Improving Accuracy and Interpretability of Clinical Decision Support Systems through Possibilistic Constrained Evolutionary Optimization

Domenico Maisto*, Massimo Esposito*

*Institute for High Performance Computing and Networking - ICAR
Italian National Council of Research
Via Pietro Castellino 111, Naples, Italy
Email: {domenico.maisto,massimo.esposito}@na.icar.cnr.it

Abstract—In this paper we propose a computable approach to represent medical rules by means of Fuzzy Clinical Decision Support Systems by preserving both accuracy and interpretability. Usually, prediction accuracy of these systems goes to overlook their linguistic interpretability and, in order to simultaneously optimize those conflicting properties, multi-objective evolutionary algorithms are adopted. Differently, our proposal relies on the alternative assumption that the interpretability-accuracy tradeoff problem can be approached as a single-objective constrained optimization problem. In this spirit, a Differential Evolution algorithm, with a selection operator suitably adapted to the aim, is used for membership function tuning by maximizing accuracy and fulfilling several constraints for linguistic distinguishability degrees — a semantic property of fuzzy sets with notable relevance for interpretability of fuzzy models — evaluated through Possibility measure. The proposed approach has been tested on the Vertebral Column Data set, a recent medical database publicly available, with results that confirm the effectiveness of our method.

Keywords—Clinical Decision Support Systems; Possibility Theory; Constrained Evolutionary Optimization

I. INTRODUCTION

Medical Informatics (MI) is a recent new interdisciplinary area quickly developed thanks to an intensive application of computer-based methods to various fields of medicine.

In the last few years, MI community’s contribution to clinical practice has been fundamental in terms of accessibility, updatability, manageability and effectiveness [1]. Among these contributions, a large part is dedicated to design Clinical Decision Support Systems (CDSSs) [2] aimed to aid physicians in clinical decision-process [3].

In general, CDSSs are computer artifacts where experts’ knowledge, expressed through standardized specifications for healthcare denoted as clinical guidelines [4], [5], is blended with several scientific evidences, e.g., biomedical data attained by clinical research or screening campaigns, to develop a formal process expected to promote more consistent, effective, and efficient medical practices and improve health outcomes when followed [6], [7].

Nevertheless, medical concepts and care decisions are often characterized by subjectivity, vagueness and inexactitude. Therefore, computable models representing medical information should be able to deal with various forms of imprecision and CDSS implementations should provide reasoning methods suitable to tolerate them [8].

With this respect, Fuzzy Logic (FL) [9] has been identified as the most suitable approach to describe vagueness and imprecision, since it enables to explicitly represent vagueness by means of a mathematical approach rather than abolishing it. Specifically, in the case of CDSSs, Fuzzy Rule Bases (FRBs) can be adopted for representing healthcare guidelines with the notable advantages that clinical rules can be encoded in a computable form similar to the natural language and clinical recommendations provided by inference are consistent with human thinking.

On the other hand, these features of FRBs could be overlooked if one considers that the prediction accuracy of the system is an objective equally important for CDSS design. In fact, in CDSSs based on FL, accuracy improvement is usually attained by tuning the parameters defining the Membership Functions (MFs) associated to the linguistic terms of the fuzzy model, but this approach modifies the shapes of the MFs and can compromise the semantic interpretability of the model. Consequently, accuracy and interpretability are conflicting properties and research about designing methods and techniques for their simultaneous optimization is still an open issue.

A very popular way to contemporarily optimize accuracy and interpretability relies on the use of multi-objective evolutionary algorithms (MOEAs) [10] by which a set of solutions with different accuracy-interpretability tradeoffs is obtained [11], [12]. However, experimental analyses have showed that lower error rates are obtained from the single-objective formulations than the multiobjective ones [13], probably because multi-objective evolutionary algorithms in large-scale combinatorial optimization problems, such as multiobjective design of fuzzy rule-based classifiers, have difficult to find a wide front of solutions [13].

In this work, an alternative approach is proposed in order to increase accuracy of FRBs representing clinical rules, trying to find a right balance with linguistic interpretability. A Differential Evolution algorithm [14], [15] is used to tune MFs while some constraint conditions introduced for distinguishability — a semantic property of fuzzy sets with notable relevance for interpretability of fuzzy models — have to be fulfilled. Particularly, Possibility Theory, intuited by Zadeh and formulated mainly by Dubois and Prade [16], is exploited to assess the degree of distinguishability of several fuzzy sets related to a single variable. Consequently, a measure of distinguishability in
possibilistic terms is proposed for FRBs as a whole.

