Medical Data Mining using Particle Swarm Optimization for Temporal Lobe Epilepsy

M. Ghannad-Rezaie, H. Soltanain-Zadeh, *Senior Member, IEEE,* M.-R. Siadat, K.V. Elisevich

Abstract—In clinical problems, numerous factors are usually involved in a medical syndrome. New advances in medicine provide a broad range of diagnosis methods to cover all aspects of a disease. However, huge amounts of raw information may confuse clinicians and decrease decision accuracy. Computerized knowledge extraction is an active area of research in medical informatics. This paper suggests a new medical data mining approach using an advanced swarm intelligence data mining algorithm. Considering medical knowledge discovery difficulties, this approach addresses common issues such as missing value management and interactive rule extraction. Here, surgery candidate selection in temporal lobe epilepsy is the main target application. However, the general idea can be applied to other medical knowledge discovery problems. Experimental results show noticeable performance improvement in the final rule-set quality while the method is flexible and fast.

I. Introduction

In recent years, generation of huge amounts of medical data as well as health care systems' need for accessing them easily and reliably have raised the interest in medical information systems and computer-aided diagnosis (CAD). These systems help physicians to improve their diagnosis quality and to select the most appropriate treatment, especially when there are many unknown parameters and large number of features, as it is common in neurological diseases. Although great progress has been made in the field of medical information processing, CAD systems have not been developed as fast as the other information systems due to certain problems.

One of the most critical problems is limited acceptance of CAD systems by the clinicians. Common data classification methods have used intelligent approaches that look like *black boxes* to the clinicians. It is our understanding that acceptance of CAD systems depend on their transparent designs and convincing evidence on the results called *reasoning*. The resulting rules describe relations between the available information and the system conclusion. Classical reasoning structures such as Bayesian Believe Networks (BBN) lack flexibility and are weak in extracting information from the raw data, i.e., data mining. However, recent innovations have introduced novel algorithms that outperform the classical methods are more likely to be accepted in the medical community. In particular, Swarm intelligence (SI) methods such as Ant Colony Optimization (ACO) and Particle Swarm Optimization (PSO) [1] have been recently applied to rule extraction and data mining problems. The dynamic essence of SI provides flexibility and robustness. With full control on the extracted rules, SI is a suitable approach to satisfy medical systems requirements.

This paper focuses on the application of PSO in a medical decision support system. It explains usage of swarm intelligence in medical diagnosis systems especially for the problem of surgery candidate selection in temporal lobe epilepsy, which is used as a test-bench to compare different approaches. It is going through fast, flexible and interactive rule extraction. The paper starts with an overview of the state-of-the-art medical diagnosis and data mining systems and the current challenges. Browsing previous work, the failure point of each approach is discussed. In the next step, PSO and its application in data mining are described and finally the combined method and direct PSO methods are compared and drawback of each approach is discussed. All testing and evaluation studies have used the real dataset from the Human Brain Image Database System (HBIDS) developed at Henry Ford Health System. Experimental results show the power of the classical and improved versions of PSO compared with the traditional reasoning algorithms such as BBN and decision tree approaches such as C4.5 [2].

II. Knowledge Discovery From Database In Medical Applications

The overall process of knowledge discovery from database (KDD) is a multistage process. The main step in KDD, Data Mining (DM), is the most commonly used name to describe the computational efforts meant to process feature space information to obtain valuable high level knowledge, which must conform to three main requisites: accuracy, comprehensibility, and interest for the user [3].

This work was supported in part by NIH grant R01-EB002450.

M. Ghannad-Rezaie is with Henry Ford Health System and Wayne State University, Detroit, MI 48202, USA, e-mail: mostafa@rad.hfh.edu.

H. Soltanian-Zadeh is with the Control and Intelligent Processing Center of Excellence, School of Electrical and Computer Engineering, University of Tehran, Tehran 14395-515, Iran, and also with the Radiology Image Analysis Laboratory, Henry Ford Health System, Detroit, MI 48202, USA, phone: (313) 874-4482, fax: (313) 874-4494, e-mails: hszadeh@ut.ac.ir, hamids@rad.hfh.edu.

M.-R. Siadat is with the Radiology Image Analysis Laboratory, Henry Ford Health System, Detroit, MI 48202, USA, e-mail: siadat@rad.hfh.edu.

K.V. Elisevich is with the Neurosurgery Department of Henry Ford Health System, Detroit, MI 48202, USA, e-mail: nskoe@neuro.hfh.edu.

DM discovers patterns among database-stored information to support special user interest such as data classification. CAD systems are the primary areas of interest in data mining [3]. Figure 1 shows the general diagram of a computerized decision system.

Designers of computer-based diagnosis systems often view the physician's primary decision-making task as a differential diagnosis. This term refers to a type of analytical task wherein the decision maker is confronted with a fixed set of diagnostic alternatives.

