Artificial Intelligence in Medicine 61 (2014) 79–88

Contents lists available at ScienceDirect

Artificial Intelligence in Medicine

journal homepage: www.elsevier.com/locate/aiim

Automatic classification of epilepsy types using ontology-based and genetics-based machine learning

Yohannes Kassahun[a,*], Roberta Perrone[b,**], Elena De Momi[b], Elmar Berghöfer[f], Laura Tassi[c], Maria Paola Canevini[d], Roberto Spreafico[e], Giancarlo Ferrigno[b], Frank Kirchner[a,f]

[a] Fachbereich 3 – Mathematics and Computer Science, University of Bremen, Robert-Hooke-Str. 5, D-28359 Bremen, Germany
[b] Department of Electronics, Information and Bioengineering (DEIB), Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milano, Italy
[c] Centro Chirurgia Epilessia "Claudio Munari", Dipartimento di Neuroscienze, Ospedale Niguarda Cà Granda, Piazza Ospedale Maggiore 3, 20162 Milano, Italy
[d] Epilepsy Center, San Paolo Hospital, Via A. Di Rudini 8, 20142 Milan, Italy
[e] Department of Experimental Neurophysiology and Epileptology, Istituto Nazionale Neurologico 'C. Besta', Via Celoria 11, 20133 Milano, Italy
[f] German Research Center for Artificial Intelligence (DFKI), Robotics Innovation Center, Robert-Hooke-Str. 5, D-28359 Bremen, Germany

ARTICLE INFO

Article history:
Received 4 September 2013
Received in revised form 24 February 2014
Accepted 7 March 2014

Keywords:
Ontology-based classification
Genetics-based classification
Data mining (knowledge discovery) from medical data
Epileptogenic zone identification

ABSTRACT

Objectives: In the presurgical analysis for drug-resistant focal epilepsies, the definition of the epileptogenic zone, which is the cortical area where ictal discharges originate, is usually carried out by using clinical, electrophysiological and neuroimaging data analysis. Clinical evaluation is based on the visual detection of symptoms during epileptic seizures. This work aims at developing a fully automatic classifier of epileptic types and their localization using ictal symptoms and machine learning methods.
Methods: We present the results achieved by using two machine learning methods. The first is an ontology-based classification that can directly incorporate human knowledge, while the second is a genetics-based data mining algorithm that learns or extracts the domain knowledge from medical data in implicit form.
Results: The developed methods are tested on a clinical dataset of 129 patients. The performance of the methods is measured against the performance of seven clinicians, whose level of expertise is high/very high, in classifying two epilepsy types: temporal lobe epilepsy and extra-temporal lobe epilepsy. When comparing the performance of the algorithms with that of a single clinician, who is one of the seven clinicians, the algorithms show a slightly better performance than the clinician on three test sets generated randomly from 99 patients out of the 129 patients. The accuracy obtained for the two methods and the clinician is as follows: first test set 65.6% and 75% for the methods and 56.3% for the clinician, second test set 66.7% and 76.2% for the methods and 61.9% for the clinician, and third test set 77.8% for the methods and the clinician. When compared with the performance of the whole population of clinicians on the rest 30 patients out of the 129 patients, where the patients were selected by the clinicians themselves, the mean accuracy of the methods (60%) is slightly worse than the mean accuracy of the clinicians (61.6%). Results show that the methods perform at the level of experienced clinicians, when both the methods and the clinicians use the same information.
Conclusion: Our results demonstrate that the developed methods form important ingredients for realizing a fully automatic classification of epilepsy types and can contribute to the definition of signs that are most important for the classification.

1. Introduction

For drug resistant focal epileptic patients, brain surgery can be the only option to cure the patient. The epileptogenic zone (EZ) is the cortical area where ictal discharges originate and which must be surgically resected or disconnected to achieve seizure freedom [1]. The definition of the EZ is usually obtained through

* Corresponding author. Tel.: +49 42125752299.
** Corresponding author.
E-mail addresses: kassahun@informatik.uni-bremen.de (Y. Kassahun), roberta.perrone@polimi.it (R. Perrone).

http://dx.doi.org/10.1016/j.artmed.2014.03.001
0933-3657/

non-invasive pre-surgical investigations, in most cases by clinical, electrophysiological and neuroimaging data analysis [2]. Clinical evaluation is based on the anamnestic data and on the visual analysis of symptoms during epileptic seizures in long term scalp video-electroencephalograms (video-EEG). Intra-cerebral invasive monitoring (stereo-EEG) may be needed for very difficult cases. Clinical manifestations can be just subjective [3] (epileptic auras without impairment of consciousness) or may progress to objective signs that can be observed and analyzed, often associated with impairment of awareness. The chronological sequence of symptoms and their relative duration might indicate the seizure onset area and the progressive involvement of adjacent brain structures and is, therefore, crucial for the definition of the EZ [4], though identical symptoms may arise from different cortical regions, leading to misleading localization of the EZ [5,6]. The most significant symptoms appear in the first 60 s, even if the post-ictal confusion can be prolonged, mainly if associated with speech disturbances, which is the case for EZ located in the dominant hemisphere.

Temporal lobe epilepsies (TLE), in which seizures originate from the temporal lobe, represent the majority of focal epilepsies (over 60%). The surgical treatment of TLE has the best results with almost 70% of the patients reported to be seizure-free after surgery [7]. The duration, the time course, the loss of consciousness and the post-ictal phenomena are different in TLE and extra-temporal seizures (ExTE). Seizures arising from the temporal lobe have a relatively gradual evolution (compared to ExTE) [8]. In TLE the subjective phenomena are more frequent, the loss of consciousness is more gradual, seizures are longer and the post-ictal confusion is more important. The ideal sequence of ictal modifications in TLE is a subjective manifestation (epigastric feeling often raising up, auditory, visual and psychic symptoms, are quite exclusively located in the temporal lobe) followed by a loss of consciousness, oro-alimentary automatisms, homolateral head orientation and a contralateral arm dystonic posturing with homolateral gestural automatisms. On the contrary, seizures in ExTE are frequently abrupt, brief, without subjective manifestations, the loss of consciousness is precocious, no oro-alimentary automatisms are present and motor modification can be bilateral and more tonic. However, not all the seizures are easily recognizable and particularly the hypermotor gestural automatisms can occur in both TLE and ExTE [5]. The state-of-the-art is the visual analysis of the clinical data performed by expert clinicians always in association with anatomical medical images and EEG data. In literature, there is not yet any automatic classification of epilepsy types based on clinical symptoms only.

It would be extremely important to provide clinicians with an automatic classifier of epilepsy types, which could help to validate the opinion of the clinicians in case of agreement between the methods and the clinicians, or would, on the contrary, force the clinicians to revise their decision or look in more detail at the case. In this paper we analyze two methods for realizing an automatic classifier of epilepsy types, starting from medical data, which is usually unstructured information. The first method is able to learn the human knowledge on how to perform diagnosis, based on clinical symptoms and the second automatically retrieves knowledge from medical data, searching for existing patterns in the data (e.g. clinical symptoms). For the first method we used an ontology-based classification approach [9], while for the second method we used genetics-based machine learning (data mining) [10]. In addition to dealing with unstructured data, both methods can also deal with heterogeneous and incomplete data. The motivation behind choosing the two methods is the view that learning by no means entails a tabula rasa view. It involves the incorporation of prior knowledge that can accelerate or otherwise improve the learning process. A combination of the above methods will allow us to incorporate domain knowledge and at the same time discover novel medical knowledge. The two methods were tested on a clinical dataset of patients and compared with the classifications carried out by expert clinicians.

The paper is organized as follows: first we review relevant work in the field of ontology-based classification and genetics-based data mining in the context of learning from medical data. Then we present the algorithms used in the paper, followed by the presentation of the actual experiments with patient data and the results obtained. Finally, we give a conclusion and an outlook based on the work presented in this paper.

