RESEARCH   **Open Access**

Particularities of data mining in medicine: lessons learned from patient medical time series data analysis

Shadi Aljawarneh[1]* , Aurea Anguera[2], John William Atwood[3], Juan A. Lara[4] and David Lizcano[4]

Abstract

Nowadays, large amounts of data are generated in the medical domain. Various physiological signals generated from different organs can be recorded to extract interesting information about patients' health. The analysis of physiological signals is a hard task that requires the use of specific approaches such as the Knowledge Discovery in Databases process. The application of such process in the domain of medicine has a series of implications and difficulties, especially regarding the application of data mining techniques to data, mainly time series, gathered from medical examinations of patients. The goal of this paper is to describe the lessons learned and the experience gathered by the authors applying data mining techniques to real medical patient data including time series. In this research, we carried out an exhaustive case study working on data from two medical fields: stabilometry (15 professional basketball players, 18 elite ice skaters) and electroencephalography (100 healthy patients, 100 epileptic patients). We applied a previously proposed knowledge discovery framework for classification purpose obtaining good results in terms of classification accuracy (greater than 99% in both fields). The good results obtained in our research are the groundwork for the lessons learned and recommendations made in this position paper that intends to be a guide for experts who have to face similar medical data mining projects.

Keywords: KDD, Data mining, Physiological signals, Medical data mining, Lessons learned, EEG, Stabilometry, Sensors

1 Introduction

Computing plays an increasingly important role within medicine. It helps to diagnose diseases and also plays a very important role in the treatment of different diseases. Computing is used in almost any healthcare field: management, diagnosis, decision support and so on.

There are now countless computer systems that help physicians to do their job. Physicians use information technologies to gather a great deal of information about patients (e.g. physiological signals). However, this information is not easy to analyse and process and requires the application of special-purpose tools. As illustrated by the success stories described by Shadabi and Sharma [1], data mining techniques have a huge potential for analysing such large volumes of stored medical data in order

to discover knowledge. Generally, the extraction of useful, tacit and previously unknown knowledge from large data volumes is a process called Knowledge Discovery in Databases (KDD), which ranges from the understanding and preparation of the data to the interpretation and use of the discovered knowledge (results of the KDD process). Data mining is the analysis step of the KDD process. The goal of this step is the extraction of patterns from data [2].

Data mining, and the KDD process generally, has been successfully applied in different branches of medicine over recent years: diagnosis of hypocellular myelodysplastic syndrome and aplastic anaemia [3], malignant mesothelioma disease diagnosis [4], analysis and detection of diabetes [5], hemodynamic prediction for abdominal aortic aneurysm disease [6], prediction of very short-term versus short-/intermediate-term post-stroke mortality [7], cancer prognosis prediction [8], or discovery of acupoints and combinations with potential to treat

vascular dementia [9] are just some examples. A comprehensive survey on the contributions of Data Mining to medical data can be found in [10].

The KDD process is divided into different phases, each of which produces different results [2]. However, the characteristics of the medical domain further complicate the way in which the KDD phases are addressed, data are processed and results are interpreted. Thus, KDD in the medical domain requires specialized treatment.

Cios and Moore [11] show the particularities of medical data in terms of their heterogeneity and the special treatment required because of the applicable ethical, legal and social constraints. In this paper, we intend to go a step further, offering guidance on how to address the different KDD process stages in the medical domain especially when analysed data are recorded as time series. This should familiarize data mining engineers with the particularities of this domain before they address a KDD project with medical data, including time series. This paper addresses the following questions:

- Understanding and modelling of medical data, addressing topics like the role of the expert physician in the phase prior to KDD, the conceptual modelling of the medical data and their specific features (heterogeneity, complexity, volume, high dimensionality).
- Selection of medical data.
- Preprocessing of medical data.
- Transformation and reduction of medical data.
- Use of data mining techniques in medicine and their application to real medical problem solving.
- Interpretation and evaluation of data mining results and models in the medical field.

The main goal of this paper is to describe the lessons learned obtained from our experience applying the KDD process to the medical domain [12–21]. These lessons are the main contribution of the paper. The satisfactory results obtained in the case study (stabilometry and electroencephalography) are the groundwork for the lessons learned and recommendations made in this paper. For this reason, the case study will be also described in this paper.

Although the lessons learned are primarily underpinned by the experience that we have gained from data mining projects in the domain of medicine, we have tried to extrapolate as much knowledge as possible to general practice in this field. To be precise, the lessons learned discussed in this article are likely to be particularly useful in projects where, like the ones described in this paper, experts play an important role in domain and medical data, especially physiological signals comprehension as the basis for building data mining models

applicable to medical decision-making. In this regard, time series are a data type of special interest in the medical domain because they have become very widespread over recent years and they are complex to process and analyse. Additionally, as we will see, many other types of medical data are often transformed into time series for analysis. Thus, although mostly learned from medical domains where time series play an important role, many of the lessons highlighted in this article may also be extrapolated to other branches of medicine.

The remainder of the paper is organized as follows. Section 2 discusses the role of computing in the medical field and gives a general description of the KDD process and existing data mining techniques. Section 3 surveys the KDD methods applied and the lessons learned from its application, stressing the particularities of the medical domain which have to be considered to successfully undertake a data mining project with medical data. Section 4 describes the case study and the experience that we have gathered from real medical data mining projects. A key result of these projects was the development of a knowledge discovery framework for medical data. Section 5 discusses the lessons learned and its application. Finally, Section 6 details the conclusions of this experience and the challenges facing data mining in the medical domain and possible solutions, as well as future research lines to be addressed in this field.

2 Background

This section sets out to give an overview of some medical decision support computer systems (Section 2.1) and a general description of the KDD process and data mining techniques (Section 2. 2) usually used as part of this process in order to discover knowledge from a dataset. Additionally, we give a general account of some issues related to the characteristics and particularities of medical data (Section 2. 3).

2.1 Decision support systems in medicine

Computers play an increasingly important role within all walks of life, as their speed at analysing huge quantities for information makes them a tremendously useful tool. Medicine is unquestionably one of the disciplines that has most benefited from computing.

As an extremely complex and sensitive field, which is constantly changing and innovating, medicine continuously spawns new specialities. Thus, computing is a great ally to medicine. Information technologies help physicians to gather and quickly and efficiently process huge quantities of data about patients. Thanks to computing, physicians are relieved of their most repetitive and time-consuming tasks and can spend more time on tasks where their experience and knowledge are more

useful. On top of the time factor, computerized tools help them to diagnose and treat patients.

Diagnosis is a procedure by means of which physicians identify a disease from patient symptoms. Diagnostic decisions place physicians in a dicey position, as the patient's health may depend on a rapid and correct diagnosis. In this respect, decision support systems are very valuable to physicians.

The computer systems traditionally used to diagnose diseases are known as expert systems. Expert systems earn their name from the fact that they emulate behaviour of an expert (in this case, a physician) in a particular domain [22]. The first expert systems used in the real world were MYCIN [23] and Dendral [24, 25]. Since then, many expert systems have been designed in the field of medicine, some of the latest of which are especially interesting [26–28].

However, expert systems are not the only computer systems used in medicine. Decision support systems also play a very important role in medicine, particularly in the diagnosis of cancer [29], infectious diseases [30], cardiovascular diseases [31], urinary and reproductive system diseases [32], dementia [33], retinal disease [34], vertigo [35], blunt head trauma [36] and so on. There is a wide range of decision support systems. Some, such as the above, are based exclusively on expert knowledge. Others, however, learn from previous problems (case-based systems) [37, 38] or are representations (e.g. decision trees) that support decision-making (model-based systems) [39]. There are also hybrid approaches, such as the systems illustrated in this article, where expert knowledge is used to gain a better understanding of the domain and KDD techniques are then applied to build models for use in decision-making (e.g. diagnosis) based on the medical data. Note that systems that include knowledge discovered using KDD techniques cannot be considered expert systems as such, because, even though both system types are built in partnership with experts, KDD systems discover most of the knowledge from the data rather than eliciting it from experts.

A comprehensive review on smart decision support systems for health care can be found in [40].

2.2 The KDD process: data mining
The KDD process aims to gather useful, tacit, and previously unknown knowledge from a dataset. To gather this knowledge, the KDD process includes the following stages (which may vary slightly from author to author) [2]:

1. Domain and data understanding. This first phase (which some authors consider to be outside the scope of the KDD process) studies the general characteristics of the data to be analysed and the source domain.
2. Data selection. This phase determines all the sources of data of interest, which are unified in a target dataset.
3. Data preprocessing. The goal of this stage is to assure the quality of the data. To do this, a series of tasks are performed on the dataset generated in the selection phase. These tasks include reducing noise, handling missing values and so on.
4. Data transformation and reduction. In this phase, the preprocessed data are subjected to a series of filters and operations in order to assure that the data format is suitable for running data mining algorithms. These filters include attribute discretization or dimensionality reduction (feature extraction, for instance). Feature extraction intends to build derived values (features) that are informative and easier to deal with than original attributes. As we will describe later, in this research, we have used a feature extraction method based on the identification and characterization of events in time series. In this respect, different feature extraction approaches in medical data have been recently reported in the literature: neuro-fuzzy approach [41], Mann-Whitney test [42], rough set and harmony search [43].
5. Data mining. A series of techniques and algorithms can be applied to the correctly formatted data in order to discover knowledge. These techniques are applied in order to solve different problem types, known as tasks. The most important are:

- Classification. Classification poses the problem of predicting the unknown value of an attribute of a particular individual. This attribute, called class attribute, has to have a finite and generally small set of values. Prominent classification techniques are decision trees, neural networks and Bayesian classifiers [44]. As we will see later, this paper describes a case study where a classification method has been used to classify patients in both the EEG and the stabilometric domains. In this respect, different classifiers have been recently used for classification on medical data by other authors: fuzzy logic and wavelets [45], neural networks [46], multikernel [47].
- Regression. Regression poses a similar problem to classification where possible attribute values are generally quantitative and therefore infinite. There are two major regression techniques depending on whether the generated regression model is linear or non-linear.

- Outlier detection. Techniques for identifying outliers aim to identify the objects within a population that have substantially different features to most of the other objects of this population.
- Association rules. Association techniques aim to find relationships specified by rules between different variables of a database register. They are generally applied to solve the shopping basket problem, which consists of identifying products that are often purchased together. The Apriori algorithm is the best known association technique [48].
- Clustering. Clustering poses the problem of finding groups of similar objects in a heterogeneous population. There are different types of clustering techniques, the most prominent being hierarchical clustering, partitioning clustering, density-based clustering and grid-based clustering.

6. Knowledge interpretation/evaluation. The last step in the KDD process aims to evaluate the resulting models and, if the assessment is positive, interpret the knowledge inferred from the models.

Clearly, KDD is a well-established process divided into phases and tasks. It generally functions as a paradigmatic framework for discovering knowledge from the data of any domain. Medicine is immune either to the unquestionable benefits of being able to access a highly standardized and widely documented framework such as the above. In fact, applying the KDD process to a branch of medicine by documenting and storing (whenever possible) the interim and final results could result in a major advance in medical research based on data analysis.

Analysing the conception of a new medical domain and its data phase by phase, provided it is well enough documented, can save other researchers intending to work on the analysis of the domain data a lot of time and effort later. There are equally important benefits to be gained as regards the data selection phase. For example, the repository storing the collected data (data warehouse or similar) can be made available to other researchers (subject, of course, to the requisite permits and undertakings). All this would lead to increased medical data reuse, thereby adding to the medical community's knowledge.

On the other hand, it is also very useful to process the data before mining the data and executing the specific techniques, as the models output may be evaluated and interpreted for adoption by the scientific community as knowledge and then applied and rated based on their use.

2.3 Medical data from patient examinations

Nowadays, patient-related data do not consist of a single-valued attributes table but adopt a more complicated structure, including complex data types, like, for example, time series.

The data mining techniques can be applied to both single-valued attributes and time series, that is, we can, for example, apply clustering to a population of individuals represented by a set of single-valued attributes and, likewise, to a population of individuals each of which is represented by a time series. Time series are structures that have certain particularities that warrant a specialized field within data mining: time series analysis. A time series can be defined as a sequence TS of time-ordered data $TS = \{TS_t, t = 1,...,N\}$, where t represents time, N is the number of observations made during that time period and TS_t is the value measured at time instant t. The results of medical examinations (electroencephalogram, electrocardiogram, electromyogram, and so on) are very often a time series [49, 50]. Such is the importance of time series in medicine today that even important data types like medical images (radiodiagnosis) are very often mapped as time series for later processing and analysis [51].

Since data mining techniques for time series first emerged, there has been a wide variety of proposals for solving different problem types. Proposals have addressed different distance measures between time series, different ways of reducing their size to remove irrelevant or redundant information, and so on. Generally, there are many types of problems with respect to knowledge discovery from time series. In particular, the authors of this paper have addressed the next ones:

1. Comparison of time series in order to determine how alike two time series are. It is very useful in many domains to determine the similarity value between two time series. In medicine, for example, it is applied to compare patient examinations at two different time points in order to discover the progress of the process of recovery from a disease.
2. Generation of a reference model that represents a set of time series. It is very useful in medicine to generate a reference model of the time series of the individuals suffering from a particular disease, as it can be used to confirm or rule out that another patient has the respective disease.

Although we have analysed different data types, time series associated with patient medical examinations are actually the most widely used data item in the case of the proposal presented in this article. As we will see later, the above time series reference models have been

used in order to classify patients (e.g. for diagnosis). In this case, they were classified on the basis of the similarity of the subject to be classified with either of the reference models (one per class) using time series comparison methods. This differs from other more conventional approaches where traditional classification methods (e.g. decision trees) are applied to tabulated single-valued data.

In this case, as we explain later, we had to work hard on time series dimensionality reduction by identifying regions of interest (events). The expert played a key role in helping the data mining team to understand the domain and its data.

Although we had the support of the expert and access to expert knowledge, the data that we used were still vulnerable to what is known as the data quality problem [52]. The main problem in this case was the missing values. There are different options for dealing with this problem in conventional data sets, where these values are basically inferred. However, it is virtually impossible to reconstruct the missing values in time series or time series snippets (e.g. where examinations are incomplete). Such time series have to be discarded.

The main lesson learned in this respect is that a KDD project involving non-conventional data has to plan for many of the data that are available at the outset possibly having to be discarded on the grounds of low quality.

This all goes to show that while the experience underlying the lessons learned outlined in this article is based on the particular project characteristics, they are useful for generalizing knowledge. In our case, for example, it turned out that the medical expert played a key role in discovering the basic knowledge upon which to establish a good understanding of the domain and the data. On the other hand, the time series structure turned out to be complex and widespread enough to be able to discover lessons with a broad scope of application. It is also true that our research focused on the task of classification.