The paper is structured as follows. In Section II we introduce the concept of interpretability for FRBs and how it can be measured through Possibility Theory. Section III illustrates in detail a novel way to effect MF tuning by preserving linguistic interpretability of the model and its application to the case of CDSS design. The proposed approach has been tested on a particular case of study, the Vertebral Column Data set recently introduced in literature, and experimental results are presented in Section IV. Finally, in Conclusions, we carry out some considerations and discuss further research issues.

II. DISTINGUISHABILITY AND LINGUISTIC INTERPRETABILITY

Interpretability is a relevant feature in fuzzy modeling as attested by numerous research contributions produced so far [17]. Generally, to design interpretable fuzzy models, a set of properties regarding various components of the model, from the fuzzy sets (sets whose elements have degrees of membership) to the rules, has to be fulfilled. One of the most studied among these is distinguishability [18], [19]. Informally speaking, two, or more, fuzzy sets defined over the same Universe of Discourse (UoD) are as more distinguishable as less overlapped. Because of that, distinguishable fuzzy sets can better represent semantic difference between their associated concepts, a desirable feature if one wants to design interpretable fuzzy models. Other advantages brought by distinguishability property satisfaction are: avoiding subjectivity in MF/linguistic terms association, strengthened consistency and reduced complexity of fuzzy models and a straightforward semantic interpretation of fuzzy sets [20].

Distinguishability has been defined in several ways. A widely adopted definition uses set-theoretic similarity measure for its characterization. In this case, distinguishability definition is inversely related to the degree of similarity between fuzzy sets. This characterization comes well suited but generally involves intensive computing — a notable drawback in automatic fuzzy modeling approaches which use soft computing algorithms to build models. This is particularly true for population-based algorithms, e.g., evolutionary algorithms, based on massive number of evaluations to implement their heuristic research.

To get over this important restriction, possibility measure can be adopted to assess distinguishability between fuzzy sets [21]. It has been introduced as an extension of Fuzzy Set theory [16] and, although it does not hold some requirements of similarity, it is able to quantify the degree of belief that two fuzzy sets are similar. Furthermore, possibility evaluation is computationally inexpensive because, with many MF shapes, the possibility measure of two fuzzy sets can be analytically expressed as function of MFs’ parameters [21].

Possibility measure between two fuzzy sets A and B is defined as:

$$\Pi(A, B) = \sup_{x \in U} \min \{\mu_A(x), \mu_B(x)\}$$  \hspace{1cm} (1)

where $U$ is the UoD of the variable $x$. It evaluates the degree of belief of the statement “$x$ is $B$” for $x = A$. Therefore, possibility measure can be understood as a degree of overlapping between $A$ and $B$ [21].

Possibility measure defined in Equation (1) can be used to estimate distinguishability measure of two fuzzy sets. In facts, distinguishability measure of $A$ and $B$ is the complement of their possibility measure:

$$\Delta(A, B) = 1 - \Pi(A, B)$$  \hspace{1cm} (2)

This definition can be immediately extended to a Frame of Cognition (FoC), an entire collection of $J$ fuzzy sets defined over the same UoD labelled with linguistic terms representing semantic concepts. Consequently, distinguishability of a FoC $\Phi_x$ defined over $x$, is evaluated by means of the following expression:

$$\Delta(\Phi_x) = 1 - \sup_{j \neq j'} \Pi(X^j, X^{j'})$$  \hspace{1cm} (3)

with $X^j$ and $X^{j'}$ being two distinct fuzzy sets in $\Phi_x$, with $j, j' = 1, \ldots, J$.

Equation (3), introduced in this work for the first time, stems directly from equation (2) and can be used to assess distinguishability of fuzzy partitions obtained by both expert design and data mining procedures with the aim to increase interpretability property in fuzzy modeling.

Finally, it is worth remarking that Equation (2) and Equation (3) are not properly measures of “dissimilarity”. For example, according to Equation (2), two fuzzy sets can have distinguishability equal to 0 even if they are normal and not perfectly overlapped. Further, similarity and possibility measures are not monotonically correlated.

However, under mild conditions, it is possible to consider the equations (2) and (3) as effective measures of distinguishability. Mencar et al. [21] have proved that if fuzzy sets are 1) normal, 2) convex and 3) continuous then possibility and similarity are related monotonically. This is especially true when examined fuzzy sets are subjected to some specific transformations, like, for example, translation and contraction, holding some general hypotheses [21].

III. IMPROVING ACCURACY AND DISTINGUISHABILITY OF FUZZY CDSSs THROUGH CONSTRAINED EVOLUTIONARY OPTIMIZATION

In fuzzy CDSS design, medical concepts and decision-making process need to be translated in terms of linguistic variables, MF linguistic values and fuzzy inference able to capture their intrinsic vagueness.