Over the past two decades, a large number of specialized procedures have been developed to assist physician in differential diagnosis of a variety of well defined clinical problems. These have been extensively reported in the medical and computing literature. In addition, algorithms to deal with a host of common medical problems, expressed by means of detailed flowcharts, have increasingly found their way into the clinical application.

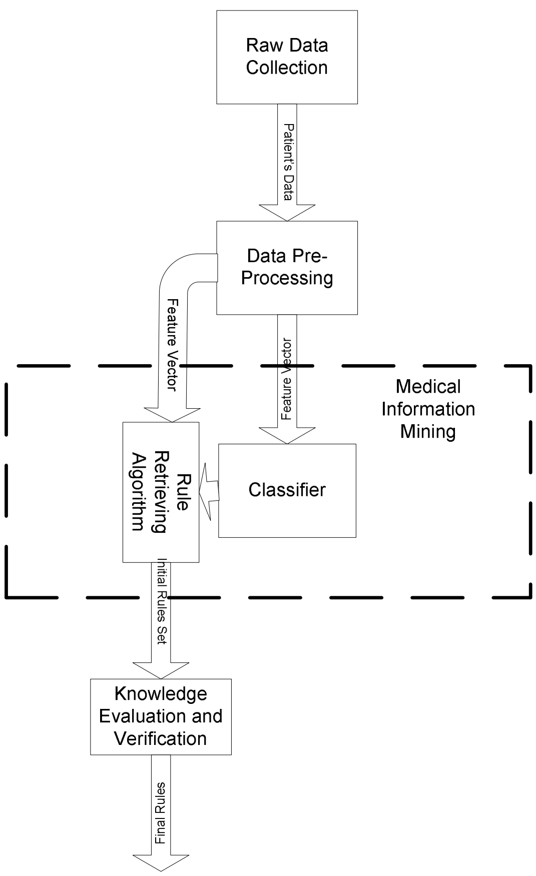

Fig. 1. Knowledge Discovery Diagram for Medical Data.

Many different techniques have been used in structuring these clinical algorithms. In some cases, special programs have been formulated to capture the logic involved in the workup of particular classes of clinical problems. In other cases, generalized procedures have been adopted that are tailored to a particular application by specification of certain parameters; for example, many diagnostic programs have been developed to use the normative models of statistical decision theory. In some complex diagnosis tasks too many parameters such as cancer staging and neurological disease, medical diagnosis systems evolve rapidly. Evaluation studies frequently show that these programs generally perform as well as experienced clinicians in their respective domains, and somewhat better than the non-specialist. It is interesting, therefore, to speculate on the reason that such programs have not had greater impact on the practice of medicine [4].

Resistance in the medical community is sometimes attributed to the natural conservatism of physicians or to their sense of being threatened by the prospect of replacement by machines. Some have argued that this can be resolved only on the basis of education and training, and that the next generation will be more comfortable with computer-based decision aids as these become routinely introduced into the medical community.

III. Epilepsy Surgery Candidate Selection Problem

This section describes the main target problem of this article, epilepsy surgery candidate selection. In the following, we will describe the importance of the problem and the challenges we faces to find a solution. The problem is new in the area of soft computing but still can be considered as a prototype of common medical diagnosis problems such as breast cancer staging or leukemia genome expression.

A. *Problem Statement*

Epilepsy is recognized as an organic process of the brain. More formally, epilepsy is an occasional, excessive, and disorderly discharge of nerve tissue, seizure, which sometimes can be detected by electroencephalographic (EEG) recording. It is a complex symptom caused by a variety of pathological processes that result in treatment selection difficulties. Pharmacotherapy or surgical treatments are the neurologist alternatives. Optimal treatment selection in the first step may change the patient’s life. Temporal lobe epilepsy is one of the most common types of known epilepsy. The main origin of seizures in this type is located in the hippocampus.

Despite optimal pharmacotherapy, about 20–30% of the patients do not become seizure-free [5]. For some of these patients, surgery is a therapeutic option. Success of resective epilepsy surgery increased from 43% to 85% during the period 1986–1999 [6]. Data from multiple sources suggest that 55–70% of patients undergoing temporal resection and 30–50% of patient undergoing extra-temporal resection become completely seizure-free [7]. A recent prospective randomized controlled trial of surgery for temporal lobe epilepsy showed that 58% of patients randomized to surgery became seizure-free compared to 8% of the medical group [7].

Surgery is considered a valuable option for medically intractable epilepsy even in the absence of a proven drug resistance. In addition, surgical outcome may be greatly influenced by the presence of selected prognostic indicators [8, 9]. However, there are still uncertainties on who are the best surgical candidates, i.e., those who most likely will present good surgical outcome.