2. Review of related work

In this paper two different algorithms are presented, which exploit different methods: the ontology-based classification (OBC) and the genetics-based data mining (GDM). We give a review of related work in the area of ontology-based classification and genetics-based data mining.

2.1. Ontology-based classification

Ontological modelling provides a description of a specific knowledge domain and it is made of classes, properties (which express relationships between classes) and instances (which are the "ground level" components of an ontology and populate classes) [11]. In medical field, ontologies can therefore be used for sharing medical knowledge in a reliable format so that it is understandable and can be processed by both humans and machines. Medical ontologies like OpenGALEN [12], SNOMED-CT [13] and UMLS [14] are used in the health care practice. The OpenGALEN ontology was used in urology to develop a decision support system for treating patients with urinary-tract infections [15]. SNOMED-CT ontology provides health care terminology with comprehensive coverage of diseases, clinical findings, etiologies and therapies in the particular field of anesthesiology. It is used in electronic health care records (EHR) and is related to critical patient-specific information, like drug allergies. SNOMED-CT makes the medical knowledge more usable and accessible [16]. EPILONT[1] is an ontology about epilepsy domain and epileptic seizures while the Epilepsy and Seizure Ontology (EpSO)[2] gives a classification of epilepsy syndromes, location, etiology and related medical conditions according to the International League Against Epilepsy (ILAE) organization standards.

Ontologies made of hierarchies and properties between classes can be useful for data aggregation and clustering. Such ontologies provide domain knowledge and support the interpretation of relations identified in dataset through data mining processes, based on linguistic or statistical techniques [17]. As an example, the foundational model of anatomy (FMA) [18] ontology is used as a source of anatomical knowledge for predicting the consequences of injury. In this application, knowledge about spatial relations between the injury and vital organs is provided by the FMA [19].

Ontological modeling was used also in other domains, like customer knowledge assessment [20], where ontologies are used to describe customers population and the ontology graph, represented with arcs and nodes, is used for classifying them according to different criteria. In [21], the authors designed surgical models of neurosurgery making use of ontology and described 106 surgical cases. Through classification trees and clustering algorithms, they extracted surgical knowledge, facilitating the surgical decision-making process and surgical planning. In [22], Lee et al. tried to overcome the limit of classical ontology dealing with uncertain knowledge and classified different diabetes syndromes using ontology-based inference rules. Highly expressive rules-based

[1] http://bioportal.bioontology.org/ontologies/EPILONT.
[2] http://prism.case.edu/prism/index.php/.

languages, like SWRL [23], fail in representing fuzzy information [24], as in case of symptoms occurrence. For this reason, we exploit ontologies to extract new probabilistic information considering the depth of concepts with respect to the root concept and the distance between symptoms.

2.2. Genetics-based data mining

Data mining (DM) is a machine learning procedure, which can be used to automatically find useful patterns from a dataset [25]. The procedure has been successfully applied in various fields and is making its way into health care systems. Data mining algorithms can learn from examples and model the non-linear relationships between variables. The most commonly used data mining algorithms for health care systems include naive Bayes classifiers, neural networks, support vector machines, logistic regression, fuzzy rules and decision trees.

There are several DM techniques developed for diagnosing diseases. For example, Soni et al. [26], and Dangare and Apte [27] presented data mining techniques for heart disease diagnosis, and Ganesan et al. [28] presented the use of artificial neural networks for cancer diagnosis.

DM techniques are also developed for prognosing diseases. In [29], a Bayesian expert system for clinically detecting coronary artery disease is given. In [30], DM techniques are used for predicting heart attacks. In [31], artificial neural networks have been applied for prognosing end stage kidney disease. The work of Floyd [32] presents the application of DM techniques for prognosis of the pancreatic cancer.

Moreover, some authors compared the performance of algorithms for the diagnosis or prognoses purposes. In [33], the discriminatory power of k-nearest neighbors, logistic regression, artificial neural networks, decision trees, and support vector machines on classifying pigmented skin lesions for diagnosis purpose is analyzed. In [34], a summary of comparison of well-performing DM algorithms used for both disease diagnosis and disease prognosis is given. Prasad et al. [35] compared auto-associative memory neural networks, Bayesian networks, iterative dichotomized 3 (ID3) and C4.5 in the diagnosis of the disease asthma.

DM techniques are furthermore applied to analyze various signals and their relationships with particular diseases or symptoms. An example is the application of support vector machines in automatic seizure detection in EEG, which can be used in diagnosing neurological disorders related to epilepsy [36].

In addition to disease diagnoses and/or prognoses, DM techniques are used to extract diagnostic rules from medical dataset [37,38]. The rules generated are similar to those created manually in expert systems and therefore can be easily validated by domain experts. For a detailed review of DM techniques and challenges in mining medical data, please refer to [39–42].

The genetics-based data mining method presented in this paper is related to the work in the area of data mining for discovering temporal hidden knowledge from medical dataset. In [43], learning a Bayesian network and rules is used to discover medical knowledge. A grammar-based genetic programming is used as a search algorithm. The method is applied in the domains of limb fracture and scoliosis. The work by Tan et al. [44] utilizes genetic programming and genetic algorithms to evolve classification rules for hepatitis and breast cancer datasets. Nuovo and Catania [45] integrated fuzzy logic and genetic algorithms in order to discover a fuzzy rule based diagnostic system from medical datasets for breast cancer and Pima-Indian diabetes data found in the UCI repository of machine learning databases [46], and medical data on aphasia found in Aachen aphasia database [47]. Dam et al. presented a neural-based learning classifier system [48] that incorporates neural networks into a learning classifier systems [49]. The method was applied on medical datasets such as the diabetes, breast cancer, heart disease and lymphography found in the UCI repository of machine learning databases. In [50] decision trees are evolved using evolutionary algorithms.

3. Methods

In this section we give an overview of the methods that we developed and used to determine the type of epilepsy of patients based on clinical data described in Section 4.1.

3.1. Ontology-based classification

The ontology-based classification method focuses on epileptic symptoms correlation. The ontological modeling describes the relationships between classes and classes and individuals.

In the "patient seizure ontology", ictal symptoms are ontological instances (i.e. the ground level objects) of class "`ictal-symptom`" (such as "visual illusions", "fear", etc. happening during seizures). The relationship between classes is expressed by predicates (such as "`subclassOf`", "`hasManifestation`", etc.). In the domain ontology, there are some general classes (i.e. collections of objects) like the "`ictal-symptom`" class which groups all the possible symptoms that can be observed during an epileptic seizure. Fig. 1 shows the seizure ontology model.

The "`patient`" class groups all the patients of the dataset and it is linked to the "`seizure`" general class by the property "`hasSeizure`". Every epileptic seizure is described by a set of manifestations, defined in the "`manifestation`" and "`noManifestation`" classes, which refer to observed and not observed symptoms during seizure by the clinicians and these two classes are linked to the "`seizure`" class by "`hasManifestation`" and "`hasNoManifestation`" relationships respectively. In addition to this, observed manifestations are characterized by their occurrence time. "`early`" and "`late`" classes are subclasses of the "`manifestation`"class. Symptoms manifested in the first 10 s belong to the "`early`" class, while later symptoms belong to the "`late`" class. "`manifestation`" and "`noManifestation`" classes are defined as "`unionOf`" objects of the class "`ictal-symptom`". As in [51], in order to translate the relationship between pairs of symptoms, considering their properties, we computed the distances in terms of graph arcs.

The C_k matrix, which expresses the correlation between each pair of symptoms (i.e. the likelihood that if during an epileptic seizure symptom i is present, symptom j is present as well) for a specific patient k ($k \in \{1, \ldots, M\}$, where M is the number of patients) is computed.

