C. R. Mecanique 347 (2019) 856–862

Contents lists available at ScienceDirect

Comptes Rendus Mecanique

www.sciencedirect.com

Data-Based Engineering Science and Technology / *Sciences et technologies de l'ingénierie basées sur les données*

Evaluation of hip fracture risk using a hyper-parametric model based on the Locally Linear Embedding technique

Enrique Nadal [a,*], David Muñoz [a], Nieves Vivó [b], Irene Lucas [c], Juan José Ródenas [a]

[a] *Centro de Investigación en Ingeniería Mecánica (CIIM), Universitat Politècnica de València, Camino de Vera s/n, 46022 Valencia, Spain*
[b] *Hospital Virgen de los Lirios, Alcoy, Spain*
[c] *Centro de Salud La Fábrica, Alcoy, Spain*

ARTICLE INFO

Article history:
Received 17 September 2019
Accepted 31 October 2019
Available online 14 November 2019

Keywords:
Machine Learning
Fracture risk
Osteoporosis
Locally Linear Embedding
Hyper-parametrization
Optimization

ABSTRACT

The hip fracture is one of the most common diseases for elder people and also, one of the most worrying one since it usually is the starting point of further complications for both, the health of the patient and their daily life. Additionally, reports shown that there exist differences between people living in different regions, thus limiting the use of global models. In this work we propose a hip fracture prediction tool for a local region, using clinical data of the population of that region. The data is processed with a dimensionality reduction tool in combination with and hyper-parametrization process and the corresponding hyper-parameter optimization process for obtaining good predictions in the diagnoses, as the results shown.

1. Introduction

One of the most worrying diseases for elder people is the hip fracture, not only because of the immediate consequences that can even lead to death, but also because, in many cases, it represents the origin of further complications that considerably affect their day to day. In many cases, these patients suffer a considerable reduction of mobility that is directly related to a reduction of their independence. Thus, hip fractures in these patients do not only represent a high cost for the public health systems but also a high "social" cost both for the patient and the relatives in their care. Therefore, any prevention tool for this disease would be of a great interest for the society. In that sense, accurate predictions of the fracture risk are of a great interest. One of the clinical reference test for this disease is the bone densitometry, usually applied to the femur neck. Then, the diagnosis is based on the so-called Bone Mineral Density (BMD), a statistical parameter that servers as an indicator of the fracture risk level. However, this indicator is not sufficient for predicting the fracture risk [1] since only takes into account the mean bone density in a specific region of the skeleton. Some authors [2,3] propose to use other clinical biomarkers and environment factors to increase the accuracy of the diagnosis. In that sense, techniques such as FRAX® have been developed. FRAX® uses patient specific data like the patient sex, age, smoking habits, steroids consumption, etc. These techniques use the statistical analysis of this indirect factors to provide predictions of the fracture risk level. However in certain circumstances, the accuracy of these techniques is considerably low [4], specially in predicting the patients who will have a fracture [5]. The purpose of this contribution is to investigate a tool capable to automatically and accurately predict

* Corresponding author.
E-mail addresses: ennaso@upvnet.upv.es (E. Nadal), damuopel@upv.es (D. Muñoz), nievior@gmail.com (N. Vivó), irenelucastomas@gmail.com (I. Lucas), jjrodena@mcm.upv.es (J.J. Ródenas).

https://doi.org/10.1016/j.crme.2019.11.010
1631-0721/

the fracture risk using available patient-specific clinical data. In this initial study we have considered a localized population, in particular patients in the region of Alcoy (Spain). This fact is important since FRAX® uses national data, but the local variations of the characteristics of the population, climate, traditions and customs can strongly affect the prediction.

