{E} = \mathbb{R}^{(p,q)} = \mathbb{R}^p \oplus \mathbb{R}^q$, which is endowed with a nondegenerate indefinite inner product $\langle \cdot, \cdot \rangle_{\mathcal{E}}$ such that this one is positive definite in $\mathbb{R}^p$ and negative definite in $\mathbb{R}^q$. Typically, $\mathcal{E}$ require either corrections on proximity data or nontrivial reformulations of the classification scheme [45].

Final remark

There is another approach to deal with these dissimilarity matrices known as the pretopological approach [41]. In this case, the nearest neighbor rule is applied directly to $\mathbf{D}(\mathcal{T}, \mathcal{P})$, considering balls around training data. The main disadvantage of this approach is the deterioration in its performance for small training sets.

3 Previous work

This chapter is about the literature review for the two practical problems mentioned in Section 1.1. Accordingly, there is a section for each research problem.

3.1. Geotechnical/natural-hazard problem: Slope/landslide safety evaluation

SPR/ML methods have very recently taken the lead on data-driven slope/landslide safety evaluation tasks using real-world datasets. Very heavily (hyper-)parameterized classifiers have been studied for slope stability prediction, including many proposals based on Artificial Neural Networks (ANN), whose learning depends on the number of layers and neurons (nodes), the number of training epochs or the selection of an appropriate activation function [103].

For the prediction of the Factor of Safety (FOS) in slope stability analysis, several trial-and-error or empirical techniques have been reported in the literature for the adjustment of parameters in the learning scheme of ANN models [140, 29], ensemble of Kohonen self-organizing maps (SOM) and Bayesian ANNs [55] as well as for the tuning process for perceptrons [1]. Das et al. [34] adjusted the hyper-parameters of two feed-forward ANNs using two techniques: a heuristic global optimization called differential evolution and a Bayesian regularization method which, in turn, require additional parameters. Also, drawbacks such as slow training speed, local minimum solution problems or overtraining have been reported [96, 66, 88].

A hybrid stacking ensemble of classifiers —11 classifiers—, using data from the Finite Element Method (FEM) and real-world cases, was proposed in [79]. In this case, the selection of classifiers is optimized by an artificial bee colony (ABC) algorithm, however, the particular hyper-parameters corresponding to each classifier were adjusted by using particle swarm optimization (PSO) algorithms, which imply a "two-level" optimization process for a large number of hyper-parameters. Similarly, hybrid ML systems based on ANN were proposed in [84], for predicting of FOS of slopes; such systems use the following optimization techniques to adjust their hyper-parameters: genetic algorithms (GA), PSO, imperialist competitive algorithm (ICA), and ABC algorithms. Rukhaiyar et al. [137] suggested an empirical approach to adjust the architecture of a PSO-ANN system for predicting the FOS of slopes.

Hoang et al. [70] made a comparative study taking three powerful SPR/ML techniques: Radial Basis Function Neural Network (RBFNN), Least Squares Support Vector Machines (LSSVM), and the Extreme Machine Learning (ELM). In the experimental setup, they mention the importance about the hyper-parameters (e.g., the regularization constant and the kernel function for the LSSVM or the number of neurons in the hidden layer for the ELM). A powerful ensemble SPR/ML system has been suggested in [131], where a grid search method was employed to adjust the individual hyperparameters and a GA optimization to build the ensemble classifiers. In order to deal with this (hyper-)parameter setting, gradient boosting machines have been applied [186].

The influence of the (hyper-)parameter setting in Kernel methods, such as kernel type or width selection, have been also pointed out. In [142, 141] the importance of a proper tuning task on classification and regression using SVM was discussed. Sensitivity issues on parameter adjustment in relevance vector machine techniques are also mentioned; see [184]. In [143] three types of kernel functions were considered, namely: polynomial, Radial Basis Function (RBF) and Spline, whose parameters are tuned by trial-and-error. Zhang et al. [182] suggested an optimized adaptive relevance vector machine (ARVM). They carried out an analysis of width hyper-parameter values for three types of kernels: Gauss, Cauchy and Laplace. On the other hand, in [89] a very advanced quantum-behaved particle swarm optimization for obtaining optimal hyper-parameters in a LSSVM system was presented. Kumar et al. [86] proposed two SPR/ML models based on the idea of Minimax Probability Machine (MPM): the Linear MPM and the Kernelized MPM; they assumed a RBF and tuned the value of width via trial and error. A GA optimization was applied in [183] for training a SVM model. Similarly, these drawbacks for an either (hyper)parameter sintonization are also highlighted in other slope stability prediction problems based on a data-driven perspective [70, 187, 95].

Likewise, in landslide susceptibility prediction [133], numerous SPR/ML systems have been proposed: statistical learning techniques, in a physics-based model parameter identification scheme, using computational models and real-world data [64] or fuzzy classifiers [164] which depend on weight values of the belief and the fuzzy membership function values. A review about landslide susceptibility mapping using SVM was carried out in [75]. In this review, it is again highlighted the need for an optimization process in any SVM model, particularly, the type of kernel and the regularization parameter. A comparative study of three ML methods —random forest (RF), boosted regression three (BRT) and SVM— was made in [4]. Each classifier, in this case, requires to adjust several hyper-parameters. Chen et al. [23] applied an adaptive neuro-fuzzy inference system combined with frequency ratio (ANFIS-FR), a SVM and a Generalized Additive Model (GAM) for landslide susceptibility; all of them being parameter-laden algorithms. Similarly, comparison and/or ensemble of SPR/ML systems [22, 124] that, in spite of their high performances, present a lot of parameters to tune such as min/max size of the group, bag size, base classifier definition etc., are also reported in the literature. A weight optimization process of a back-propagation architecture for ANN [168], as well as bagging ensembles of classifiers [123], have been considered.

Recently, deep-learning applications have turned into a popular topic in the literature about landslide susceptibility prediction. In [73], the authors was considered a fully connected spare autoencoder (FC-SAE) optimized with a dropout approach; in this case, a large data set was required for a proper adjusting. Hua et al. [72] used a deep-ANN initialized by a Dynamic Bayesian Network (DBN) with the ReLU activation function for detecting the occurence of landslides, while Yang et al. [175] developed a dynamic displacement model using time series decomposition information to feed a long short-term memory (LSTM) neural network. A grid search method was applied to search optimal hyper-parameters of the LSTM NN model.

3.2. Assessment of structural safety and performance problem: Structural Health Monitoring

The following related work on data-based SHM emphasizes the anomaly detection focus from SPR/ML in which the mainstream research about data-based SHM relies on a statistical model development for feature discrimination from measured information [52], which is typically non-vectorial such as images, acceleration time-series, etc.; therefore, the usage of many feature extraction methods, from different families, has been reported: early studies using (un-)damaged features extraction from chaos theory and nonlinear dynamics modelling were proposed in data-based SHM [110, 105, 115, 171]. Conventional dimensionality reduction techniques such as Principal Component Analysis (PCA), kernel PCA and Locally Linear Embedding (LLE), among others, are also combined to obtain highly (un-)damage-features using a damage index scheme [98, 99].

Feature extraction techniques from (non-)stationary time series are hihlighted in the literature due to their importance to capture most discriminant information from nonlinear and (non-)stationary acceleration time series [53, 157]. In [17] it is discussed a support vector classifier trained in a feature-vector space which, in turn, is built with autoregressive (AR) patterns extracted from each acceleration time-series; furthermore, the autocorrelation function is employed to choose the order of the AR model. Similarly, in [67] the authors studied a distance-based anomaly detection using AR features for a sensorized W8 $\times$ 13 I-beam in steel. In [150] Shi et al. made a cointegration analysis using a decomposition of the SHM signals for environmental and operational variations where a large number of parameters were estimated, including the ARMA coefficients, the smoothing parameter, among others. In [179], the sensitivity of autorregressive (AR) model properties is used as a damage detection index that is validated through a large-scale bridge slab experiment; in this case, parameter adjustments were also reported. An ensemble based anomaly detection using three levels of the Mahalanobis-squared distance (MSD) was conducted in [146], along with one-class nearest neighbor (NN) rule for damage identification. Four deep learning algorithms were applied in [33]: MultiLayer Perceptron, Long Short Term Memory network, 1D Convolutional Neural Network, and Convolutional Neural Networks (CNN) to SHM. The results of such study point out to the need of an appropriate tuning of the algorithm (hyper-)parameters.

Regarding feature extraction in time series, a number of practical issues with respect to the adjustment of a set of (hyper-)parameters have been mentioned, ranging from typical model order selection [58, 57, 144], to adjustments depending on the method, such as the need of user-defined thresholds [156], the window size for moving PCA and the number of environmental and operational variability with similar response [85] or stationary assumptions [16]. Some approaches to solve these difficulties have been reported in [46, 47].

Advanced (hyper-)parameterized SPR/ML systems for SHM have also been proposed, such as deep-SVM [176], where the feature set is extracted with an encoder network which, in turn, requires to adjust the number of hidden layers, the set of weights and the activation functions, apart from the kernel function of the support vector data descriptor. An unsupervised damage diagnosis framework was proposed in [77], whose main contribution is the representation of the structural time series by means of a deep auto-encoder. In [83] a brief reference about data-based SHM was reported using CNN for one-dimensional time series. On the other hand, in [160], an experiment was carried out using the time-frequency images from the set of time series for the purpose of detecting the damage under an anomaly detection scheme. In this experiment, information from both damaged and undamaged states was used during the training phase.

Similarly, a time-frequency image classification from time series using wavelet analysis and CNN was studied in [24]. The authors of [20] carried out a comparative study about the performance of one-class SVM along several deep learning-based architectures (autoencoder, robust autoencoder, one-class neural networks, soft-bound one-class deep SVDD and hard-bound one-class deep SVDD) where parameters such as the encoder and decoder network layers are tunned for the best classifier. Also, a CNN application on real-time damage detection using damaged and undamaged data during training is found in [3]. Likewise, in [2], an enhanced CNN for data-based SHM using time series was developed, with the aim of reducing the amount of measurements required for a proper training.

Less attention has been paid to non-conventional data representations in the scientific literature. A symbolic data representation of time series which is characterized by statistical quantities such as histograms or interquartile intervals is proposed in [32, 145]. In this case, a clustering analysis was applied. This representation has been fed to three classification methods: Bayesian Decision Trees, ANN and SVM [6]. Soon later, an extension of such a representation was developed in [35], where a time-frequency Interquartile Range (TF-IQRM) is achieved from time series measured for real-time unsupervised data-based SHM. Other extension is studied in [36], where an augmented IQRM was applied in a novelty detection context. Additionally, frequency representations are encountered [178]: frequency domain decomposition [102], coherence function based damage detection [130, 136] or selected modal frequencies [152, 147].

Future research trends on data-based SHM using feature extraction processing are discussed in [166, 8]. Recently, an alternative is emerging: the Population-based approach to SHM (PBSHM) [63] which, in essence, is an application of transfer learning to SHM by using dissimilar configurations of civil, mechanical or aerospace structures.

4 Contributions

This chapter summarizes each published article and provides a brief mention about the main contribution of each of them, without going into technical details in order to avoid potential conflicts related to the copyright agreement with the publishers as well as to minimize unnecesary repetition with the content that is already included in the published versions. Papers 1 and 2 (Sections 4.1 and 4.2) address the third objective of the thesis, while Paper 3 (Section 4.3) addresses the first two objectives. All papers imply a validation process using computational simulations.

4.1. Paper 1: Parameter-light classifiers for vectorial slope stability data[1]

This paper addresses the slope stability problem, from geotechnical engineering, as a data-driven task using real-world data. The premise around this contribution is supported by the Occam's razor principle i.e., to choose the simplest SPR/ML method whose results are equal to or, even, better than the results reported by complex methods for the same problem. In this respect, classical but very effective SPR/ML methods were applied, assuming an incremental evaluation during the design cycle and achieving, finally, very sound results.

Two publicly available data sets of real-world slope stability were considered: the first one is a typhoon-induced slope collapsing from a region in Taiwan, called here *Taiwan* data set [25]. A description of the the main properties is shown in Table **4-1**. Also, all the data from this set are presented in Table **4-3**. It is composed by 76 instances of which 55 examples belong to stable slopes (label Y=0) and 21 ones belong to collapsed slopes (label Y=1); each instance has 15 features.

The last one is a collection from several countries of earth slope stability assessment data, called here *Multinational* data set [71]. It consists of 6 features for 84 instances of stable slopes (label Y=1) and 84 of unstable ones (label Y=0). A description of the the main properties is shown in Table **4-2** and all the data from this set are presented in Table **4-4**.

[1]This section was published as: Y.M. Ospina-Dávila and Mauricio Orozco-Alzate. Parsimonious design of pattern recognition systems for slope stability analysis in Earth Science Informatics, Springer Nature, Vol. 13, 2020, pp. 523-536. https://doi.org/10.1007/s12145-019-00429-5

Table 4-1: Main properties of the Taiwan data set. Source [25]

Factor	Feature	Min.	Max.	Average	SD
IF1	Slope aspect (°)	0.0	345.0	174.0	103.7
IF2	Slope gradient (°)	30.0	90.0	61.3	12.1
IF3	Slope height (m)	5.0	60.0	17.7	10.7
IF4	Slope form (°)	-44.9	47.7	2.4	13.8
IF5	Formation type	1.0	5.0	4.5	0.9
IF6	Angle between slope aspect and depositional trend (°)	0.0	180.0	92.9	31.3
IF7	Angle between slope gradient and stratigraphic inclination (°)	-10.0	80.0	49.9	19.2
IF8	Rock mass size (m)	0.2	2.5	0.6	0.4
IF9	Rock mass volume (%)	20.0	100.0	71.2	14.2
IF10	Vegetation coverage percentage (%)	5.0	95.0	74.1	17.0
IF11	Vegetation coverage thickness (m)	0.5	4.0	2.2	1.0
IF12	Catchment area (m2)	406.0	132,901.0	12,495.0	25,927.0
IF13	Excavation height of slope toe (m)	2.0	30.0	5.5	4.0
IF14	Variance in gradient (°)	0.0	35.0	9.6	11.1
IF15	Maximum accumulated typhoon rainfall (mm)	941.1	1947.3	1728.9	391.4

Table 4-2: Main properties of the Multinational data set. Source [71]

Factor	Feature	Min.	Max.	Average	SD
IF1	Unit weight (kN/m^3)	12.0	31.30	21.76	4.13
IF2	Soil cohesion (kPa)	0.00	300.0	34.12	45.82
IF3	Internal friction angle (°)	0.00	45.0	28.72	10.58
IF4	Slope angle (°)	16.00	59.0	36.10	10.22
IF5	Slope height (m)	3.60	511.0	104.19	132.68
IF6	Pore pressure ratio	0.00	45.0	0.48	3.45

Table 4-3: Taiwan data set. Source [25].