Far from being special, the above circumstances are an increasingly common feature in the branches of medicine on which KDD projects are applied. These branches, some of which are very new, require important knowledge that only experts (of which there are, in some cases, not many) in such disciplines can provide. Additionally, time series are, as already mentioned, a more and more frequent structure as many medical examinations are based on monitoring certain parameters over a time interval. In fact, other types of data that do not have a specific time component are also processed as time series. Finally, being the most used data mining task for medical decision-making where it can discriminate between alternatives (class attribute values), classification is also a key task in the field of KDD applied to medicine.

3 Methods for knowledge discovery: particularities of mining medical examination data

One of the goals of this paper is to show the particularities to be taken into account when undertaking a data mining project on data from patient medical examinations in order to provide data mining engineers that have never worked in this field with guidance. This guidance should make them better able to correctly understand and model medical data (Section 3.1); select, preprocess and transform data (Sections 3.2, 3.3 and 3.4); apply useful data mining techniques in the field of medicine (Section 3.5); and interpret and evaluate the models output by applying these techniques (Section 3.6).

3.1 Understanding and modelling medical data

In order to gather and handle medical data, data mining engineers must be mindful of their nature. They are normally very sensitive and private data about disease sufferers. An ethical responsibility to preserve the confidentiality of those data and, ultimately, the privacy of the patients to which they refer is uppermost. There should be a standardized mechanism regulating how such data have to be processed. Data security should be guaranteed at all times: during elicitation, transfer to storage media, processing and, of course, publication of scientific results obtained from references such as [11]. Fundamental premises are that patients should give their consent to and be informed about what is being done with their data and possible results.

In this respect, note that it is essential for medical KDD project participants to be acquainted with and declare that they accept the established company or national regulations and procedures, and so on, in order to guarantee data privacy and security. Such regulations are often subject to frequent amendments, and retraining will be required.

In the particular case of our project, the applicable regulations are the European Parliament Directive 95/46/EC [53] and the Spanish Data Protection Act [54]. Directive 95/46/EC establishes that data (particularly medical data) shall be processed fairly and lawfully, legitimately, adequately, accurately and with safeguards. This regulation also stipulates that subjects have to give their consent to data processing by means of an agreement that protects their interests and guarantees that data processing is in the public interest. On the other hand, Spanish law establishes that medical data are subject to special protection and shall not be disclosed by the processing institutions which shall also have to gain the consent of the subjects concerned. Failure to comply with such provisions may lead to penalties of up to 600,000 euros in the particular case of Spain.

The regulations clearly depend on the country in which the research is undertaken, with which national researchers must therefore be acquainted. In this case, a protocol, designed in accordance with the regulations, was put into place. It was based on subjects giving their consent to the use of their data and on confidentiality agreements with project staff to assure data non-disclosure. We also implemented the respective technical and legal safeguards in the computer systems used to store and process the data.

As a previous step in the KDD process in medicine, it is necessary to understand the specific field of application and aims of the physician. All this requires some interaction between the physician and the data mining engineer who should ultimately be familiar with the goals (for example, production of a decision support system), target performance criteria (target response times), the use to which any results will be put (for example, disease diagnosis in a patient group) and so on.

In order to correctly understand the medical domain and the characteristics of medical data, we recommend that data mining engineers should seek out medical experts in the respective domain. Our experience suggests that it is highly advisable to consult a group of experts, known as a panel of experts, rather than relying on a single expert [12–21]. The premises of this panel-based approach are as follows:

1. There are two or more individuals, each characterized by his or her own perceptions, attitudes, motivations, and personalities,
2. Who recognize the existence of a common problem, and
3. Attempt to reach a collective decision.

A panel of experts often participates in different decision-making rounds. The decisions made by each particular member are used as input for new decision-making rounds involving the whole panel. The Delphi method is an example of an expert panel technique. Using techniques like this, experts have access to the decisions of their peers, which can lead them to change or add to the decisions that they made based on the viewpoints of other experts [55].

On this point, it is important to select qualified experts that have a command of the data domain under analysis. Relying on physicians that are not specialists in the respective field may lead to rather unrepresentative results. As regards the number of experts, experience suggests that there should be an odd number of experts (to be able to make a majority decision) totaling at least five [12–21]. A smaller number can result in very biased decisions, whereas too many experts can slow down the process or make it unworkable.

Another important aspect that any data mining engineers should know before they undertake a KDD project with medical data is that, in the more recent medical disciplines, there are relatively few experts with whom they will be able to enter into partnership. Consequently, what few experts there are in a specific field often have limited availability. Failure to take this circumstance into account can lead to deviations from the planned deadlines and schedule, in which case the data mining project will not be a success.

The knowledge elicited form expert panels is used to gain a better understanding of the data. To do this, we have found in our experience that the conceptual modelling of domain data is a fundamental task in order to grasp a particular domain [12–21]. A conceptual model is useful for representing data, their relationships, their meaning and the constraints that have to be met to be able to maintain data integrity. In other words, data models are able to describe the real-world elements that are involved in a particular problem and how they relate to each other. Data models must losslessly and non-redundantly translate the meaning of the real world that they represent.

Therefore, before trying to discover knowledge from any medical dataset, it is fundamental and absolutely necessary to have a profound understanding of these data. On this ground, it is important to provide resources to model and represent datasets. The conceptual modelling of a dataset has many benefits:

- It is useful for clearly establishing relationships among different dataset entities, especially when the dataset contains different levels or hierarchies.
- It is useful for representing the entity attributes, as well as the possible attribute value types.
- A visual data representation is useful for giving a rapid and intuitive overview of the dataset.
- Conceptual modelling is often the basis for later data storage in databases.
- Additionally, conceptual modelling is the potential starting point for automating other tasks such as the comparison of individuals or the generation of reference models.
- Modelling specifies and standardizes data and is the starting point for their transformation to other models of different levels of abstraction [56].

Apart from understanding the data of a particular domain, data mining engineers must be acquainted with the difficulties posed by the particular characteristics of some medical data, especially their high dimensionality, their volume and their impreciseness. These points have been exhaustively addressed by other authors [57] and, on this ground, are not detailed here.

3.2 Data selection

After modelling the data and understanding the fundaments of the domain under study, we are ready to address the early stages of the KDD process. To do this, data mining engineers have to acquire the data to be analysed. In this case, the ideal thing would be a data warehouse, which is helpful for establishing different views of the data to be data mined. But this is not an option in the field of medicine, as not all the medical data recorded at hospitals today are entirely digitized, and those that are have other complications like data source heterogeneity, data redundancy, and so on [58]. In our experience, data acquisition is a mostly ad hoc and manual process [12–21].

As already mentioned, the provision of well-documented medical data repositories with responsible and controlled access would encourage data reuse and, therefore, increase the quantity of discovered knowledge. In this respect, the use of big data methods (based on efficient distributed information storage frameworks) and open data approaches (which would, of course, be anonymous in this case) could be a major incentive for applying the KDD process in medicine. In this respect, more information on the challenges and future research lines is given at the end of this paper.

In order to speed up data selection, we recommend the use of the XML standard as support for the storage of medical data. This recommendation is based on the fact that there are many tools that are able to infer the data structure of each domain which can then be used to automate the data preparation process. Figure 1 shows a fragment of an XML document generated as a result of a patient examination in the field of stabilometry, one of the reference domains used in this research. We have used this and other similar documents as a data source from which to extract useful knowledge.

A final noteworthy issue with respect to data selection is that datasets in some, and especially the more modern, fields of medicine are likely to have been drawn from very small samples. Because of the dearth of data in such cases, it is really hard to discover useful knowledge, and data mining engineers therefore have to opt for the use of techniques that work well with very few samples. In order to illustrate this circumstance, suppose that we want to test a sample for normality. The usual procedure would be a χ^2 test, but, if we have a set with few elements, we have to use other approaches, like, for example, the Shapiro-Wilk test. The Shapiro-Wilk test is considered to be one of the most powerful tests of normality, especially for small samples of size less than 30. This circumstance also applies when validating classification models requiring the use of training and test sets, for example. Simple classification model validation techniques set aside a subset of the original data as a test set, whereas the other data are used to build the classification model. If there are few data, we may not able to afford to set aside some of the data for the test stage, and it is much better to use advanced validation techniques (for example, cross validation, which is discussed later).

3.3 Data preprocessing

Before applying data mining, the selected data generally have to be cleaned and preprocessed. Data preparation is a fundamental stage within the KDD process whose main aim is to manipulate and transform the data so that they are can be easily displayed and accessed. The success of later tasks within the KDD process very much depends on this phase. Moreover, data preparation is the most time-consuming and one of the trickiest phases of a KDD project.

Data preparation is necessary for several reasons:

1. Real data can be corrupt and therefore lead to the extraction of useless models, possibly because data are incomplete, noisy or inconsistent.
2. Data preparation can reduce the dimensionality of the original data and thereby improve the efficiency of the data mining techniques.
3. Data preparation generates quality data from which it is possible to generate representative and useful models.

The data preparation phase includes very different tasks. In our experience, two have proved to be necessary and useful for successfully undertaking a data mining project in medicine [12–21]. The first is normalization. In medicine, each patient has many attributes, each of which is specified on a particular scale. During the data mining phase, it may be necessary, for example, to calculate the distance between two patients. If the data are not normalized, attributes that are on a higher scale may bias the final result of the comparison. Normalization is therefore absolutely essential. In our case, normalization involved subtracting from each attribute the mean attribute value and dividing it by the standard deviation. The second preprocessing task that we found to be very useful was outlier filtering. Before building a reference model of a set of individuals, for example, it is necessary to locate and remove outliers, as their inclusion may bias the resulting model. This is an example of how a data mining technique (outlier detection, in this case) can be used as a preprocessing technique whose output is the input for another data mining technique. Other filters, of course, may also be applicable and necessary, especially depending on the data mining technique to be applied in later phases.

```
<?xml version="1.0" standalone="yes">
<stabilometry>
    <exploration>
        <explorationId>12745</explorationId>
        <date>2013-09-02</date>
        <patientName>Anonymous</patientName>
        <patientAge>28</patientAge>
        ...
        <rws>
            <directionalControl>71.2</directionalControl>
            ...
        </rws>
        <uni>
            <balanceTest>75.6</balanceTest>
            ...
            <l-c>
                <balanceCond>73.2</balanceCond>
                <stab-TS>
                    <timestamps>1000</timestamps>
                    <lf>
                        <timestamp>1</timestamp>
                        <value>41</value>
                    </lf>
                    <lf>
                        <timestamp>2</timestamp>
                        <value>41</value>
                    </lf>
                    ...
                    <lf>
                        <timestamp>999</timestamp>
                        <value>92</value>
                    </lf>
                    <rf>
                        <timestamp>1</timestamp>
                        <value>93</value>
                    </rf>
                    ...
```

Fig. 1 XML snippet containing patient stabilometric data. The element <stabilometry>, the name of reference domain, represents the root of the document. Then, it appears the element <exploration>, common to any domain. Afterwards, some general data are included such as ID, exploration date, patient name, age, and so on. Finally, there are elements for each test (e.g. <rws>, <uni>) and feature measured (e.g. < directionalControl>), including the recorded time series (<stab-TS>), containing the number of timestamps (<timestamps>) and values (<value>) for each timestamp (<timestamp>)

3.4 Data transformation and reduction of medical data including time series

For data mining techniques to be able to operate on the preprocessed data, they have to be transformed and their complexity reduced. In the particular case of medicine, one of the main tasks for transforming and reducing the data is related to the analysis of time series. As already mentioned, this type of complex data emerge in uncountable medical disciplines. A number of different data mining techniques have been proposed to solve the many open problems regarding time series analysis. Most time series analysis techniques scrutinize whole time series [59, 60]. However, there are many domains, like medicine, where it is requisite to focus on certain regions of interest, known as *events*, rather than analysing the whole time series [61]. This applies to areas where the focus is on the analysis of momentary events. Electrocardiogram time series, where the regions of interest denote

a heartbeat, are a good example. The typical electrocardiogram tracing recording a normal heartbeat is a P wave, a QRS complex, and a T wave (Fig. 2[1]).

The main data transformation and reduction task that we considered necessary for medical data within our research was precisely to identify and characterize the events in medical time series.

In order to be able to discover useful knowledge from such time series, it is first necessary to identify the events as they are the only regions of the time series that provide information of interest. Event identification is an open problem that is not addressed by most existing techniques, which are either only applicable to particular domains for which they propose ad hoc mechanisms or propose the identification of time series breakpoints that do not adopt expert knowledge and

---

[1]Taken from http://perarduaadastra.eu/2010/01/el-electrocardiamaese-garabato-con-picos-y-curvas/
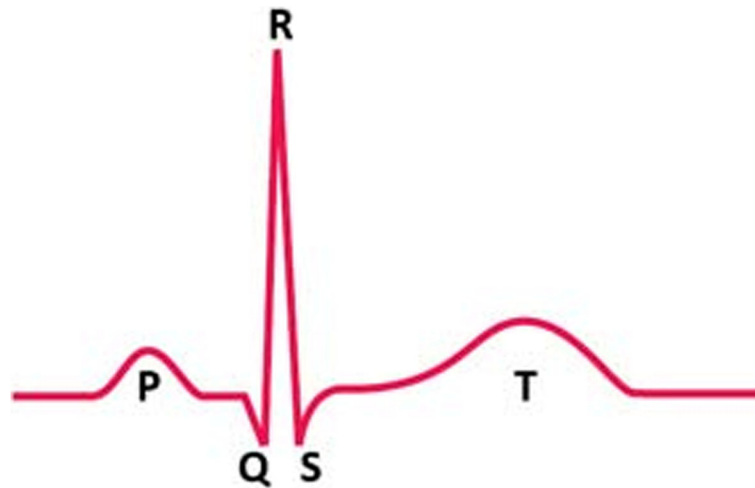
Fig. 2 Section of an electrocardiogram. The P wave on an electrocardiogram represents atrial depolarization (which results in atrial systole), T wave represents the repolarization of the ventricles and QRS complex is the combination of graphical deflections occurring on a typical electrocardiogram between P and T waves. Taken from http://perarduaadastra.eu/2010/01/el-electrocardiograma-ese-garabato-con-picos-y-curvas/

are therefore meaningless. To solve this problem, we recommend the use of general high-level tools like, for example, the events definition language for multidimensional time series proposed as part of our research [18]. This language is designed to be general enough to be applied in any domain. A translation tool developed as part of our research is responsible for automatically generating high-level source code from the defined events. This source code can be instantiated using specific time series returning a list of events in the time series. This tool is thus capable of reducing the dimensionality of the time series, which it converts into a sequence of events that make sense to physicians.

Finally, exactly which transformation techniques are to be applied will generally depend on the needs imposed by the algorithms that will be executed later in the data mining phase.