Several Artificial Intelligence methods are developed for data classification using a supervised learning process [6,7] like decision trees or random forest [8] for tumour classification, among others. This kind of data assimilation algorithms have a good performance for large amount of data but are not that accurate for a small number of data. On the other hand, Artificial Intelligence algorithms based of non-linear dimensionality reduction such as the Locally Linear Embedding (LLE) [9–11] or the kernel-PCA [12,13] have been successfully used to extract the internal structure of the data. More recently the Topological Data Analysis (TDA) has been used to generate low-dimensional representations and applied to many different disciplines ranging from astrophysics to medicine [14]. These kind of algorithms are sensitive to the measure of distances, metric, in the high-dimensional space. The distance in this context indicates how similar two individuals are. Some authors consider the normalized data before applying the dimensionality reduction techniques, thus giving the same importance/weight to all the parameters that define the individuals and modifying somehow the latent neighbourhood. However it is possible to add a new set of parameters (hyper-parameters), weighting the different kind of input data, with an *a priori* unknown value, thus modifying the metric in the high-dimensional space. Then, the value of the hyper-parameters is sought in order to minimize a certain cost function of interest for a given application. In other words, the optimization of the hyper-parameters customizes the metric for an specific objective, tailoring the definition of the neighbour individuals for a certain purpose.

The method proposed in this paper uses a Genetic Algorithm (GA) [15] as optimizer to find the best values of the hyper-parameters for a given objective function. In the case of study, since the objective is to predict the hip failure, the objective function will try to separate (cluster) from the rest the patients with a high risk of osteoporosis fracture. The data used is a cohort of 144 patients from the region of Alcoy. The objective of this work is to develop a tool applicable to a small region, taking into account the local characteristics of the population.

After this introduction, the manuscript is organized as follows: in section 2 the combination of the LLE and the optimization algorithm is described. Section 3 presents the results obtained for the local population and finally section 4 draws the main conclusions.

2. Methodology

In this section the Locally Linear Embedding [9] and also the GA formalism are introduced. Additionally, the combination of both of them is also detailed in this section.

2.1. Locally linear embedding

In data analysis, techniques as the Singular Value Decomposition (SVD), also known as the Principal Component Analysis (PCA) [16], try to extract a reduced number of dimensions for describing the data represented in a high-dimensional space. These tools consist in finding a new reference system aligned with the data structure. However, in some situations there is not a clear global data structure, therefore these techniques fail in reducing the number of dimensions. Some more advanced techniques falls in the group of non-linear dimensionality reduction tools. Among them we can highlight the kernel-PCA [17] and the LLE [9]. The kernel-PCA is based on the use of the "kernel trick" in order to move between the high dimension space and the reduced one. On the other hand the LLE works similar to a patch-wise SVD, as it will be explained later.

The LLE is a dimensionality reduction approach. Data can be represented in a high-dimensional space. For instance the image provided by a high-resolution digital camera of 12 Mpixels has 12 millions of data per channel (i.e. red, green, etc.). That is, any picture taken by the camera can be represented in a $12 \cdot 10^6$-dimensional space (for the sake of simplicity only one channel is considered). However, imagine the case in which all pictures represent faces of people from a certain region of a country. Every image lives in a high-dimensional space intractable for humans. However, for us, the humans, it is easy to differentiate the main features of each individual, i.e. nose, eyes, mouth, etc. They could represent the main features of the images, in other words, the intrinsic dimensions. Humans have the inherent ability to differentiate known shapes. All of us did this at least once when looking at the clouds and identify faces or familiar shapes. It seems that humans has a set of hyper-parameters tuned for rapidly differentiate shapes.

The LLE works in a similar way. In the learning step LLE discriminates those individual that are close, in a given metric, in the high-dimensional space, and extract the smallest number of common features among them. That is, from the high-dimensional space, LLE tries to extract the main dimensions that are enough to describe the characteristics of the data, i.e. its internal structure, leading to a considerable reduction of the dimensionality. For instance authors in [11] were able to represent any human liver with only two independent (intrinsic) parameters.