IF1	IF2	IF3	IF4	IF5	IF6	IF7	IF8	IF9	IF10	IF11	IF12	IF13	IF14	IF15	Y
55	65	10	33.7	3	20	0	0.6	80	70	1	1422	6	15	1947.3	0
25	65	12	9.4	2	90	65	0.7	70	85	3	1829	8	15	1947.3	0
190	50	15	-4	1	90	50	0.3	50	80	1	2134	2	0	1947.3	0
20	70	12	-2.8	5	90	70	0.8	75	85	1	5182	3	0	1947.3	0
300	55	12	-7.2	5	90	55	0.6	80	80	3.5	5690	3	0	1947.3	0
265	60	15	2.4	5	90	60	0.9	75	90	4	1422	3	0	1947.3	0
315	55	12	2.7	5	90	55	0.7	75	85	2	1290	3	0	1947.3	0
30	65	10	7.2	5	90	65	0.7	70	90	3	1475	3	0	1947.3	0
55	60	15	9.4	5	90	60	0.8	75	95	3	2120	3	0	1947.3	0
260	70	13	-7	5	90	70	0.6	75	80	2.5	16036	3	0	1947.3	0
310	65	8	13	5	90	65	0.5	70	85	3	737	4	25	1947.3	0
310	50	12	9.7	5	90	50	0.5	75	95	3	1659	2	0	1947.3	0
60	90	12	2.3	5	90	60	1.2	95	75	1.5	922	4	30	1947.3	0
25	70	12	47.7	5	180	40	1.1	95	80	1	1106	4	20	1947.3	0
100	60	11	28.3	5	90	60	0.5	70	75	2.5	1016	3	0	1947.3	0
335	65	15	8.3	5	90	35	1.2	90	80	1	2032	4	15	1947.3	0
45	70	15	-4.1	5	90	20	1.1	90	75	2	2337	5	0	1947.3	0
85	60	17	-8.1	5	90	60	0.6	70	80	2.5	5284	3	0	1947.3	0
345	70	11	8.2	5	100	40	0.8	85	80	0.5	2032	7	15	1947.3	0
90	85	11	-6.6	5	100	65	0.9	90	60	1.5	5995	5	20	1947.3	0
0	70	15	8	5	90	70	0.7	75	80	3	1626	4	0	1947.3	0
310	50	23	-18.5	5	90	50	0.8	75	80	4	3455	2	0	1947.3	0
290	65	18	11.4	5	90	65	0.9	75	85	4	1524	4	0	1947.3	0
70	65	25	-5.6	5	150	35	0.7	80	95	4	813	4	0	1947.3	0
105	65	25	-3.7	5	140	20	0.5	75	70	1.5	3861	4	0	1947.3	0
80	60	12	20.5	5	90	60	0.7	80	85	1.5	813	3	0	1947.3	0
295	70	15	9	5	150	40	0.8	80	80	1.5	2337	4	10	1947.3	0
280	60	17	-9.6	5	90	60	0.8	80	80	2	4369	4	0	1947.3	0
280	85	15	12.5	5	170	70	0.9	90	75	1	4267	9	15	1947.3	0
250	70	20	15.6	4	100	55	0.8	85	85	1	2337	9	15	1947.3	0
245	30	18	4.6	5	90	30	0.3	50	70	2.5	2032	3	0	1947.3	0
230	30	17	-7.1	5	90	30	0.3	50	70	2.5	2845	3	0	1947.3	0
50	50	15	9.8	5	90	50	0.2	50	90	3	3861	4	0	1947.3	0
130	50	12	9	5	0	-10	0.7	80	70	0.5	1931	5	0	1947.3	0
130	60	15	-4.3	5	90	60	0.5	75	80	1	12802	5	0	1947.3	0
260	55	7	17.4	5	90	55	0.5	70	85	2.5	813	4	15	1947.3	0
125	50	17	-6.6	5	90	50	0.3	70	85	3	3658	4	0	1947.3	0
290	65	6	4.5	5	90	65	0.3	60	80	2.5	1422	4	20	1947.3	0
200	90	20	2.1	5	125	80	2.3	97	60	1	2337	10	15	1947.3	0
125	70	13	5.5	5	110	55	0.7	75	80	1	2134	8	15	1947.3	0
80	70	10	15.9	5	90	70	0.5	70	70	3	1626	2	35	1947.3	0
290	50	13	9.1	5	90	50	0.4	70	80	1	1219	5	15	1947.3	0
245	70	7	-4.9	5	90	70	0.5	75	65	1	1727	6	30	1947.3	0
190	55	15	-8.4	5	90	55	0.6	75	65	1	2134	10	20	1947.3	0
225	60	6	12.9	5	90	60	0.7	75	60	1	2743	5	20	1947.3	0
200	60	12	13.8	5	90	60	0.7	75	30	0.5	1727	7	20	1947.3	0
325	60	6	9.1	5	120	10	1.2	85	70	3.5	1931	4	35	1947.3	0
335	60	5	8	5	90	60	0.6	75	90	4	3455	3	30	1947.3	0
215	60	9	19.4	5	90	60	0.6	75	65	3	914	8	25	1947.3	0
215	90	7	-9.7	5	90	60	2.5	100	55	1.5	711	5	30	1947.3	0
225	60	6	17.4	5	90	60	0.7	75	70	3	2642	4	25	1947.3	0
255	60	5	-16.7	5	90	60	0.5	70	80	3	38204	4	25	1947.3	0
265	45	5	29	5	90	45	0.8	75	55	3	914	4	25	1947.3	0
270	45	10	-22	5	90	45	0.5	70	80	3	24081	7	15	1947.3	0
120	60	20	1.5	3	0	10	0.3	65	80	2	12396	5	0	952.4	1

40	50	20	-9.8	3	90	50	0.4	65	80	2	20118	4	0	942.4	1
35	65	25	-5.2	3	90	65	0.2	55	75	1.5	3150	15	15	950.6	1
110	60	25	-8.6	3	10	10	0.5	90	80	2.5	9754	3	0	941.1	1
165	70	20	16.3	3	60	20	0.3	40	75	1.5	406	12	10	968	1
95	70	45	-6.3	5	170	60	0.3	65	60	1	12599	8	5	1153	1
95	75	60	-44.9	5	120	55	0.3	60	30	1	55695	13	0	1383.9	1
160	50	50	-13.2	5	90	50	0.2	50	80	3.5	74797	5	0	1393.8	1
190	70	50	-3	5	90	65	0.3	75	60	1	1626	30	10	1505.7	1
240	40	25	-22	5	90	40	0.6	50	90	4	97373	4	0	1733	1
0	70	25	-26.6	5	90	70	0.7	60	80	3	112682	9	30	1918.8	1
185	80	23	15	5	180	55	0.7	80	5	0.5	3048	14	20	1947.3	1
330	70	30	0	3	90	70	0.3	65	90	2.5	27332	2	0	958.7	1
100	40	25	0	3	50	0	0.3	70	80	3.5	15343	2	0	967.9	1
40	60	30	0	3	90	60	0.3	60	80	2	3353	5	10	967.9	1
30	60	20	0	3	90	60	0.3	65	80	2	2134	5	10	967.9	1
175	60	25	0	3	50	10	0.3	50	80	2	11380	7	0	949.2	1
0	60	25	0	2	90	60	0.2	50	85	3.5	7925	8	0	953.5	1
215	45	30	0	3	90	45	0.2	20	20	1.5	80574	6	0	987.8	1
150	50	40	0	5	90	50	0.2	40	20	2	47552	3	0	987.7	1
275	40	20	0	5	90	40	0.6	75	80	3	132901	4	0	984.8	1

Table 4-4: Multinational data set. Source [71].

IF1	IF2	IF3	IF4	IF5	IF6	Y
18.68	26.34	15	35	8.23	0	0
16.5	11.49	0	30	3.66	0	0
18.84	14.36	25	20	30.5	0	1
18.84	57.46	20	20	30.5	0	1
28.44	29.42	35	35	100	0	1
28.44	39.23	38	35	100	0	1
20.6	16.28	26.5	30	40	0	0
14.8	0	17	20	50	0	0
14	11.97	26	30	88	0	0
25	120	45	53	120	0	1
26	150.05	45	50	200	0	1
18.5	25	0	30	6	0	0
18.5	12	0	30	6	0	0
22.4	10	35	30	10	0	1
21.4	10	30.34	30	20	0	1
22	20	36	45	50	0	0
22	0	36	45	50	0	0
12	0	30	35	4	0	1
12	0	30	45	8	0	0
12	0	30	35	4	0	1
12	0	30	45	8	0	0
23.47	0	32	37	214	0	0
16	70	20	40	115	0	0
20.41	24.9	13	22	10.67	0.35	1
19.63	11.97	20	22	12.19	0.41	0
21.82	8.62	32	28	12.8	0.49	0
20.41	33.52	11	16	45.72	0.2	0
18.84	15.32	30	25	10.67	0.38	1
18.84	0	20	20	7.62	0.45	0
21.43	0	20	20	61	0.5	0
19.06	11.71	28	35	21	0.11	0
18.84	14.36	25	20	30.5	0.45	0

21.51	6.94	30	31	76.81	0.38	0
14	11.97	26	30	88	0.45	0
18	24	30.15	45	20	0.12	0
23	0	20	20	100	0.3	0
22.4	100	45	45	15	0.25	1
22.4	10	35	45	10	0.4	0
20	20	36	45	50	0.25	0
20	20	36	45	50	0.5	0
20	0	36	45	50	0.25	0
20	0	36	45	50	0.5	0
22	0	40	33	8	0.35	1
24	0	40	33	8	0.3	1
20	0	24.5	20	8	0.35	1
18	5	30	20	8	0.3	1
26.49	150	33	45	73	0.15	1
26.7	150	33	50	130	0.25	1
26.89	150	33	52	120	0.25	1
26.57	300	38.7	45.3	80	0.15	0
26.78	300	38.7	54	155	0.25	0
26.81	200	35	58	138	0.25	1
26.43	50	26.6	40	92.2	0.15	1
26.7	50	26.6	50	170	0.25	1
26.8	60	28.8	59	108	0.25	1
22.4	10	35	45	10	0.4	0
20	20	36	45	50	0.5	0
20	0	36	45	50	0.25	0
20	0	36	45	50	0.5	0
22	0	40	33	8	0.35	1
20	0	24.5	20	8	0.35	1
27	40	35	43	420	0.25	0
27	50	40	42	407	0.25	1
27	35	35	42	359	0.25	1
27	37.5	35	37.8	320	0.25	1
27	32	33	42.6	301	0.25	0
27	32	33	42.4	289	0.25	1
27.3	14	31	41	110	0.25	1
27.3	31.5	29.7	41	135	0.25	1
27.3	16.8	28	50	90.5	0.25	1
27.3	26	31	50	92	0.25	1
27.3	10	39	41	511	0.25	1
27.3	10	39	40	470	0.25	1
25	46	35	47	443	0.25	1
25	46	35	44	435	0.25	1
25	46	35	46	432	0.25	1
26	150	45	30	200	0.25	1
18.5	25	0	30	6	0.25	0
18.5	12	0	30	6	0.25	0
22.4	10	35	30	10	0.25	1
21.4	10	30.34	30	20	0.25	1
25	46	35	46	393	0.25	1
25	48	40	49	330	0.25	1
31.3	68.6	37	47	305	0.25	0
25	55	36	45.5	299	0.25	1
31.3	68	37	47	213	0.25	0
18.66	26.41	14.99	34.98	8.2	0	0
28.4	29.41	35.01	34.98	100	0	1
25.96	150.05	45	49.98	200	0	1

18.46	25.06	0	30	6	0	0
21.36	10.05	30.33	30	20	0	1
15.99	70.07	19.98	40.02	115	0	0
20.39	24.91	13.01	22	10.6	0.35	1
19.6	12	19.98	22	12.2	0.41	0
21.78	8.55	32	27.98	12.8	0.49	0
20.39	33.46	10.98	16.01	45.8	0.2	0
19.03	11.7	27.99	34.98	21	0.11	0
17.98	4.95	30.02	19.98	8	0.3	1
20.96	19.96	40.01	40.02	12	0	1
20.96	34.96	27.99	40.02	12	0.5	1
19.97	10.05	28.98	34.03	6	0.3	1
18.77	30.01	9.99	25.02	50	0.1	1
18.77	30.01	19.98	30	50	0.1	1
18.77	25.06	19.98	30	50	0.2	0
20.56	16.21	26.51	30	40	0	0
16.47	11.55	0	30	3.6	0	0
18.8	14.4	25.02	19.98	30.6	0	1
18.8	57.47	19.98	19.98	30.6	0	1
28.4	39.16	37.98	34.98	100	0	1
13.97	12	26.01	30	88	0	0
24.96	120.04	45	53	120	0	1
18.46	12	0	30	6	0	0
22.38	10.05	35.01	30	10	0	1
21.98	19.96	36	45	50	0	0
18.8	15.31	30.02	25.02	10.6	0.38	1
18.8	14.4	25.02	19.98	30.6	0.45	0
21.47	6.9	30.02	31.01	76.8	0.38	0
13.97	12	26.01	30	88	0.45	0
17.98	24.01	30.15	45	20	0.12	0
22.38	99.93	45	45	15	0.25	1
22.38	10.05	35.01	45	10	0.4	0
19.97	19.96	36	45	50	0.25	0
19.97	19.96	36	45	50	0.5	0
20.96	45.02	25.02	49.03	12	0.3	1
20.96	30.01	35.01	40.02	12	0.4	1
19.97	40.06	30.02	30	15	0.3	1
17.98	45.02	25.02	25.02	14	0.3	1
18.97	30.01	35.01	34.98	11	0.2	1
19.97	40.06	40.01	40.02	10	0.2	1
18.83	24.76	21.29	29.2	37	0.5	0
18.83	10.35	21.29	34.03	37	0.3	0
18.77	25.06	9.99	25.02	50	0.2	0
18.77	19.96	9.99	25.02	50	0.3	0
19.08	10.05	9.99	25.02	50	0.4	0
18.77	19.96	19.98	30	50	0.3	0
19.08	10.05	19.98	30	50	0.4	0
21.98	19.96	22.01	19.98	180	0	0
21.98	19.96	22.01	19.98	180	0.1	0
20.41	33.52	11	16	45.7	0.2	0
18.84	0	20	20	7.62	0.45	0
19.06	11.7	28	35	21	0.11	0
18.84	14.36	25	20	30.5	0.45	0
14	11.97	26	30	88	0.45	0
18	24	30.15	45	20	0.12	0
22.4	10	35	45	10	0.4	0
20	20	36	45	50	0.5	0

22.4	100	45	45	15	0.25	1
27	50	40	42	407	0.25	1
31.3	68	37	46	366	0.25	1
27	35	35	42	359	0.25	1
27	37.5	35	38	320	0.25	1
27	32	33	42	289	0.25	1
27	14	31	41	110	0.25	1
27	31.5	29.7	41	135	0.25	1
27	16.8	28	50	90.5	0.25	1
27	26	31	50	92	0.25	1
27	10	39	41	511	0.25	1
27	10	39	40	470	0.25	1
25	46	35	47	443	0.25	1
20	20	36	45	50	0.25	0
19.63	11.97	20	22	21.19	0.4	0
25	55	36	44	299	0.25	1
27.3	10	39	40	480	0.25	1
25	46	35	46	393	0.25	1
16.5	11.49	0	30	3.66	0	0
25	120	45	53	120	0	1
19.06	11.75	28	35	21	0.11	0
18.84	14.36	25	20.3	50	45	0

4.1.1. Aim

Most of state-of-art methods applied in data-driven slope stability analysis, from a SPR/ML perspective, are highly parameter-laden algorithms which imply, in many cases, unclear and knowledge-dependent tunning processes. Examples are (deep-)neural networks or (deep-)Kernel methods whose tunning processes often require a large number of parameters to adjust. In contrast, this paper proposes an incremental parsimonious path, starting with a previous data visualization step in order to check data distribution in a two-dimensional space; techniques such as the Principal Component Analysis (PCA) and the Multidimensional Scaling (MDS) were applied to the original data and to the distance matrix computed from them, respectively. The main goal of the data visualization is to define a roadmap for the learning and generalization phase in the design of a SPR/ML system (see Fig.**1-4**), such that the designer is able to choose a proper classification technique according to the problem requirements that is, to be sure that the solution is not more complex than the problem. The aim of the classifier is then to categorize the slope state in either stable or collapsed.