Usually, this fuzzy representation is obtained by formalizing the medical condition-action process as a FRB, i.e., a collection of fuzzy if-then rules defined as implication statements with an antecedent and a consequent. In particular, in order to guarantee comprehensibility of our fuzzy representation, we assume that antecedents are fuzzy-logic expressions in Conjunctive Normal Form (CNF), i.e., a conjunction of more clauses composed of a disjunction of propositions referring to a unique linguistic variable, and
consequently are composed of a single clause assigning a value to an action variable [22].

Furthermore, we have introduced another assumption regarding the kind of inference process chosen to assess the collection of fuzzy rules composing the FRB, so as to carry out a medical action. We assume that the fuzzy rule base has an implicit interface interpretation, i.e., rules are conjuncted and their aggregation is realized through t-norms [23], and we apply the Mamdani’s Center-of-Gravity (CoG) defuzzification method [24] to obtain crisp values as results of reasoning.

At this point, once specified which is the fuzzy representation of examined medical rules, one needs to determine the MFs corresponding to the linguistic terms introduced for clinical variables.

In this ticklish phase, known in the FL literature as fuzzy partitioning, a series of fuzzy sets are generated for every FoC with the aim to obtain a right balance between interpretability and accuracy, i.e., a classification rate assessed with respect to a database taken as reference.

Generally this problem is approached as a multi-objective optimization problem and often solved by means of evolutionary algorithms using Pareto dominance [10] to find solutions with a good accuracy-interpretability tradeoff [11], [12].

In this paper, we present an alternative approach by which the accuracy-interpretability balance problem is transformed in a multi-constrained single-objective optimization problem with real values where MF tuning solutions are real vectors encoding fuzzy partitions satisfying specific constraints related to linguistic interpretability.

A. A Formulation In Terms of Multi-Constrained Optimization

Before introducing our multi-constrained optimization formulation for the considered problem, it is necessary to premise some remarks about the constraints we use and the vectorial representation adopted in the optimization procedure.

The constraints we consider are based on the following consideration: a fuzzy partition can be reckoned as “interpretable” if 1) all fuzzy sets remain acceptably disjoint allowing in such a way to associate a semantic meaning to their own linguistic values; 2) any UoD element belongs to at least one of the fuzzy sets defined for linguistic values.

These requirements can be formalized by using the distinguishability measure given in (3). A fuzzy partition is linguistically interpretable if, for each FoC $\Phi_x$, defined for each variable $x_u$, holds the following statement:

$$\Delta_{\text{min}} < \Delta(\Phi_{x_u}) < \Delta_{\text{MAX}}$$

where $\Delta_{\text{min}}$ and $\Delta_{\text{MAX}}$ denote respectively the minimum and the maximum values allowed for the distinguishability measure with values determined in order to comply with the aforementioned requirements. By setting $\Delta_{\text{min}} = 0.5$ and $\Delta_{\text{MAX}} = 1$, we assume that adjacent fuzzy sets belonging to each FoC overlap just partially, preserving their semantic interpretation, and at least in a point so that in such a way each UoD results always totally covered.

For what concerns the vectorial representation of the fuzzy partitions, we fix a number of fuzzy sets per variable in accordance with the concepts introduced in the medical rules and we establish the shape of MFs by choosing it among those ones fulfilling the properties of normality, convexity and continuity. These hypotheses are quite general and are fulfilled by the most used MF shapes such as the triangular, trapezoidal or gaussian one.

The MF shape we have chosen is the trapezoidal one because it preserves the intervals specified by the clinical rules and provides for the imprecision of medical concepts in a intelligible way.

According to the chosen shape, a MF can be encoded by a set of real parameters. For instance, a trapezoidal MF related to a fuzzy set $A^x_j$, $j = 1, \ldots, J_u$ with $J_u$ the number of fuzzy sets composing the FoC associated to the variable $x_u$, is represented as an ordered set of four real values $(a_{u1}^j, a_{u2}^j, a_{u3}^j, a_{u4}^j)$, where the first represents the starting value of the leading edge, the second its ending value, the third the starting value of the trailing edge and the fourth its final value. Hence, the trapezoid will be determined by the four points $(a_{u1}^j, 0), (a_{u2}^j, 1), (a_{u3}^j, 1), (a_{u4}^j, 0)$.