The algorithm for the LLE reads as follows. Given a set of individuals $\mathbf{y}_i$, being i an individual and $\mathbf{y}_i \in V^D \subset \mathbb{R}^D$, living in the high-dimensional space and equipped with the Frobenius' norm, i.e. $\|\cdot\|_{2(V^D)}$, for distance measurements. Let us suppose that each individual i has its counterpart in a low-dimensional space. Let $\mathbf{x}_i$ represent the individual i in the low-dimensional space: $\mathbf{x}_i \in V^d \subset \mathbb{R}^d$, being $d << D$. Space V^d is also equipped with the Frobenius' norm $\|\cdot\|_{2(V^d)}$. We admit that any individual i in the high-dimensional space can be interpolated with its neighbours $\mathbf{y}_i = \sum_j W_{ij}\mathbf{y}_j$, where $\mathbf{W}$ represents the weight matrix. Then the LLE works first solving the following problem to obtain the best weight matrix:

$$\min_{\mathbf{W}} \epsilon(\mathbf{W}) = \sum_i |\mathbf{y}_i - \sum_j W_{ij}\mathbf{y}_j|^2 \tag{1}$$

where $W_{ij} = 0$ if $\|\mathbf{y}_j - \mathbf{y}_i\|_{2(V^d)} > \tau$ (being τ a user defined value) or when the limit of the number of neighbours is exceeded. This equation allows us to extract the shape of the data structure in the manifold. The second step is to extract the d main dimensions on which the manifold lives. In order to do that, the following minimization:

$$\min_{\mathbf{x}} \Phi(\mathbf{x}) = \sum_i |\mathbf{x}_i - \sum_j W_{i,j}\mathbf{x}_j|^2 \tag{2}$$

yields to a eigenvalues and eigenvectors problem [9]. In this case, the smallest eigenvalues indicate the uncorrelated dimensions that define the reduced number of dimensions d and their eigenvectors represent the coordinates of each individual in the reduced-dimensional space.

2.2. Genetic algorithm

The GA are heuristic optimization algorithms inspired by the evolution of species. It is assumed that one individual is represented by k parameters, that is individual l can be represented by $\mathbf{t}_l \in \mathbb{R}^k$. A population is a set of N_p individuals of the same generation. GA start by assuming a random, or quasi-random, initial population and evaluating the cost function $f(\mathbf{t})$ for each of the N_p individuals of the initial population. Those best performing are called the elite. The next generation is obtained as follows:

- some new individuals are obtained by coping the elite;
- crossover; some new individuals are obtained from the combination of the best individuals;
- mutation; some new individuals are obtained from random mutations of combinations of individuals.

The different proportions between the three different methods for the generation of the new population are defined by the user as well as the size of the population. There is not a strict rule in this context and the experience of the user for each problem is needed to define the configuration of the GA algorithm. The algorithm generates as many generations as required until a stopping criterion is reached. For example, it is common to stop the algorithm when the variation of the cost function has been below a predefined threshold for a user-specified number of generations.

2.3. Hyper-parametrization

One individual in the high dimensional space is defined by a vector, being each one of the components a feature that defines the individual. For instance, a person can be defined with the age, height, weight, salary, etc. However, non of these parameters has the same units neither scale. Thus, some authors propose to scale/normalize all variables to the same range. This implies that all variables in $\mathbf{y}$ will have the same importance/weight on the results extracted from the data. Let us consider that we have a number of variables in $\mathbf{y}$ (sex, height, glucose level, smoking habits, length of little finger, ...) that we are going to use to diagnose a certain illness. It is clear that not all variables will have the same relevance. Hence, in the case of the use of the LLE, the naive approach consisting on considering all the variables with the same weight/importance most probably will not lead to the desired clustering. Therefore, the definition of the neighbourhood to be used by the LLE is not arbitrary. Thus, for this work we redefine the vector $\mathbf{y}$ as follows:

$$\tilde{\mathbf{y}} = \boldsymbol{\alpha} \odot \mathbf{y} \tag{3}$$

where $\odot$ is the Hadamard product and $\boldsymbol{\alpha} \in \mathbb{R}^D$ is the vector containing the hyper-parameters that act as weighting factors of the variables. Note that modifying $\boldsymbol{\alpha}$ will produce variations on the results produced by the LLE technique because each vector $\boldsymbol{\alpha}$ will define a different neighbourhood. Hence we propose to find the vector $\boldsymbol{\alpha}$ that makes the LLE technique to produce the most adequate results for the classification process defined as an objective function defined by the analyst. Hence, for obtaining the hyper-parameters, given an objective function $f(\boldsymbol{\alpha})$, an optimization problem is established:

$$\min_{\boldsymbol{\alpha}} f(\boldsymbol{\alpha}) \tag{4}$$

3. Results

The proposed method has been tested against a cohort from the region of Alcoy (Spain). The objective is to classify the patients into sane or fractured according their characteristics. The population, previously anonimized by the corresponding authorities and fulfilling all ethical requirements, has the following characteristics:

- hospital influence area: 140,000 inhabitants;
- population size: 144 patients;

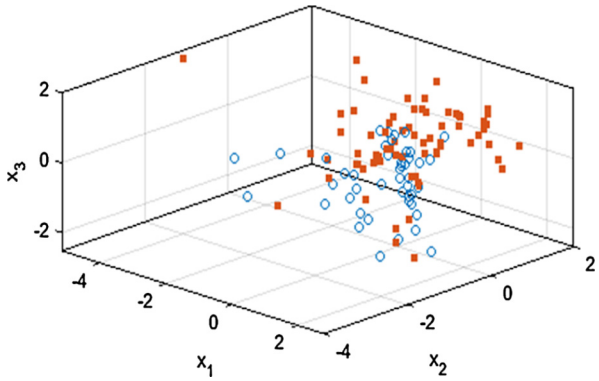

Fig. 1. Coordinates in the low-dimensional space of the patients. Orange points represent the fractured patients and blue balls represent the sane patients.

- 60 patients without fracture;
- 84 patients with osteoporotic fracture;
- training: 80% of the population;
- test: 20% of the population.

The variables defining each patient are the following:

- age;
- sex; female 1, male 0;
- weight;
- height;
- previous fracture; yes = 1, no = 0;
- parent fractured hip; yes = 1, no = 0;
- current smoking; yes = 1, no = 0;
- glucocorticoids; yes = 1, no = 0;
- rheumatoid arthritis; yes = 1, no = 0;
- secondary osteoporosis; yes = 1, no = 0;
- alcohol issues; yes = 1, no = 0.

As you can observe, many of the variables are dichotomous. For those variables, we adopted a binary approach being 0 the negative response and 1 the affirmative one. In the case of the patient's sex, we choose 0 for male and 1 for female. As you can observe, the dichotomous variables in the high-dimensional space have no evident metric, since the assigned values are completely arbitrary.

Without loss in generality, suppose we take two individuals i and j with the same components in the high dimensional space but one dichotomous variable different. The distance, according the norms defined in section 2.1, is 1. This could be generalized for any dichotomous variable. However, when the hyper-parametrization is used, since each component is multiplied by a constant, the metric varies according to values of the parameters $\boldsymbol{\alpha}$. This allows us to accurately evaluate which is the best metric for the proposed objective function. However, in case the that categorical variables are used, which include more than two states without possibility of having any kind of ordinarity between them, more advanced approaches should be investigated. In any case, by increasing the number of variables, it will be always possible to convert any categorical variable into a set of dichotomous variables.

When applying the LLE to the data, without considering any kind of hyper-parametrization, i.e. $\boldsymbol{\alpha} \equiv 1$, and considering the dimensionality of the reduced space $d = 3$ and the number of neighbours 7 the results are represented in Fig. 1. As it can be appreciated in the figure, there is no possibility in finding any pattern nor any cluster for classifying the patients. Therefore the naive approach is not a valid method.

In order find the optimum values of the hyper-parameters $\boldsymbol{\alpha}$ the first step consist in finding an appropriate objective function for clustering the patients. This will depend on the problem at hand. In order to obtain the objective function for this problem the following approach is followed. First the distance between all sane samples and fractured samples is evaluated:

$$L = \|\bar{\mathbf{x}}_{\text{sane}} - \bar{\mathbf{x}}_{\text{fractured}}\|_{2(V^d)} \tag{5}$$

where $\bar{\mathbf{x}}_{\text{sane}}$ represents the mean value of the coordinates of the sane patients in the low-dimensional space and $\bar{\mathbf{x}}_{\text{fractured}}$ the mean value of the coordinates of the fractured patients. The second step consists in evaluating the mean radius of all

Table 1
Value of the objective function and of the hyper-parameters for the three optimization procedures. Age is set to 1 as a reference value.