According to this motivation, this paper seeks a *parameter-light* approach, considering that most geotechnical and geological engineers do not have practical knowledge on statistical/machine learning, particularly regarding validation or tunning issues, and using simple and classical SPR/ML methods would require very few (hyper-)parameters to adjust.

4.1.2. Methods

In this paper, three steps can be distinguished: first, a data normalization, next, a data visualization and, finally, the training phase of the classifiers. For the data normalization, we select the z-score normalization procedure, where for each feature, all the instances are offset-corrected by the mean and scaled by the standard deviation. This mean and standard deviation are computed from the training set and applied to both, the training set and the test set.

After that, two different dimensionality reduction techniques are used: PCA which is a linear method and MDS which is a non-linear method that works with the matrix of pairwise distances. This data visualization process enables us a bottom-up design of a parameter-light SPR/ML system. In particular, the following classifiers were trained: the nearest neighbor rule (1-NN), the k-nearest neighbor rule (k-NN), the linear discriminant (Bayes) classifier (LDC), the quadratic discriminant (Bayes) classifier (QDC), the Parzen classifier and an automatic neural network classifier (AANN). Here, 1-NN does not require any free-parameter tunning, while the AANN requires at least three of them.

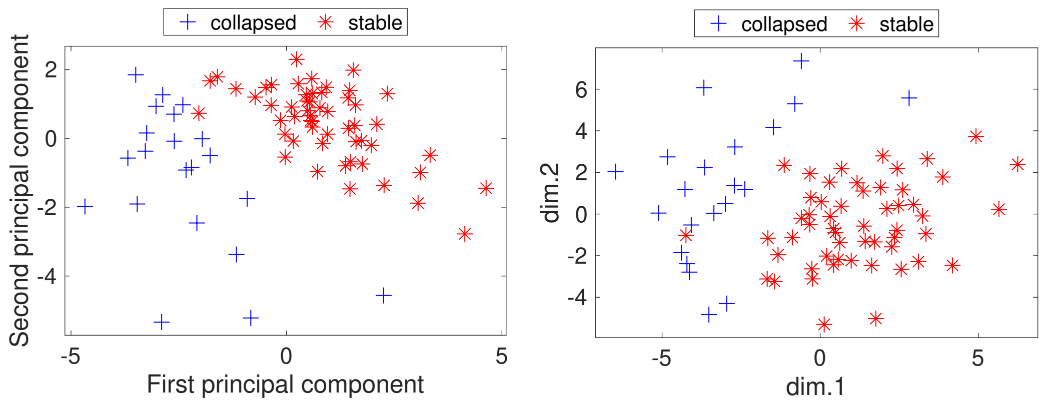
Furthermore, for all the experiments, the very same setup that was used in [25] and [71] was considered.

4.1.3. Results

The experiments show that the Taiwan data set has a sharped separability between stable and collapsed data, see for example the data distribution in the two-dimensional PCA and MDS space (Figure 4-1). For this data set, high accuracies were obtained for both, the original 15-dimensional space and the two-dimensional PCA space, which is consistent with the visualization results (see Figure 4-2).

In other words, this data set indicates that parameter-light classifiers should, at first, address the classification problem without sophisticated and (hyper-)parameter SPR/ML systems. Notice also that, in Figure 4-2(a), the accuracy of a very advanced parameter-laden classifier was included, which uses an artificial bee colony optimization for SVM, called BeeSVC, developed in [25].

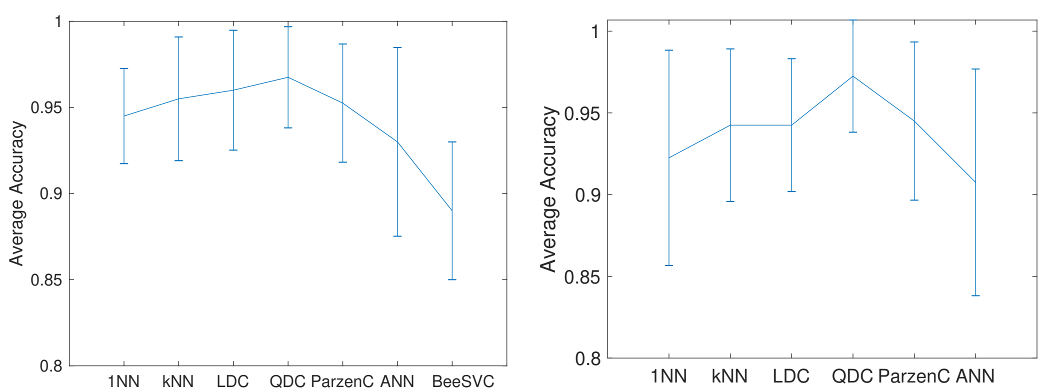
On the other hand, the Multinational data set does not show a sharped separability between stable and collapsed data (Figure 4-3), however, parameter-light classifiers such as 1-NN and k-NN achieve competitive results compared to state-of-art SPR/ML methods proposed in the literature. For this experiment, the accuracy of a very advanced parameter-laden classifier was included, which uses a metaheuristic-optimization tunning for a least squares support vector classifier (MO-LSVC), proposed in [71].



(a) Projections onto the first two principal components.

(b) Multidimensional scaling projections.

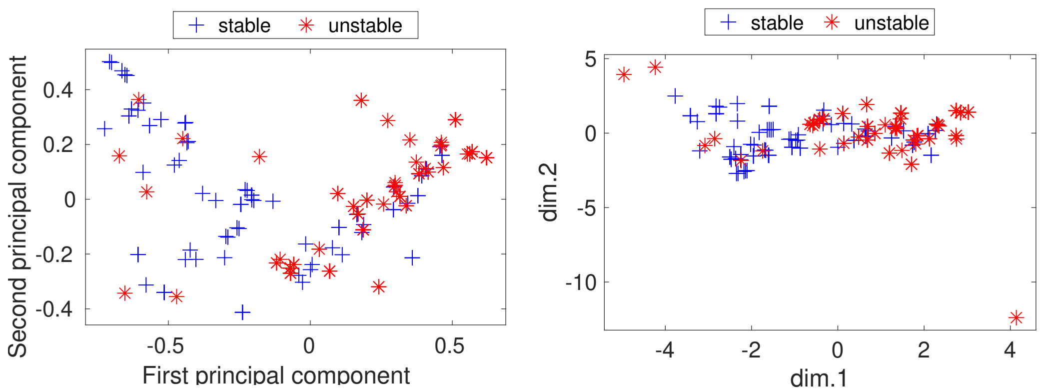
Figure 4-1: Taiwan data set. Data visualizations using PCA and MDS.



(a) Accuracies in the original 15-dimensional space.

(b) Accuracies in a two-dimensional space (after applying PCA).

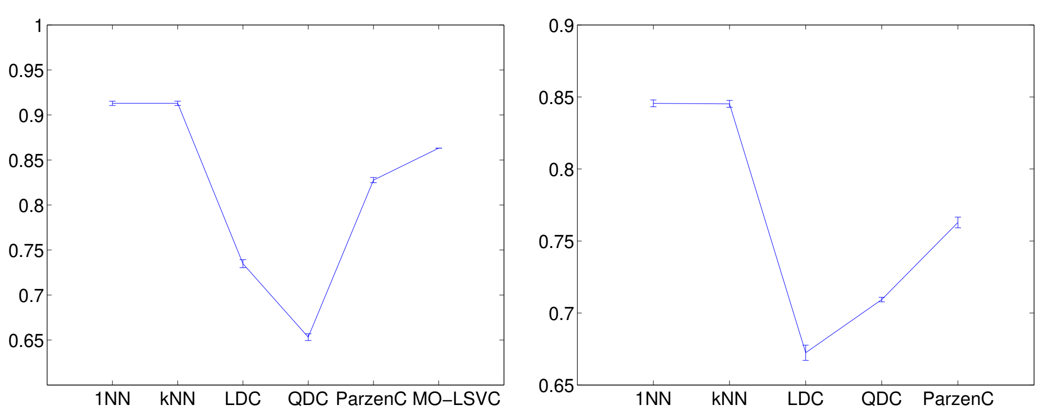
Figure 4-2: Taiwan data set. Classification accuracies, along with standard deviations for 20 runs.



(a) Projections onto the first two principal components.

(b) Multidimensional scaling projections.

Figure 4-3: Multinational data set. Data visualizations using PCA and MDS.



(a) Accuracies in the original 15-dimensional space.

(b) Accuracies in a two-dimensional space (after applying PCA).

Figure 4-4: Multinational data set. Classification accuracies, along with standard deviations for 20 runs.

4.2. Paper 2: Enriched and enhanced parameter-light classifiers for vectorial slope/landslide stability data[2]

This paper is an extension of the previous study referenced in Section 4.1 but still preserving the main goal is *of keeping the models as simple as possible* in the design of SPR/ML systems applied to data-based slope stability analysis and landslide prediction. With this in mind, and supported by the visualization findings from the previous study, this paper explores the importance of enriching the data representation —instead of overcomplicating the classification rule itself via (non-)metric distance learning— for vectorial slope/landslide stability data.

The *Taiwan* data set [25] and the *Multinational* data set [71] were used in this paper. In addition, a rainfall-induced landslide susceptibility data set, called the *Yongxin* data set, was included in order to extend the scope of the computational experiments. It is released on a companion repository (see the URL in Ref. [48]), which is located at the western part of Jiangxi Province, China (see Fig. **4-5**). In this data set, 16 factors were included whose graphical distribution over *Yongxin* area can be seen in Fig. **4-6** and **4-7**, among which are the following: altitude (m), aspect, distance to faults (m), land use, lithology, Normalized Difference Vegetation Index (NDVI), plan curvature, profile curvature, rainfall (mm/y), distance to rivers (m), distance to roads (m), slope, type of soil, Stream Power Index (SPI), Sediment Transport Index (STI) and Topographic Wetness Index (TWI); for further explanation and details, see [48]. Also, this data set is composed by 728 examples, where one half corresponds to landslide case (label Y=1) and the other half corresponds to non-landslide case (label Y=0).

4.2.1. Aim

Originally, the data that involve a data-based slope/landslide prediction analysis are given as measured parameters, that means, they require a SPR/ML system design directly based on features. With this *feature-based representation* given in advance, two purposes related to computational learning could be considered in order to boost the classification performance: (1) to enhance the data representation or (2) to adapt the metric employed. In this paper, these goals are subject to a *interpretable* view of point, which enables potential experts —from geosciences or geotechnical areas— to evaluate the results without tricky tunning processes.

With the aim of designing parameter-light classifiers, this paper encloses the experimental

[2]This section is under review in the Mathematical Geosciences journal, Springer Nature, as: Y.M. Ospina-Dávila and Mauricio Orozco-Alzate. Enriching representation and enhancing nearest-neighbor classification of slope/landslide data by using rectified-feature-line segments and hypersphere-based scalings: A reproducible experimental comparison.

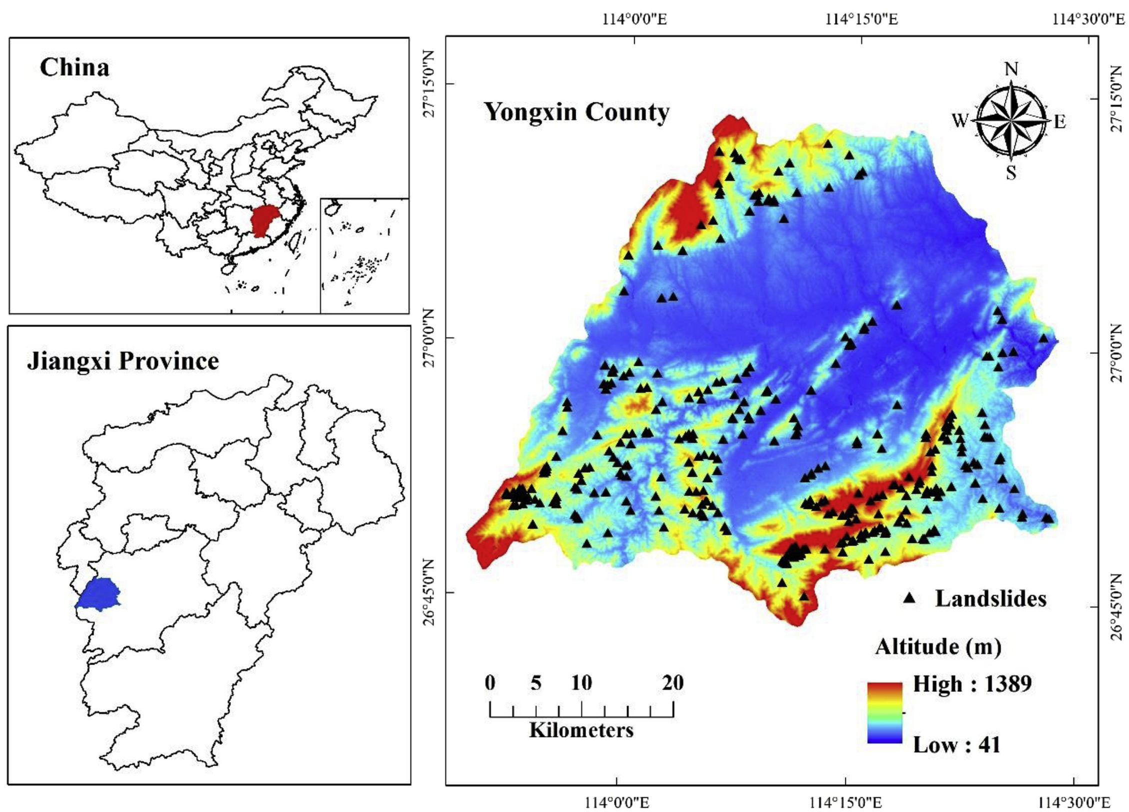


Figure 4-5: Yongxin area. Source [48] (permission granted by Elsevier Ltd. by order number 5333851293554)