3.5 Data mining techniques on medical examination data for decision-making support

Over recent years, different data mining techniques have been proposed to discover useful knowledge from medical data, yielding very satisfactory results as classification models based on logical diagnostic rules [62], neural networks [63], k-nearest neighbour algorithm [64], decision trees [65], Bayesian networks [66] and so on. In the following we list what, according to our experience, we believe to be the data mining techniques that are most applicable to medicine:

- Classification. Classification is unquestionably one of the most applicable techniques to problem-solving in the medical field. Being able to classify a particular patient within one category or another is very

useful for medical experts. From our experience of application domains, we have found that it is very useful to be able to categorize a new patient into the {*ill, healthy*} classes because this constitutes an important decision support tool for physicians in the process of diagnosing a particular disease for that patient [12–21]. As a lesson learned, a very useful approach to classification is to build reference models (M_{ill} and $M_{healthy}$) that specify the features of the patients in each {*ill, healthy*} class by means of a learning process. In order to determine whether or not another patient suffers from a particular disease, the patient data are compared with each of the above two models. The physician can tell whether a patient might be suffering from the respective disease if the patient is more similar to the model of the ill patients (M_{ill}) than to the model of the healthy patients ($M_{healthy}$). A pair of models has to be built for each of the diseases to be studied {($M_{ill\_disease\_1}$ and $M_{healthy\_disease\_1}$), ($M_{ill\_disease\_2}$ and $M_{healthy\_disease\_2}$), …, ($M_{ill\_disease\_n}$ and $M_{healthy\_disease\_n}$)}.

- Above, we mentioned a process of classification based on the construction of reference models and the calculation of similarity of patients to models. Generally, the best way of going about building a reference model from a set of patients is to identify the features exhibited by most patients that are members of this group. For example, if we have a group of patients each represented by a time series with events, the recommended mechanism would be to identify those events that are present in a sufficient number of time series. If they have single-valued attributes, it is advisable to use statistics like

the mean (for continuous quantitative attributes) or the mode (for other attributes).

- On the other hand, the comparison of a patient with a model requires the use of a metric specifying how good a match there is between the patterns in the model and the patient data. If attributes are single valued, classical distance measures (city-block, Euclidean or Minkowski) can be used. However, the distance between time series, especially if they contain events, is not easy to calculate, as it requires a previous step to identify which events of a series occur at the same time as others.

- If each patient is represented by one single-valued attribute table, other more conventional classification techniques can be used, such as decision trees [67] (e.g. C5.0, CART or CHAID) or case-based methods [68] (e.g. k-nearest-neighbour).

- Regression. Regression techniques are also useful in medicine. In some cases, it is impossible to conduct a certain examination to gain a specific medical sign because of the patient's condition. It can be very useful to have regression models to estimate this sign. Besides this direct application, regression techniques can be very useful as part of the KDD process. Different model fits can be used for regression, ranging from the most basic linear models to other more advanced (polynomial, logarithmic, exponential, and so on) models.

- Outlier detection. We consider that it is very beneficial to detect and filter out outliers as part of the preprocessing stage as the resulting models will be more accurate and offer a better quality representation of the original data to be data mined.

- Outlier detection techniques are also very useful in the field of medicine. Thanks to these techniques, it is possible to detect patients with out-of-range parameters. In order to determine the respective ranges, it suffices to build a reference model of healthy patients where the respective ranges of normality are inferred from certain statistical estimators like the mean and standard deviation. If such ranges represent the normality of healthy patients, the fact that a patient is out of range can be an important sign of the existence of a particular disease.

- There are different approaches for identifying outliers: proximity-based, clustering-based, density-based methods and statistics-based methods. It is usual practice to generate an outlier factor (OF) for each object. This factor indicates the degree of certainty that the object is an outlier. Then, a threshold ∂ is established for that outlier factor above which the objects are considered outliers. We believe that all the existing techniques have

strengths and weaknesses. Therefore, a proper understanding of the different techniques is required to select one or more applicable techniques to be used. A major drawback of the techniques is that they do not consider the natural variance between domain objects. In the field of the sports medicine, for instance, consider a rugby team. A rugby team includes many, quite different clusters of players (a cluster of corpulent and ungainly players, other very athletic and nimble individuals, and so on). In this type of domains, the data mining engineers have to implement rather *permissive* outlier detection techniques as the fact that there are isolated objects does not necessarily mean that they are outliers. In our particular case, we propose the use of a variance factor θ that indicates the degree of natural variance that is useful in the data of each domain [19]. The variance factor should be used to establish the threshold by means of a function f, such that $\partial = f(\theta, OF_i)$ i, where OF_i is the outlier factor calculated for each object i. This function should be monotonic decreasing with respect to the variable θ such that the threshold value ∂ is inversely proportional to the variance factor value θ.

- Association. Association techniques are also widely applicable in the analysis of medical data. For example, they can study relationships between variables observed in the patient and others that mark the existence of a particular disease. Also, they can be used to locate symptoms that, if they appear together, are a sign of disease proneness. The study of the effectiveness of treatments applied together is another possible application of association techniques in data mining. *Apriori* is the main association algorithm, of which there are several variants [69].

- Clustering. Our experience suggests that clustering techniques tend to be less often used than other data mining techniques for discovering models that are directly applicable to medical decision-making [12–21]. However, they are in our view an indispensable intermediate technique in this process. Clustering techniques are a powerful tool whose results provide the input for data mining algorithms. The identification of groups of similar objects speeds up other later tasks like, for example, outputting the most representative objects of a set.

- There are different types of clustering techniques: hierarchical, partition, density-based methods and grid-based methods. In our experience, one of the most useful techniques for the medical domain is the hierarchical clustering technique as there is no need to specify the target number of clusters, a task that physicians who are unfamiliar with this

terminology find hard to do [12–21]. Hierarchical clustering techniques are based on the generation of ordered series (hierarchies) of clusters. The hierarchical structure is represented as a tree called a dendogram (Fig. 3).

- The advantage of this clustering algorithm is that it is semi-automatic. It automatically determines the number of clusters without the intervention of any user.
- A key issue to be considered for clustering techniques is the choice of distance measure. There are many distance measures. Looking at existing data mining applications, the three most commonly used distances are the city-block, Euclidean and Minkowski distances. Our recommendation to data engineers is to take into account that each distance metric has specific properties that make them applicable or not applicable to their problem. In our case, the distance measures yielded similar results in terms of precision. However, we preferred the city-block distance as it was easier to calculate during the process of outputting clusters from the dendogram.
- Another aspect worth considering when calculating the distance between a pair of objects is the data

type. If data are quantitative, the calculation is immediate. However, there are added complications if we are using qualitative data because it is hard to apply mathematical operations to this class of data.
- Taking the city-block distance as a reference, we recommend the use of formula (1) in order to calculate the distance value between two individuals X and 7 with n attributes (x_i and y_i, with $i = 1..n$).

$$\mathrm{DIST}(X, Y) = \sum_{i=1}^{n} \mid x_i - y_i \mid, \qquad (1)$$

- Equation (1) is fully applicable to continuous and discrete quantitative attributes. However, if the attributes are qualitative, the above formula must be modified slightly. In the case of ordinal qualitative attributes, we recommend the use of a transform function (F_{transf}) that assigns 0 to the first possible value of this attribute and 1 to the last possible value of the attribute, whereas the other possible attributes are assigned equidistant values within the
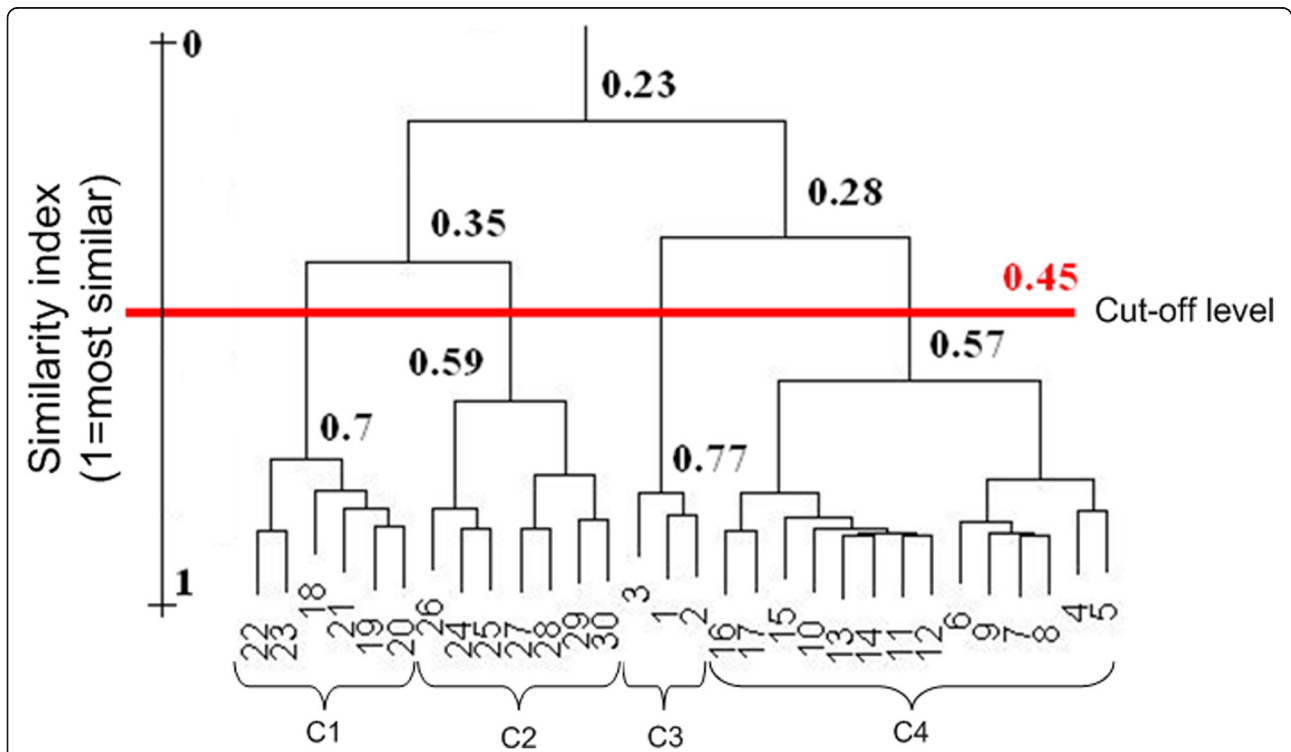

Fig. 3 Specimen dendogram. Example of a dendogram built from 30 objects whose root stores the value 0.23. This means that the minimum similarity value between any pair of objects is 0.23. The value of 0.35 in the left-hand subtree means that the minimum similarity value between any couple of objects underneath this value (specifically objects 18 to 30) is 0.35. The value of 0.28 in the right-hand subtree means that the similarity value between any pair of objects underneath this value (specifically objects 1 to 17) is 0.28. Once the dendogram has been built, we have to determine how to output the clusters. The usual practice is to establish a cut-off level. For the dendogram in this example, the cut-off level was set at $T = 0.45$, outputting four clusters, one for each of the branches cut by the line specified by threshold T

interval [0,1]. For example, if we have an ordinal qualitative attribute with four possible values (first, second, third and fourth), this transform would assign the value 0 to first, 1/3 to second, 2/3 to third and 1 to fourth. The distance value for this type of attributes would be calculated as specified in Eq. (2).

$$\mathrm{DIST}(X, Y) = \sum_{i=1}^{n} \mid F_{\mathrm{transf}}(x_i) - F_{\mathrm{transf}}(y_i) \mid, \qquad (2)$$

- For nominal qualitative attributes, where there is no established order between the elements, we recommend defining, with the help of a domain expert, a complete undirected graph where the nodes are the possible values and the edges are labelled with the distance value between each pair of possible values. The value of the distance between two individuals with this type of attributes is the sum of the normalized distance values between x_i and y_i that are labelled on the edges linking the respective nodes.

On top of the above techniques, specific time series analysis techniques are highly applicable field in medicine. Time series are data structures that, as already mentioned, are very common in medicine. Iconographic time series, like electroencephalographic and electrocardiographic time series, are increasingly common in medicine. Our experiments revealed that an unsuspected quantity of useful knowledge can be gathered from this type of structures. As already mentioned, some of the most applicable time series analysis problems are the comparison between time series and the generation of reference models that represent a set of time series.

In both cases, techniques existing in the literature offer solutions for analysing the time series as a whole. However, the relevant information in most medical time series is confined to regions of interest, called events. Therefore, data mining engineers have to be careful about the selection of the technique so as not to discover superfluous knowledge about regions of time series that are of no interest. To do this, it is essential to rely on mechanisms capable of automatically identifying such events, like, for example, an events definition language.

Another lesson learned on data mining techniques applied to the medical domain is that, irrespective of the techniques used for knowledge discovery, any system developed to implement knowledge discovery techniques for use by medical experts should have the fewest possible input parameters as physicians do not usually have data mining knowledge and are very uncomfortable about having to establish the values of such parameters. A typical case are thresholds, like, for example, the maximum distance threshold defining the boundary for determining whether or not two objects belong to the same cluster (0.45 in the dendogram illustrated in Fig. 3). Another typical parameter is the number of clusters to be output by the clustering method. It is very advisable to minimize the number of parameters and, if there are any, rely on mechanisms that infer them automatically whenever possible.

The last lesson learned on data mining techniques is that there is no one technique that works best for any particular data type. Each data set will have specific features and be of a particular type. This may call for the application of many assorted data mining techniques to discover the target knowledge.

3.6 Models: interpretation and evaluation

The data mining models that we consider most useful are output by applying classification techniques (decision trees, diagnostic rules, neural networks), outlier detection (each individual's outlyingness), regression (regression models), clustering (patient groups) and time series analysis (similarities between time series, reference models, presence or absence of a sequence in a time series and so on).

Sometimes it is not easy for medical experts to interpret these models as they are usually unfamiliar with their structures (trees, Cartesian plots, time series, and so on). Physicians, on the other hand, are very much used to seeing the results in the form of data summary tables, possibly accompanied by support charts, if at all. Therefore, KDD frameworks require a built-in layer that transforms the resulting models into a visual representation that meets physicians' expectations. In most cases, such models should be easily exported to medical reports in formats like PDF.

On the other hand, we have found that it is extremely difficult to validate models output by applying medical data mining. The best option might appear to be to measure the match between the results output by these models and those specified by experts performing the same operations.

However, there are several problems with this approach:

1. It introduces a component of expert subjectivity into the validation process.
2. It overburdens the expert as thousands of operations would have to be performed to validate real systems. To lighten the expert's workload, we would have to resort to a sizeable group of experts, each having their own preferences. Using several

experts brings in different viewpoints that can lead to discrepancies.