Parameter	Optimization 1	Optimization 2	Optimization 3
Age	10^{0}	10^{0}	10^{0}
Sex	$10^{3.77}$	$10^{2.91}$	$10^{3.05}$
Weight	$10^{2.25}$	$10^{0.45}$	$10^{0.18}$
Height	$10^{1.83}$	$10^{-1.29}$	$10^{-2.22}$
Previous fracture	$10^{3.91}$	$10^{3.95}$	$10^{3.61}$
Parent fractured hip	$10^{-0.31}$	$10^{0.78}$	$10^{-2.73}$
Smoking	$10^{1.32}$	$10^{-2.63}$	$10^{0.10}$
Glucocorticoids	$10^{2.42}$	$10^{0.53}$	$10^{-0.72}$
Rheumatoid arthritis	$10^{-1.51}$	$10^{1.15}$	$10^{-2.26}$
Secondary osteoporosis	$10^{-1.93}$	$10^{0.48}$	$10^{1.70}$
Alcohol issues	$10^{2.03}$	$10^{-2.59}$	$10^{0.39}$
Objective function	55.63	37.31	35.80

the sane patients $\bar{R}_{\text{sane}}$, with respect to $\bar{\mathbf{x}}_{\text{sane}}$, and the fractured patients $\bar{R}_{\text{fractured}}$, with respect to $\bar{\mathbf{x}}_{\text{fractured}}$. Note that the coordinates in the low-dimensional space implicitly depends on $\boldsymbol{\alpha}$, but for shake of simplicity this will be omitted in the text. Finally the objective function is written as follows:

$$f(\boldsymbol{\alpha}) = \frac{1}{L} + \frac{\bar{R}_{\text{fractured}} + \bar{R}_{\text{sane}}}{L} \tag{6}$$

The first term of the objective function (6) tries to separate the set of sane patients from the set of the fractured ones. Additionally the second term tries to reduce, in a relative manner, the size of the balls containing the fractured patients and the sane patients, respectively.

In order to minimize the objective function, the GA tool implemented in Matlab® is used. It has been configured considering a crossover fraction of 0.8 and a population of 200 individuals per generation and a tolerance of 10^{-12} for the variation of the cost function as stopping criterion. The algorithm stops when the objective function does not change more than this tolerance after 50 generations. Any other parameter is left to its standard configuration. Since the global minimum is not guaranteed in this kind of algorithms, the optimization has been carried out three times and the best performing one has been considered as valid. The training population has been randomly chosen for each optimization procedure. Table 1 shows the values of the hyper-parameters for the three optimizations and the value of the objective function. The best performing hyper-parameter combination corresponds to the third optimization since the objective functions takes the smallest value. Additionally, it is also observed that some hyper-parameters, such as those corresponding to “sex” and “previous fracture” systematically take high values dominating among the others. On the other hand, those with smaller values suffers from high variations. This fact could be due to two reasons: *i*) the GA does not find the global minimum or, more likely, *ii*) the richness of the data is not enough to discriminate these variables.

Fig. 2 shows the representation of the patients in the reduced-dimensional space, when considering the hyper-parameters obtained by the third optimization. It is clearly observed that the coordinate x_2 discriminates the sane and fractured patients. Thus, it will be used as diagnoses function. It is also worth mentioning that the two isolated points corresponds to the patients with alcoholic problems.

3.1. Diagnosis function

Once the hyper-parameters are obtained, the coordinates of a new patient to be diagnosed can be interpolated from its neighbours, since the LLE preserves the interpolation weights between the high-dimensional space and the reduced-dimensional space. Using this procedure the coordinates of the new patient can be obtained in the reduced-dimensional space without re-evaluating the hyper-parameters. In Fig. 2, the reduced coordinates of the new patient (orange ball) were obtained following this procedure.