Each correlation matrix element $c_k(i, j)$ between pairs of symptoms is calculated using

$$c_k(i,j) = \frac{(d(s_i) + d(s_j))\alpha + \beta}{((D(s_i, s_j) * \alpha) + 2\max(d(s_i), d(s_j))) + \beta} \quad (1)$$

$i, j \in \{1, \ldots N\}$, where N is the total number of symptoms, s_i and s_j are the considered symptoms (e.g. "head deviation" and "visual illusion"), $d(s_i)$ and $d(s_j)$, are the distances of s_i and s_j from the seizure class in the ontology tree as shown in Fig. 2. $D(s_i, s_j)$ represents the length of the shortest path between s_i and s_j in the ontology tree (number of arcs between s_i and s_j) as shown in Fig. 3a and b. $max(d(s_i), d(s_j))$ represents the longest distance from the "`seizure`" class in the domain ontology tree of the two symptoms s_i and s_j. The variables α and β range from 0 to 1. α is used to balance the proportion between d and D values, β makes the denominator not to approach zero.

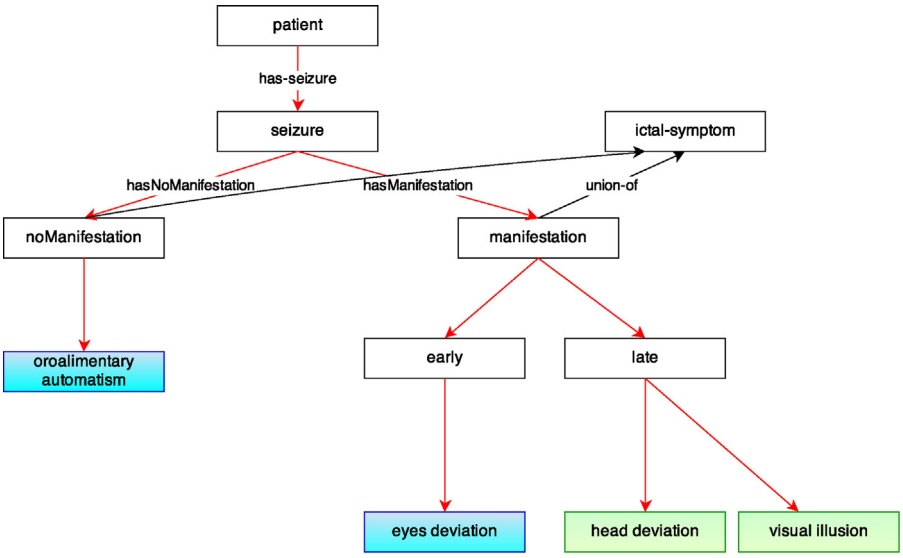

Fig. 1. Ontology model of a patient seizure with early and late symptoms manifestations. "eyes deviation" is observed in the first 10 s of the registration, while "head deviation" and "visual illusion" appear after the first 10 s.

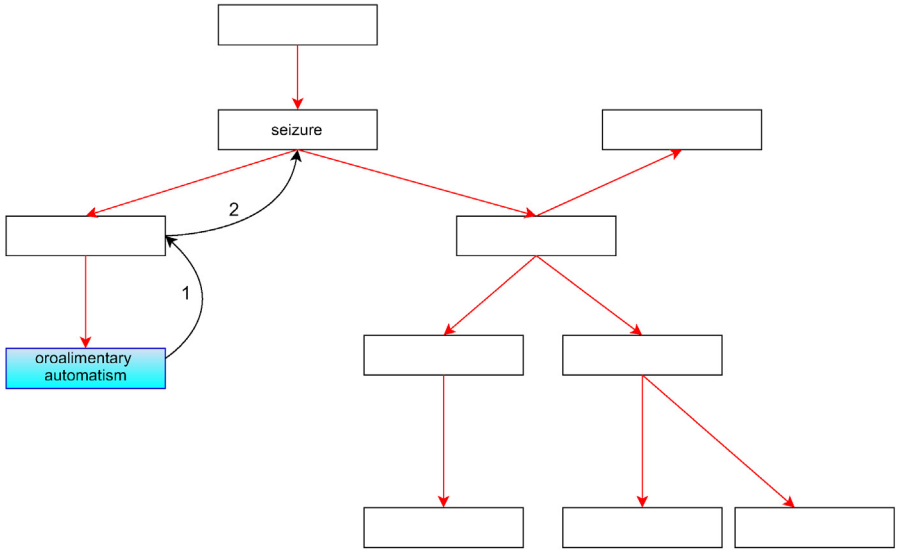

Fig. 2. The distance of symptom s_i from the "`seizure`" class in the ontology tree is calculated as the number of arcs in between and it is equal to 2.

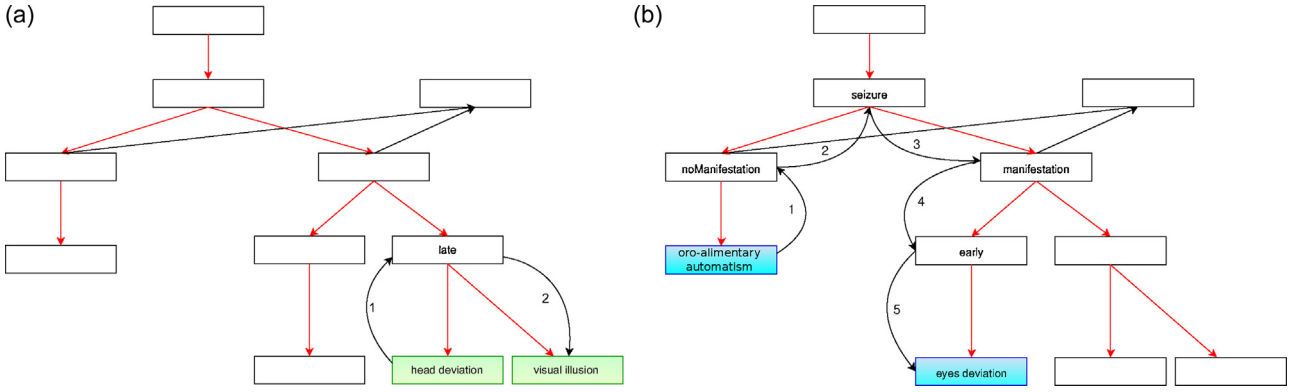

Fig. 3. The distance between two symptoms (D) is calculated as the number of arcs between them: e.g. the distance between "head deviation" and "visual illusion" is 2 arcs (a), while the distance between "eyes deviation" and "oroalimentary automatism" (a non-observed symptom) is 5 (b).

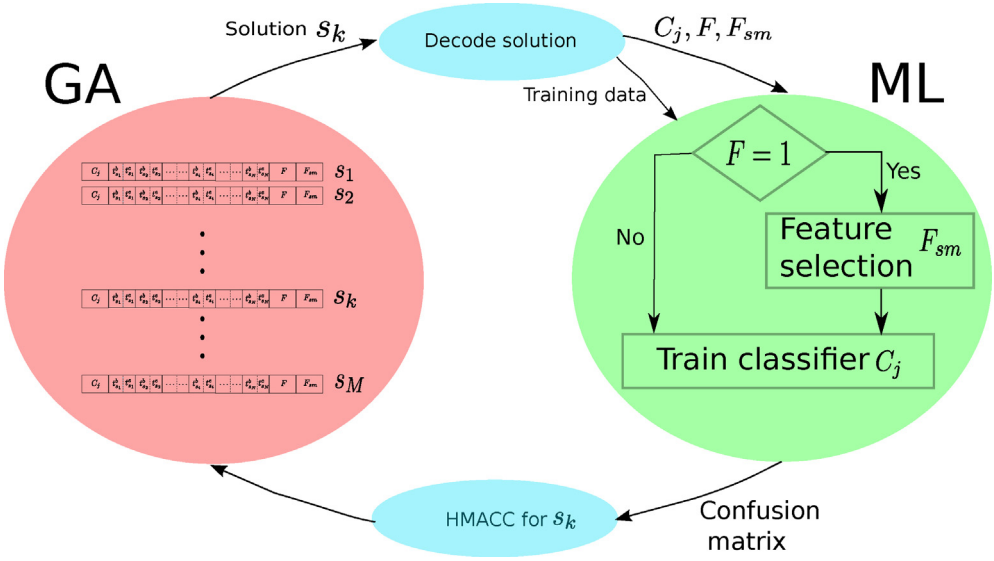

Fig. 4. Interaction between the GA and the ML tools. In the figure M indicates the population size of the genetic algorithm.