In a recent narrative literature review of temporal resections, good surgical outcome was associated with a number of factors (hippocampal sclerosis, anterior temporal localization of interictal epileptiform activity, absence of preoperative generalized seizures, and absence of seizures in the first post-operative week) [5]. However, the published results were frequently confusing and contradictory, thus preventing inferences for clinical practice. Methodological issues (e.g., sample size, selection criteria, and methods of analysis) were indicated as the most likely explanation of the conflicting literature reports [9]. For this reason, a quantitative review of the available literature has been undertaken in [9] to assess the overall outcome of the epilepsy surgery and to identify the factors better correlating to seizure outcome. The aim of the study was to perform a meta-analysis of the results of published observational studies and assess the prognostic significance of the selected variables outlining the characteristics of the clinical condition, the correlations between the epileptogenic and functional lesion, and the type of surgical procedure.

B. *Database for Epilepsy Patients*

Human brain image database system (HBIDS) is under development for epilepsy patients at Henry Ford Health System, Detroit, MI [10, 11]. It will examine surgical candidacy among temporal lobe epilepsy patients based on their brain images and other data modalities. Moreover, it can discover relatively weak correlations between symptoms, medical history, treatment planning, outcome of the epilepsy surgery, and brain images. Its data include MRI and SPECT along with patient's personal and medical information and EEG study [10]. The data is de-identified according to HIPPA regulations [11].

For the first phase of the EEG study, the non-visual feature extractor is an expert or specialist. The experts do this routinely in the clinic based on well-defined standards. For un-structured text information, the wrapper is the expert or trained nurse. The structured data such as patient's personal information do not need to be analyzed by the wrapper, so they are directly stored in the database [11].

C. *Candidate Selection Problem*

Most of data mining methods are designed to work on a huge amount of data. Thus, KDD problem with small sample size does not broadly browse in the data mining literature. The most successful approach is to add classification or modeling stage before the rule extraction. A smart classifier can recover the patterns inside dataset and minder can recover the classification rules. In this approach, thoughtful selection of both steps is critical.

Candidate selection in epilepsy, more generally in medical diagnosis, is a hard pattern recognition problem. As well as many current bioinformatics problems, the main challenge in candidate selection problem is to find an optimal point in a very large-dimensional data-space with few samples. As an example of other problems with the same challenge, functional gene classification problem [12] has a reduced feature space with 200 dimensions while usually less than 50 samples are available in each case. Our epilepsy problem has a 40-dimensional space and around 35 samples. Common soft computing tools such as neural networks are efficiently applicable only on large datasets. The longer feature vector, the larger database is required. Overtraining problem is always a threat for small samples machine learning. On the other hand, conventional feature space dimension reduction algorithms such as principle component analysis (PCA) are based on statistical computations that can not be applied to small number of samples. Other difficulties such as missing data, large variety of data types, feature disturbances, and prior knowledge make the problem more complicated. Furthermore, knowledge recovery in this problem is not straightforward.

IV. Knowledge Discovery Algorithms

Many KDD algorithms have been proposed in the literature. However, most of them do not satisfy limitations of medical applications and thus medical data mining is still a hot research topic. The first common KDD algorithm class is based on decision tree classifiers. Decision trees are a standard tool in data mining, and many are available in packages such as C4.5 [13].

Another alternative for a popular classifier is Bayesian network [14, 15]. Here, all prior assumptions are made explicit and the weights and hyper plane parameters are determined by applying the Bayes theorem to map the prior assumptions into posterior knowledge after having observed the training data.

V. Soft Computing For Knowledge Discovery

In recent years, various soft computing methodologies have been applied to address data mining [16]. Generally, there is no universal best data mining algorithm. Choosing appropriate data mining algorithm utterly depends on the application. Social and SI algorithms are well-known alternatives of soft computing tools that can be used for retrieving information from raw data [17]. *Distributed Genetic Algorithm*, *ACO* and *PSO* are the most commonly used evolutionary algorithms in this domain. The most interesting contribution of these methods is in flexible rule extraction where we are facing with *incomplete and inaccurate* datasets [18].

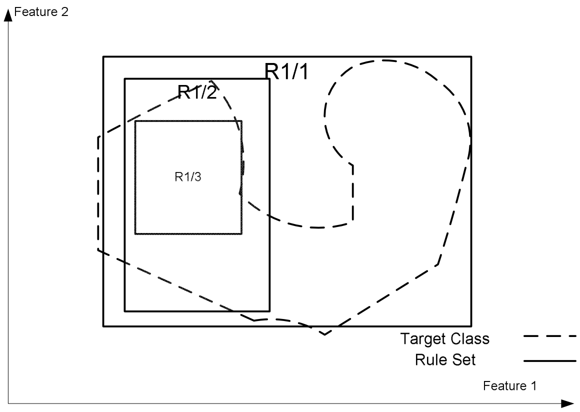

Fig. 2. Rectangular Tree Based Rule Set.