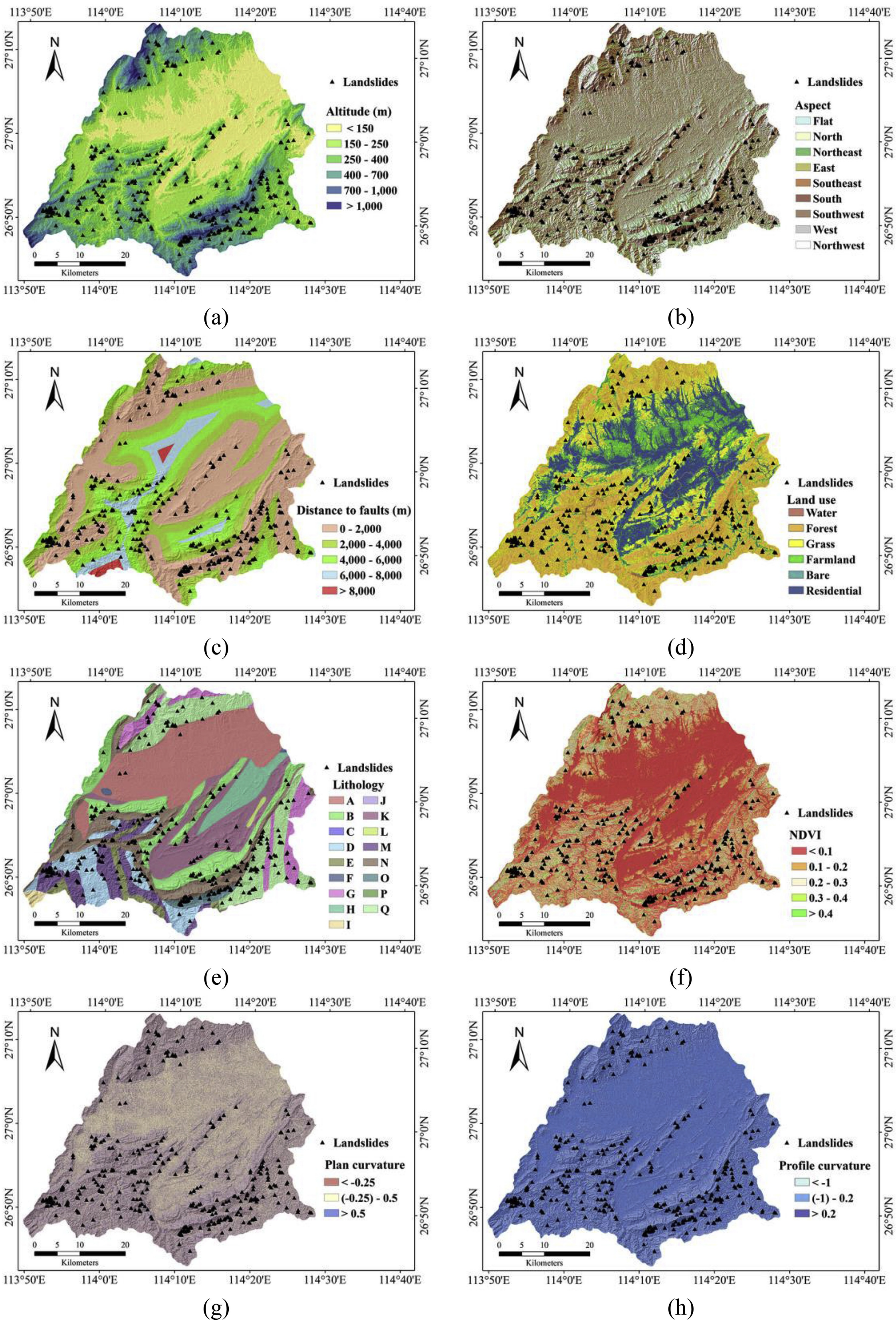


Figure 4-6: Yongxin area: landslide factors. Source [48] (permission granted by Elsevier Ltd. by order number 5333851293554)

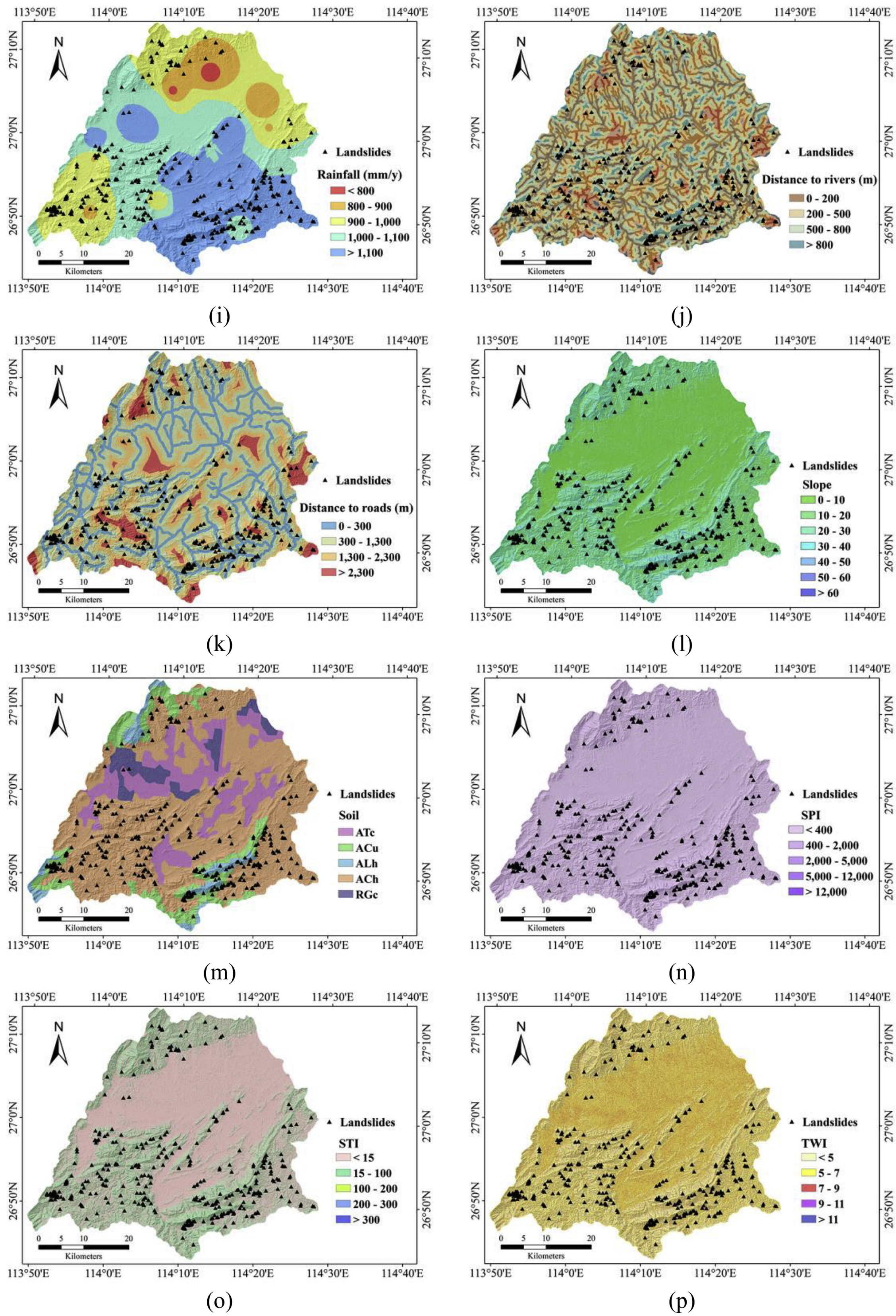


Figure 4-7: Yongxin area: landslide factors (Continuation). Source [48] (permission granted by Elsevier Ltd. by order number 5333851293554)

comparison into the NN rule family i.e., starting from its basic formulation until using advanced NN classifiers which, in general, require far few (hyper-)parameters to adjust. Thus, just classification methods from this family are compared each other. Moreover, the experimental setup and results are under a very reproducible framework in order to enable comparisons between methods or algorithms.

4.2.2. Methods

Considering the NN classifier as a parameter-light classifier, this paper relies on a simple, but powerful, extension of this one that belongs to the family of the so-called nearest feature classifiers known as RNFLS [37] (see Section 2.2) and that pursues enriching the representation by using linear interpolations and extrapolations. Conversely, adaptive (non-)metric distance-learning strategies, such as the Hypersphere classifier (HC) [97] or the Adaptive Nearest Neighbor (ANN) classifier [169] are applied when trying to enhance the classification via (non-)metric distance learning.

4.2.3. Results

The results are based on two slope stability data sets and another one that belongs to landslide susceptibility. In addition, as explained above, the experiments were restricted to the NN familiy of classifiers, namely: 1-NN, ANN, HC and RNFLS. According to the results, the most important insightful conclusion is that, for this slope/landslide data, it is more useful to focus computational learning strategies on enhancing data representation than on adaptive (non-)learning distance procedures. In this aspect, the RNFLS method achieves the best accuracy performance for all the data sets and the most noticeable improvement over the baseline 1-NN classifier; see Table **4-5**.

Table 4-5: Classification accuracies. Estimations for *Taiwan* and *Multinational* were made with leave-one-out; 5-fold cross-validation was used for *Yongxin*.

	Classifiers			
Dataset	1-NN	ANN	HC	RNFLS
Taiwan	0.9333	0.9467	0.9467	**0.9600**
Multinational	0.9286	0.9286	0.9167	**0.9405**
Yongxin	0.6689	0.7033	0.7005	**0.7225**

From a SPR/ML perspective, these results is consistent with many research literature in geotechnical/natural-hazard engineering, where is pointed out that most of data sets are scarce; in other words, these problems can be considered as *learning from small-data* (see Section 1.2).

4.3. Paper 3: Parameter-light classifiers for non-vectorial proximity SHM data[3]

Data-based Structural Health Monitoring (SHM), from a SPR/ML perspective, commonly involves sophisticated and, in many cases, hyper-parameterized classifiers which require a domain-dependent and, quite often, non-trivial feature extraction preprocessing step, on which the classifier performance is very dependent. Consequently, when the features extracted are not discriminating enough, on the one hand, advanced state-of-the-art machine learning methods, such as ensemble learning methods, deep neural networks, etc., are proposed or, on the other hand, unrealistic settings are adopted, such as classification frameworks which use simultaneously data from damaged and undamaged conditions. An alternative linked to the notion of proximity —when non-vectorial data is originally measured— is the so-called dissimilarity representation. This paper is dedicated to this approach when considered for SHM.

In fact, this proposal offers a suitable framework for the design of highly competitive data-based SHM methods, from a SPR/ML perspective, considering a bottom-up strategy subject to real-world settings. Furthermore, this one can be considered as an important step towards a *featureless parameter-free* data-based SHM approach.

Two well-known publicly available data sets are used in order to validate the results: the three-storey building structure behaving nonlinearly from Los Alamos Laboratory [59] and a large-scale grandstand simulator from Qatar University [3].

4.3.1. Aim

The data measured in SHM are, in general, non-vectorial, e.g., images from (un)damaged conditions or time series of structural accelerations or displacements, among others. In this sense, this paper seeks a proper representation of these non-vectorial vibration data and their right embedding in vector spaces. In particular, a novel framework was developed for building dissimilarity-vector spaces from spectral/time-frequency information that, in turn, comes from data-based SHM. This framework includes several alternatives of spectral estimation, in particular the multitaper spectral density estimator, coupled with the DTW distance, resulting in a real-world SPR pipeline that in its last step makes use of classification techniques to predict the structural damage. In addition, we used only undamaged data during training which is not just more realistic but also highlights a challenging SHM setting. This was addressed by using an one-class classification approach i.e., classifiers which learn —during training— a decision boundary around the undamaged class in absence of damaged

[3]This section was published as: Y.M. Ospina-Dávila and Mauricio Orozco-Alzate. Dissimilarity-vector spaces based on Dynamic Time Warpings of spectral/time-frequency information for structural health monitoring in Computers & Structures, Elsevier, Vol. 263, 2022, pp. 106754. https://doi.org/10.1016/j.compstruc.2022.106754

data, enclosing the first ones by means of (hyper-)spheres or (hyper-)planes, among others shapes, in order to properly detect unseen (un-)damaged data.

It is important to note that building dissimilarity-based vector spaces is not a pre-designed task, moreover, capturing the most discriminative information by using an appropriate dissimilarity or distance measure neither is self-evident. As mentioned above, this problem turns even more challenging by the fact that only undamaged data is utilized during the training phase, resulting in a more real-world context for SHM.

4.3.2. Methods

The dissimilarity representation addressed in this paper, for the data-based SHM problem, is mainly based on the following steps: from the spectral/time-frequency information obtained for each one of the measured signals —using the Multitaper power spectral density estimation, the Coherence function, the Short-Time Fourier Transform (STFT) and the Multitaper spectrograms— its corresponding dissimilarity-vector space is built up for each spectral/time frequency method and using the DTW distance as dissimilarity measure. In these vector spaces, the following one-class classifiers are trained: the nearest neighbor based data description (NNDD), the k-nearest neighbor data description (k-NNDD), the Parzen density data description (parzenDD), the support vector data description (SVDD) and the minimum spanning tree data description (mstDD).

4.3.3. Results

As mentioned before, two data sets were used in this paper. The first one, the three-storey building structure behaving nonlinearly from Los Alamos Laboratory [59] (see Figure **4-8**), is a benchmark vibration test that comprises 17 different structural state conditions, each of them composed by 50 measurements as listed in Table **4-6**. The second one, the Qatar University (QU) grandstand simulator [3, Figure **4-9**], is a large laboratory stadium structure with 30 accelerometers located in its joints.

These two publicly available data sets are used to validate the *featureless parameter-light* data-based SHM approach proposed here, which is endowed with a powerful discriminant ability using one-class classifiers, obtaining thus sound results, under very challenging conditions. Consider, for the Three-storey building structure data set, the Multitaper power spectral density along with the DTW distance results (see Figure **4-10**). If you choose, for example, a fraction of 15 % (or 100 %) representation objects extracted from the training set, the minimum spanning tree (`mst`) classifier trained in the dissimilarity space achieves an error of 0,0 % for both, target rejected error and the outlier accepted error.

A similar situation occurs in the Qatar University (QU) grandstand data set case, for the same classifier and for the following fractions of representation objects from the training set: $10\,\%, 15\,\%, 25\,\%, 45\,\%, 80\,\% - 95\,\%$ (see Figure **4-11**). Notice that the accuracy of our

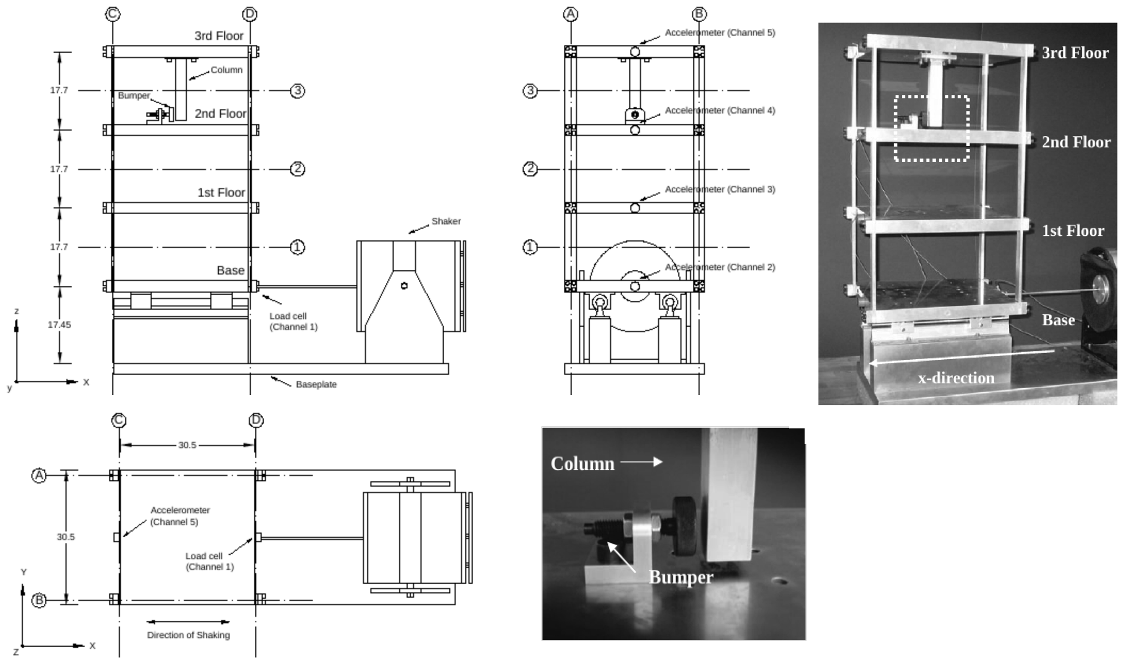


Figure 4-8: Three-storey building structure and shaker. Source [59].