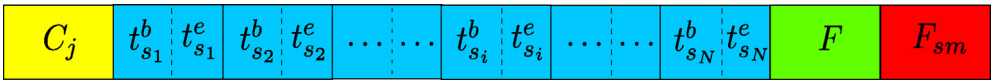

Fig. 5. A chromosome encoding a solution proposed by the GA. The gene C_j encodes the jth classifier to train and use. The first N genes following the gene encoding the classifier represent the time intervals to consider. For symptom s_i the time interval is given by $t^e_{s_i} - t^b_{s_i}$. The gene F is a flag encoding whether a feature selection will be employed. If its value is one, then feature selection will be employed, and if its value is zero, no feature selection will be used. The last gene F_{sm} encodes the feature selection method to apply.

Each element $mc_{TLE}(i,j)$ and $mc_{ExTE}(i,j)$ of the mean of the matrices that express the correlation between symptoms for each group (TLE and ExTE patients, MC_{TLE} and MC_{ExTE} matrices), is computed as

$$mc(i,j)_{TLE} = \frac{1}{W}\sum_{w=1}^{W} C_w(i,j) \tag{2}$$

$$mc(i,j)_{ExTE} = \frac{1}{V}\sum_{v=1}^{V} C_v(i,j), \tag{3}$$

where $W+V=M$, W and V are the number of patients of the dataset with temporal lobe epilepsies and extratemporal lobe epilepsies respectively.

In order to classify the seizure types, each seizure is assigned to the group whose distance from MC is minimum, where the distances, D_{TLE} and D_{ExTE}, are computed by

$$D_{TLE} = \sqrt{\sum_{i=1}^{N}\left(\sum_{j=1}^{N}(c_h(i,j) - mc(i,j)_{TLE})^2\right)} \tag{4}$$

$$D_{ExTE} = \sqrt{\sum_{i=1}^{N}\left(\sum_{j=1}^{N}(c_h(i,j) - mc(i,j)_{ExTE})^2\right)} \tag{5}$$

$h \in \{1, \ldots H\}$, where H is the total number of patients to be tested. We optimized α and β using the tenfold cross-validation on the dataset [52] by continuously changing their values. Varying the values of α and β, does not affect the classification power, even if the correlation between symptoms is changed. We therefore set both α and β to 0.5, as the sum of the variables has to be one [51].

The epileptic seizures ontology is built in ontology web language (OWL), the popular ontological language, using Protégé 3.5, an open-source tool, for creating, editing and visualizing ontologies [53]. We followed the proposed guideline [54], collecting specific domain knowledge from experts, textbooks and papers integrating existing ontologies and defining classes and class hierarchy.

3.2. Genetics-based data mining

The GDM algorithm is a combination of a genetic algorithm (GA), Pyevolve [55], and machine learning (ML) tools for supervised learning WEKA [56] and Orange [57], which implement different classifiers (e.g. decision trees, multilayer perceptrons, support vector machines, etc.). The flow chart of GDM is shown in Fig. 4.

The parameters that are optimized by the genetic algorithm are:

1 the classifier, C_j, to train,
2 the instant of time in which a symptom s_i starts $t^b_{s_i}$ and the instant of time when it ends $t^e_{s_i}$,
3 the feature selection method, F_{sm}, that is used to determine symptoms which play an important role in classifying patients into TLE and ExTE groups.

The GA starts by generating random solutions (i.e. the chromosomes $s_1, s_2, \ldots, s_k, \ldots, s_M$, where M is the number of solutions). Fig. 5 shows a chromosome encoding properties of a solution. Each solution, s_k, will be decoded into the classifier to train C_j, the feature selection flag, F, the feature selection method to apply, F_{sm}, and the corresponding training data. After training, the machine learning tools give back the confusion matrix [58] that results after the tenfold cross-validation [52] is performed. From the confusion matrix, the fitness of the individual (quality of a solution) is calculated using the harmonic mean accuracy (HMACC), which is given by

$$HMACC(s_k) = \frac{2}{(1/(1 - FP(s_k))) + (1/(1 - FN(s_k)))}, \tag{6}$$

Table 1
GA parameters.

Population size	50
Mutation probability	0.05
Crossover probability	0.8
Maximum number of generations	100
Selection method	Rank based selection [59]

where $FP(s_k)$ stands for the false positive rate and $FN(s_k)$ stands for the false negative rate corresponding to a solution s_k. The values for the HMACC lie between zero and one. A higher value of HMACC means a better solution. If a solution classifies correctly all the TLE patients but misclassifies the ExTE patients, it will result in a low HMACC value.

The optimization process of the genetic algorithm is run for some generations and stopped if the maximum HMACC so far obtained does not improve over the last 20 generations or the maximum number of generations is overcome. Table 1 shows the parameters of the genetic algorithm for results obtained using genetics-based data mining algorithm. The parameters were set manually and kept the same for all experiments reported in this paper.

4. Experiments and results

To evaluate machine learning methods, one has to take care of a few general problems. For categorizing the epilepsy types into temporal (TLE) and extra-temporal (ExTE), the methods presented in this paper are trained on a training dataset and tested on a test dataset. These two datasets were selected to be disjoint because the algorithms are optimized on the training dataset for instance using cross validation. Since the test dataset is not included in the training dataset at all, it is completely unknown to the methods when it is presented to the methods to test their performance. Hence the responses of the methods give an idea on how the methods would perform on a new data dataset (new patients in our case).

In this section we describe first the clinical data of patients we used for analyzing the performance of the methods in Section 4.1 and then the results obtained with it in Sections 4.2, 4.3 and 4.4. In contrast to the statistical analysis given in Section 4.4, in Sections 4.2 and 4.3 we deal with the comparison between the methods and clinicians. For clinicians re-sampling of test datasets is not possible because the clinicians would remember the examples in the datasets. Therefore, different "trials" can only be done by different clinicians. In the first test (Section 4.2), we provided a clinician, whose level of expertise is very high, with three test datasets of patients, and we asked her to classify each patient. In the second test (Section 4.3) we asked seven different clinicians, whose expertise is high/very high, to classify 30 patients that did not belong to any training, testing or validation dataset previously used.

4.1. Description of dataset used

The clinical dataset of patients that we used in our work was obtained from three clinical epileptological centers: Carlo Besta Neurological Institute, Niguarda Ca Granda Hospital and San Paolo Hospital, all in Milan, Italy. All patients used in the experiments reported in this paper are drug-resistant epileptic patients, who do not present any other pathology (e.g. malignant tumors), and they were operated on and seizure-free for at least one year. This fact is used as a ground-truth for the generation of the labels of the patients, which are TLE and ExTE. All of the patients have signed an informed consent to the treatment of their anonymized data. The dataset for the experiments consists of 99 labeled patients, of which 60 suffer from TLE and 39 from ExTE. Furthermore, an additional test set of 30 patients, of which 10 suffer from TLE and 20 from ExTE, is used for the experiment described in Section 4.3. The dataset is generated as follows. Once clinicians have a collection of video recordings of epileptic seizures, they create a table for each patient in which 29 symptoms are present and which gives the duration in seconds of the seizures. The first 14 rows of the table correspond to the subjective symptoms and the next 15 to the objective symptoms. For each patient, clinicians fill in the table, assigning to the symptoms observed a '1', and a '0' to manifestation not observed (Fig. 6). The patient from which the data is shown in Fig. 6 has three different symptoms ("epigastric" sensation, "visual illusion simple" and "motor dystonic monolateral leg"). The occurrence of symptoms in patients dataset used for this work is shown in Table 2.