A. Rule Extraction

According to previous discussions, finding a rule-based classifier and reasons behind the decision making process are essential parts of CAD systems. Also, they are key parts of KDD. This section discusses the mathematical modeling of rule extraction process and application of PSO to find the rule set describing a support vector machine classifier.

Assume S is the search space and θi is a data point inside S and $yi=V(\theta i)$ is its class. A classifier is define by $yi^{1}= U(\theta_i)$. The target of rule extraction is to find a rule set $R^{U}{}_{1..n}$ that describes the U classifier.

Each rule is an "IF-THEN" statement with two clauses. In the simplest case, the former clause is a condition on the search space and the latter clause is the target class. The structure of these two clauses is called rule set grammar limiting phrases in the rule clauses. The simpler the grammar, the more comprehensive statements can be retrieved. Rectangular grammar limits the IF-clause to intervals of each individual feature. Fig. 6 shows an example of the rectangular rules. Decision tree is another alternative of rule set topology (Fig. 2) which is quite common in medical applications because of better interpretability and higher searching speed. In this structure, the IF-part may also include another rule in addition to intervals but the number of intervals is limited.

B. Rule-Set Evaluation

The value of a rule is evaluated using different parameters:

– *Accuracy*: percentage of data points correctly classified.

$$A_R = \frac{\#\theta : V(\theta) = R(\theta)}{\#\theta} = \frac{TP+TN}{TP+TN+FP+FN} \qquad (1)$$

– *Quality*: obtained from the ROC curve by:

$$Q_R = \text{sensitivity} \cdot \text{specificity} = \frac{TP \cdot TN}{(TP+FN)\cdot(TP+FN)} \qquad (2)$$

– *Coverage*: percentage C_R of specific class data points that are covered by the rule set.

– *Simplicity*: the number of terms in the rule clauses and the number of intervals on each term condition representation (S_R). Generally, a very complex rule can describe any classifier and achieve very high Coverage and Accuracy spontaneously. Comprehensibility as a critical parameter in medical data mining is measured by this term. The number of rules in a rule set, and the number of terms in a rule represent the complexity phenomena. Simplicity is defined as *1/Complexity*.

– *Rule interference*: Adding to the individual rule evaluating parameters, the final rule set is admirable when it can cover the entire search space while the conflict among rules is kept as low as possible. Interference parameter (I_R) is particularly considered for reasoning process and reliable decision making.

Finally, the rule set evaluation is $Eval(R) = \alpha A_R + \beta C_R + \gamma S_R - \delta I_R$. The accuracy measure can be replaced by the quality when the trade-off between sensitivity and specificity is highly interested. Finding the best rule set is a complex multi-objective process.

C. Swarm Intelligence Rule Extraction

Previous work in the literature shows power of PSO in solving rule extraction problems in medical diagnosis systems. Many recommended modifications of PSO for providing a flexible approach to address common difficulties of medical information retrieval from databases. Here, we present hybrid approaches to overcome problems presented in earlier sections.

ACO and ant miners is the first swarm intelligent data mining algorithm. The purpose of their algorithm, Ant-Miner is to use ants to create rules describing an underlying data set. The overall approach of Ant-Miner is a separate-and-conquer one as same as C4.5 [17]. It starts with a full training set, creates a best rule that covers a subset of the training data, adds the best rule to its discovered rule list, removes the instances covered by the rule from the training data, and starts again with a reduced training set. This goes on until only a few instances are left in the training data (fewer than max number allowed) or the fitness function meets the target, at which point a default rule is created to cover remaining instances. Once an ant has stopped building a rule antecedent a rule consequent is chosen. This is done by assigning to the rule consequent the class label of the majority class among the instances covered by the built rule antecedent. The rule is then pruned in order to improve its quality and comprehensibility. The basic idea is to iteratively remove one term at a time from the rule while this process improves the rule quality as defined by a fitness function. In an iteration, each term in turn is temporarily removed from the rule antecedent, a new rule consequent is assigned, and the rule quality is evaluated. At the end of the iteration, the term that has actually been removed is the one that improves the rule quality the most.

VI. Particle Swarm Optimization for Rule Indication

Rule discovery process can be done using a Particle Swarm Intelligence Algorithm. PSO imitates the intelligent behavior of beings as part of a group to experience sharing in a society. In contrast to conventional learning algorithms with individual reaction to environment or searching space, PSO is based on adaptive social behaviors. The basic idea of the PSO model is constructed on three ideas: evaluation, comparison and imitation. Evaluation is the kernel part of any intelligent algorithm measuring quality of the result in the environment and usefulness inside the community. Some metrics are defined to represent the particle superiority. This evaluation is pointless without well-defined comparison process which is a sequential relationship in the particle space. The improvement of particles is made by imitating best solution up to now. Looking to best solution in the neighborhood, a particle decides where to move in the next step. There are many alternatives for implementation of neighborhood and distance between particle concepts.