Table 4-6: Structural state conditions from three-storey building. Source [59].

Label	Condition	Description
State01	Undamaged	Added mass of 1,2 kg at the 1st floor
State02	Undamaged	Added mass of 1,2 kg at the base
State08	Damage	Gap (0,13 mm)
State09	Damage	Gap (0,10 mm)
State10	Damage	Gap (0,05 mm)
State11	Damage	Gap (0,15 mm)
State12	Damage	Gap (0,20 mm)
State13	Undamaged	Baseline condition
State14	Damage	Gap (0,20 mm) and mass of 1,2 kg at the 1st floor
State15	Damage	Gap (0,10 mm) and mass of 1,2 kg at the 1st floor
State16	Damage	Gap (0,20 mm) and mass of 1,2 kg at the base
State17	Undamaged	Stiffness reduction in column 1BD
State18	Undamaged	Stiffness reduction in column 1AD and 1BD
State21	Undamaged	Stiffness reduction in column 3BD
State22	Undamaged	Stiffness reduction in column 3AD and 3BD
State23	Undamaged	Stiffness reduction in column 2AD and 2BD
State24	Undamaged	Stiffness reduction in column 2BD

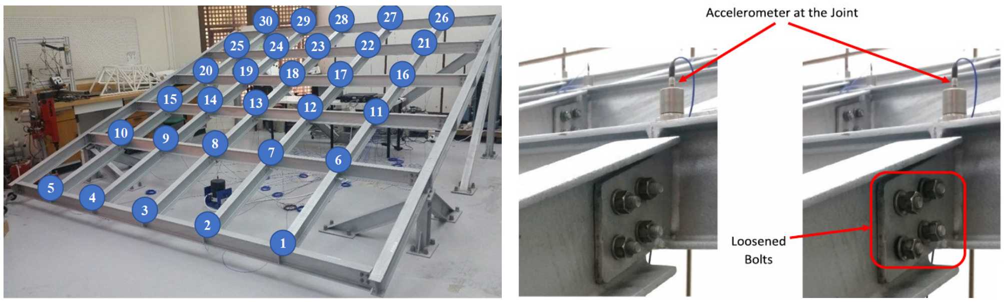
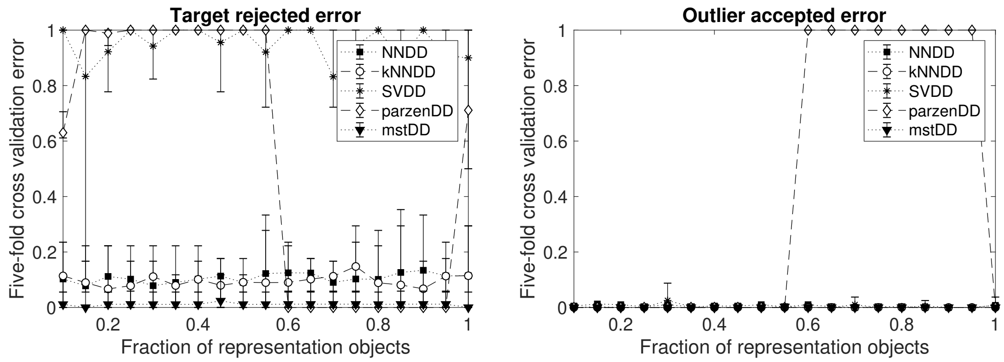
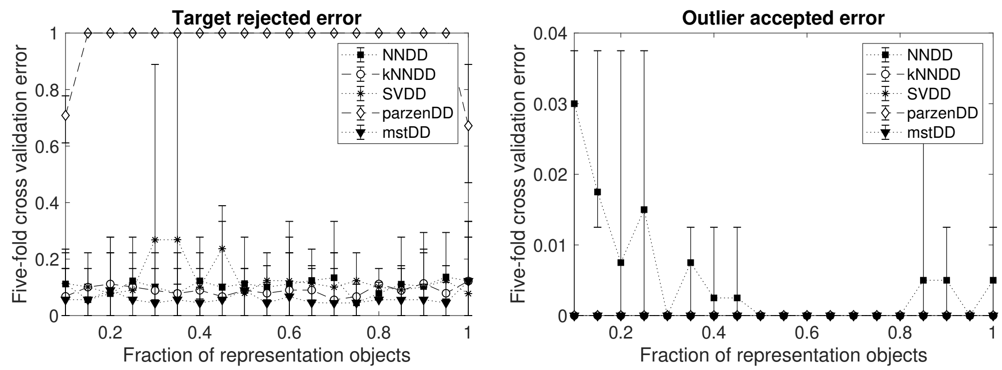


Figure 4-9: The Qatar University (QU) grandstand simulator. Source https://www.structuraldamagedetection.com/benchmark/damage/

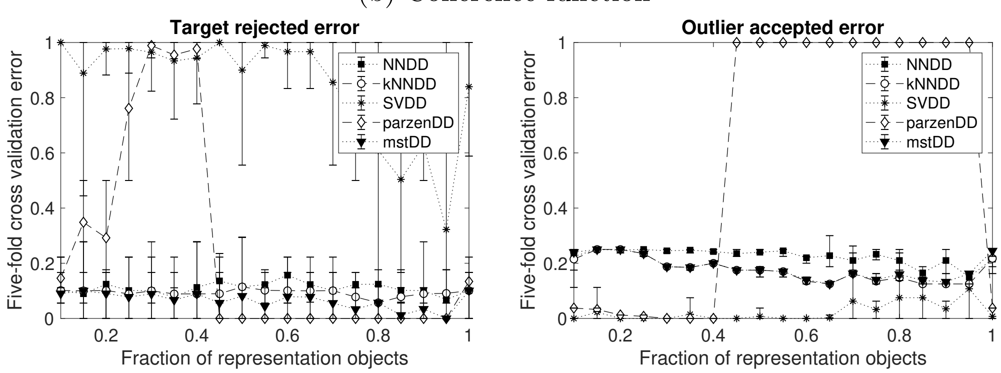
framework is competitive with respect to those of the state-of-art methods.



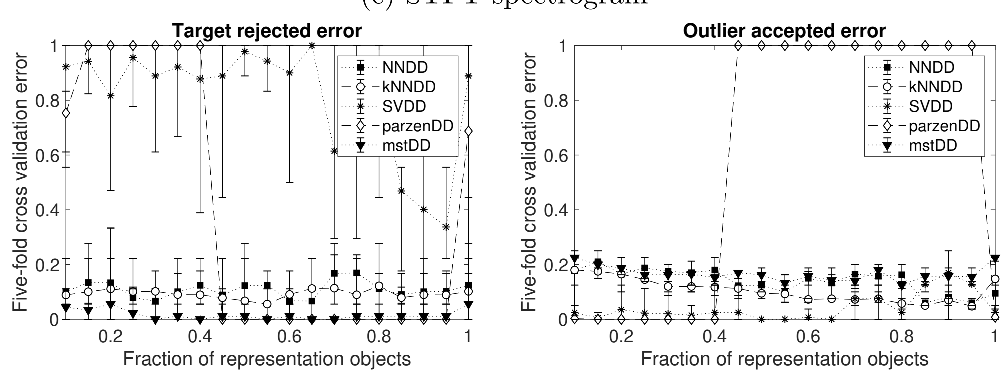
(a) Multitaper power spectral density



(b) Coherence function

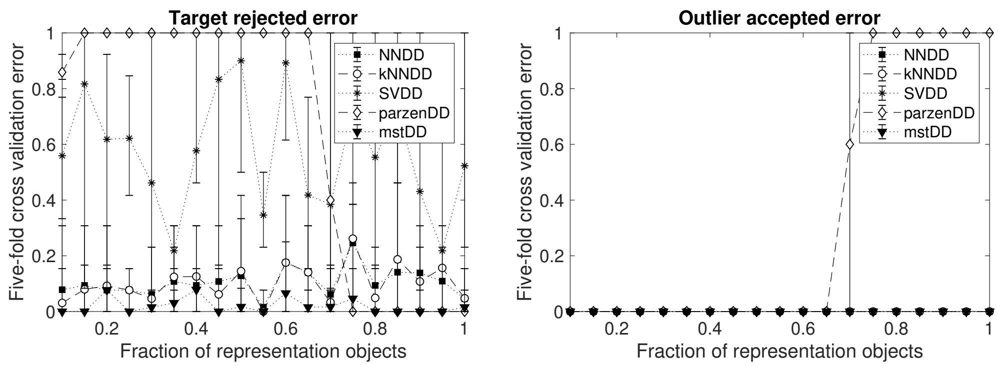


(c) STFT spectrogram

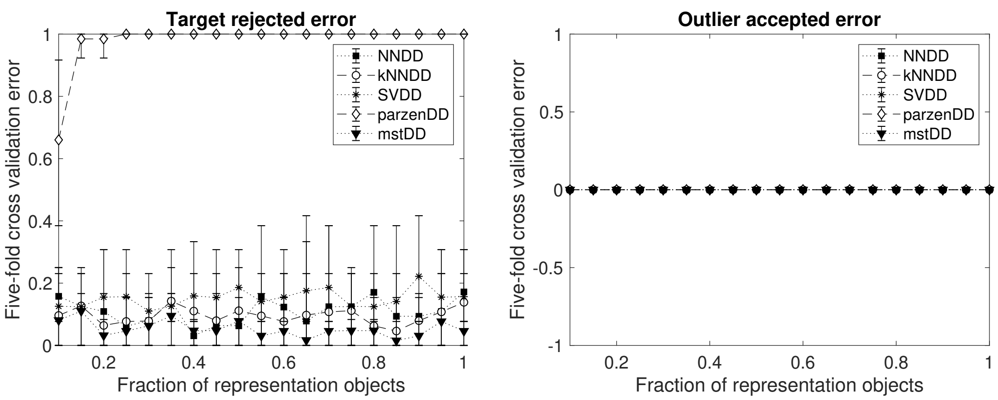


(d) Multitaper spectrogram

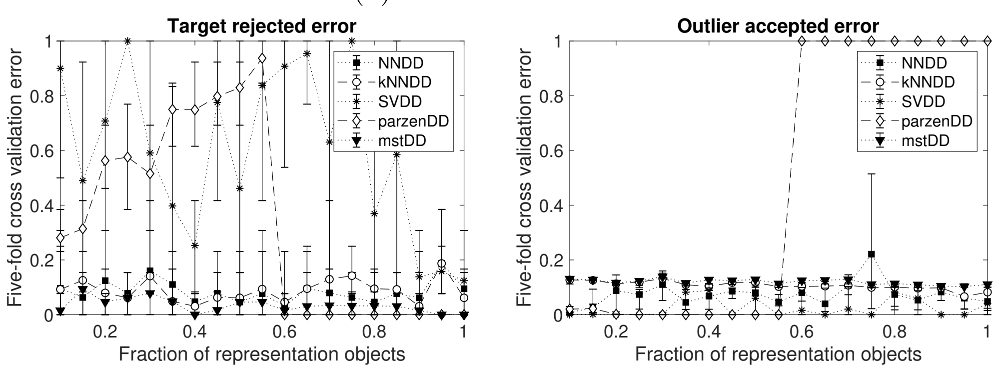
Figure 4-10: Three-storey building structure data set. Performance of several one-class classifiers.



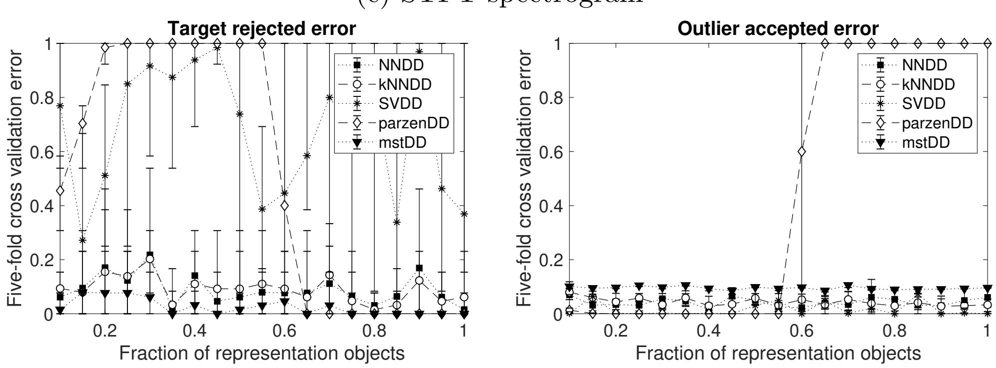
(a) Multitaper power spectral density



(b) Coherence function



(c) STFT spectrogram



(d) Multitaper spectrogram

Figure 4-11: The Qatar University (QU) grandstand data set. Performance of several one-class classifiers.

5 Conclusions and future work

This thesis addressed, on the one hand, the slope stability analysis and the landslide susceptibility assessment and, on the other hand, the SHM task, both from a data-driven perspective using SPR/ML techniques. Recent studies confirm the importance of this perspective on geotechnical/natural-hazard engineering, recognizing it as a distinctive research field under the name of 'data-centric geotechnics' [127, 125]. Likewise, for the SHM case, a 'learning from data' approach is highly appreciated [19]. This perspective is of major importance when assessments and predictions for complex engineering systems should be executed in a on-line or real-time framework. Furthermore, the results obtained in this thesis deliver state-of-art performance dealing with highly complex and uncertain engineering systems. In spite that advanced simulation methods are available for these kind of engineering decisions, such as the Finite Element Method (FEM) or meshless methods, these ones are time consuming, expensive and difficult to calibrate.

In the following, the concluding remarks and insights in each of these directions will be presented, where relatively low and imbalanced sample sizes are the common restriction.

5.1. Concluding Remarks

- In contrast to advanced but heavily (hyper-)parameter-laden SPR/ML systems such as (deep-)NN/Kernel methods, the ones designed in this thesis are parameter-light. In general, this framework allows simple and interpretable systems that can be useful in a practical context. Indeed, this parameter-light approach enables comparisons across methods or to reproduce results.

- Even though (deep-)NN/Kernel methods are powerful tools for designing data-driven models in engineering, they require large volume of training data for each class, in order to guarantee the accuracy of the SPR/ML system. In contrast, for a real-world context, slope/landslide and SHM are —typically— sparse, noisy, and poorly balanced. Under these conditions, the SPR/ML systems proposed in this thesis are appropriate and competitive.

- For the slope stability analysis, a bottom-up design is crucial for SPR/ML implementations; in fact, the parsimonious design of SPR/ML systems for slope stability proposed here showed that classical classifiers could properly solve the problem. In particular, better results than state-of-art SPR/ML systems can be achieved using parameter-light classifiers for two well-known data sets used in the scientific literature (*Taiwan* and *Multinational* data sets).

- This bottom-up design in data-driven slope/landslide analysis helped identify the importance about a feature-based enrichment of data. This suggests that classifiers which generate some kind of subspace are recommended instead of adaptive techniques which are useful in noisy data sets.