3. It is sometimes impossible for experts to come up with a result against which to compare the output of a real system.

In view of this, our recommendation is to test the resulting models as objectively as possible, using indicators that are clearly able to measure the goodness of the resulting models. We believe that the descriptive techniques, particularly clustering, are the hardest to validate, as it is very hard to objectively measure the goodness of the generated clusters. Likewise, the comparison and outlier detection techniques also have the drawback of having to be assessed by comparing the results against domain expert decisions, which, as explained above, is troublesome. In order to check the soundness of an outlier detection method, for example, we have to compare the individuals specified as outliers by the method and by an expert, which may have a different viewpoint from another expert in the same field. The false-positive and false-negative rates are usually used as estimators of the goodness of this type of methods.

On the other hand, predictive regression and classification techniques are easier to evaluate as the success rate of such models can be measured objectively by means of existing validation methods. Of these, the most useful in our view is k-fold cross-validation. In k-fold cross-validation, sometimes called rotation estimation, a dataset D is randomly split into k mutually exclusive subsets $D_1, D_2, ..., D_k$ of approximately equal size. The classifier is trained and tested k times; each time $t \in \{1, 2, ..., k\}$, it is trained on DD_t and tested on D_t. The cross-validation estimate of accuracy is the overall number of correct classifications, divided by the number of instances in the dataset. The most common value is generally $k = 10$. However, when there are few data (a common occurrence in the medical field), a smaller value of k may be used (e.g. $k = 5$) so as to reduce the number of iterations and size of the test set (too large a test set would undermine the capacity of the training set to build representative models).

Apart from the precision of the resulting models, a fundamental indicator of the goodness of the model is the consistency between the model and the knowledge elicited from experts. Although there are approaches for doing this, this is a very troublesome test.

Although there is a wider variety of technique choice for other tasks, the model evaluation techniques discussed in this section are the most commonly used in data mining.

4 Experimental case study: EEG and stabilometry

The view of the KDD process applied to medicine described in Section 3 is based on our experience in the use of data from two branches of medicine: EEG and stabilometry. Section 4.1 generally describes the characteristics of these domains. Section 4.2, on the other hand, describes the KDD process that we enacted on those data to output satisfactory results, which are the groundwork for the lessons learned, presented in this paper.

4.1 Reference domains
4.1.1 The EEG field
Electroencephalography (EEG) is a branch of medicine responsible for studying electrical brain activity. To do this, it uses an electroencephalogram machine, which is able to graphically represent this activity. Electroencephalography is used among other things to diagnose disorders like epilepsy and brain injuries or tumours. The signals generated by an electroencephalogram are time series, whose analysis has brought major advances in the medical domain [70–72].

Electroencephalography has been a tool used exclusively by physicians in the past. Recently, different methods from the intelligent systems field have been applied to discover knowledge from electroencephalographic time series [73–80]. This was a major opportunity to specify medical knowledge and standardize different diagnostic procedures.

Electroencephalographic devices generate time series that record electrical activity (voltage) generated by brain structures along the scalp. EEG signals contain a series of waves characterized by their frequency and amplitude. In EEG time series, it is possible to find certain types of special waves that are characteristic of some neurological pathologies, like epilepsy. Those waves are known as paroxysmal abnormalities and can be considered as events.

During this research, we have taken into account three kinds of events:

- Spike wave: It is a wave whose amplitude is relatively higher than the other waves in the signal. It has a period of between 20 and 70 ms.
- Sharp wave: It is a wave whose amplitude is relatively higher than the other waves in the signal. It has a period of between 70 and 200 ,ms.
- Spicule: It is a sharp wave with an abrupt change of polarity.

The features characterizing these events are the duration and amplitude of the wave.

4.1.2 Stabilometry field
Stabilometry is a branch of medicine responsible for studying human postural control [81, 82]. Postural

control is a key element for understanding a person's ability to perform their routine activities.

Postural control is measured by means of a device called a posturograph. To do this, patients take a series of tests, designed to single out the major sensory, motor and biomechanical components that contribute to people's balance [83]. Figure 4 shows a patient performing a posturographic test.

Although stabilometry was originally devised merely as a technique for assessing a patient's postural control and balance, it is now considered to be a useful tool for diagnosing and treating balance-related disorders [85–90]. Some recent examples of the application of computing techniques on stabilometric data can be found in references such as [91, 92].

Throughout this research, we have used a posturography device called Balance Master, manufactured by NeuroCom® International [84]. The device is composed of a metal plate placed on the floor and divided into two interconnected longitudinal plates. The metal plate is surrounded by a wooden platform, whose sole mission is to prevent patients from stumbling and falling. The patient stands on the metal plate and completes different types of tests, called US (or UNI), LOS, BIS, RWS and WBS [15].

These tests generate time series that measure patient balance. For example, the aim of the US test is to measure how well able patients are to keep their balance standing on one foot with either eyes open or eyes closed. Ideally, patients should remain perfectly static with no sway throughout the test. An interesting event type for this test is located at times when patients lose their balance and put their raised foot down on the platform. This event type is known in the domain as a *fall*.

4.2 Experimentation in reference domains

We have created a knowledge discovery framework for medical data to undertake KDD projects using data from the last two domains. The design of the framework was a troublesome process beset with complications that we had to address. One of the main handicaps that we came up against was the dearth of experts in the reference domains, especially stabilometry, a relatively new discipline and their temporary unavailability. The project would have failed if it had had a demanding deliverables and milestones schedule. Therefore, a lesson learned is that, when dealing with specialist medical experts, the schedule has to be flexible.

Regarding medical data, we soon learned that they are very sensitive data whose acquisition is governed by sometimes very slow protocols. Additionally, there were very often not enough samples because of the complexity of the medical tests and the need to gain the patients' consent to use the medical data.

4.2.1 Understanding the domain and the data

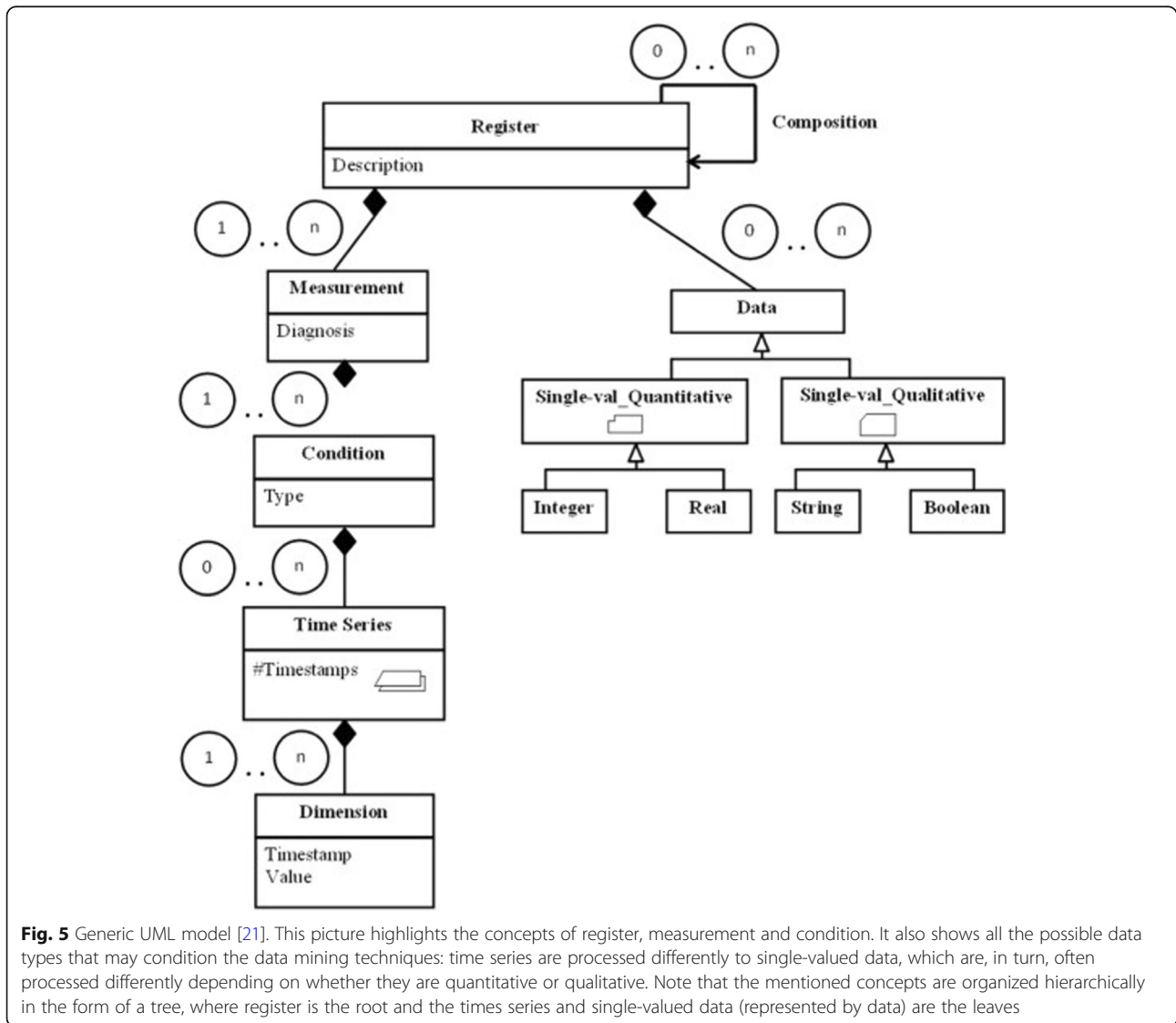In the first place, we addressed the data understanding phase for each of the two domains. Domain conceptual

Fig. 4 Patient performing a test on a stabilometric platform [84]. A patient can be seen on a sensor-based platform (portugraph) which records the pressure exerted on the sensors when carrying out a stabilometric test. The platform is connected by a cable to a computer station which store data for its latter visualization, printing, analysis or exportation

modelling was very useful for this purpose. Looking at these and other areas, we found that, in all the studied cases, there is a central entity or *register* that represents the analysed object (in this case, a patient). Other lower-level data entities including different *measurements* of the object under analysis (for example, a patient EEG) usually depend on the central entity. Some *conditions* are usually altered when these measurements are taken in order to check the behaviour with different parameters (for example, an EEG of an epileptic patient could be repeated immediately after a seizure or a long time after the last seizure). The data collected from each of the measurements under each particular condition may be single valued or adopt more complex structures, like, for example, time series. Data engineers that undertake a project in the field of medicine must be aware that they will come across large volumes of complex, high-dimensional data that cannot be analysed by hand. For example, patient stabilometric data are composed of several tens of time series and several tens of single-valued attributes, and a patient's stabilometric data total around 3 MB of information.

Following the above structure, common to any branch of medicine, we have proposed a general-purpose procedure for conceptually modelling data in UML2 (Unified Modelling Language), as illustrated in Fig. 5. UML is a language used in software engineering that is intended to provide a standard way to visually describe the structure and design of a system. As we will see later, this generic model is able to automate the medical data preprocessing phase. The proposed model includes stereotypes, a mechanism for extending UML2 whereby it is endowed with more meaningful conceptual representations using icons and constraints based on a UML mechanism called *profile*, that can

Fig. 5 Generic UML model [21]. This picture highlights the concepts of register, measurement and condition. It also shows all the possible data types that may condition the data mining techniques: time series are processed differently to single-valued data, which are, in turn, often processed differently depending on whether they are quantitative or qualitative. Note that the mentioned concepts are organized hierarchically in the form of a tree, where register is the root and the times series and single-valued data (represented by data) are the leaves

be represented graphically as shown in Fig. 5. For an exhaustive description of the above stereotypes, see [16].

This generic notation has to be tailored to each domain of experimentation. For example, Fig. 6 shows the model tailored for stabilometry domain data.

The proposed notation has been used as a major support tool for understanding the analysed data and domains, as well as reducing the workload necessary for developing the other tasks. As reported by Lara et al. [16], the domain and data understanding phase can be performed about 1.6 times faster using the proposed notation in the studied domains.

4.2.2 Data selection
During our research, we used several electroencephalographic and stabilometric data sources.

With respect to the electroencephalographic domain, we have used the publicly available data described by Andrzejak et al. [93] that includes data from real patients. The complete dataset consists of five sets (denoted A–E), each containing 100 single-channel EEG segments. These segments were selected and cut out from continuous multi-channel EEG recordings after visual inspection for artefacts, e.g. due to muscle activity or eye movements. Sets A and B consisted of segments taken from surface EEG recordings that were carried out on five healthy volunteers. Volunteers were relaxed in an awake state with eyes open (A) and eyes closed (B), respectively. Sets C, D and E originated from an EEG archive of pre-surgical diagnosis.