Table 2
Patients dataset's occurrence table. The rows are for subjective and objective symptoms, the columns are for the two groups of patients (TLE and ExTE). The first 14 symptoms are subjective, the last 15 symptoms are objective.

No.	Symptom	TLE	ExTE
1	Epigastric	20	1
2	Auditory illusion	0	0
3	Auditory hallucinations	0	2
4	Visual illusion simple	1	0
5	Visual illusion complex	0	1
6	Visual hallucination simple	0	1
7	Visual hallucinations complex	0	1
8	Olfactive hallucinations	1	0
9	Gustatory hallucinations	1	0
10	Psychic manifestation	10	2
11	Fear	2	3
12	Autonomic (tachicardia, dyspnoea, etc.)	11	8
13	Motor-sensitive manifestation monolateral	1	2
14	Motor-sensitive manifestation bilateral	1	0
15	Oroalimentary automatisms	36	10
16	Gestural automatisms monolateral	26	14
17	Gestural automatisms bilateral	16	12
18	Head deviation	35	38
19	Eyes deviation	32	29
20	Motor clonic monolateral arm	5	6
21	Motor clonic bilateral arms	1	2
22	Motor clonic monolateral leg	1	3
23	Motor clonic bilateral legs	1	4
24	Motor clonic monolateral mouth	3	5
25	Motor dystonic monolateral arm	27	23
26	Motor dystonic bilateral arms	14	10
27	Motor dystonic monolateral leg	2	5
28	Motor dystonic bilateral legs	6	7
29	Autonomic (flushing, shialorrhea, vomiting)	11	3

4.2. Test 1. Comparison of OBC and GDM with a single clinician

The comparison of OBC and GDM diagnosis with a single clinician, expert in epilepsy type classification, is performed on three test datasets generated from the training set we obtained. The single clinician's level of expertise is very high and she is one of the 7 clinicians who participated in the second study (see Section 4.3). The test datasets are generated by selecting random entries from the training set. After selecting the test datasets, the rest of the data is used in training OBC and GDM. Please note that while the members of the test datasets do not change once selected, the members of the training and validation will change during cross-validation. Table 3 gives the experimental setup used for Test 1.

Table 4 shows the performance with respect to accuracy, that is the number of correctly classified patients for the clinician, OBC and GDM on the test datasets. The accuracy we used is given by the equation

$$\text{Accuracy} = \frac{\#\text{ correctly classified TLEs} + \#\text{ correctly classified ExTEs}}{\#\text{ TLEs} + \#\text{ ExTEs}} \quad (7)$$

where the symbol # stands for "number of". Cochran test was performed ($p \leq 0.05$) [60].

		0	1	2	3	4	5	6	...	58	59	60	(seconds)
Semiology	epigastric	1	1	1	1	1	1	1	...	0	0	0	
	auditory illusion	0	0	0	0	0	0	0	...	0	0	0	
	auditory hallucination	0	0	0	0	0	0	0	...	0	0	0	
	visual illusion simple	0	0	0	0	0	0	0	...	1	1	1	
	visual illusion complex	0	0	0	0	0	0	0	...	0	0	0	
	visual hallucination simple	0	0	0	0	0	0	0	...	0	0	0	
	visual hallucination complex	0	0	0	0	0	0	0	...	0	0	0	
	...												
	...												
	motor dystonic monolateral leg	1	1	1	1	1	1	1	...	0	0	0	
	motor dystonic bilateral legs	0	0	0	0	0	0	0	...	0	0	0	
	autonomic	0	0	0	0	0	0	0	...	0	0	0	

Fig. 6. Example of patient symptoms. Symptoms are listed in the first column. A '1' stands for a presence of a symptom, and a '0' for the absence of a symptom at a particular time.

As can be seen from Table 4, the performance of OBC and GDM is better than the clinician for dataset 0 and comparable performance for dataset 1 and dataset 2 even if there is no statistical significance ($p_0 = 0.2592$, $p_1 = 0.5292$, $p_2 = 1$). Table 5 shows the reliability of the agreements using the Fleiss' kappa values [61] among the clinician, OBC and GDM. For dataset 0 and dataset 2, OBC and GDM have the highest reliability of agreements, while for dataset 1 GDM has the highest reliability agreement with the clinician. For dataset 0, GDM has the least reliability of agreement with the clinician.

4.3. Test 2. Comparison of OBC and GDM with seven clinicians

Seven clinicians, experts in the field of epilepsy diagnosis, were asked to classify epilepsy types on the basis of a patient's seizure symptoms. Please note that the level of expertise of all of the seven clinicians is high/very high. Accuracy was computed as in Eq. (7) and the Cochran test was performed ($p \leq 0.05$). Table 6 shows the experimental setup for Test 2.

As can be seen from Table 7 both OBC and GDM perform at the level of experts ($p = 0.4139$). Table 8 shows the collaborative decision made by the clinicians, the algorithms, and the algorithms and the clinicians. A majority of votes is used to determine the decision of the clinicians and the algorithms. For the clinicians, if more than three clinicians voted in favor of an epilepsy type, this will be considered as the decision of the clinicians. For OBC and GDM, only cases where both voted in favor of an epilepsy type is considered as the decision of the algorithms. If both disagree, this case will be considered as a tie since there is no majority voting in favor of an epilepsy type. Another interesting analysis is the agreement levels of the clinicians, the algorithms and the algorithms and the clinicians. Table 9 shows the Fleiss' kappa values for the clinicians, the algorithms and the algorithms and the clinicians.

From the table, one can see that the reliability of agreements of the clinicians is better than the reliability of agreements of the algorithms. Moreover, one can see that the reliability of agreements of the algorithms with the clinicians is comparable with GDM slightly better than OBC. Following Landis and Koch [62] there is a fair agreement between OBC and GDM and a moderate agreement among the clinicians. One can also see that there is a moderate agreement among OBC, GDM and the clinicians.

Table 3
Test 1. Experimental setup used for comparing OBC and GDM classification with a single clinician.

	Dataset 0		Dataset 1		Dataset 2	
	TLE	ExTE	TLE	ExTE	TLE	ExTE
Training	39	19	39	29	49	32
Validation	4	5	7	3	5	4
Testing	17	15	14	7	6	3

Table 4
Test 1. Accuracy obtained for the clinician, OBC and GDM on three test datasets.

	Dataset 0	Dataset 1	Dataset 2
Clinician	18/32	13/21	7/9
OBC	24/32	14/21	7/9
GDM	21/32	16/21	7/9

Table 5
Test 1. Fleiss' kappa values.

	Dataset 0	Dataset 1	Dataset 2
OBC and clinician	0.13	0.25	0.13
GDM and clinician	0.02	0.57	0.45
OBC and GDM	0.60	0.49	0.59
Clinician, OBC and GDM	0.55	0.36	0.31

Table 6
Test 2. Experimental setup.

	TLE	ExTE
Training	45	26
Validation	5	3
Testing	10	20

Table 7
Test 2. Accuracy obtained by OBC, GDM and seven clinicians on 30 testing patients. The symbols Cl-1, Cl-2, . . ., Cl-7 stand for clinician 1, clinician 2, . . . and clinician 7 respectively.

OBC	GDM	Cl-1	Cl-2	Cl-3	Cl-4	Cl-5	Cl-6	Cl-7
18/30	18/30	15/30	16/30	19/30	19/30	19/30	20/30	22/30

Table 8
Accuracy obtained by OBC and GDM and seven clinicians on 30 novel patients using majority voting.

	Clinicians	OBC and GDM	All
Majority voting	20/30	12/30	20/30
Tie	–	11/30	–

Table 9
Test 2. Fleiss' kappa values for the clinicians, OBC and GDM, and OBC, GDM and the clinicians.