- Data-based SHM was addressed by using a time series anomaly detection approach. In this sense, recent studies —from data mining community— indicate that much simpler and existing SPR/ML methods may be so competitive as (deep-)NN/Kernel methods [173]. Sound results obtained in this thesis for data-based SHM, supported by a dissimilarity pattern recognition and proximity learning approach, could be considered as an example of this claim. It is important to note that this 'proximity' or 'dissimilarity' data-based SHM has the following interesting properties: It is featureless, fast and more intuitive.

5.2. Future Work

The results obtained in this thesis, for the data-driven geotechnical/natural-hazard engineering approach, motivate the following future works:

- Applying other classifiers that enrich the feature space. This category includes affine and convex hull based classifiers and the NFP rule, among others. Moreover, in the literature, there are research studies that associate the SVM classifier and the convex hulls for classification that might be of practical interest.

- Geotechnical reliability of soil or rock slopes as a surrogate modelling task by using the above mentioned classifiers and including uncertainty propagation techniques such as Karhunen-Loéve expansions.

- One of the most important factors affecting the landslide susceptibility is rainfall. Therefore, finding proper patterns in rainfall time series (e.g., using shapelets or motifs) could be a key element in an early warning scheme along with an unsupervised or one-class classification.

On the other hand, potential extensions for a dissimilarity SHM research can be explored considering:

- A combination of object representations, such as Symbolic Data Object and spectral/time-frequency decompositions, in order to build robust dissimilarity spaces. Under this rationales, (non-)elastic measures such as the DTW, the Derivative DTW (DDTW), the

Weighted DTW (WDTW), the Time Warp Edit Distance (TWED), among others, can be useful.

- An unsupervised scheme (e.g., clustering) in dissimilarity spaces. In this direction, real-time strategies could be proposed using the whole information contained in the structural signals in a proximity learning approach.

Bibliography

[1] Jamal A. Abdalla, Mousa F. Attom, and Rami Hawileh. Prediction of minimum factor of safety against slope failure in clayey soils using artificial neural network. *Environmental Earth Sciences*, 73(9):5463–5477, May 2015.

[2] Osama Abdeljaber, Onur Avci, Mustafa Serkan Kiranyaz, Boualem Boashash, Henry Sodano, and Daniel J. Inman. 1-D CNNs for structural damage detection: Verification on a structural health monitoring benchmark data. *Neurocomputing*, 275:1308–1317, 2018.

[3] Osama Abdeljaber, Onur Avci, Serkan Kiranyaz, Moncef Gabbouj, and Daniel J. Inman. Real-time vibration-based structural damage detection using one-dimensional convolutional neural networks. *Journal of Sound and Vibration*, 388:154–170, 2017.

[4] Yacine Achour and Hamid Reza Pourghasemi. How do machine learning techniques help in increasing accuracy of landslide susceptibility maps? *Geoscience Frontiers*, 11(3):871–883, 2020.

[5] Ethem Alpaydin. *Introduction to Machine Learning*. MIT Press, 3rd edition, 2014.

[6] Vinicius Alves, Alexandre Cury, Ney Roitman, Carlos Magluta, and Christian Cremona. Structural modification assessment using supervised learning methods applied to vibration data. *Engineering Structures*, 99:439 – 448, 2015.

[7] Rania Rizki Arinta and Emanuel Andi W.R. Natural Disaster Application on Big Data and Machine Learning: A Review. In *2019 4th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*, pages 249–254, 2019.

[8] Onur Avci, Osama Abdeljaber, Serkan Kiranyaz, Mohammed Hussein, Moncef Gabbouj, and Daniel J. Inman. A review of vibration–based damage detection in civil structures: From traditional methods to Machine Learning and Deep Learning applications. *Mechanical Systems and Signal Processing*, 147:107077, 2021.

[9] Yuequan Bao and Hui Li. Machine learning paradigm for structural health monitoring. *Structural Health Monitoring*, 20(4):1353–1372, 2021.

[10] Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador Garcia, Sergio Gil-Lopez, Daniel Molina, Richard Benjamins, Raja Chatila, and Francisco Herrera. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58:82–115, 2020.

[11] Robert M. Bell and Yehuda Koren. Lessons from the Netflix Prize Challenge. *SIGKDD Explorations Newsletter*, 9(2):75–79, dec 2007.

[12] Manuele Bicego and Mauricio Orozco-Alzate. PowerHC: non linear normalization of distances for advanced nearest neighbor classification. In *25th International Conference on Pattern Recognition (ICPR)*, pages 1205–1211, 2021.

[13] Michael Biehl, Barbara Hammer, and Thomas Villmann. Prototype-based models in machine learning. *WIREs Cognitive Science*, 7(2):92–111, 2016.

[14] Alain Biem. A model selection criterion for classification: application to HMM topology optimization. In *Proceedings of the Seventh International Conference on Document Analysis and Recognition - ICDAR 2003*, pages 104–108, August 2003.

[15] Christopher Bishop. *Pattern Recognition and Machine Learning*. Springer-Verlag, NY, 1st edition, 2006.

[16] Luke Bornn, Charles R. Farrar, David Higdon, and Kevin P. Murphy. Modeling and diagnosis of structural systems through sparse dynamic graphical models. *Mechanical Systems and Signal Processing*, 74:133–143, 2016. Special Issue in Honor of Professor Simon Braun.

[17] Luke Bornn, Charles R. Farrar, Gyuhae Park, and Kevin Farinholt. Structural Health Monitoring With Autoregressive Support Vector Machines. *Journal of Vibration and Acoustics*, 131(2), 02 2009.

[18] Steven L. Brunton and J. Nathan Kutz. *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control*. Cambridge University Press, 2019.

[19] Lawrence A. Bull, Paul Gardner, Timothy J. Rogers, Elizabeth J. Cross, Nikolaos Dervilis, and Keith Worden. Probabilistic inference for structural health monitoring: New modes of learning from data. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*, 7(1):03120003, 2021.

[20] Raghavendra Chalapathy and Nguyen Lu Dang Khoa. Comparison of unsupervised shallow and deep models for structural health monitoring. *Proceedings of the Institution of Civil Engineers - Bridge Engineering*, pages 1–11, 2021.

[21] Muhammad Shahzad Cheema, Abdalrahman Eweiwi, and Christian Bauckhage. High dimensional low sample size activity recognition using geometric classifiers. *Digital Signal Processing*, 42:61–69, 2015.

[22] Wei Chen, Hamid Reza Pourghasemi, Aiding Kornejady, and Ning Zhang. Landslide spatial modeling: Introducing new ensembles of ANN, MaxEnt, and SVM machine learning techniques. *Geoderma*, 305:314–327, 2017.

[23] Wei Chen, Hamid Reza Pourghasemi, Mahdi Panahi, Aiding Kornejady, Jiale Wang, Xiaoshen Xie, and Shubo Cao. Spatial prediction of landslide susceptibility using an adaptive neuro-fuzzy inference system combined with frequency ratio, generalized additive model, and support vector machine techniques. *Geomorphology*, 297:69–85, 2017.

[24] Zuoyi Chen, Yanzhi Wang, Jun Wu, Chao Deng, and Kui Hu. Sensor data-driven structural damage detection based on deep convolutional neural networks and continuous wavelet transform. *Applied Intelligence*, 51(8):5598–5609, 2021.

[25] Min-Yuan Cheng and Nhat-Duc Hoang. Typhoon–induced slope collapse assessment using a novel bee colony optimized support vector classifier. *Natural Hazards*, 78(3):1961–1978, September 2015.

[26] Vladimir Cherkassky and Filip M. Mulier. *Learning from Data: Concepts, Theory, and Methods*. Wiley-IEEE Press, 2007.

[27] Jen-Tzung Chien and Chia-Chen Wu. Discriminant waveletfaces and nearest feature classifiers for face recognition. *IEEE Trans. Pattern Anal. Machine Intell.*, 24(12):1644–1649, 2002.

[28] Jianye Ching and Kok-Kwang Phoon. *Characterizing Unknown Trend Using Sparse Bayesian Learning*, chapter Geo-Risk 2017: Geotechnical Risk Assessment and Management, pages 22–31. 2017.

[29] A. J. Choobbasti, F. Farrokhzad, and A. Barari. Prediction of slope stability using artificial neural network (case study: Noabad, Mazandaran, Iran). *Arabian Journal of Geosciences*, 2(4):311–319, Nov 2009.

[30] Paul Cilliers. *Complexity and Postmodernism. Understanding Complex Systems.* Routledge, 1st edition, 2002.

[31] T. Cover and P. Hart. Nearest neighbor pattern classification. IEEE *Transactions on Information Theory*, 13:21–27, 1967.

[32] Alexandre Cury and Christian Crémona. Pattern recognition of structural behaviors based on learning algorithms and symbolic data concepts. *Structural Control and Health Monitoring*, 19(2):161–186, 2012.

[33] Hung V. Dang, Mohsin Raza, Tung V. Nguyen, T. Bui-Tien, and Huan X. Nguyen. Deep learning-based detection of structural damage using time-series data. *Structure and Infrastructure Engineering*, 17(11):1474–1493, 2021.

[34] Sarat Kumar Das, Rajani Kanta Biswal, N. Sivakugan, and Bitanjaya Das. Classification of slopes and prediction of factor of safety using differential evolution neural networks. *Environmental Earth Sciences*, 64(1):201–210, Sep 2011.

[35] R. De Almeida Cardoso, Alexandre Cury, and Flavio Barbosa. Automated real–time damage detection strategy using raw dynamic measurements. *Engineering Structures*, 196:109364, 2019.

[36] Rharã de Almeida Cardoso, Alexandre Cury, Flavio Barbosa, and Carmelo Gentile. Unsupervised real-time SHM technique based on novelty indexes. *Structural Control and Health Monitoring*, 26(7):e2364, 2019.

[37] Hao Du and Yan Qiu Chen. Rectified nearest feature line segment for pattern classification. *Pattern Recognition*, 40(5):1486 – 1497, 2007.

[38] Robert P. W. Duin. The origin of patterns. *Frontiers in Computer Science*, 3, 2021.

[39] Robert P. W. Duin, Manuele Bicego, Mauricio Orozco-Alzate, Sang-Woon Kim, and Marco Loog. Metric learning in dissimilarity space for improved nearest neighbor performance. In Pasi Fränti, Gavin Brown, Marco Loog, Francisco Escolano, and Marcello Pelillo, editors, *Structural, Syntactic, and Statistical Pattern Recognition*, pages 183–192, Berlin, Heidelberg, 2014. Springer Berlin Heidelberg.

[40] Robert P. W. Duin and E. Pękalska. The dissimilarity space: Bridging structural and statistical pattern recognition. *Pattern Recognition Letters*, 33(7):826–832, 2012. Special Issue on Awards from ICPR 2010.

[41] Robert P. W. Duin, E. Pękalska, and Marco Loog. *Non-Euclidean Dissimilarities: Causes, Embedding and Informativeness*, pages 13–44. Springer London, London, 2013.

[42] Robert P. W. Duin and Elżbieta Pekalska. *The Science of Pattern Recognition. Achievements and Perspectives*, chapter 10, pages 221–259. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007.

[43] Robert P.W. Duin, E. Pękalska, P. Paclík, and D.M.J. Tax. The dissimilarity representation, a basis for a domain-based pattern recognition? In J Goldfarb, editor,

Proceedings of a satellite workshop of the 17th international conference on pattern recognition, pages 43–56. s.n., 2004.

[44] R.P.W. Duin and E. Pękalska. Non–Euclidean dissimilarities: Causes and informativeness. In *Structural, Syntactic, and Statistical Pattern Recognition, Joint* IAPR *International Workshop,* SSPR–SPR, Cesme, Izmir, Turkey, 2010.

[45] R.P.W. Duin and E. Pękalska. The dissimilarity representation for structural pattern recognition. In C. San-Martin and Sang-Woon Kim, editors, *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, volume 7042 of *Lecture Notes in Computer Science*, pages 1–24. Springer Berlin Heidelberg, 2011.

[46] Alireza Entezami, Hashem Shariatmadar, and Abbas Karamodin. Data–driven damage diagnosis under environmental and operational variability by novel statistical pattern recognition methods. *Structural Health Monitoring*, 18(5-6):1416–1443, 2019.

[47] Alireza Entezami, Hashem Shariatmadar, and Stefano Mariani. Early damage assessment in large-scale structures by innovative statistical pattern recognition methods based on time series modeling and novelty detection. *Advances in Engineering Software*, 150:102923, 2020.

[48] Zhice Fang, Yi Wang, Ling Peng, and Haoyuan Hong. Integration of convolutional neural network and conventional machine learning classifiers for landslide susceptibility mapping. *Computers & Geosciences*, 139:104470, 2020. Dataset available at: `https://bit.ly/3rFxqFw`.

[49] C. R. Farrar and H. Sohn. Pattern recognition for structural health monitoring. In *Workshop on Mitigation of Earthquake Disaster by Advanced Technologies*, pages 1–6, Las Vegas, NV, USA, Nov–Dec 2000.

[50] C. R. Farrar and K. Worden. An introduction to structural health monitoring. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 365(1851):303–315, December 2007.

[51] Charles Farrar, Mayuko Nishio, Francois Hemez, Chris Stull, Gyuhae Park, Phil Cornwell, Eloi Figueiredo, D. J. Luscher, and Keith Worden. Feature Extraction for Structural Dynamics Model Validation. Technical Report LA-UR-16-20151, Los Alamos National Laboratory, 2016.

[52] Charles R. Farrar and Keith Worden. *Structural Health Monitoring: A Machine Learning Perspective*. John Wiley & Sons, West Sussex, UK, 2013.

[53] Spilios D Fassois and John S Sakellariou. Time-series methods for fault detection and identification in vibrating structures. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 365(1851):411–448, 2007.

[54] Qingxiang Feng, Lijun Yan, and Jeng-Shyang Pan. A novel classifier based on nearest feature line. In *International Conference on Computing, Measurement, Control and Sensor Network (CMCSN)*, pages 265–268, July 2012.

[55] M.D. Ferentinou and M.G. Sakellariou. Computational intelligence tools for the prediction of slope performance. *Computers and Geotechnics*, 34(5):362–384, 2007. Special Issue on Biologically Inspired and Other Novel Computing Techniques in Geomechanics.

[56] Manuel Fernández-Delgado, Eva Cernadas, Senén Barro, and Dinani Amorim. Do we need hundreds of classifiers to solve real world classification problems? *Journal of Machine Learning Research*, 15(90):3133–3181, 2014.

[57] Eloi Figueiredo, Joaquim Figueiras, Gyuhae Park, Charles R. Farrar, and Keith Worden. Influence of the Autoregressive Model Order on Damage Detection. *Computer-Aided Civil and Infrastructure Engineering*, 26(3):225–238, 2011.

[58] Eloi Figueiredo, Gyuhae Park, Charles R Farrar, Keith Worden, and Joaquim Figueiras. Machine learning algorithms for damage detection under operational and environmental variability. *Structural Health Monitoring*, 10(6):559–572, 2011.

[59] Eloi Figueiredo, Gyuhae Park, Joaquim Figueiras, Charles Farrar, and Keith Worden. Structural Health Monitoring Algorithm Comparisons Using Standard Data Sets. Technical Report LA-14393, Los Alamos National Laboratory, 2009.

[60] Qing-Bin Gao and Zheng-Zhi Wang. Center–based nearest neighbor classifier. *Pattern Recognition*, 40(1):346–349, 2007.

[61] Salvador Garcia, Joaquin Derrac, Jose Cano, and Francisco Herrera. Prototype selection for nearest neighbor classification: Taxonomy and empirical study. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(3):417–435, 2012.