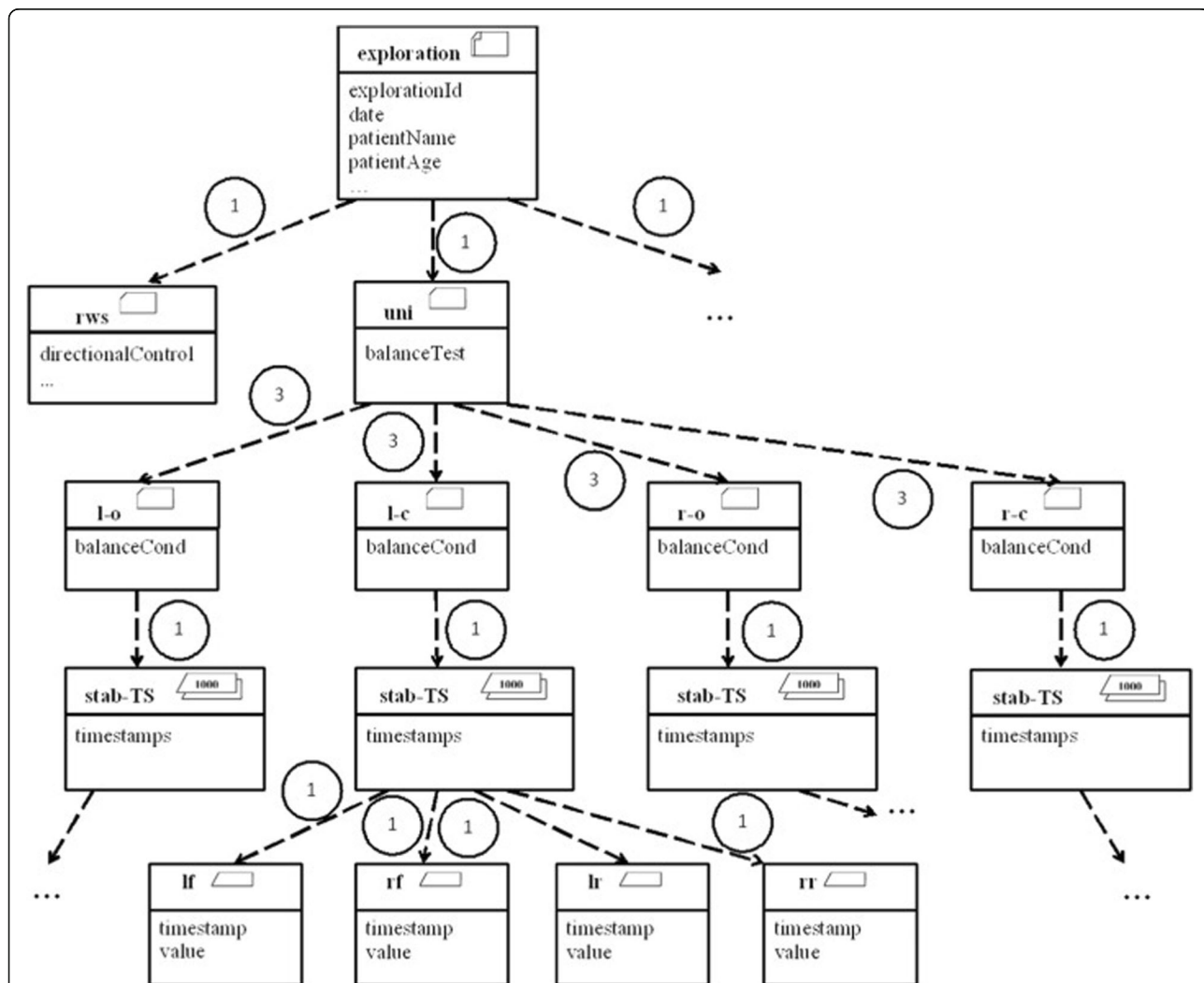As regards the stabilometry domain, we used different stabilometric datasets: (a) data from real top athletes,

Fig. 6 Part of the conceptual model of the stabilometry domain [21]. In this case, the element register is represented by the stabilometric exploration of a patient and the different tests (rws, uni) carried out during that exploration. The element condition is represented by the different alternatives for each test (l-o, l-c, …). Finally, for each condition, there is a time series (stab-TS) containing the recorded values for each dimension of the time series (lf, rf, lr and rr)

including professional basketball players and elite ice skaters and (b) data from real individuals making up a control group of healthy patients that are not professional sportspeople.

4.2.3 Data preprocessing

Based on this generic conceptual model common to both the analysed areas, we devised a mechanism for automatically transforming the data of any medical field to an equivalent format on which data mining techniques can operate directly. This is really a mechanism for automatic data preparation based on the use of description logic that is able to check that a particular XML data source is not ambiguous and is consistent with the generic UML2 data model used as a baseline, automatically finding a finite set of XSLT transformations for preparing data for the application of data mining [17]. The fact that there are XML data sources (see example in Fig. 1) is useful for quickly inferring the domain data structure and using automated mechanisms such as this during the data processing phase.

The architecture supporting this automatic data preparation mechanism uses the proposed UML2 model, which is mapped to description logic by means of a series of transformations. A tool called MOFLON is used to automatically build a rule box called ABox, "AssertionComponent", from the output description logic. The description logic is also used to build a terms box called TBox, "TerminologicalComponent", which contains a description of the terms used (*register*, *measurement*, *condition*, and so on). These two components feed an existing tool, called RACER, whose input is the XML data of any domain and their respective XML schema definition (XSD). RACER outputs two Boolean values: subsumption and instance. Subsumption indicates whether the input component model is a subsumption of the generic model, and instance indicates whether the component syntax is an instance of the generic model and a new ABox' component that contains a series of XSLT mappings. As they are applied to the initial XML data and XSD, the XSLT mappings transform the data into other equivalent data structured to conform to the proposed generic UML2 model.

The automatic data preprocessing mechanism is capable, according to experiments, of reducing the error rate in the preprocessing phase to at most 2%. Besides the low error rate, automatic preprocessing saves time and effort. In any case, the time taken to apply the proposed mechanism is, according to the results, linearly correlated to the size (number of lines) of the generated XML and XSD data files of 0.99 and 0.56, respectively. This linear behaviour evidences the scalability of our proposal.

4.2.4 Data transformation and reduction

Within the proposed framework, after preprocessing the data automatically, it is necessary to apply filters in order, for example, to reduce data dimensionality.

The principal filter for reducing data applied in our framework is the identification of time series events, if only certain fragments rather than the whole time series are of interest. The identification of events in times series is a complex task and requires costly ad hoc methods for each domain. Therefore, we proposed the time series events definition language [18]. This language enables the domain experts to simply and naturally define any events appearing in the time series of each domain.

For example, Fig. 7 shows an excerpt from the event definition for one of the stabilometric domain tests. To do this, we used the notation proposed in our time series event definition language.

The experiments conducted using our language in the reference domains revealed a 98.1% match between the results for the same test dataset returned by the proposed mechanism and a panel of experts used to gauge the quality of this mechanism.

4.2.5 Data mining

The next step after transforming and reducing data is to apply data mining techniques to discover useful models. To be able to apply classical data mining techniques (regression, classification, and so on), we proposed a series of techniques before data analysis. Specifically, these techniques are:

- A method for comparing two patients in order to output a measure of similarity between the two [19]. This similarity measure indicates how alike patients are or how a patient evolves over time. It is the baseline for solving other problems like outlier detection or reference model generation. Therefore, we propose a method for comparing individuals. This method is composed of several algorithms including one for comparing of two time series [12].

- Taking the above comparison method as a starting point, we proposed a method for generating reference models from two or more patients [19]. Note that the algorithm for generating reference models for time series based on the cluster analysis of events [13, 14] using clustering techniques is an important part of this method. In this clustering process, we chose to use bottom-up hierarchical clustering techniques, as they obviate the need to specify the target value for the number of clusters and are also very efficient. The selected distance measure was the city-block distance, as it offered similar results to other distance measures, and was

```
// Time series statement
lf series;
lr series;
rf series;
rr series;

set cand1 {x in lf such that
        (x in max(lr))
        && (∃ y in max(lf) such that abs(x-y) < TIME_AXIS_THLD)
        && (∃ z in min(rf) such that abs(x-z) < TIME_AXIS_THLD)
        && (∃ w in min(rr) such that abs(x-w) < TIME_AXIS_THLD)};
// The TIME_AXIS_THLD is elicited from the expert

set cand2 {y in cand1 such that
        ((lf.value(y) + lr.value(y)) − (mod(lf) + mod(lr))) > INTENSITY_THLD };
        // The INTENSITY_THLD threshold is elicited from the expert

set intersec {z in lf such that
        lf.value(z) == mod(lf) };

event stepping {break_point in cand2,
        start in intersec,
        end in intersec
        such that
            (previous(break_point,intersec) == start)
            && (next(break_point,intersec) == end)};
```

Fig. 7 Definition of events for the US stabilometric test [21]. In this definition, we first state the different dimensions of the stabilometric time series. Then, we define different interesting sets of timestamps based on a series of conditions established by the experts of the domain of application (e.g. *cand1*, *cand2* or *intersec*). Finally, the event type is defined by experts (in this case, *stepping*) using the elements of previously defined sets

more directly calculable as part of the dendogram construction process (see Fig. 3 for an example) than the other distance measures. In order to calculate the distance between objects, the usual distance formulas had to be modified for the purpose of dealing with non-quantitative attribute types. The event attributes had to be normalized during preprocessing in order to output representative clusters.

In order to assure that outliers do not distort the resulting reference models, the reference model generation method also includes an outlier detection and filtering algorithm [19]. The outlier detection method is based on four criteria that are designed to emulate how human beings identify outliers within a set of objects after analysing the clusters containing those objects. This has an advantage over other clustering-based outlier detection techniques that are founded on a purely numerical analysis of clusters.

In all of our proposals, we tried to devise algorithms in which experts had to define the smallest possible number of input parameters, as we found that physicians are not at all happy about rating these parameters with which they are mostly unfamiliar.

The above contributions are combined to solve the problem of classification of individuals, which is a disease diagnosis support tool.

The process of classifying individuals is based on a strategy combining the use of the method of comparing two patients and a method for generating reference models from a set of patients. The strategy followed to classify patients is as follows:

1. Generate, for each class C_i ($i = 1, 2,…, K$), a reference model (M_i) from a training set of individuals.

2. Compare the new patient to be classified (P_{NEW}) with each previously generated reference model M_i (i = 1, 2,..., K).

3. Select the class C_j whose reference model M_j is most similar to the new patient P_{NEW} such that C_j = C_i | similarity(P_{NEW},M_j) = min(similarity(P_{NEW},M_i)) i = 1, 2,..., K.

The results of applying the proposed techniques have been satisfactory, as shown in the following section on the evaluation of the discovered knowledge.

4.2.6 Interpretation/evaluation of discovered knowledge

4.2.6.1 Interpretation A fundamental design premise of the proposed framework was that the resulting models should be easily interpretable considering the type of expert targeted by the framework. Additionally, we had a mind to use graphical elements, like tables, figures and diagrams, at all times in order to help experts with the task of interpretation.

The provision of a stereotyped conceptual model of the data, as shown in Fig. 6, was a great help in this respect. For example, the result of the comparison of the stabilometric data of the two individuals is a tree with the same structure as illustrated in Fig. 6, labelled with the similarity among the individuals at each level of the tree. The physician can browse the tree to study the similarities and differences between the two individuals under comparison. The result of the outlier detection process was a list of outlying values, sorted in top-down order. This list is accompanied by a chart plotting the outlying values against the threshold values, whereby the outlier individuals can be located at just a glance.

On the other hand, the generation of reference models results in an archetypal patient that represents a patient group. Using the proposed standard notation (see Fig. 6 for the stabilometric domain), the archetype has the same structure as any patient. This makes the model a lot easier for the medical expert to understand. Our proposal has a sizeable added value compared to other classical classification proposals, like neural networks for example, as it shows the resulting models in a manner that is easy to interpret and justify. In this respect, we recommend using data mining methods whose output can be easily represented and interpreted by the physicians. In other words, the experts are quite likely not to feel comfortable when taking a decision based on a model that they cannot interpret according to their previous expert knowledge. Therefore, methods like decision trees, archetypal reference models or case-based techniques are preferable to other less interpretable approaches such as neural networks.
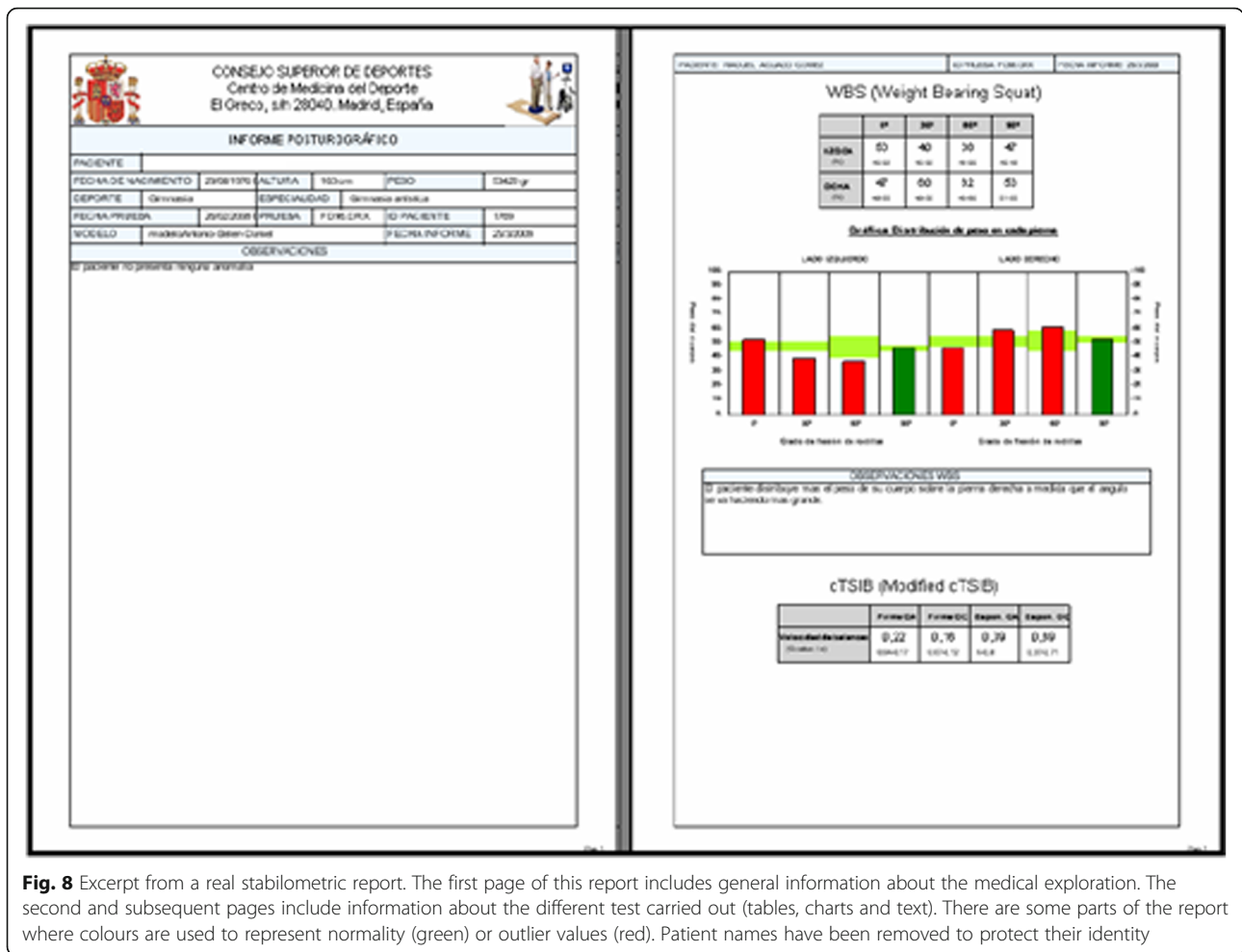
Medical experts are dynamic professionals who are always on the go. They have to travel from one institution to another, visit patients at home or athletes at training facilities, and so on. Therefore, not only do the models have to be displayed by the application, they also have to be exportable to manageable and printable formats. Figure 8 shows an example of the medical report in PDF exported from our framework, showing the summary data tables and support charts. As shown in Fig. 8, the visualization of this type of information (summary table, support charts, and so on) is essential for physicians to be able to correctly interpret the results of applying the data mining techniques and make the right decision.

The qualitative studies that we have conducted as part of this research gathering the impressions of different experts from Spain's National Sports Council about the developed framework reveal that the acceptance and rating the use of the described KDD techniques and tools in the field of medicine are high.

4.2.6.2 Evaluation Domain experts were used to evaluate most of the data mining techniques proposed as part of this framework. In this case, the validation involved comparing the models yielded by applying our framework against those generated by experts. This poses the problems of differing criteria and subjectivity because it requires the participation of more than one expert.