A. *PSO Algorithm*

PSO is a set of individual agents simply searching for an optimal point in their neighborhood. The movement of agents depends on behavior of other agents in the vicinity and the best visited nodes. During PSO training particle's best met position (BPi) and the best solution met by neighbors (BP_{Ni}) is updated. A position vector and a velocity vector in the feature space are assigned to each particle. The standard PSO parameters update formulation is:

$$\begin{cases} v_i(t) = v_i(t-1) + \varphi_{1i}[BP_i - x_i(t-1)] \\ \qquad\qquad + \varphi_{2i}[BP_{Ni} - x_i(t-1)] \\ x_i(t) = x_i(t-1) + v_i(t) \end{cases} \tag{3}$$

In the new version of PSO, a weight term is applied to prevent divergence of the velocity vector:

$$v_i(t) = \alpha(v_i(t-1) + \varphi_{1i}[BP_i - x_i(t-1)] + \varphi_{2i}[BP_{Ni} - x_i(t-1)]) \tag{4}$$

For a more general definition of the distance aspect [16], a more general form of position update equation is:

$$v_i(t) = \alpha(v_i(t-1) + \varphi_{1i}[Dis(BP_i, x_i(t-1))] + \varphi_{2i}[Dis(BP_{Ni}, x_i(t-1))]) \tag{5}$$

General optimization algorithm is summarized in Table 1.

Table 1. PSO pseudo-code.

```
Assume K particles
Distribute particles in the searching area

While (Fitness < TARGET_FITNESS && Epoch < MAX_EPOCH)
{
evaluateTotalFitness();
For any particle p
{
  Evaluate(p);
  UpdateBestPosition(p);
  UpdateNeighborhoodList(p);
  UpdateBestPositionInNeighborhood(p);

  UpdatePositionInAllDimensions(p);

}
}
```

B. *PSO for Database Rule Extraction*

Rule extraction process usually contains two stages: rule set generation and pruning. Rule generation is a forward selection algorithm adding new rules to current rule-set. In contract, pruning or cleaning process is a backward elimination algorithm omitting extra rules form rule-set. PSO could efficiently apply in the rule generation process. While PSO is a strong optimization algorithm in the large search space, it could obtain rule-set with maximum fitness function, no mater how complex it is. Sousa et al [17] proposed a particle swarm based data mining algorithm. Comparing different PSO implementations with C4.5 and other evolutionary algorithms, they concluded that PSO can obtain competitive results against other alternative data mining algorithms, although it needs a bit more computational effort.

C. *Neighborhood Structure Effect on Rule Set*

Neighborhood and social networks phenomena are new issues proposed in swarm intelligence by PSO. The influence of different social network structures on the performance of PSO has been studied arguing that a manipulation affecting the topological distance among particles might affect the rate and degree to which individuals are attracted towards a particular solution. Four different types of neighborhood topologies have been proposed in the literature. In circles structure, each individual is connected to a number of its immediate neighbors. In wheels structure, one individual is connected to all others, and these are connected only to that one. In star neighborhood, every individual is connected to every other individual. Finally, random topology is used for some individuals and random symmetrical connections are assigned between pairs of individuals.

Structure of the neighborhood may affect the data mining process. Also, the rule-set fitness and convergence depend on the neighborhood structure. Experimental results have shown that the neighborhood topology of the particle swarm has a significant effect on its ability in finding the optimal rules. Best pattern of connectivity among the particles depends on the fitness function. In the single rule extraction, the wheel structure improves the convergence speed of the data mining process. However, starting with a random neighborhood is a better choice for multi-rule extraction.

D. Decision Tree Rule Indication Using Structured PSO

Decision trees are powerful classification structures. Standard techniques such as C4.5 can produce structured rules for the decision tree. These techniques follow divide and conquer strategy on the data set to obtain two subsets. The algorithm is applied to subsets recursively. Each intermediate node tests a feature. Following the path between leafs and root could be taken as a simple rectangular rule. The PSO neighborhood concept can be designed to extract tree based rules directly. In this structure, each agent has a single decision node. Neighborhood definition could force tree rule indication. Adjacent agents are defined as similar limitations on all features but one. The best point in the vicinity is the solution that satisfies neighborhood condition while having lowest limitation. From the decision tree point of view, the best node in the neighborhood is the root of its sub-tree and of course the best solution is the decision tree root. The final solution is presented by all agents together. Also, the fitness function in this application is different from the original target function and depends on the size of the rule-set as well as the accuracy parameters as described in the previous chapter.