[62] P. Gardner, L.A. Bull, J. Gosliga, J. Poole, N. Dervilis, and K. Worden. A population-based SHM methodology for heterogeneous structures: Transferring damage localisation knowledge between different aircraft wings. *Mechanical Systems and Signal Processing*, 172:108918, 2022.

[63] P. A. Gardner, L. A. Bull, N. Dervilis, and K. Worden. On the Application of Heterogeneous Transfer Learning to Population-Based Structural Health Monitoring. In Ramin Madarshahian and Francois Hemez, editors, *Data Science in Engineering, Volume 9*, pages 87–98, Cham, 2022. Springer International Publishing.

[64] Jason N. Goetz, Richard H. Guthrie, and Alexander Brenning. Integrating physical and empirical landslide susceptibility models using generalized additive models. *Geomorphology*, 129(3):376–386, 2011.

[65] Lev Goldfarb. A unified approach to pattern recognition. *Pattern Recognition*, 17(5):575–582, 1984.

[66] Behrouz Gordan, Danial Jahed Armaghani, Mohsen Hajihassani, and Masoud Monjezi. Prediction of seismic slope stability through combination of particle swarm optimization and neural network. *Engineering with Computers*, 32(1):85–97, Jan 2016.

[67] M. Gul and F. N. Catbas. Statistical pattern recognition for structural health monitoring using time series modeling: Theory and experimental verifications. *Mechanical Systems and Signal Processing*, 23(7):2192–2204, 2009.

[68] B. Haasdonk and D. Keysers. Tangent distance kernels for support vector machines. In *2002 International Conference on Pattern Recognition*, volume 2, pages 864–868 vol.2, 2002.

[69] Ralf Herbrich. *Learning Kernel Classifiers. Theory and Algorithms*. MIT Press, 2001.

[70] Nhat-Duc Hoang and Dieu Tien Bui. Chapter 18 - Slope Stability Evaluation Using Radial Basis Function Neural Network, Least Squares Support Vector Machines, and Extreme Learning Machine. In Pijush Samui, Sanjiban Sekhar, and Valentina E. Balas, editors, *Handbook of Neural Computation*, pages 333–344. Academic Press, 2017.

[71] Nhat-Duc Hoang and Anh-Duc Pham. Hybrid artificial intelligence approach based on metaheuristic and machine learning for slope stability assessment: A multinational data analysis. *Expert Systems with Applications*, 46:60–68, 2016.

[72] Ye Hua, Xianmin Wang, Yongwei Li, Peiyun Xu, and Wenxiang Xia. Dynamic development of landslide susceptibility based on slope unit and deep neural networks. *Landslides*, 18(1):281–302, 2021.

[73] Faming Huang, Jing Zhang, Chuangbing Zhou, Yuhao Wang, Jinsong Huang, and Li Zhu. A deep learning algorithm using a fully connected sparse autoencoder neural network for landslide susceptibility prediction. *Landslides*, 17(1):217–229, 2020.

[74] Yang H. Huang. *Slope Stability Analysis by the Limit Equilibrium Method*. American Society of Civil Engineers, 2014.

[75] Yu Huang and Lu Zhao. Review on landslide susceptibility mapping using support vector machines. *CATENA*, 165:520–529, 2018.

[76] A. K. Jain, R. P. W. Duin, and Jianchang Mao. Statistical Pattern Recognition: A Review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1):4–37, 2000.

[77] Kejie Jiang, Qiang Han, Xiuli Du, and Pinghe Ni. A decentralized unsupervised structural condition diagnosis approach using deep auto-encoders. *Computer-Aided Civil and Infrastructure Engineering*, 36(6):711–732, 2021.

[78] Piotr Juszczak, David M.J. Tax, E. Pękalska, and R. P. W. Duin. Minimum spanning tree based one-class classifier. *Neurocomputing*, 72(7):1859–1869, 2009. Advances in Machine Learning and Computational Intelligence.

[79] Navid Kardani, Annan Zhou, Majidreza Nazem, and Shui-Long Shen. Improved prediction of slope stability using a hybrid stacking ensemble method based on finite element analysis and field data. *Journal of Rock Mechanics and Geotechnical Engineering*, 13(1):188–201, 2021.

[80] George Em Karniadakis, Ioannis G. Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440, 2021.

[81] Eamonn Keogh, Stefano Lonardi, and Chotirat Ann Ratanamahatana. Towards parameter-free data mining. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '04, pages 206–215, New York, NY, USA, 2004. Association for Computing Machinery.

[82] Eamonn Keogh, Stefano Lonardi, Chotirat Ann Ratanamahatana, Li Wei, Sang-Hee Lee, and John Handley. Compression-based data mining of sequential data. *Data Mining and Knowledge Discovery*, 14:99–129, 2007.

[83] Serkan Kiranyaz, Onur Avci, Osama Abdeljaber, Turker Ince, Moncef Gabbouj, and Daniel J. Inman. 1D convolutional neural networks and applications: A survey. *Mechanical Systems and Signal Processing*, 151:107398, 2021.

[84] Mohammadreza Koopialipoor, Danial Jahed Armaghani, Ahmadreza Hedayat, Aminaton Marto, and Behrouz Gordan. Applying various hybrid intelligent systems to evaluate and predict slope stability under static and dynamic conditions. *Soft Computing*, May 2018.

[85] Kundan Kumar, Prabir Kumar Biswas, and Nirjhar Dhang. Time series-based SHM using PCA with application to ASCE benchmark structure. *Journal of Civil Structural Health Monitoring*, 10(5):899–911, 2020.

[86] Manoj Kumar, Pijush Samui, and Ajay Kumar Naithani. Determination of stability of epimetamorphic rock slope using Minimax Probability Machine. *Geomatics, Natural Hazards and Risk*, 7(1):186–193, 2016.

[87] David J. Lary, Gebreab K. Zewdie, Xun Liu, Daji Wu, Estelle Levetin, Rebecca J. Allee, Nabin Malakar, Annette Walker, Hamse Mussa, Antonio Mannino, and Dirk Aurin. *Machine Learning Applications for Earth Observation*, pages 165–218. Springer International Publishing, Cham, 2018.

[88] A.J. Li, S. Khoo, A.V. Lyamin, and Y. Wang. Rock slope stability analyses using extreme learning neural network and terminal steepest descent algorithm. *Automation in Construction*, 65:42–50, 2016.

[89] Bo Li, Duanyou Li, Zhijun Zhang, Shengmei Yang, and Fan Wang. Slope stability analysis based on quantum–behaved particle swarm optimization and least squares support vector machine. *Applied Mathematical Modelling*, 39(17):5253–5264, 2015.

[90] Dian-Qing Li, Dong Zheng, Zi-Jun Cao, Xiao-Song Tang, and Kok-Kwang Phoon. Response surface methods for slope reliability analysis: Review and comparison. *Engineering Geology*, 203:3–14, 2016. Special Issue on Probabilistic and Soft Computing Methods for Engineering Geology.

[91] Jiang Li and Can-Yi Lu. A new decision rule for sparse representation based classification for face recognition. *Neurocomputing*, 116:265–271, 2013. Advanced Theory and Methodology in Intelligent Computing.

[92] Stan Z. Li and Juwei Lu. Face recognition using the nearest feature line method. *IEEE Transactions on Neural Networks*, 10(2):439–443, 1999.

[93] S.Z Li and J. Lu. Generalizing capacity of face database for face recognition. In *Third IEEE International Conference on Automatic Face and Gesture Recognition*, pages 402–406, Apr 1998.

[94] Xueyou Li, Limin Zhang, and Shuai Zhang. Efficient bayesian networks for slope safety evaluation with large quantity monitoring information. *Geoscience Frontiers*, 9(6):1679–1687, 2018. Reliability Analysis of Geotechnical Infrastructures.

[95] Y. Lin, K. Zhou, and J. Li. Prediction of slope stability using four supervised learning methods. *IEEE Access*, 6:31169–31179, 6 2018.

[96] Zaobao Liu, Jianfu Shao, Weiya Xu, Hongjie Chen, and Yu Zhang. An extreme learning machine approach for slope stability evaluation and prediction. *Natural Hazards*, 73(2):787–804, Sep 2014.

[97] Noel Lopes and Bernardete Ribeiro. Incremental Hypersphere Classifier (IHC). In *Machine Learning for Adaptive Many-Core Machines - A Practical Approach*, volume 7 of *Studies in Big Data*, chapter 6, pages 107–123. Springer International Publishing, Cham, 2015.

[98] I. Lopez and N. Sarigul-Klijn. A novel dimensional reduction approach for structural damage diagnosis using feature similarity. In Tribikram Kundu, editor, *Health Monitoring of Structural and Biological Systems 2009*, volume 7295, pages 511 – 522. International Society for Optics and Photonics, SPIE, 2009.

[99] Israel Lopez and Nesrin Sarigul-Klijn. Distance similarity matrix using ensemble of dimensional data reduction techniques: Vibration and aerocoustic case studies. *Mechanical Systems and Signal Processing*, 23(7):2287 – 2300, 2009.

[100] Zhengjing Ma, Gang Mei, and Francesco Piccialli. Machine learning for landslides prevention: a survey. *Neural Computing and Applications*, 33(17):10881–10907, 2021.

[101] Ramin Madarshahian and Francois Hemez, editors. *Data Science in Engineering, Proceedings of the 39th IMAC, A Conference and Exposition on Structural Dynamics*, volume 9 of *Conference Proceedings of the Society for Experimental Mechanics Series book series (CPSEMS)*, Berlin, Germany, 2021. Springer.

[102] Mehrisadat Makki Alamdari, Ali Anaissi, Nguyen L. D. Khoa, and Samir Mustapha. Frequency domain decomposition-based multisensor data fusion for assessment of progressive damage in structures. *Structural Control and Health Monitoring*, 26(2):e2299, 2019.

[103] Hossein Moayedi, Mansour Mosallanezhad, Ahmad Safuan A. Rashid, Wan Amizah Wan Jusoh, and Mohammed Abdullahi Muazu. A systematic review and meta-analysis of artificial neural network application in geotechnical engineering: theory and applications. *Neural Computing and Applications*, 32(2):495–518, 2020.

[104] Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar. *Foundations of Machine Learning*. MIT Press, 2nd edition, 2018.

[105] L. Moniz, J.M. Nichols, C.J. Nichols, M. Seaver, S.T. Trickey, M.D. Todd, L.M. Pecora, and L.N. Virgin. A multivariate, attractor-based approach to structural health monitoring. *Journal of Sound and Vibration*, 283(1):295–310, 2005.

[106] Maliki Moustapha, Stefano Marelli, and Bruno Sudret. Active learning for structural reliability: Survey, general framework and benchmark. *Structural Safety*, 96:102174, 2022.

[107] Maximilian Münch, Christoph Raab, and Frank-Michael Schleif. Encoding of Indefinite Proximity Data: A Structure Preserving Perspective. In Maria De Marsico, Gabriella Sanniti di Baja, and Ana Fred, editors, *Pattern Recognition Applications and Methods*, pages 112–137, Cham, 2020. Springer International Publishing.

[108] Maximilian Münch, Michiel Straat, Michael Biehl, and Frank-Michael Schleif. Complex-valued embeddings of generic proximity data. In Andrea Torsello, Luca Rossi, Marcello Pelillo, Battista Biggio, and Antonio Robles Kelly, editors, *Structural, Syntactic, and Statistical Pattern Recognition*, pages 14–23, Cham, 2021. Springer International Publishing.

[109] W. James Murdoch, Chandan Singh, Karl Kumbier, Reza Abbasi-Asl, and Bin Yu. Definitions, methods, and applications in interpretable machine learning. *Proceedings of the National Academy of Sciences*, 116(44):22071–22080, 2019.

[110] J. M. Nichols, M. D. Todd, M. Seaver, and L. N. Virgin. Use of chaotic excitation and attractor property analysis in structural health monitoring. *Phys. Rev. E*, 67:016209, Jan 2003.

[111] Mayuko Nishio, Francois Hemez, Keith Worden, Gyuhae Park, Nobuo Takeda, and Charles Farrar. Feature extraction for structural dynamics model validation. In Tom Proulx, editor, *Linking Models and Experiments, Volume 2*, pages 153–163, New York, NY, 2011. Springer New York.

[112] Mauricio Orozco-Alzate, Sisto Baldo, and Manuele Bicego. Relation, transition and comparison between the adaptive nearest neighbor rule and the hypersphere classifier. In Elisa Ricci, Samuel Rota Bulò, Cees Snoek, Oswald Lanz, Stefano Messelodi, and Nicu Sebe, editors, *Image Analysis and Processing – ICIAP 2019*, pages 141–151, Cham, 2019. Springer International Publishing.

[113] Mauricio Orozco-Alzate and Manuele Bicego. A cheaper rectified-nearest-feature-line-segment classifier based on safe points. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 2787–2794, 2021.

[114] Mauricio Orozco-Alzate, Robert P. W. Duin, and César Germán Castellanos-Domínguez. A generalization of dissimilarity representations using feature lines and feature planes. *Pattern Recognition Letters*, 30(3):242–254, feb 2009.

[115] L.A. Overbey and M.D. Todd. Analysis of Local State Space Models for Feature Extraction in Structural Health Monitoring. *Structural Health Monitoring*, 6(2):145–172, 2007.

[116] E. Pękalska. *Dissimilarity representations in pattern recognition. Concepts, theory and applications*. PhD thesis, Delft University of Technology, 2005.

[117] E. Pękalska and R. P.W. Duin. *Dissimilarity measures*, volume 64 of *Machine Perception and Artificial Intelligence*, chapter 5, pages 215–252. World Scientific, Singapore, 2005.

[118] E. Pękalska and R. P.W. Duin. *The Dissimilarity Representation for Pattern Recognition: Foundations and Applications*, volume 64 of *Machine Perception and Artificial Intelligence*. World Scientific, Singapore, 2005.

[119] E. Pękalska and B. Haasdonk. Kernel discriminant analysis for positive definite and indefinite kernels. IEEE *Transactions on Pattern Recognition and Machine Intelligence*, 31(6):482–492, 2009.

[120] Elzbieta Pękalska, Pavel Paclik, and Robert P. W. Duin. A generalized kernel approach to dissimilarity-based classification. *J. Mach. Learn. Res.*, 2:175–211, March 2002.

[121] Jin-Song Pei, Dean F. Hougen, Sai Teja Kanneganti, Joseph P. Wright, Eric C. Mai, Andrew W. Smyth, Sami F. Masri, Armen Derkevorkian, François Gay-Balmaz, and Ludian Komini. *Interpretable Machine Learning for Function Approximation in Structural Health Monitoring*, pages 369–388. Springer International Publishing, Cham, 2022.

[122] E. Pekalska and R. P. W. Duin. Beyond Traditional Kernels: Classification in Two Dissimilarity-Based Representation Spaces. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 38(6):729–744, Nov 2008.

[123] Binh Thai Pham and Indra Prakash. A novel hybrid model of Bagging-based Naive Bayes Trees for landslide susceptibility assessment. *Bulletin of Engineering Geology and the Environment*, 78(3):1911–1925, 2021.

[124] Binh Thai Pham, Indra Prakash, Sushant K. Singh, Ataollah Shirzadi, Himan Shahabi, Thi-Thu-Trang Tran, and Dieu Tien Bui. Landslide susceptibility modeling using Reduced Error Pruning Trees and different ensemble techniques: Hybrid machine learning approaches. *CATENA*, 175:203–218, 2019.