- EEG

In order to validate our proposal in the EEG domain, we used a set of data divided into five subsets (denoted A, B, C, D and E), each containing 100 electroencephalographic times series [93]. This experiment focused on sets C (healthy patients with open eyes) and E (epileptic patients during an episode). It is precisely the wealth of these data and their availability that led me to explore this medical domain in order to validate the proposed model. First, we applied the event definition language in order to discover events (see Section 4.2.4) from a total of more than 200 time series. To evaluate the accuracy of our event identification proposal on the above 200 time series, we asked an EEG domain expert to identify the events in those series and we then applied our technique to do the same thing. For each time series, we measured the accuracy of our proposal using Eq. (3) that measures the degree of similarity (SIM_Exp_Lang) between the number of events identified by the expert ($\#EV_{Exp}$) and by our language ($\#Ev_{Lang}$). Note that this formula offers a normalized result in the

Fig. 8 Excerpt from a real stabilometric report. The first page of this report includes general information about the medical exploration. The second and subsequent pages include information about the different test carried out (tables, charts and text). There are some parts of the report where colours are used to represent normality (green) or outlier values (red). Patient names have been removed to protect their identity

interval [0,1], where 1 indicates a total coincidence between the number of events identified by the expert and by the language. The worst case is when the expert locates events in a series and our system does not identify any:

$$SIM_{\text{Exp}_{\text{Lang}}} = 1 n \frac{\left| \#Ev_{\text{Exp}} - \#Ev_{\text{Lang}} \right|}{\#Ev_{\text{Exp}}}, \tag{3}$$

From a global analysis of the 200 time series, we find that there is a good match between the expert and proposed language, as shown by the mean similarity, which is greater than 96.5%, between the expert and language (Table 1). This value is very close to the ideal.

Based on the events identified previously using the event definition language, we applied and evaluated the outlier detection method. To do this, we applied our time series comparison technique on data from different patients to perform pairwise comparison. This produced a similarity matrix for each pair of individuals. We then ran the outlier detection algorithm on the above matrix. This algorithm returns a list of the outlier individuals from the input matrix. On the other hand, we asked the experts consulted in our research to use conventional techniques to identify the individuals that they considered to be outliers. Table 2 shows the confusion matrix comparing the method and expert criteria.

Table 1 Overall results of the application of the event definition language to the EEG domain

#Series	$\overline{\#Ev_{\text{Exp}}}$	$\overline{\#Ev_{\text{Lang}}}$	SIM_Exp_Lang
200	7.23	7.48	0.965

Table 2 Confusion matrix of the application of the outlier detection method to the EEG domain

		Language	
	Outlier	Yes	No
Expert	Yes	11	1
	No	3	185

Table 3 Overall results of the application of the outlier detection method to the EEG domain

Indicator	Value
Precision	0.79
Recall	91.7
Specificity	98.4
Accuracy	98.0

The above data provide some indicators of the goodness of the outlier detection method (precision, recall, specificity and accuracy[2]), as shown in Table 3. In particular, the accuracy of the above results for the outlier detection method that we propose is 98%.

After filtering out the outliers, the remainder were used to evaluate the classification method based on the generation of reference models against which the individuals to be classified are compared. To evaluate the mechanism, we carried out a series of experiments using the tenfold cross-validation technique, which is a particular case of *k*-fold cross-validation. This is a clearly defined standard technique for validating classification techniques. The goal of this evaluation is to determine the quality of the classifications using the framework in terms of accuracy. The accuracy of a classifier *CF* is the probability of correctly classifying a randomly selected instance $<P_{\text{NEW}}, C_i>$, i.e. accuracy = Pr $(CF(P_{\text{NEW}}) = C_i)$ [94].

First, we generated two reference models, one for each class (M_{healthy} and $M_{\text{epileptic}}$). The first model (M_{healthy}) was created from a training set composed of 90% of time series of the set of healthy patients (*C*). The other 10% of patients were part of the test set. The second model ($M_{\text{epileptic}}$) was generated from a training set composed of 90% of the time series in the epileptic patient set (*E*). The other 10% of patients were part of the test set. The patients in the test sets were chosen randomly.

Both generated models were evaluated to check whether M_{healthy} properly represents the group of healthy patients and $M_{\text{epileptic}}$ is representative of the group of epileptic patients. To do this, we classified the individuals in the test sets according to their similarity with the two generated models (this similarity value was determined using the time series comparison method). This entire process was repeated ten times, varying the training and test sets.

2

		Predicted label	
		Positive	Negative
Known label	Positive	True positive (TP)	False negative (FN)
	Negative	False positive (FP)	True negative (TN)

Table 4 Comparison of the classification of patients by different methods in the EEG domain

Patient type	Knowledge discovery framework in time series (%)	AFINN (%)	Multilayer perceptron (%)
Epileptic	99.86	96.26	96.61
Healthy	98.11	95.12	93

Table 4 reports a comparison of the results of classifying individuals of sets *C* (healthy) and *E* (epileptics) using the proposed knowledge discovery framework, the AFINN system (a fuzzy neural network) and a multilayer perceptron. We find that our proposal outperforms other state-of-the-art systems previously applied to the data set.

From the medical viewpoint, the reference models output by our proposal are, according to the above results, a promising option for epilepsy diagnosis from electroencephalographic examinations. With a classification accuracy greater than 99.8%, our method is capable of correctly classifying patients suffering from epilepsy based on their EEG time series. Note that the proposed method and the resulting models are not designed as a medical diagnosis tool but as a medical decision support tool, in this case, for EEG.

- Stabilometry

In order to validate the proposal on this domain, we used stabilometric data from a total of 33 elite sportspeople, of which 15 are professional basketball players and 18 are elite ice skaters. The studies focused on the US test, a test that provides interesting balance-related information. The events of interest occur when patients lose their balance and step on the platform (see Section 4.1.2). This test has four trials that are each repeated three times during a stabilometric examination. Therefore, we had access to a total of 33(subjects) × 4(trials) × 3(repetitions) = 396 time series.

We repeated the validation procedure on the 33 sportspeople. First, we applied the time series event identification method and compared the results with the events discovered by the expert using Eq. (3). The results are shown in Table 5, revealing a match of almost to 99%.

We then applied the time series outlier detection method as explained above. Table 6 illustrates the

Table 5 Overall results of the application of the event definition language to the stabilometric domain

#Series	$\#EV_{\text{Exp}}$	$\#EV_{\text{Lang}}$	SIM_Exp_Lang
396	2.38	2.41	0.987

Table 6 Confusion matrix of the application of the outlier detection method to the stabilometric domain

	Outlier	Language	
		Yes	No
Expert	Yes	13	2
	No	4	377

confusion matrix highlighting the comparison between the method and the experts.

Based on this matrix, we calculated the outlier detection performance indicators, which are shown in Table 7. Worthy of special note is the overall accuracy value of 98%.

After filtering out the outlier elements, we again enacted a classification process with the two problem classes ($M_{\text{basketball}}$ and M_{skating}). We performed this process using the same validation technique (tenfold cross validation). The results are shown in Table 8, where the classification accuracy for our method is greater than 99%, illustrating that our method outperforms the other analysed methods.

From the medical viewpoint, these models reveal that balance is a variable related to the practised sport. In this case, there is a 99% likelihood of sportspeople being classified in their respective sport. These and other possible models for other sports have potential in the field of sports medicine, as balance (especially of young athletes) can help to classify sportspeople within the discipline for which they are best suited according to their postural control. This would help to point young sportspeople in the direction of the disciplines at which they are most likely to be proficient during early-age sports talent recruitment and possibly increase their future success as professional athletes.

5 Discussion: lessons learned and its application

Table 9 lists the lessons learned and their scope of application, endorsed by the elicited knowledge and satisfactorily validated as explained above. Additionally, in order to clarify their applicability, Table 9 includes two columns representing the following measures:

- Generality. The lessons learned are based on our previous experience in medical KDD projects. We have studied the lessons learned in detail in order to establish how generally applicable we think they are. To do this, each lesson is labelled with a numerical generality rating from 1 (lesson specific to our project) to 5 (lesson generally applicable to any other medical KDD project).
- TS. This field specifies whether the respective lesson learned is (1) not at all specific to projects with time series, (2) not specific to but especially significant for projects with time series, or (3) specific to projects with time series.

In order to establish the values of the above measures, we used a simplified version of the Delphi method in which several researchers related to this research participated. At the end of Table 9, we added the mean value for each of the two fields, and we found that the generality of the lessons learned is high (mean generality of 4.49/5). We also found that, even though time series are an especially relevant data type for our lessons learned (the TS values for 16 out of 37 lessons, or 43.2%, are 2 or 3), many of the lessons learned (21 out of 37, or 56.8%) can also be extrapolated to branches of medicine that use other data types.

It is important to circumscribe the lessons learned reported in this article to a reference field of potential use for researchers working or about to work on knowledge discovery from medical data.

In this respect, we will first list the key characteristics of the project that we carried out to discover the described lessons learned:

a. This is a project in which experts played a key role in providing a preliminary understanding of the domain and the medical data.
b. Based on this knowledge, the following stages were led by the medical data gathered from patient medical examinations.
c. Such medical data were mostly time series.

Table 7 Overall results of the application of the outlier detection method to the stabilometric domain

Indicator	Value
Precision	0.76
Recall	0.87
Specificity	0.99
Accuracy	0.98

Table 8 Comparison of the classification of patients by different methods in stabilometric domain

Patient type	Knowledge discovery framework in time series	AFINN	Multilayer perceptron
Basketball	99.4	98.8	97.7
Skating	99.1	98.1	97.1

Table 9 Lessons learned

		Lesson learned	Scope of application	Generality	S
1. Domain and data understanding	1.1	It is important to have standardized mechanisms regulating medical data processing because of their sensitivity.	General	5	1
	1.2	Ethically, patients should consent to and be informed about the use of their data and any results.	Especially patients with serious diseases	5	1
	1.3	The participants in any KDD project in the medical domain must be acquainted with the regulations regarding data security and privacy and adopt the established procedures for compliance.	General	5	1
	1.4	Expert negotiation techniques, like the Delphi method, should be used.	General	4	1
	1.5	Knowledge elicitation could be biased if there are only one or not many experts. The elicitation process is slow if there are too many experts. We recommend the use of a panel of about five experts.	General	4	2
	1.6	The availability of medical experts may be limited, as a result of which realistic deadlines should be set and scheduling should be flexible.	Especially in new fields of medicine	3	1
	1.7	The use of conceptual modelling tools is very useful for understanding the medical domain data.	General. Also applicable outside medicine	5	2
2. Data selection	2.1	Data are usually acquired manually, and we suggest that this process should be automated with sensor-based instrumentation.	General	4	2
	2.2	Standard data storage formats should be deployed. We recommend the use of XML (and the domain-specific XSD).	General, but especially in domains with structurally complex data	4	2
	2.3	Medical data are sometimes composed of few samples. Data engineers should be proficient in techniques specialized in processing small samples.	Especially in new fields of medicine or for rare diseases	3	1
3. Data preprocessing	3.1	Normalization is very useful given the wide range of orders of magnitude of medical data.	General	4	1
	3.2	Outlier filtering is usually a very useful preprocessing task in medicine, as the outliers can degrade the resulting models.	General	5	2
4. Data transformation and reduction	4.1	The location of regions of interest (events) is one of the most useful reduction tasks in time series analysis.	Medical domains that generate time series	4	3
	4.2	The use of high-level tools akin to medical language is useful for the task of transformation and reduction.	Especially in domains with high-dimensional data	5	2
5. Data mining	Classification				
	5.1	The use of the *ill* and *healthy* classes for each disease is useful for applying diagnosis support classification techniques.	General	5	1
	5.2	Reference models, composed of elements common to patients, are very useful for classification.	General	5	2
	Regression				
	5.3	Regression techniques can be used to estimate values that cannot be calculated if a patient is unable to undergo a particular examination.	Patients with serious diseases or for extremely expensive tests	4	1
	Outlier detection				
	5.4	Outlier detection can be used as a decision support tool for disease diagnosis. The manifestation of characteristics unlike a healthy majority can be a sign that a patient is ill.	General	4	1
	5.5	An object that is isolated, i.e. a long way away from the others, does not necessarily have to be an outlier.	Especially in sports medicine for sports with very different player profiles	4	1
	Association				

Table 9 Lessons learned *(Continued)*

		Lesson learned	Scope of application	Generality	S
	5.6	Association techniques can be used to study several symptoms that, if present together, can be a sign of a disease, or for evaluating the effectiveness of combined treatments.	General	5	1
	Clustering				
	5.7	Clustering techniques are usually applied in advance of other data mining techniques.	General	4	2
	5.8	Hierarchical clustering techniques are very advisable, as it is not necessary to specify the number of clusters K.	General	4	1
	5.9	It is important to try out different distance measures and choose the one that yields the best result in each case.	General	5	1
	5.10	It is hard to calculate distances for qualitative data. It is a good idea to consult an expert if they are nominal as well as qualitative.	General	5	1
	Time series analysis				
	5.11	The generation of time series reference models is useful for outputting archetypes of a particular group of individuals (for example, ill and healthy patients).	General	5	3
	5.12	It is useful to compare time series in order to measure the similarity between a patient and a reference model. This is handy for measuring the evolution of a patient during treatment, for example.		4	3
	General				
	5.13	Data mining methods applied to the medical domain should have the fewest possible input parameters because medical experts are unfamiliar with the technique applied and the input parameters.	General	5	1
	5.14	The methods to be applied in the data mining stage and their input data characteristics decide the preprocessing techniques to be applied in previous stages.	General	5	2
	5.15	The arithmetic mean may be a good estimator for modelling continuous quantitative attributes, whereas the mode is usually the best option for the other attributes.	General	4	1
6. Knowledge interpretation/ evaluation	Interpretation				
	6.1	It is important to display the resulting models in formats with which physicians are familiar, like tables, figures, and so on, rather than data mining formats, like trees, time series, Cartesian diagrams, and so on.	General	5	2
	6.2	Project participants must try to clarify the potential benefits of the knowledge elicited in the medical domain.	General	5	1
	6.3	Mechanisms for exporting models and results that should be able to be easily printed and transferred from one piece of equipment to another.	Very dynamic physicians that work in different environments	4	2
	Evaluation				
	6.4	Evaluating the results of a KDD system with experts is a subjective and, sometimes tedious, process for the expert. On this ground, we recommend the use of an expert panel.	Especially in medical fields where there are few experts	4	2
	6.5	Experts may not be able to evaluate intermediate data mining techniques, as they do for the final result. A flexible schedule reduces the possible negative effects of this circumstance.	General	5	1
	6.6	Descriptive methods, particularly clustering, are usually very hard to evaluate objectively with the help of medical experts.	General	5	1
	6.7	Predictive methods are usually easier to evaluate as they can objectively measure the success rate against historical data. Techniques like tenfold cross validation are very useful.	General, provided that historical data are available	5	1
	6.8	Even if a KDD model is good according to classical data mining criteria, it should not exhibit inconsistencies with the elicited knowledge from experts.	General	5	2
Mean value				4.49	.51

d. The results were gathered by means of experimentation in the EEG and stabilometry domains.