	Fleiss' kappa
Clinicians	0.55
OBC and clinicians	0.44
GDM and clinicians	0.48
OBC and GDM	0.27
Clinicians, OBC and GDM	0.41

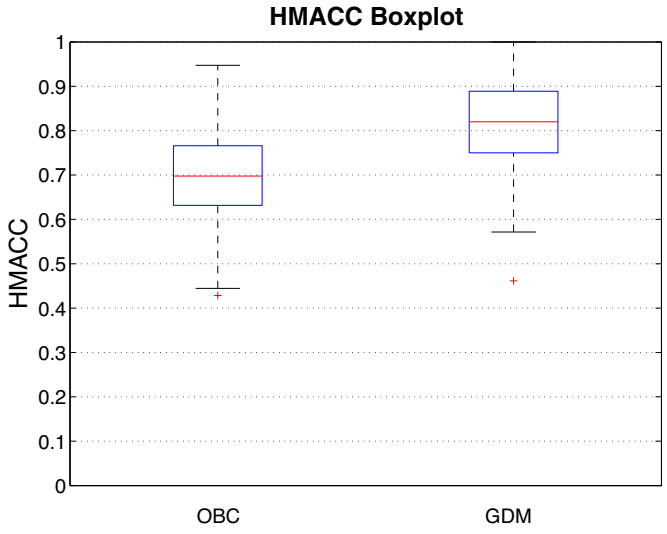

Fig. 7. A boxplot diagram of the HMACC of 100 values from 100 repetitions of the training and testing of the algorithms. The boxplot of the HMACC is only for the test accuracy.

4.4. Test 3. Statistical analysis of the methods

To get an idea of the performance that we can expect from the algorithms we used standard measures from statistics, such as expected value (mean) and the standard deviation. In order to be able to calculate such values we have to repeat the training phase and the testing of the algorithms independently and multiple times. Since the data comes from real patients it is not possible to generate a large number of test samples. Therefore, the only possibility is to generate new training and test datasets by re-sampling the available data.

We decided to use 20% of the available data as the test dataset and the remaining 80% as the training dataset. For each trial the test dataset is drawn from the data by random selection, 20% of the temporal and 20% of the extra-temporal examples and the remaining examples are used as the training dataset. Selecting the samples in this way guarantees that the distribution of the two classes is preserved for the test and the training dataset. The algorithms are trained on the training dataset and after they are optimized for their performance they are evaluated on the test dataset. When the test performance is calculated it is important that the algorithms are completely reset so that they do not have any information about the prior trial. The training and testing of the algorithms is repeated 100 times.

After the test performance of the algorithms is measured 100 times, the mean value and the standard deviation of all trials are calculated. The mean value is now the expected value of the HMACC that an algorithm would make on new patient data and the standard deviation gives a measure of how stable the algorithm is. Therefore, a high expectation value of the HMACC as well as a small standard deviation is desired.

The results obtained by the algorithms are shown in a boxplot diagram in Fig. 7. The plot shows the median accuracy of the algorithms, which gives a good idea of the expected likelihood of correctly classifying new data. The box around the mean value shows the 25th and 75th percentiles of the accuracy values during the 100 repetitions and the whiskers show the minimum and maximum ranges which are not detected as outliers, while outliers are marked as crosses. This gives an idea how sensitive the algorithm is to the selection of the training dataset. In a real application, after training the algorithm with all of the available data, this is the range in which one can expect the accuracy to lie. In general, it is expected that the expectation of the accuracy will increase and the standard deviation will decrease with additional training examples. Therefore, if the algorithms are continuously trained on new data, their performance will increase over time.

Table 10
The mean (μ) and the standard deviation (σ) of the HMACC for both methods on the 100 test runs.

	μ	σ
OBC	0.7125	0.1080
GDM	0.8072	0.1065

In Fig. 7 one can see that the median accuracy for the GDM is about *82%* and for the OBC it is *70%*. The values for the mean accuracies and the standard deviations for the two algorithms are given in Table 10. It is obvious that both algorithms show a much better result than random guessing, which would be 50%.

4.5. Computational efficiency of OBC and GDM

The average training time needed for GDM for training on 79 patients takes on average 3 h, while the resulting classifier takes on average 0.1 s for delivering the output of the classification on a standard personal computer. Since genetic algorithms can be easily parallelized, in principle it should be possible to reduce the training time of GDM drastically. For OBC the time that is needed to train the classifier on 79 patients is on average 10 min, and once the classifier is computed it takes 0.3 ms to classify a novel patient. This shows that once the algorithms are trained they can be applied in practical settings.

5. Conclusion and outlook

Clinical evaluation of epileptic patients is based on careful visual detection of symptoms during epileptic seizures in long term Video-EEG recordings. We proposed two machine learning methods to help the clinicians during the diagnosis. The first method (OBC) can incorporate the knowledge of expert clinicians directly, while the second method (GDM) can mine or discover medical knowledge in implicit form from medical data. Our ontology-based classification method (OBC) is not based on inference rules since it is not possible to define deterministic rules correlating the presence or the absence of a particular symptom to epilepsy types. Both methods consider the timing of occurrence of symptoms since early symptoms are more significant than late symptoms (i.e. typical behavioral patterns are closely related to the seizure initial spread location). In addition to this, the genetics-based data mining algorithm has a feature selection step, where a feature is a particular symptom. Through feature selection it is possible to identify which feature plays an important role in the classification of the epilepsy types. It is possible to conclude that the methods can be applied to discover interesting spatial (symptoms playing an important role) and temporal (sequence of symptoms) patterns in medical data for a particular disease.

When comparing the performance of the algorithms with that of a single clinician, both OBC and GDM show a slightly better performance (even if not significant) than the clinician. When compared with the performance of the whole population of clinicians (in our case seven clinicians), the accuracy of the machine learning algorithms (18/30) is slightly worse than the average accuracy of the clinicians (18.5/30) (see Section 4.3). Results of the two comparisons show that the performance of the machine learning algorithms is approaching the level of experts for the case where both the clinicians and the machine learning methods exploit only the clinical data used in this paper (strings of '1's and '0's in time). When comparing OBC with GDM (see Section 4.4), we have found

that GDM performs slightly better than OBC, and both algorithms obviously perform much better than random guessing. The accuracy values obtained by the methods are at first glance not expected by persons normally working with other machine learning benchmark problems, since in these benchmark problems the accuracy values obtained are usually higher than the results reported in this paper. Therefore, one would consider that the results are not satisfactory. However, when comparing the performance of the algorithms with that of the clinicians, it can be easily seen from the results that the algorithms approach the level of expertise of very experienced clinicians, who participated in the experiments reported in the paper.

Considering the Fleiss' kappa index, computed values are strictly dependent on the considered datasets (the agreement between OBC and GDM can be substantial or fair in case of different dataset considered, even if the datasets have the same number of subjects). Kappa values interpretation is somewhat controversial, the reference value for our analysis is the agreement among all the clinicians, which is 0.55. From the results, it appears that the developed methods have the same performance in comparison to the pool of considered clinicians, when all are analyzing the same data.