[125] Kok-Kwang Phoon, Jianye Ching, and Zijun Cao. Unpacking data-centric geotechnics. *Underground Space*, 2022.

[126] Kok-Kwang Phoon, Jianye Ching, and Takayuki Shuku. Challenges in data-driven site characterization. *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, 0(0):1–13, 2021.

[127] Kok-Kwang Phoon and Wengang Zhang. Future of machine learning in geotechnics. *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, 0(0):1–16, 2022.

[128] Luca Piciullo, Michele Calvello, and José Mauricio Cepeda. Territorial early warning systems for rainfall-induced landslides. *Earth-Science Reviews*, 179:228–247, 2018.

[129] Daniele Posenato, Francesca Lanata, Daniele Inaudi, and Ian F.C. Smith. Model-free data interpretation for continuous monitoring of complex structures. *Advanced Engineering Informatics*, 22(1):135–144, 2008. Intelligent computing in engineering and architecture.

[130] Miguel A. Prada, Janne Toivola, Jyrki Kullaa, and Jaakko Hollmén. Three-way analysis of structural health monitoring data. *Neurocomputing*, 80:119 – 128, 2012. Special Issue on Machine Learning for Signal Processing 2010.

[131] Chongchong Qi and Xiaolin Tang. A hybrid ensemble method for improved prediction of slope stability. *International Journal for Numerical and Analytical Methods in Geomechanics*, 42(15):1823–1839, 2018.

[132] Thanawin Rakthanmanon, Bilson Campana, Abdullah Mueen, Gustavo Batista, Brandon Westover, Qiang Zhu, Jesin Zakaria, and Eamonn Keogh. Addressing Big Data Time Series: Mining Trillions of Time Series Subsequences Under Dynamic Time Warping. *ACM Transactions on Knowledge Discovery from Data*, 7(3), September 2013.

[133] Paola Reichenbach, Mauro Rossi, Bruce D. Malamud, Monika Mihir, and Fausto Guzzetti. A review of statistically-based landslide susceptibility models. *Earth-Science Reviews*, 180:60–91, 2018.

[134] Hannah Ritchie and Max Roser. Natural disasters. *Our World in Data*, 2014. It was last revised in November 2021.

[135] G. Roberti, J. McGregor, S. Lam, D. Bigelow, B. Boyko, C. Ahern, V. Wang, B. Barnhart, C. Smyth, D. Poole, and S. Richard. INSPIRE standards as a framework for artificial intelligence applications: a landslide example. *Natural Hazards and Earth System Sciences*, 20(12):3455–3483, 2020.

[136] T.J. Rogers, K. Worden, R. Fuentes, N. Dervilis, U.T. Tygesen, and E.J. Cross. A Bayesian non-parametric clustering approach for semi-supervised structural health monitoring. *Mechanical Systems and Signal Processing*, 119:100–119, 2019.

[137] S. Rukhaiyar, M. N. Alam, and N. K. Samadhiya. A PSO–ANN hybrid model for predicting factor of safety of slope. *International Journal of Geotechnical Engineering*, 12(6):556–566, 2018.

[138] J. S.-Taylor and N. Cristianini. *Kernel Methods for Pattern Analysis*. Cambridge University Press, Cambridge, UK, 2004.

[139] Seyedomid Sajedi and Xiao Liang. Dual Bayesian inference for risk-informed vibration-based damage diagnosis. *Computer-Aided Civil and Infrastructure Engineering*, 36(9):1168–1184, 2021.

[140] M. G. Sakellariou and M. D. Ferentinou. A study of slope stability prediction using neural networks. *Geotechnical & Geological Engineering*, 23(4):419, Aug 2005.

[141] P. Samui and D.P. Kothari. Utilization of a least square support vector machine (LSSVM) for slope stability analysis. *Scientia Iranica*, 18(1):53–58, 2011.

[142] Pijush Samui. Slope stability analysis: a support vector machine approach. *Environmental Geology*, 56(2):255, Feb 2008.

[143] Pijush Samui. Support vector classifier analysis of slope. *Geomatics, Natural Hazards and Risk*, 4(1):1–12, 2013.

[144] Adam Santos, Eloi Figueiredo, M.F.M. Silva, C.S. Sales, and J.C.W.A. Costa. Machine learning algorithms for damage detection: Kernel-based approaches. *Journal of Sound and Vibration*, 363:584 – 599, 2016.

[145] João Pedro Santos, Christian Cremona, André D. Orcesi, and Paulo Silveira. Early Damage Detection Based on Pattern Recognition and Data Fusion. *Journal of Structural Engineering*, 143(2):04016162, 2017.

[146] Hassan Sarmadi, Alireza Entezami, Behzad Saeedi Razavi, and Ka-Veng Yuen. Ensemble learning-based structural health monitoring by Mahalanobis distance metrics. *Structural Control and Health Monitoring*, 28(2):e2663, 2021.

[147] Hassan Sarmadi and Abbas Karamodin. A novel anomaly detection method based on adaptive Mahalanobis-squared distance and one-class kNN rule for structural health monitoring under environmental effects. *Mechanical Systems and Signal Processing*, 140:106495, 2020.

[148] W. J. Scheirer, M. J. Wilber, M. Eckmann, and T. E. Boult. Good recognition is non–metric. *Pattern Recognition*, 47:2721–2731, 2014.

[149] Frank-Michael Schleif and Peter Tino. Indefinite Proximity Learning: A Review. *Neural Computation*, 27(10):2039–2096, 10 2015.

[150] Haichen Shi, Keith Worden, and Elizabeth J. Cross. A cointegration approach for heteroscedastic data based on a time series decomposition: An application to structural health monitoring. *Mechanical Systems and Signal Processing*, 120:16 – 31, 2019.

[151] Moisés Silva, Adam Santos, and Elói Figueiredo. *Damage Detection for Structural Health Monitoring of Bridges as a Knowledge Discovery in Databases Process*, pages 1–24. Springer Singapore, Singapore, 2019.

[152] Moisés Silva, Adam Santos, Eloi Figueiredo, Reginaldo Santos, Claudomiro Sales, and João C.W.A. Costa. A novel unsupervised approach based on a genetic algorithm for structural damage detection in bridges. *Engineering Applications of Artificial Intelligence*, 52:168 – 180, 2016.

[153] Osvaldo Simeone. *A Brief Introduction to Machine Learning for Engineers*, volume 12. Now Foundations and Trends, 2018.

[154] H Sohn, C R Farrar, F M Hemez, and J J Czarnecki. A Review of Structural Health Monitoring Literature: 1996-2001. Technical Report LA-UR-02-2095, Los Alamos National Lab. (LANL), 1 2002.

[155] Hoon Sohn and Chang Kook Oh. *Statistical Pattern Recognition*, chapter 30. John Wiley & Sons, Ltd., 2009.

[156] Nikos A Spanos, John S Sakellariou, and Spilios D Fassois. Vibration-response-only statistical time series structural health monitoring methods: A comprehensive assessment via a scale jacket structure. *Structural Health Monitoring*, 19(3):736–750, 2020.

[157] M.D. Spiridonakos and S.D. Fassois. Adaptable functional series TARMA models for non-stationary signal representation and their application to mechanical random vibration modeling. *Signal Processing*, 96:63 – 79, 2014.

[158] Ingo Steinwart. *Empirical Inference. Festschrift in Honor of Vladimir N. Vapnik*, chapter Some Remarks on the Statistical Analysis of SVMs and Related Methods, pages 25–36. Springer-Verlag Berlin Heidelberg, 2013.

[159] Han Sun, Henry V. Burton, and Honglan Huang. Machine learning applications for building structural design and performance assessment: State-of-the-art review. *Journal of Building Engineering*, 33:101816, 2021.

[160] Zhiyi Tang, Zhicheng Chen, Yuequan Bao, and Hui Li. Convolutional neural network-based data anomaly detection method using multiple information for structural health monitoring. *Structural Control and Health Monitoring*, 26(1):e2296, 2019.

[161] David Tax. *One-class classification: Concept-learning in the absence of counter-examples.* Doctoral thesis, Delft University of Technology, Netherlands, June 2001.

[162] David M.J Tax and Robert P.W Duin. Support vector domain description. *Pattern Recognition Letters*, 20(11):1191–1199, 1999.

[163] Sergios Theodoridis and Konstantinos Koutroumbas. *Pattern Recognition.* Academic Press, 2009.

[164] Dieu Tien Bui, Biswajeet Pradhan, Owe Lofman, Inge Revhaug, and Oystein B. Dick. Spatial prediction of landslide hazards in Hoa Binh province (Vietnam): A comparative assessment of the efficacy of evidential belief functions and fuzzy logic models. *CATENA*, 96:28–40, 2012.

[165] Dieu Tien Bui, Biswajeet Pradhan, Haleh Nampak, Quang-Thanh Bui, Quynh-An Tran, and Quoc-Phi Nguyen. Hybrid artificial intelligence approach based on neural fuzzy inference model and metaheuristic optimization for flood susceptibility modeling in a high-frequency tropical cyclone area using GIS. *Journal of Hydrology*, 540:317–330, 2016.

[166] Maria Valero, Fangyu Li, Liang Zhao, Chi Zhang, Jose Garrido, and Zhu Han. Vibration sensing-based human and infrastructure safety/health monitoring: A survey. *Digital Signal Processing*, 114:103037, 2021.

[167] Chaofeng Wang, Qian Yu, Kincho H. Law, Frank McKenna, Stella X. Yu, Ertugrul Taciroglu, Adam Zsarnóczay, Wael Elhaddad, and Barbaros Cetiner. Machine learning-based regional scale intelligent modeling of building information for natural hazard risk management. *Automation in Construction*, 122:103474, 2021.

[168] H.B. Wang, J.M. Li, B.Zhou, Y.Zhou, Z.Q. Yuan, , and Y.P. Chen. Application of a hybrid model of neural networks and genetic algorithms to evaluate landslide susceptibility. *Geoenvironmental Disasters*, 4(15):1–12, 2017.

[169] Jigang Wang, Predrag Neskovic, and Leon N. Cooper. Improving nearest neighbor rule with a simple adaptive distance measure. *Pattern Recognition Letters*, 28(2):207 – 213, 2007.

[170] Xiaoyue Wang, Abdullah Mueen, Hui Ding, Goce Trajcevski, Peter Scheuermann, and Eamonn Keogh. Experimental comparison of representation methods and distance measures for time series data. *Data Mining and Knowledge Discovery*, 26(2):275–309, 2013.

[171] Keith Worden, Charles R. Farrar, Jonathan Haywood, and Michael Todd. A review of nonlinear dynamics applications to structural health monitoring. *Structural Control and Health Monitoring*, 15(4):540–567, 2008.

[172] Keith Worden and Graeme Manson. The application of machine learning to structural health monitoring. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 365(1851):515–537, 2007.

[173] Renjie Wu and Eamonn Keogh. Current Time Series Anomaly Detection Benchmarks are Flawed and are Creating the Illusion of Progress. *IEEE Transactions on Knowledge and Data Engineering*, pages 1–9, 2021.

[174] Xindong Wu, Vipin Kumar, J. Ross Quinlan, Joydeep Ghosh, Qiang Yang, , Hiroshi Motoda, Geoffrey J. McLachlan, Angus Ng, Bing Liu, Philip S. Yu, Zhi-Hua Zhou, Michael Steinbach, David J. Hand, and Dan Steinberg. Top 10 algorithms in data mining. *Knowledge and Information Systems*, 14(1):1–37, 2008.

[175] Beibei Yang, Kunlong Yin, Suzanne Lacasse, and Zhongqiang Liu. Time series analysis and long short-term memory neural network to predict landslide displacement. *Landslides*, 16(4):677–694, 2019.

[176] JianXi Yang, Fei Yang, Likai Zhang, Ren Li, Shixin Jiang, Guiping Wang, Le Zhang, and Zeng Zeng. Bridge health anomaly detection using deep support vector data description. *Neurocomputing*, 2021.

[177] Xin-She Yang, Siamak Talatahari, Amir Hossein Gandomi, and Amir Hossein Alavi, editors. *Metaheuristics in Water, Geotechnical and Transport Engineering*. Elsevier, 2013.

[178] Yang Yang, Zhike Peng, Wenming Zhang, and Guang Meng. Parameterised time–frequency analysis methods and their engineering applications: A review of recent advances. *Mechanical Systems and Signal Processing*, 119:182 – 221, 2019.

[179] Ruigen Yao and Shamim N. Pakzad. Autoregressive statistical pattern recognition algorithms for damage detection in civil structures. *Mechanical Systems and Signal Processing*, 31:355–368, 2012.

[180] Zaher Mundher Yaseen, Ahmed El-shafie, Othman Jaafar, Haitham Abdulmohsin Afan, and Khamis Naba Sayl. Artificial intelligence based models for stream-flow forecasting: 2000–2015. *Journal of Hydrology*, 530:829–844, 2015.

[181] Saleh Yousefi, Hamid Reza Pourghasemi, Sayed Naeim Emami, Soheila Pouyan, Saeedeh Eskandari, and John P. Tiefenbacher. A machine learning framework for multi-hazards modeling and mapping in a mountainous area. *Scientific Reports*, 10(1):12144, 2020.

[182] F Zhang, Zaobao Liu, Lifeng Zheng, and Yu Zhang. Development of an adaptive relevance vector machine approach for slope stability inference. *Neural Computing and Applications*, 25(7):2025–2035, December 2014.

[183] Yin Zhang, Miaolin Dai, and Zhimin Ju. Preliminary Discussion Regarding SVM Kernel Function Selection in the Twofold Rock Slope Prediction Model. *Journal of Computing in Civil Engineering*, 30(3):04015031, 2016.

[184] Hongbo Zhao, Shunde Yin, and Zhongliang Ru. Relevance vector machine applied to slope stability analysis. *International Journal for Numerical and Analytical Methods in Geomechanics*, 36(5):643–652, 2012.

[185] Wenming Zheng, Li Zhao, and Cairong Zou. Locally nearest neighbor classifiers for pattern classification. *Pattern Recognition*, 37(6):1307–1309, 2004.

[186] Jian Zhou, Enming Li, Shan Yang, Mingzheng Wang, Xiuzhi Shi, Shu Yao, and Hani S. Mitri. Slope stability prediction for circular mode failure using gradient boosting machine approach based on an updated database of case histories. *Safety Science*, 118:505–518, 2019.

[187] Lim Yi Zhou, Fam Pei Shan, Kunio Shimizu, Tomoaki Imoto, Habibah Lateh, and Koay Swee Peng. A comparative study of slope failure prediction using logistic regression, support vector machine and least square support vector machine models. *AIP Conference Proceedings*, 1870(1):060012, 2017.

[188] Yonglei Zhou, Changshui Zhang, and Jingchun Wang. Extended nearest feature line classifier. In Chengqi Zhang, H. W. Guesgen, and Wai-Kiang Yeap, editors, *PRICAI 2004: Trends in Artificial Intelligence*, volume 3157 of *Lecture Notes in Computer Science*, pages 183–190. Springer Berlin Heidelberg, 2004.

[189] Z. Zhou, S.Z. Li, and K.L. Chan. A theoretical justification of nearest feature line method. In *Proc. 15th ICPR International Conference on Pattern Recognition*, volume 2, pages 759–762, 2000.

[190] Zonglin Zhou and Chee Keong Kwoh. The pattern classification based on the nearest feature midpoints. In *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004*, volume 3, pages 446–449, Aug 2004.