From the above, it is obvious that, as medicine is such a broad field, the lessons learned are not totally applicable to all fields of medicine. The reasons (related to each of the points listed above) are as follows:

a. There are very well-known and widely addressed medical disciplines, where experts may not play such a key role as in our case. Hence, some of the lessons learned in this article may be less applicable to such domains.
b. There are medical data related to many other different fields apart from medical examinations: data regarding treatments, medical resource allocation, hospital management, and so on.
c. Not all domains necessarily generate time series.
d. There are many other medical domains, apart from the branches of medicine used as a benchmark in this article.

Although we admit that they may not be fully applicable to any medical field, we do believe that the lessons learned may be very valuable in many cases on the following grounds (again related to the points listed above):

e. Although some medical disciplines are very well known, it is always a good idea to seek out medical experts. This article outlines lessons learned that may be useful regarding the relationship with experts, especially in the newer branches of medicine.
f. It is true that there are many other fields of medicine apart from the one addressed in this article: patient medical examinations. However, diagnosis-focused decision-making is still one of the most important and most researched fields in the medical domain.
g. While not all the domains generate time series, the lessons learned in this article are very general and inclusive and may therefore be useful for domains with other data types (medical imaging, for example).
h. It is impossible to conduct a study considering all branches of medicine. However, the selected domains have characteristics that make them especially attractive and valid with respect to lessons learned. For example, EEG is one of the branches of medicine that data mining engineers have targeted most in recent years. Stabilometry, on the other hand, is a more recent domain about which there is therefore less knowledge than in other fields and which also generates structurally highly complex data.

6 Conclusions and future lines

The medical domain is very different to other domains for many reasons described throughout this paper. Briefly, they include:

- Problems with medical data acquisition, which are often are not digitized or centralized in a proper structure for analysis.
- Special features of medical data, especially the confidentiality requirement and other important factors such as volume, high dimensionality, complexity and heterogeneity.
- Applicability of the different data mining techniques in the field of medicine.
- Problems with the validation of the data mining models to be used as decision support tools. A correct diagnosis can be a matter of life or death, meaning that quality models are a must in the decision support process for disease diagnosis.

The two specific examples reported in this paper show that is possible to discover knowledge that is useful for medical experts in the routine decision-making process. According to the results obtained, it is possible to conclude that the reference models generated can be used for diagnosis of epileptic disorders (EEG domain) and balance-related sports talent recruitment and mapping (stabilometric domain). The good results are the groundwork for the lessons learned mentioned in this paper. These lessons can be useful for researchers intending to work on similar medical data mining projects.

While medicine is possibly one of the richest domains for data mining engineers, it is definitely the toughest. To overcome this, we think that the scientific community needs to address the following challenges in the medical data mining field:

1. The implementation of mechanisms to guarantee patient confidentiality throughout the KDD process from the collection of the data to the publication of the results of the study. As mentioned throughout the paper, different countries are subject to different medical data processing regulations. However, all these laws aim to safeguard data security and privacy. Therefore, it would be a good idea to set up a forum in which the scientific community participates in order to establish a code of good conduct and a mechanism to guarantee compliance with the law. In this respect, it might be worthwhile considering, as an initial proposal, a

protocol that (a) obliges everyone involved to be acquainted with such regulations, (b) establishes that they should enter into an agreement on data confidentiality and adherence to the regulations and to the established code of conduct, (c) holds the institutions concerned responsible for enforcing the regulations and the code of good conduct and for reporting any abuse to the authorities and (d) demands that patients give their consent for data processing and are informed about the results and potential benefits derived from the research.

2. The design of tools to automate some KDD process tasks that consume a lot of resources, such as data preparation. We have stated at various points of the paper that some phases of the KDD process are rather manual. To automate these phases, it looks like it would be necessary to propose or improve methodologies in order to clearly establish what to do and how to do it. The use of standards could be very useful in this respect, as they encourage the reuse of previously defined techniques and procedures. A good alternative might be to use XML to support data warehousing and to standardize XSLT mappings for different tasks (e.g. preprocessing). This idea could also be extended to other KDD process phases.

3. The proposal of models of representation that are able to capture all the singularities of medical data, their heterogeneity and their structural complexity. Medical data with more complex structures and characteristics are becoming more and more common. We have given some examples throughout this paper. However, other domains might have different particularities that have to be represented using existing mechanisms. An interesting alternative in this respect might be UML, which offers a mechanism based on UML profiles for extending UML notation with stereotypes. These stereotypes visually and/or textually enrich representation models. A good recommendation would be for different researchers to propose such UML extensions for processing data in their reference domains. The result would be a comprehensive body of data representations for different medical examinations. Other researchers might find this very useful as it would reduce the time it takes to understand the domain at the beginning of the KDD process.

4. The specification of secure models for medical big data storage and publication with the aim of increasing efficient data reuse and processing. As in many other walks of life, the amount of data generated in medicine is growing. This raises a number of problems related mainly to data accessibility and processing efficiency. In order to solve this problem, we propose the use of big data frameworks such as the now increasingly popular Hadoop and alike. These frameworks are able to store the information in a distributed manner. This increases data processing efficiency without compromising data security. There is more and more information on such frameworks and many related technologies enabling their integration with special-purpose data mining tools.

5. The proposal of mechanisms to objectively and correctly validate the data mining models generated in the field of medicine. This is perhaps one of the toughest of all the challenges stated in this article to address. There are techniques for validating data mining models from a purely technical viewpoint, but the same does not apply on the more medical side. On this ground, it looks as if the physician who is ultimately going to use the model will play a fundamental role. One solution may be to use expert panels to validate the models, thereby eluding the subjectivity of a single expert. The problem, as discussed in this article, is expert unavailability often motivated by a heavy workload. The institutions employing the experts should be aware of the importance of their reallocation to such tasks, which are crucial for deploying the elicited knowledge.

6. The implementation of visual support tools for the medical expert. Throughout this paper, we have also discussed the fact that medical experts are better acquainted and therefore more at home working with visual representations (of data, models or process) than with more technical representations (source code, engineering notations, and so on). In order to solve this problem, it is necessary to work on implementing middleware to connect the medical expert with the lower level technologies that are often present in medical decision support systems. Precisely, one of our future lines of research aims to address this challenge.

The main line of future research that we intend to undertake is related to the last of the challenges described above. In particular, we intend to devise visual tools that will be built into and improve expert interaction with our framework:

7. A tool for visually defining events in time series which is currently a text-based process (see Fig. 7). We are now working on a visual tool to enable experts to identify events in time series. This tool is

composed of an interface that displays graphs of different time series for experts. Experts can use the mouse to point to the regions that they consider of interest (events). The proposed system infers the conditions that the identified regions meet (analyses aspects such as time series maximums or minimums, changes of trend, and so on), which it maps to the event definition language code. Clearly, this tool acts an intermediary between the experts and the event definition language (which is rather complex for experts who have no experience in using programming languages or similar).

8. A visual tool for managing panels of experts and applying the Delphi method [55, 95]. The system described above is rounded out by another tool that considers the opinion of several rather than just one medical expert. Expert collaboration via the Delphi method renders the elicited expert knowledge more objective and the events more accurate. However, as already mentioned, expert availability is low, for which reason we are working on a tool capable of applying the Delphi method remotely and asynchronously. It is actually a web application that manages the different rounds of the Delphi method by sending out warnings and reminders to the email addresses of the participating experts according to an established schedule. The preliminary results are satisfactory with respect to both lines of research.

Abbreviations
EEG: Electroencephalography; KDD: Knowledge Discovery in Databases; PDF: Portable Document Format; UML: Unified Modelling Language; XML: eXtensible Markup Language

Availability of data and materials
EEG data are publicly available in [93]. Regarding stabilometry, the authors cannot share the data, due to the legal requirements of the foundations that provided the data for this research.

Author details
[1]Faculty of Computer and Information Technology, Jordan University of Science and Technology, P.O.Box 3030, Irbid 22110, Jordan. [2]School of Computer Science, Campus de Montegancedo, Technical University of Madrid, s/n - 28660, Boadilla del Monte, Madrid, Spain. [3]High Speed Protocols Laboratory, Concordia University, 1455 De Maisonneuve Blvd. West, Montréal, Québec H3G 1M8, Canada. [4]UDIMA, School of Computer Science, Madrid Open University, Carretera de La Coruña, KM. 38,500, Vía de Servicio, no 15, 28400 Collado Villalba, Madrid, Spain.

References
1. F. Shadabi, D. Sharma, Artificial intelligence and data mining techniques in medicine – success stories. Int Conf BioMedical Eng Inform **1**, 235 (2008)
2. U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, From data mining to knowledge discovery: an overview, advances in knowledge discovery and data mining. eds. U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy. AAAI Press/The MIT Press. 1-34 (1996).
3. J. Wu, L. Zhang, S. Yin, H. Wang, G. Wang, J. Yuan, Differential diagnosis model of hypocellular myelodysplastic syndrome and aplastic anemia based on the medical big data platform. Complexity **2018** (2018). https://doi.org/10.1155/2018/4824350
4. S. Mukherjee, Malignant mesothelioma disease diagnosis using data mining techniques. Appl Artif Intell **32**(3), 293–308 (2018). https://doi.org/10.1080/08839514.2018.1451216
5. B. G. Ma Bai, B. M. Nalini, J. Majumdar, Analysis and detection of diabetes using data mining techniques—a big data application in health care. In: Shetty N., Patnaik L., Nagaraj H., Hamsavath P., Nalini N. (eds) Emerging Research in Computing, Information, Communication and Applications. Advances in Intelligent Systems and Computing. **882** (2019)
6. V. Paramasivam, T. S. Yee, S. K. Dhillon, A. S. Sidhu, A methodological review of data mining techniques in predictive medicine: an application in hemodynamic prediction for abdominal aortic aneurysm disease. Biocybernetics and Biomedical Engineering. Elsevier. **34**(3), 139-145 (2014).
7. J. F. Easton, C. R. Stephens, M. Angelova, Risk factors and prediction of very short term versus short/intermediate term post-stroke mortality: A data mining approach. Comput Biol Med, Elsevier. **54**, 199-210 (2014)
8. J.S. Saleema, P.D. Shenoy, K.R. Venugopal, L.M. Patnaik, Cancer prognosis prediction model using data mining techniques. Int J Soft Comput Artif Intell Appl (IJSCAI) **3**(1), 9–18 (2014)
9. S. Feng, Y. Ren, S. Fan, M. Wang, T. Sun, F. Zeng, P. Li, F. Liang, Discovery of acupoints and combinations with potential to treat vascular dementia: a data mining analysis. Evidence-Based Complementary and Alternative Medicine, Hindawi Publishing Corporation, in press (2015)
10. M. N. Sohail, R. Jiadong, M. M. Uba, M. Irshad, A comprehensive looks at data mining techniques contributing to medical data growth: a survey of researcher reviews. In: Patnaik S., Jain V. (eds) Recent Developments in Intelligent Computing, Communication and Devices. Advances in Intelligent Systems and Computing. **752** (2019)
11. K.J. Cios, G.W. Moore, Uniqueness of medical data mining. Artif Intell in Med J. **26**(1-2), 1–24 (2002)
12. J. A. Lara, G. Moreno, A. Pérez, J. P. Valente, A. López-Illescas, Comparing posturographic time series through events detection. 21st IEEE International Symposium on Computer-Based Medical Systems, CBMS '08. 293-295 (2008)
13. J. A. Lara, A. Pérez, J. P. Valente, A. López-Illescas, Modelling stabilometric time series. Proceedings of the 3rd International Conference on Health Informatics – HEALTHINF. 485-488 (2010)
14. J.A. Lara, A. Pérez, J. P. Valente, A. López-Illescas, Generating time series reference models based on event analysis. 19th European Conference on Artificial Intelligence - ECAI 2010. 1115-16 (2010).
15. J.A. Lara, Marco de Descubrimiento de Conocimiento para Datos Estructuralmente Complejos con Énfasis en el Análisis de Eventos en Series Temporales. Technical University of Madrid. PhD Thesis (2011)
16. J.A. Lara, D. Lizcano, M.A. Martínez, J. Pazos, T. Riera, A UML Profile for the conceptual modelling of structurally complex data: easing human effort in the KDD process. Inf Software Technol **56**(3), 335–351 (2014)

17. J.A. Lara, D. Lizcano, M.A. Martínez, J. Pazos, Data preparation for KDD through automatic reasoning based on description logic. Inf Syst **44**, 54–72 (2014)

18. A. Anguera, J.A. Lara, D. Lizcano, M.A. Martínez, J. Pazos, Sensor-generated time series events: a definition language. Sensors **12**(9), 11811–11852 (2012)

19. F. Alonso, J.A. Lara, L. Martínez, J.P. Valente, Generating reference models for structurally complex data: application to the stabilometry medical domain. Methods Inf Med **52**, 441–453 (2013)

20. J.A. Lara, D. Lizcano, A. Pérez, J.P. Valente, A general framework for time series data mining based on event analysis: Application to the medical domains of electroencephalography and stabilometry. J Biomed Inf **51**, 219–241 (2014). https://doi.org/10.1016/j.jbi.2014.06.003

21. A. Anguera, J. M. Barreiro, J. A. Lara, D. Lizcano, Applying data mining techniques to medical time series: an empirical case study in electroencephalography and stabilometry. Comput Struct Biotechnol J. 14, 185-199 (2016). Doi: https://doi.org/10.1016/j.csbj.2016.05.002.

22. F. Puppe, Systematic introduction to expert systems: knowledge representations and problem-solving methods. Ed. Springer-Verlag (1993)

23. E. H. Shortliff, Computer based medical consultations: MYCIN. American Elsevier (1976)

24. B. G. Buchanan, E. A. Feigenbaum, DENDRAL and Meta-DENDRAL: their applications dimension. Technical Report. Artificial Intelligence. **11** (5-2) (1978).