E. Rule Injection and Rejection

The clinicians are involved in the rule-set construction by rejecting an existing rule inside the database or injecting new rules into the dataset. After the injection or rejection process, other rules inside the database may be affected. In the PSO algorithm, there are two *absorption* points, best local solution and best global solution. New injection rule could model as a new absorption point. Injected rule may affect other solutions in the vicinity. However, the training process is not applied to this rule. On the other hand, rejected rules model penalty terms in the fitness function. By adjusting mandatory conditions on the rules, the target function changes to produce more realistic rule-set.

VII. Experimental Results

A. Classifier Evaluation

Medical classification accuracy studies often yield continuous data based on predictive models for treatment outcomes. Evaluation of the classifier efficiency is computed with regard to the true or false classification results. True positive (TP), true negative (TN), false positive (FP), and false negative (FN) values are the basic evaluation measures for a classifier. The sensitivity and specificity of a diagnostic test depends on more than just the "quality" of the test -- they also depend on the definition of what constitutes an abnormal test. A popular method for evaluating the performance of a diagnostic test is the receiver operating characteristic (ROC) curve analysis [19]. This is a plot of the true positive rate against the false positive rate for the different possible cut-points of the classifier. Each point of the ROC curve is obtained by finding the true positive rate when the decision threshold is selected based on a specific false alarm rate.

The area under the ROC curve represents accuracy of a classifier. In medical problems, false alarm rate as well as false rejection rate should be lower than pre-specified limits. The trade off between false alarm rate and false rejection rate is problem specific. In our surgery decision-making problem, both rates must be considered. However, false alarm rate (doing surgery for a patient who does not need it) is more likely to be of concern.

B. Cross Validation Training

Because of a very low number of samples, complete separation of the test and train sets is not appropriate. Cross-validation is used to reuse train information in test process to measure the generalization error [20]. Assume $F=\{(\vec{\theta}_1,y_1),(\vec{\theta}_2,y_2),\ldots,(\vec{\theta}_l,y_l)\}$ is a set with cardinality of l and an algorithm maps F to V_F in the results space. We would like to measure the generalization error. Cross-validation uses l-p samples to find the function of V_{l-p} where the generalization error is measured by:

$$e_1=\sum_{x_i\in F_p} Eval(V_{l-p}(\vec{\theta}_i),y_i) \tag{6}$$

This process is repeated M times and the final error expectation is:

$$\hat{e}=\frac{1}{M}\sum_{i=1}^{M} e_i \tag{7}$$

which is expected to be the generalization error of V_l. When $p=1$, it can be shown that the generalization error estimation is unbiased. Although this validation is time consuming, it significantly increases the power of the training process. For most efficient use of the data, training and test sets are not separated. In each training epoch, 4/5 of the patients are randomly selected to train the classifier. The rest of the patients (1/5) are used to test. The final classifier is the average of many training processes. This training strategy provides maximum database usage efficiency at the cost of higher computational complexity. In this experiment, more than 50 train-test sets are used. The training process terminates when the classifier's mean squared error of the test-set increases in the last two epochs.

C. Performance

Here, we present experimental results of comparison of the four algorithms: structured decision tree training as the representative of classic mining algorithms, Ant colony miner as an evolutionary algorithm pioneer in medical data mining, previously proposed PSO database miner, and the hybrid approach. Common train and test datasets have been used for all data miners. Tables 2 and 3 compare performance of different algorithms.

The performance of data mining algorithms is compared from different points of view. Performance of the generated rule sets has been compared using evaluation functions proposed in the previous parts. Relatively, C4.5 generates the most accurate solution. Actually, it misses very few test cases but the overall score of this approach is quit low. Having a close look at the simplicity metric and the number of rules, it is obvious that the high accuracy of C4.5 is the result of a more complex rule set and lower generalization.

C4.5 is a very fast algorithm compared with the SI algorithm due to its iterative and divide/conquer strategy, thus it cannot be fairly compared with the evolutionary data miners (Fig. 4). It is especially designed to handle huge amount of data, thus, we expect a very fast convergence. Among the evolutionary algorithms, PSO shows a very good convergence speed. Simulations show that even with an additional classification learning process, PSO is faster than the conventional ACO miner while having the same performance.

Altogether, C4.5 shows to be a powerful method but the resulting rule is too complex to use. Fast convergence of C4.5 is impressive but for small databases, it is not recommended. The PSO applied after the classification process outperforms the direct PSO knowledge recovery from the database but it is comparable with ACO in some aspects. Generally, simulation results show that the proposed hybrid process has the best overall evaluation while is still somewhat faster than the previous evolutionary algorithms. Also, a bit more memory usage can count as a drawback of the new approach compared to ACO and simple PSO. Effect of rule injection and rejection process is shown in Fig. 6.