From the results shown in Fig. 2, the variable x_2 can be used as the diagnosis function, as follows:

$$\Upsilon(x_2) = \begin{cases} x_2 \geq \varsigma \text{ then sane} \\ x_2 < \varsigma \text{ then fractured} \end{cases} \tag{7}$$

where ς is a threshold value. By varying ς it is possible to obtain the Receiver Operating Characteristic (ROC) curve represented in Fig. 3. As it can be observed, the ROC curve has an appropriate shape, showing an Area Under the Curve (AUC) of 0.76. Authors in [4] report that for the population studied in their contribution with FRAX® obtains an AUC of 0.75 with the same variables as input data.

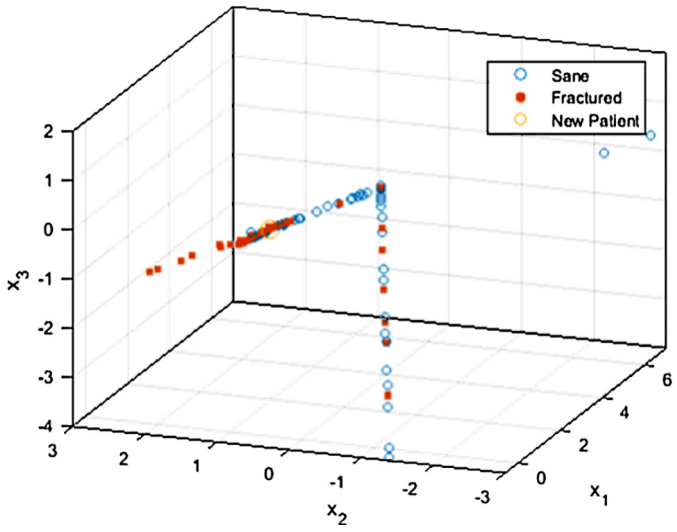

Fig. 2. Coordinates of the patients in the low-dimensional space after using the hyper-parameters of the third optimization. The orange points represent the fractured patients, the blue balls represent the sane patients and the orange ball represent a patient (called external patient) not used in the training set.

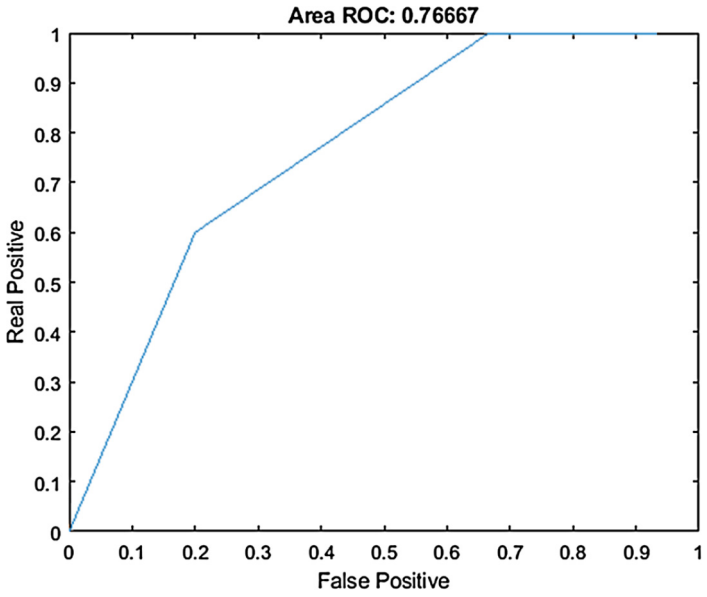

Fig. 3. ROC curve for the proposed diagnoses function.

4. Conclusions

The contribution proposes a methodology, based on non-linear dimensionality reduction techniques combined with hyper-parametrization and GA, for obtaining a tool for the diagnosis of the hip fracture risk in the population of a local region, considering only clinical data. The results of the method allow us to identify the most important variables and are of similar accuracy to the established techniques used for global population.

Due to the variability of the results of the three optimization processes of the hyper-parameters affecting certain variables, the improvement on the training set of patients will improve the results as a result of the increase in the robustness of the optimization algorithm. In any case, with the available data, an small set of variables is identified as the most relevant for the diagnoses, obtaining a AUC of 0.76.