25. J. Lederberg, How dendral was conceived and born. ACM Symposium on the History of Medical Informatics. Rockefeller University, New York: National Library of Medicine (1987)

26. C.-S. Lee, M.-H. Wang, A fuzzy expert system for diabetes decision support application. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics. **41**(1), 139 – 153 (2011)

27. A. Keleş, A. Keleş, U. Yavuz, Expert system based on neuro-fuzzy rules for diagnosis breast cancer. Expert Systems with Applications. **38**(5), 5719–5726 (2011)

28. C. Mahesh, E. Kannan, M.S. Saravanan, Generalized regression neural network based expert system for hepatitis b diagnosis. J. Comput. Sci. **10**, 563–556 (2014)

29. Z.-G. Zhou, F. Liu, L.-L. Li, L.-C. Jiao, Z.-J. Zhou, J.-B. Yang, Z.L. Wang, A cooperative belief rule based decision support system for lymph node metastasis diagnosis in gastric cancer. Knowledge-Based Systems **2015** (in press)

30. L. Ge, A.R. Kristensen, M.C. Mourits, R.B. Huirne, A new decision support framework for managing foot-and-mouth disease epidemics. Ann Oper Res **219**(1), 49–62 (2014)

31. A. Raghu, D. Praveen, D. Peiris, L. Tarassenko, G. Clifford, Lessons from the evaluation of a clinical decision support tool for cardiovascular disease risk management in rural India, Technologies for Development. Ed. Springer International Publishing. **Part V**, 199-209 (2015).

32. D. Gil, A. Soriano, D. Ruiz, C. A. Montejo, Embedded systems for diagnosing dysfunctions in the lower urinary tract. Proceedings of the 22nd Annual ACM Symposium on Applied Computing (2007)

33. S. Waring, M. Sharland, J. Bianco, M. Boyce, S. Quinlan, PS2-8: Development and implementation of clinical decision support tools in epic to standardize dementia diagnosis and care at essentia health. Clin Med Res **12**(1-2), 88 (2014)

34. A. Bourouis, M. Feham, M.A. Hossain, L. Zhang, An intelligent mobile based decision support system for retinal disease diagnosis. Decision Support Systems. **59**, 341–350 (2014)

35. N. Tavakoli, A. Vahdat, Designing a clinical decision support system for managing and treating patients with the chief complaint of vertigo. J Isfahan Med School. **35**(460), 1806–1811 (2018)

36. D.W. Ballard, N. Kuppermann, D.R. Vinson, E. Tham, J.M. Hoffman, M. Swietlik, S.J.D. Davies, E.A. Alessandrini, L. Tzimenatos, L. Bajaj, D.G. Mark, S.R. Offerman, U.K. Chettipally, M.D. Paterno, M.H. Schaeffer, R. Richards, T.C. Casper, H.S. Goldberg, R.W. Grundmeier, P.S. Dayan, Implementation of a clinical decision support system for children with minor blunt head trauma who are at nonnegligible risk for traumatic brain injuries. Annals of Emergency Medicine **73**(5), 440–451 (2019). https://doi.org/10.1016/j.annemergmed.2018.11.011

37. H. Rosenblum, N. Radcliffe, Case-based approach to managing angle closure glaucoma with anterior segment imaging. Can J Ophthalmol **49**(6), 512–518 (2014)

38. A. Siva, C. Lampl, Case-based diagnosis and management of headache disorders, Ed. Springer International Publishing (2015)

39. M. Hor, I. Glauche, M.C. Müller, R. Hehlmann, A. Hochhaus, M. Loeffler, I. Roeder, Model-based decision rules reduce the risk of molecular relapse after cessation of tyrosine kinase inhibitor therapy in chronic myeloid leukemia. Blood J **121**(2), 378–384 (2013)

40. M.W.L. Moreira, J.J.P.C. Rodrigues, V. Korotaev, J. Al-Muhtadi, N. Kumar, A comprehensive review on smart decision support systems for health care. IEEE Syst J (2019). https://doi.org/10.1109/JSYST.2018.2890121

41. A.T. Azar, A.E. Hassanien, Dimensionality reduction of medical big data using neural-fuzzy classifier. Soft Comput **19**, 1115–1127 (2015)

42. N. Pérez, M.A. Guevara, A. Silva, I. Ramos, Improving the Mann–Whitney statistical test for feature selection: an approach in breast cancer diagnosis on mammography. Art Intell Med **63**(1), 19–31 (2015)

43. H.H. Inbarani, M. Bagyamathi, A.T. Azar, A novel hybrid feature selection method based on rough set and improved harmony search. Neural Comput Appl **26**(8), 1859–1880 (2015)

44. T. N. Phyu. Survey of classification techniques in data mining. Proceedings of the International MultiConference of Engineers and Computer Scientists. Vol 1 (2009).

45. T. Nguyen, A. Khosravi, D. Creighton, S. Nahavandi, Medical data classification using interval type-2 fuzzy logic system and wavelets. Appl Soft Comput **30**, 812–822 (2015)

46. S.K. Nayak, S.C. Nayak, H.S. Behera, Evolving low complex higher order neural network based classifiers for medical data classification. Adv Intell Syst Comput **411**, 415–425 (2015)

47. F. Segovia, J.M. Gorriz, J. Ramirez, J. Levin, M. Schuberth, M. Brendel, A. Rominger, G. Garraux, C. Phillips, Analysis of 18F-DMFP PET data using multikernel classification in order to assist the diagnosis of Parkinsonism. IEEE MIC (2015)

48. R. Agrawal, R. Srikant. "Fast algorithms for mining association rules." Proc. 20th int. conf. very large data bases, VLDB. Vol. 1215 (1994)

49. F. Alonso, L. Martínez, A. Pérez, A. Santamaría, J.P. Caraça-Valente, Integrating expert knowledge and data mining for medical diagnosis. Expert Syst Res Trends **3**, 113–137 (2007)

50. W. A. Chaovalitwongse, Y. Fan, R. C. Sachdeo, On the time series K-nearest neighbor classification of abnormal brain activity. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans **1** (2007).

51. S. P. K. Karri, H. Garud, D. Sheet, J. Chatterjee, D. Chakraborty, A. K. Ray, M. Mahadevappa, Learning scale-space representation of nucleus for accurate localization and segmentation of epithelial squamous nuclei in cervical smears, IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI). 772 – 775 (2014)

52. J.A. Martin, E.C. Wilson, M.J. Osterman, E.W. Saadi, S.R. Sutton, B.E. Hamilton, Assessing the quality of medical and health data from the 2003 birth certificate revision: results from two states, National Vital Statistics Reports : From the Centers for Disease Control and Prevention, National Center for Health Statistics. Natl Vital Stat Syst **62**(2), 1–19 (2013)

53. European Parliament, 95/46/EC data protection directive, Council of 24 October 1995.

54. Office of the Spanish Head of State, Personal Data Protection Act 15/1999, of 13 December, Official State Gazette No. 298 of 14 December 1999 [amended as of 6 March 2011]) [in Spanish].

55. D. R. Anderson, D. J. Sweeney, T. A. Williams, Quantitative methods for business. Seventh ed. International Thomson Publishing (1998)

56. A. G. Kleppe, J. Warmer, J.B. Warmer, W. Bast, MDA explained: the model driven architecture: practice and promise. Addison-Wesley Professional. (2003)

57. Q. Ang, W. D. Wang, B. Y. Zhao, J. Li, K. Y. Li, Application of data mining based on clinical medicine database. 2nd International Conference on Signal Processing Systems (2010)

58. C. Groselj, Data mining problems in medicine. Proc. 15th IEEE Symposium on Computer-Based Medical Systems (2002)

59. R. Agrawal, C. Faloutsos, A. Swami. Efficient similarity search in sequence databases, FODO (1993)

60. K. Chan, A. W. Fu, Efficient time series matching by wavelets. ICDE. 126-133 (1999)

61. R. Povinelli, Time Series Data Mining: identifying temporal patterns for characterization and prediction of time series. PhD. Thesis. Milwaukee (1999)

62. J.A. Sanz, M. Galar, A. Jurio, A. Brugos, M. Pagola, H. Bustince, Medical diagnosis of cardiovascular diseases using an interval-valued fuzzy rule-based classification system. Appl Soft Comput **20**, 103–111 (2014)

63. A. E. Hassanien, H. M. Moftah, A. T. Azar, M. Shoman, MRI breast cancer diagnosis hybrid approach using adaptive ant-based segmentation and multilayer perceptron neural networks classifier, Applied Soft Computing. **14**-Part A, 62-71 (2014)

64. C.-H. Chen, W.-T. Huang, T.-H. Tan, C.-C. Chang, Y.-J. Chang, Using K-nearest neighbor classification to diagnose abnormal lung sounds. Sensors J **15**, 13132–13158 (2015)

65. G. Sahu, R.K. Khare, Decision tree classification based decision support system for derma disease. Int J Comput Appl **94**(17), 21–26 (2014)

66. F.L. Seixas, B. Zadrozny, J. Laks, A. Conci, D.C. Muchaluat Saade, A Bayesian network decision model for supporting the diagnosis of dementia, Alzheimer's disease and mild cognitive impairment. Computers in Biology and Medicine. **51**, 140–158 (2014)

67. J.R. Quinlan, Induction of decision trees. Machine Learn **1**(1), 81–106 (1986)

68. J. Kolodner, Case-based reasoning. Ed. Morgan Kaufmann (1993)

69. R. Agrawal, R. Srikant, Fast algorithms for mining association rules in large databases. Proceedings of the 20th International Conference on Very Large Data Bases, VLDB. 487-499 (1994)

70. U. Rajendra, S. Vinitha, G. Swapna, R.J. Martis, J.S. Suri, Automated EEG analysis of epilepsy: a review. Knowledge Based Syst **45**, 147–165 (2013)

71. U. Rajendra, H. Fujita, V.K. Sudarshan, S. Bhat, J.E.W. Koh, Application of entropies for automated diagnosis of epilepsy using EEG signals: a review. Knowledge Based Syst **88**, 85–96 (2015)

72. R.J. Barry, A.R. Clarke, S.J. Johnstone, A review of electrophysiology in attention-deficit/hyperactivity disorder: 1 Qualitative and quantitative electroencephalography 2. Event-related potentials. Clin Neurophysiol **114**, 171–198 (2003)

73. D. Kundra, B. Pandey, Classification of EEG based diseases using data mining. Int J Comput Appl **90**(18), 11–15 (2014)

74. J. Chen, B. Hu, P. Moore, X. Zhang, X. Ma, Electroencephalogram-based emotion assessment system using ontology and data mining techniques. Appl Soft Comput **30**, 663–674 (2015)

75. S. A. Hosseini, Epilepsy recognition by higher order spectra analysis of EEG signals. Encyclopedia of Information Science and Technology, Third Edition (2015)

76. R.J. Barry, F.M. De Blasio, E.M. Bernat, G.Z. Steiner, Event-related EEG time-frequency PCA and the orienting reflex to auditory stimuli. Psychophysiol **52**(4), 555–561 (2015)

77. F. Riaz, A. Hassan, S. Rehman, I.K. Niazi, K. Dremstrup, EMD-based temporal and spectral features for the classification of EEG signals using supervised learning. IEEE Trans Neural Syst Rehabil Eng **24**(1), 28–35 (2016)

78. P.G. Kanmani Prince, R.R. Hemamalini, S. Kumar, Seizure detection by classification of EEG signals based on DWT reconstruction error and CWT using a novel wavelet. Biomed Res **26**(3), 530–533 (2015)

79. O. Fausta, U. Rajendra Acharyab, H. Adelic, A. Adelig, Wavelet-based EEG processing for computer-aided seizure detection and epilepsy diagnosis. Seizure **26**, 56–64 (2015)

80. A. Bijoy Das, M. I. Hassan Bhuiyan, Discrimination and classification of focal and non-focal EEG signals using entropy-based features in the EMD-DWT domain. **29**: 11-21 (2016)

81. P. Barigant, P. Merlet, J. Orfait, C. Tetar, New design of E.L.A. Statokinesemeter, Agressol. **13**(C), 69-74 (1972)

82. R. Boniver, Posture et posturographie. Rev Med Liege **49**(5), 285–290 (1994)

83. H. Chaudhry, B. Bukiet, Z. Ji, T. Findley, Measurement of balance in computer posturography: comparison of methods—a brief review. J Bodywork Mov Ther **15**(1), 82–91 (2011)

84. Neurocom® International. Balance Master Operator's Manual v8.2. www.onbalance.com (last accessed in December 2014).

85. D. Song, F. Chung, J. Wong, S. Yogendran, The assessment of postural stability after ambulatory anesthesia: a comparison of desflurane with propofol. Anesth Analg **94**(1), 60–64 (2002)

86. D. Nguyen, C. Pongchaiyakul, J.R. Center, J.A. Eisman, T.V. Nguyen, Identification of high-risk individuals for hip fracture: a 14-year prospective study. J Bone Miner Res **20**(11), 1921–1928 (2005)

87. V. Raiva, W. Wannasetta, S. Gulsatitporn, Postural stability and dynamic balance in Thai community dwelling adults. Chula Med J. **49**(3), 129–141 (2005)

88. M. Sinaki, R.H. Brey, C.A. Hughes, D.R. Larson, K.R. Kaufman, Significant reduction in risk of falls and back pain in osteoporotic-kyphotic women through a Spinal Proprioceptive Extension Exercise Dynamic (SPEED) program. Mayo Clin **80**(7), 849–855 (2005)

89. J.-H. Park, S. Youm, Y. Jeon, S.-H. Park, Development of a balance analysis system for early diagnosis of Parkinson's disease. Int J Ind Ergon **48**, 139–148 (2015)

90. H. Sucuoglu, S. Tuzun, Y. Akbaba, M. Uludag, H. H. Gokpinar, Effect of whole-body vibration on balance using posturography and balance tests in postmenopausal women, American Journal of Physical Medicine & Rehabilitation. In press (2005)

91. T.P. Exarchos, G. Rigas, A. Bibas, D. Kikidis, C. Nikitas, F.L. Wuyts, B. Ihtijarevic, L. Maes, M. Cenciarini, C. Maurer, N. Macdonald, D.-E. Bamiou, L. Luxon, M. Prasinos, G. Spanoudakis, D.D. Koutsouris, D.I. Fotiadis, Mining balance disorders' data for the development of diagnostic decision support systems. Comput Biol Med **77**, 240–248 (2016)

92. L. H. G. Marrega, S. M. Silva, E. F. Manffra, J. C. Nievola, Comparison between decision tree and genetic programming to distinguish healthy from stroke postural sway patterns. Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE (2015)

93. R. G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, C. E. Elger, Indications of nonlinear deterministic and finite dimensional structures in time series of brain electrical activity: dependence on recording region and brain state. Phys Rev E Stat Nonlin Soft Matter Phys. **64**(6), 061907:1-8 (2001)

94. R. Kohavi, A study of cross-validation and bootstrap for accuracy estimation and model selection. International Joint Conference on Artificial Intelligence, IJCAI (1995)

95. G. Scott, Strategic planning for high-tech product development. Technol Anal Strat Manag **13**(3), 343–364 (2010)

Publisher's Note

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