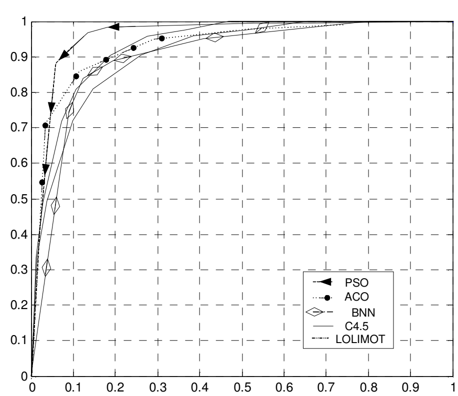

Fig. 3. Receiver operating characteristic (ROC) curves for different algorithms.

VIII. Conclusion

We discussed an integration of a classifier and PSO rule extraction methods to mine patient data for surgery candidate selection in epilepsy. We have demonstrated that support vector machines can accurately classify patients into suitable and inappropriate candidates for surgery. Among the techniques we examined, the regularized classifier using a radial basis kernel function provided the best performance.

References

[1] J. Kennedy, "The behavior of particles," Proc. of Evolutionary Programming Conf., pp. 32-37, 1998.

[2] J. R. Quinlan, "C4.5: programs for machine learning," Morgan Kaufmann Series in Machine Learning, Morgan Kaufmann Inc., London, 1993.

[3] H. Mannila, "Theoretical frameworks for data mining." SIGKDD Explorations J., vol. 1, issue 2, pp. 30-32, 2000.

[4] H. E. Pople, "Heuristic methods for imposing structure on ill-structured problems: the structuring of medical diagnostics: artificial intelligence in medicine," Westview Press, Boulder, Colorado, 1982.

[5] J. Janszky, R. Schulz, A. Ebner, "Clinical features and surgical outcome of medial temporal lobe epilepsy with a history of complex febrile convulsions," Elsevier Trans. on Epilepsy Research, vol. 55, issue 1, pp. 1-8, 2003.

[6] C. Kilpatrick, T. O'Brien, Z. Matkovic, "Preoperative evaluation for temporal lobe surgery." J. of Clinical Neuroscience, vol. 10, issue 5, pp. 535-539, 2003.

[7] P. Kwan, M. J. Brodie, "Early identification of refractory epilepsy." Neurology J. in England, vol. 342, issue 5, pp. 314-319, 2000.

[8] C. Tonini, E. Beghi, A. T. Berg, "Predictors of epilepsy surgery outcome: a meta- analysis," Elsevier Trans. on Epilepsy Research, vol. 62, pp. 75-87, 2004.

[9] J. Janszky, H.W. Pannek, A. Fogarasi, B. Bone, R. Schulz, F. Behne, A. Ebner, "Prognostic factors for surgery of neocortical temporal lobe epilepsy," Seizure J., vol. 15, issue, pp. 125-132, 2006.

[10] M. G. Rezaie, H. Soltanian-Zadeh, M. Siadat, R. A. Zoroofi, K. V. Elisevich, "MRI-SPECT data fusion for temporal lobe epilepsy surgery candidate selection," Proc of Symp. of Signal and Image Processing (SIP), Hawaii, 2005.

[11] M. G. Rezaie, H. Soltanian- Zadeh, M. Siada, K. V. Elisevich, "Soft computing approaches to computer aided decision making for temporal lobe epilepsy," Proc. of IEEE International Conf. on Fuzzy Systems (NAFIPS), Ann Arbor, Michigan, 2005.

[12] P. J. S. Silva, R. F. Hashimoto, S. Kim, J. Barrera, L. O. Brandao, E. Suh, E. R. Dougherty, "Feature selection algorithms to find strong genes," Pattern Recognition Letters, vol. 26, issue 10, pp. 1444-1453, 2005.

[13] J. Li, L. Wong, "Using rules to analyse biomedical data: comparison between C4.5 and PCL." Technical Report, Institute for Infocomm Research, Singapore, 2003.

[14] P. Larranaga, "Learning Bayesian networks from data: some applications in biomedicine," Technical Report, Intelligent System Group, Department of Computer Science and Artificial Intelligence, University of Basque, 2002.

[15] J. I. Kazi, P. N. Furness, M. Nicholson, E. Ahmed, F. Akhter, A. Naqvi, A. Rizvi, "Interinstitutional variation in the performance of Baysian belief network for the diagnosis of acute renal graft rejection ," Transplantation Proc., vol. 31, issue 8, pp. 3152, 1999.

[16] M. Galea, "Applying swarm intelligence to rule induction." MS Thesis, University of Edinburgh, Scotland, 2002.

[17] T. Sousa, A. Silva, A. Neves, "Particle Swarm based data mining algorithms for classification tasks," Parallel Computing J., vol. 30, pp. 767-783, 2004.

[18] Z Zheng, BT Low, "Classifying unseen cases with many missing values," Proc. of Methodologies for Knowledge Discovery and Data Mining Conf., vol. 2, pp. 370, 1999.