In order to improve the accuracy of the method, the authors propose to add, in a future work, some mechanical variables to characterize each patient. The additional variables could include a stress analysis of the femur neck of the patient. This stress analysis will implicitly include the morphology of the femur neck and the bone density distribution which are, from a mechanical point of view, of a high interest for describing the mechanical behaviour of the femur neck.

Acknowledgements

The authors wish to thank the Spanish “Ministerio de Economía y Competitividad”, the “Generalitat Valenciana” and the “Ministerio de Educación” for their financial support received through the projects DPI2017-89816-R, Prometeo 2016/007 and FPU016/07122, respectively.

References

[1] S. Gnudi, E. Sitta, E. Pignotti, Prediction of incident hip fracture by femoral neck bone mineral density and neck-shaft angle: a 5-year longitudinal study in post-menopausal females, Br. J. Radiol. 85 (1016) (2012) 467–473.
[2] J.A. Kanis, F. Borgstrom, C. De Laet, H. Johansson, O. Johnell, B. Jonsson, A. Oden, N. Zethraeus, B. Pfleger, N. Khaltaev, Assessment of fracture risk, Osteoporos. Int. 16 (jun 2005) 581–589.
[3] R. Marks, Hip fracture epidemiological trends, outcomes, and risk factors, 1970-2009, Int. J. Gen. Med. 3 (apr 2010) 1–17.
[4] E. Sornay-Rendu, F. Munoz, P.D. Delmas, R.D. Chapurlat, The FRAX tool in French women: how well does it describe the real incidence of fracture in the OFELY cohort, J. Bone Miner. Res. 25 (10) (2010) 2101–2107.
[5] X. Jiang, M. Gruner, F. Trémollieres, W. Pluskiewicz, E. Sornay-Rendu, P. Adamczyk, P.F. Schnatz, Diagnostic accuracy of FRAX in predicting the 10-year risk of osteoporotic fractures using the USA treatment thresholds: a systematic review and meta-analysis, Bone 99 (2017) 20–25.
[6] F. Perbet, B. Stenger, A. Maki, Random forest clustering and application to video segmentation, in: British Machine Vision Conference, BMVC 2009 – Proceedings, 2012, pp. 100.1–100.10.
[7] C. Aggarwal, Data Classification: Algorithms and Applications, CRC Press, 2015.
[8] T. Shi, D. Seligson, A.S. Belldegrun, A. Palotie, S. Horvath, Tumor classification by tissue microarray profiling: random forest clustering applied to renal cell carcinoma, Mod. Pathol. 18 (4) (2005) 547–557.
[9] S.T. Roweis, L.K. Saul, Nonlinear dimensionality reduction by locally linear embedding, Science 290 (5500) (2000) 2323–2326.
[10] O. Kayo, Locally Linear Embedding Algorithm - Extensions and Applications, PhD thesis, University of Oulu, 2006.
[11] D. González, E. Cueto, F. Chinesta, Computational patient avatars for surgery planning, Ann. Biomed. Eng. 44 (1) (2016) 35–45.
[12] B. Scholkopf, a.J. Smola, K.R. Muller, Kernel principal component analysis, in: Computer Vision and Mathematical Methods in Medical and Biomedical Image Analysis, vol. 1327, 2012, pp. 583–588.
[13] D. González, J.V. Aguado, E. Cueto, E. Abisset-Chavanne, F. Chinesta, kPCA-based parametric solutions within the PGD framework, Arch. Comput. Methods Eng. (2016) 1–18.
[14] L.M. Chen, Topological data analysis, in: Mathematical Problems in Data Science: Theoretical and Practical Methods, 2015, pp. 101–124.
[15] M. Mitchell, Elements of Generic Algorithms, Massachusetts Institute of Technology, 1998.
[16] S. Wold, K. Esbensen, P. Geladi, Principal component analysis, Chemom. Intell. Lab. Syst. 2 (1987) 37–52.
[17] B. Schölkopf, A. Smola, K.-R. Müller, Nonlinear component analysis as a kernel eigenvalue problem, Neural Comput. 10 (jul 1998) 1299–1319.