An interesting application of the machine learning methods developed in this paper would be in the area of training young clinicians in the classification of epilepsy types. It would make the learning process for the young clinicians easier. In case of less experienced clinicians, automatic classification algorithms can be used to teach how a symptom, rather than another one, could be more important and significant for a specific group of epilepsy. An investigation of this topic would be interesting for future work. In addition to this, the machine learning methods could be applied to generate an alternative opinion for classifying epilepsy types. This would help to validate the opinion of the clinician in case of agreement between the methods and the clinicians, or would, in case of divergence, force the clinician to revise his/her decision or look at the case in more detail. A further application of the methods lies in formalizing medical knowledge and best practices. The advantage of developing an ontology of epileptic seizure is the symptoms classification standard, so that all clinical centers would be able to represent an epileptic seizure in the same way. The use of an ontology to define symptoms of an epileptic seizure will also overcome the problem of terminology used by different clinical centers for symptoms description. Thus, it is important to have a glossary of the symptoms: our ontology will face this discrepancy (i.e. "`owl:sameAs`" statement indicates that two individuals actually refer to the same thing: the individuals have the same "identity", so they inherit all properties or "`rdfs:seeAlso`" statement used to indicate a resource that might provide additional information about the subject resource). This is obviously a difficult task and requires a detailed analysis of the structure and the concepts of medical terminologies. It can be achieved by constructing medical domain ontologies for representing medical terminology systems. In this work, the realization of an ontology (OBC) that represents the domain of epilepsy diagnosis will help in formalizing the symptoms observable during an epileptic seizure and to correctly classify the two types of epilepsy. Moreover, a further subclassification should be obtained in ExTE concerning the different lobes (parietal, insular, occipital and frontal).

An important bottleneck that always arises when machine learning is transferred to practical implementation is the knowledge transfer from the expert to the agent. In this case, clinicians are able to classify patients observing the seizure using their sensitivity acquired by experience, which is difficult to translate into the artificial intelligence algorithm. It has to be noticed that in standard clinical routine, clinicians base their evaluation also on anamnestic and neuroimaging data. In order to improve the performance of automatic classification algorithms, more information should be encoded and subsequently provided to the classifiers.

In health care, data is frequently not ideal or incomplete. In our case, a problem can arise if clinicians make errors in compiling the table of symptoms occurrence during epilepsy seizure. These mistakes can be corrected because we have a second reliable kind of data that is the video recordings of epilepsy seizures. Also, health care records are unstructured information. Our ontological approach is also suited for such kind of data, since ontologies can be structured to automatically classify unstructured text, once correspondences are defined. Future work could add new clinical information of a patient, like a new symptom, "staring" (a motionless state with the interruption of all the previous activities), when patients stare at seizure onset, and which is more frequent during temporal lobe epilepsies. An additional variable could improve the classifier, even more if that variable had higher occurrence in a set of population patients. The electroencephalogram (EEG) detects electrical activity in the brain and it is a fundamental tool for the analysis in the diagnosis of epilepsy because electrical alterations, often very indicative, can also be present in the absence of other symptoms. Other diagnostic tests include magnetic resonance imaging and laboratory tests, which can verify or exclude specific causes. In this sense, with the help of the proposed algorithms, different types of information could be unified and then processed to obtain a more accurate diagnosis. Since it is known that clinicians use different sources of information other than the coded clinical data used in this paper for decision, it remains to be seen how the algorithms improve their performance if additional information is available. Moreover, the reliability of the methods to correctly classify co-morbid patients could be evaluated in order to see if the presented methods are able to localize the epileptic focus although some symptoms may be produced by secondary pathology.

The paper is a feasibility study on the application of machine learning techniques for the automatic localization of the epileptogenic zone. Although the epilepsy types analyzed are characterized by distinctive signs during the seizure manifestations, patients diagnosis based on the visual inspection of symptoms during the first 60 s can be subjective, as shown by clinician disagreements on the 30 patients analyzed.

The OBC and GDM methods presented can be applied in mining medical knowledge as shown in the paper, or even be used to discover a novel idea in the area of medical diagnosis of a particular disease. It is worth underlining that this makes the automatic classification systems creative and interactive, giving back the discovered knowledge to the clinicians. This in turn improves the existing knowledge of classifying epilepsy types.

Acknowledgements

This work is supported in part within the ACTIVE European project (FP7-ICT-2009-6-270460) and the EuRoSurge European project (FP7-ICT-2011-7-288233). We would like to thank the two anonymous reviewers for their constructive suggestions and comments.

References

[1] Kahane P, Landré E, Minotti L, Francione S, Ryvlin P. The Bancaud and Talairach view on the epileptogenic zone: a working hypothesis. Epileptic Disord 2006;8:16–26.
[2] Miserocchi A, Cascardo B, Piroddi C, Fuschillo D, Cardinale F, Nobili L, et al. Surgery for temporal lobe epilepsy in children: relevance of presurgical evaluation and analysis of outcome. J Neurosurg Pediatr 2013;11(3):1–12 [Clinical article].
[3] Lüders H, Acharya J, Baumgartner C, Benbadis S, Bleasel A, Burgess R, et al. Semiological seizure classification. Epilepsia 1998;39(9):1006–13.