[19] K. H. Zou, F. S. Resnic, I. Talos, "A global goodness-of-fit test for receiver operating characteristic curve analysis via the bootstrap method," J. of Biomedical Informatics, available online on www. sciencedirect. com, 2005.

[20] N. T. Merwe, A. J. Hoffman, "Developing an efficient cross validation strategy to determine classifier performance," Proc. of Neural Network Conf. (IJCNN), vol. 3, pp. 1663-1668, 2001.

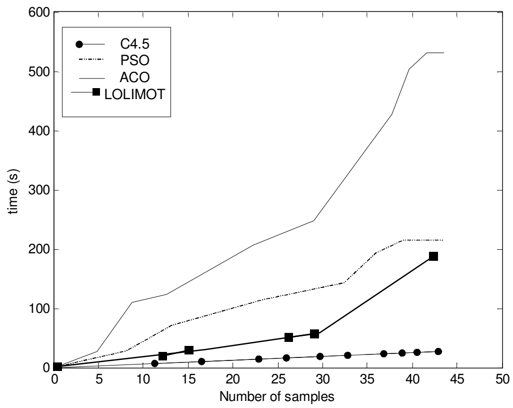

Fig. 4. Effect of the number of sample points on the convergence time.

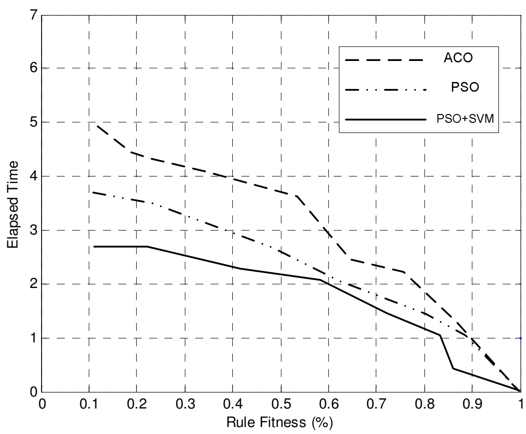

Fig. 5. Effect of the target fitness function on the convergence time.

Table 2. Overall performance of different data miners.

	Number of Total Terms in Rules	Accuracy (%)	Overall Evaluation (%)
C4.5 (J48)	9	92.9	81.7
ACO	8	76.5	88.3
PSO on Database	10	87.7	84.2
PSO on a Regularized Classifier	6	89.1	91.7

Table 3. Classification parameters for the rule sets obtained.

	Sensitivity (%)	Specificity (%)	Simplicity (%)
C4.5 (J48)	0.8421	0.9911	64.7
ACO	0.9521	0.9821	74.2
PSO on Database	0.8192	0.9541	53.8
PSO on a Regularized Classifier	0.9821	0.9781	87.5

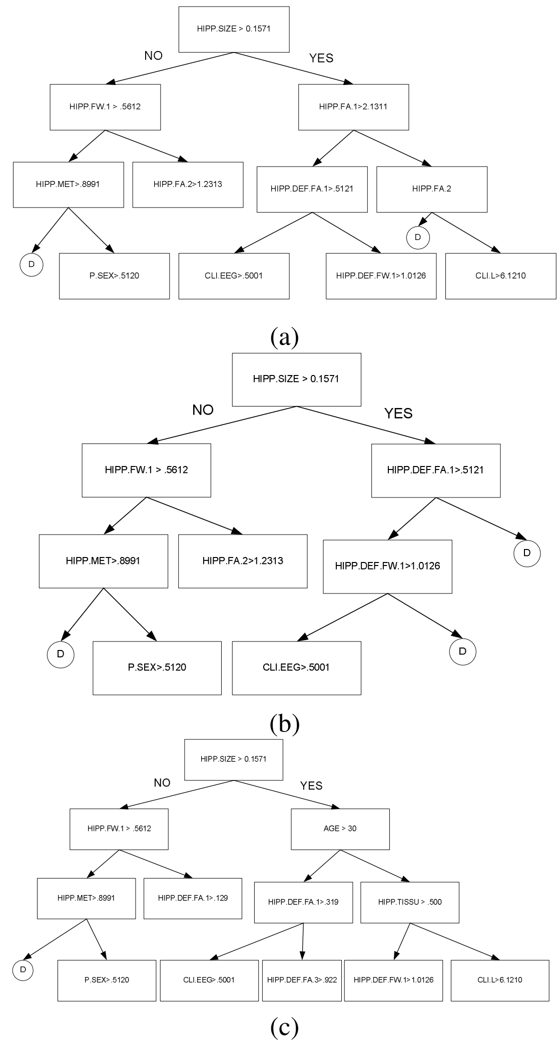

Fig. 6. a) Rules extracted in the decision tree form, b) rule rejection (R22), c) Rule injection.