[4] Nobili L, Francione S, Mai R, Cardinale F, Castana L, Tassi L, et al. Surgical treatment of drug-resistant nocturnal frontal lobe epilepsy. Brain 2007;130(2):561–73.
[5] Tassi L, Meroni A, Deleo F, Villani F, Mai R, Russo GL, et al. Temporal lobe epilepsy: neuropathological and clinical correlations in 243 surgically treated patients. Epileptic Disord 2009;11(4):281–92.
[6] Mai R, Sartori I, Francione S, Tassi L, Castana L, Cardinale F, et al. Sleep-related hyperkinetic seizures: always a frontal onset? Neurol Sci 2005;26(3):220–4.
[7] Jackson GD, Briellmann RS, Kuzniecky RI. Temporal lobe epilepsy. In: Kuzniecky RI, Jackson GD, editors. Magnetic resonance in epilepsy. 2nd ed. Burlington: Academic Press; 2005. p. 99–176 [chapter 4].
[8] Gupta A, Jeavons P, Hughes R, Covanis A. Aura in temporal lobe epilepsy: clinical and electroencephalographic correlation. J Neurol Neurosurg Psychiatry 1983;46(12):1079–83.
[9] Burger S, Stieger B. Ontology-based classification of unstructured information. In: The fifth international conference on digital information management (ICDIM). IEEE; 2010. p. 254–9.
[10] Kovacs T. Genetics-based machine learning. In: Rozenberg G, Bck T, Kok J, editors. Handbook of natural computing. Berlin, Heidelberg: Springer; 2012. p. 937–86.
[11] Sowa J. Ontology, metadata, and semiotics. In: Ganter B, Mineau G, editors. Conceptual structures: logical, linguistic, and computational issues, vol. 186 of Lecture Notes in Computer Science. Berlin, Heidelberg: Springer; 2000. p. 55–81.
[12] Rector AL, Rogers JE, Pole P. The GALEN high level ontology. Stud Health Technol Inform 1996;34:174–8.
[13] Bos L, Donnelly K. SNOMED-CT: the advanced terminology and coding system for eHealth. Stud Health Technol Inform 2006;121:279–90.
[14] Bodenreider O. The unified medical language system (UMLS): integrating biomedical terminology. Nucleic Acids Res 2004;32(suppl. 1):D267–70.
[15] Karlsson D. A design and prototype for a decision-support system in the field of urinary tract infections-application of OpenGALEN techniques for indexing medical information. Stud Health Technol Inform 2001;84:479–83.
[16] Elevitch FR. SNOMED-CT: electronic health record enhances anesthesia patient safety. AANA J 2005;73(5):361–6.
[17] Bodenreider O. Biomedical ontologies in action: role in knowledge management, data integration and decision support. Yearb Med Inform 2008;47:67–79.
[18] Rosse C, Mejino Jr JL. A reference ontology for biomedical informatics: the foundational model of anatomy. J Biomed Inform 2003;36(6):478–500.
[19] Rubin DL, Dameron O, Bashir Y, Grossman D, Dev P, Musen MA. Using ontologies linked with geometric models to reason about penetrating injuries. Artif Intell Med 2006;37(3):167–76.
[20] Zhang L, Wang Z. Ontology-based clustering algorithm with feature weights. J Comput Inform Syst 2010;6(9):2959–66.
[21] Jannin P, Morandi X. Surgical models for computer-assisted neurosurgery. Neuroimage 2007;37(3):783–91.
[22] Lee C-S, Wang M-H, Acampora G, Loia V, Hsu C-Y. Ontology-based intelligent fuzzy agent for diabetes application. In: IEEE symposium on intelligent agents, 2009 (IA'09). IEEE; 2009. p. 16–22.
[23] Horrocks I, Patel-Schneider PF, Boley H, Tabet S, Grosof B, Dean M, et al. SWRL: a semantic web rule language combining OWL and RuleML. W3C Member Submission 2004;21:79.
[24] Pan J, Stoilos G, Stamou G, Tzouvaras V, Horrocks I. f-SWRL: a fuzzy extension of SWRL. In: Spaccapietra S, Aberer K, Cudr-Mauroux P, editors. Journal on Data Semantics VI, vol. 4090 of Lecture Notes in Computer Science. Berlin, Heidelberg: Springer; 2006. p. 28–46.
[25] Gorunescu F. Data mining: concepts, models and techniques. Berlin, Heidelberg: Springer-Verlag; 2011.
[26] Soni J, Ansari U, Sharma D, Soni S. Predictive data mining for medical diagnosis: an overview of heart disease prediction. Int J Comput Appl 2011;17(8):43–8.
[27] Dangare CS, Apte SS. Improved study of heart disease prediction system using data mining classification techniques. Int J Comput Appl 2012;47(10):44–8.
[28] Ganesan N, Venkatesh K, Rama MA, Palani AM. Application of neural networks in diagnosing cancer disease using demographic data. Int J Comput Appl 2010;1(26):76–85.
[29] Chu C-M, Chien W-C, Lai C-H, Bludau H-B, Tschai H-J, LuPai S-MH, et al. A Bayesian expert system for clinical detecting coronary artery disease. J Med Sci 2009;4:187–94.
[30] Srinivas K, Rani BK, Govrdhan A, Karimnagar J. Applications of data mining techniques in healthcare and prediction of heart attacks. Int J Comput Sci Eng 2010;2(2):250–5.
[31] Di Noia T, Ostuni VC, Pesce F, Binetti G, Naso D, Schena FP, et al. An end stage kidney disease predictor based on an artificial neural networks ensemble. Expert Syst Appl 2013;40(11):4438–45.
[32] Floyd S. Data mining techniques for prognosis in pancreatic cancer [Master's thesis]. USA: Worcester Polytechnic Institute; 2007.
[33] Dreiseitl S, Ohno-machado L, Kittler H, Vinterbo S, Binder M. A comparison of machine learning methods for the diagnosis of pigmented skin lesions. J Biomed Inform 2001;34:28–36.
[34] Kolçe E, Frasheri N. A literature review of data mining techniques used in healthcare databases. In: Markovski S, Gusev M, editors. Proceedings of ICT innovations. 2012. p. 577–82.
[35] Prasad B, Prasad P, Sagar Y. A comparative study of machine learning algorithms as expert systems in medical diagnosis (asthma). In: Meghanathan N, Kaushik B, Nagamalai D, editors. Advances in computer science and information technology, vol. 131 of communications in computer and information science. Berlin, Heidelberg: Springer; 2011. p. 570–6.
[36] Fan J, Shao C, Ouyang Y, Wang J, Li S, Wang Z. Automatic seizure detection based on support vector machines with genetic algorithms. In: Wang T-D, Li X, Chen S-H, Wang X, Abbass H, Iba H, et al., editors. Simulated evolution and learning, vol. 4247 of Lecture Notes in Computer Science. Berlin, Heidelberg: Springer; 2006. p. 845–52.
[37] Meamarzadeh H, Khayyambashi M, Saraee M-H. Extracting temporal rules from medical data. In: International conference on computer technology and development (ICCTD 09), vol. 1. 2009. p. 327–31.
[38] Mena LJ, Orozco EE, Felix VG, Ostos R, Melgarejo J, Maestre G. Machine learning approach to extract diagnostic and prognostic thresholds: application in prognosis of cardiovascular mortality. In: Computational and mathematical methods in medicine; 2012, 6 pp.
[39] Freitas A. A survey of evolutionary algorithms for data mining and knowledge discovery. In: Ghosh A, Tsutsui S, editors. Advances in evolutionary computing, natural computing series. Berlin, Heidelberg: Springer; 2003. p. 819–45.
[40] Wasan SK, Bhatnagar V, Kaur H. The impact of data mining techniques on medical diagnostics. Data Sci J 2006;5:119–26.
[41] Hosseinkhah F, Ashktorab H, Veen R, Owrang M. Challenges in data mining on medical databases. IGI Global; 2009. p. 1393–404 [chapter 83].
[42] Satyanandam N, Satyanarayana Ch, Riyazuddin Md, Shaik A. Data mining machine learning approaches and medical diagnose systems: a survey. Int J Comput Org Trends 2012;2(3):53–60.
[43] Ngan PS, Wong ML, Lam W, Leung KS, Cheng JCY. Medical data mining using evolutionary computation. Artif Intell Med 1999;16(1):73–96.
[44] Tan K, Yu Q, Heng C, Lee T. Evolutionary computing for knowledge discovery in medical diagnosis. Artif Intell Med 2003;27(2):129–54.
[45] Di Nuovo A, Catania V. Genetic tuning of fuzzy rule deep structures for efficient knowledge extraction from medical data. IEEE Int Conf Syst Man Cybernet 2006;6:5053–8.
[46] Bache K, Lichman M. UCI machine learning repository. URL: http://archive.ics.uci.edu/ml [accessed 25.11.13].
[47] Axer H, Jantzen J, von Keyserlingk DG. An aphasia database on the internet: a model for computer-assisted analysis in aphasiology. Brain Lang 2000;75(3):390–8.
[48] Dam HH, Abbass HA, Lokan C, Yao X. Neural-based learning classifier systems. IEEE Trans Knowl Data Eng 2008;20(1):26–39.
[49] Urbanowicz RJ, Moore JH. Learning classifier systems: a complete introduction, review, and roadmap. Artif Evol Appl 2009;2009:1–25.
[50] Kokol P, Pohorec S, tiglic G, Podgorelec V. Evolutionary design of decision trees for medical application. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 2012;2(3):237–54.
[51] Siregar P, Toulouse P. Model-based diagnosis of brain disorders: a prototype framework. Artif Intell Med 1995;7(4):315–42.
[52] Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: Proceedings of the 14th international joint conference on Artificial intelligence – vol. 2, IJCAI-95. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.; 1995. p. 1137–43.
[53] Knublauch H, Fergerson RW, Noy NF, Musen MA. The Protégé OWL plugin: an open development environment for semantic web applications. In: The Semantic Web-ISWC 2004. Springer; 2004. p. 229–43.
[54] Noy NF, McGuinness DL. Ontology development 101: a guide to creating your first ontology [Tech. rep.]. Stanford Knowledge Systems Laboratory (KSL-01-05) and Stanford Medical Informatics (SMI-2001-0880); 2001.
[55] Perone CS. Pyevolve: a python open-source framework for genetic algorithms. SIGEVOlution 2009;4(1):12–20.
[56] Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. The WEKA data mining software: an update. SIGKDD Explorations 2009;11(1):10–8.
[57] Curk T, Demar J, Xu Q, Leban G, Petrovic U, Bratko I, et al. Microarray data mining with visual programming. Bioinformatics 2005;21:396–8.
[58] Ting K. Confusion matrix. In: Sammut C, Webb G, editors. Encyclopedia of machine learning. USA: Springer; 2010. p. 209.
[59] Whitley D. The GENITOR algorithm and selection pressure: why rank-based allocation of reproductive trials is best. In: Proceedings of the third international conference on genetic algorithms. Morgan Kaufmann; 1989. p. 116–21.
[60] Cochran WG. The 2 test of goodness of fit. Ann Math Stat 1952;23(3):315–45.
[61] Fleiss JL. Statistical methods for rates and proportions, 2nd edition, Wiley series in probability and mathematical statistics. New York: John Wiley & Sons; 1981.
[62] Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33(1):159–74.