Obviously, the value of each parameter $a_{uli}^j$ ranges between the minimum $x_{u\text{min}}$ and the maximum $x_{u\text{max}}$ value of the related variable $x_u$. Furthermore, to preserve the semantic meaning of the encoded fuzzy partition, we assume a boundary constraint: for $j = 1$ and $j = J_u$

$$\begin{align*}
    a_{u1}^j &= x_{u\text{min}}^j \\
    a_{u4}^j &= x_{u\text{max}}^j
\end{align*}$$

Practically, Equation (5) entails that the MF trapezoidal shape of both the most left and most right fuzzy sets is scalene with value 1 assigned at the boundary points of UoD.

Therefore, because of what we said above, a fuzzy partition defined over $n$ variable $x_u$ involved in a FRB can be encoded as a vector $a \in \mathbb{R}^D$ with length equal to

$$D = 4 \cdot \sum_{j=1}^{J_u} J_u$$

and satisfying the boundary constraints expressed in Equation (5).

Finally, we are able to formulate MF tuning with accuracy-interpretability balance, an intrinsic multi-objective task, in terms of a single-objective multi-constrained minimization problem solved by a vector $a \in \mathbb{R}^D$ representing a fuzzy partition such that:

i) minimizes $f(a)$

ii) subject to 0.5 $\leq \Delta(\{a_{uli}^j\}) < 1$, $u = 1, \ldots, n$

iii) with $\forall u$

$$\begin{align*}
    x_{u\text{min}}^j &< a_{uli}^j < x_{u\text{max}}^j, \quad j \neq 1, J_u \\
    a_{u1}^j &= a_{u2}^j = x_{u\text{min}}^j, \quad j = 1 \\
    a_{u3}^j &= a_{u4}^j = x_{u\text{max}}^j, \quad j = J_u
\end{align*}$$

where $\{a_{uli}^j\}$ indicates the components of a encoding $\Phi_{x_u}$ and the function objective $f$ is defined as

$$f(a) = \sqrt{(1 - S(a, \{y_k\}))^2 + (1 - \Delta(a))^2}$$
with \( S(\mathbf{a}, \{y_k\}) \in [0, 1] \) being the relative classification accuracy evaluated on the set of examples \( y_k \) with cardinality \( K \) as follows

\[
S(\mathbf{a}, \{y_k\}) = \frac{1}{K} \sum_k \text{diag}_{w_k}(\mathbf{a}, y_k) \tag{8}
\]

with \( \text{diag}_{w_k}(\mathbf{a}, y_k) \) equal 1 if the FRB output coincides with the output of \( y_k \) and equal to 0 otherwise, while \( \Delta(\mathbf{a}) \) is the distinguishability measure associated to the whole fuzzy partition:

\[
\Delta(\mathbf{a}) = \frac{1}{n} \sum_{u=1}^{n} \Delta(\{a_{jk}^u\}_u) \tag{9}
\]

with \( \Delta(\{a_{jk}^u\}_u) \) computed through Equation (3) and ranging in \([0, 1]\).

Equation (7), as it can be immediately noticed, is the euclidean distance between the point \((S(\mathbf{a}, \mathbf{y}), \Delta(\mathbf{a}))\) and the point \((1, 1)\) which represents the pair of ideal values of the relative classification accuracy and the distinguishability measure achievable in the best case.

**B. The DE Algorithm With Constraint Handling**

We adopt an evolutionary procedure, whose pseudocode is reported in Figure 1, based on Differential Evolution (DE) [14], a stochastic evolutionary optimization algorithm, to find a vector \( \mathbf{a} \) representing a set of MFs involved in the FRB able to solve, as best as possible, the problem defined in Equation (6).

Initially DE starts with a set, called population, of \( NP \) randomly generated solution vectors, and successively, at each generation until a maximum number \( G \) of generations is reached, creates new individuals by combining other individuals randomly chosen from the current population and mutated according to predefined schemes [14].

The point is that these schemes generate also “infeasible” individuals, i.e., individuals not fulfilling the items ii) and iii) of the Equation (6). These individuals represent solutions that do not satisfy the interpretability assumptions we have imposed and, therefore, cannot be considered as “suitable”.

Among a wide variety of approaches proposed for handling infeasible individuals [25], we have employed that one introduced by Lampinen [26] in which the commonly used greedy selection operator is substituted with the following one:

1) The DE/best/1/bin scheme is adopted: for each individual \( \mathbf{a} \), a trial vector \( \mathbf{v} \equiv (v_1, \ldots, v_D) \) is generated by summing the individual representing the best solution in the current population to a difference vector made by two individuals randomly chosen, both different from the best and modulated by amplitude factor usually denoted as \( F \in [0, 1] \). Successively, \( \mathbf{v} \) is combined with \( \mathbf{a} \) according to a crossover probability \( CR \in [0, 1] \);

2) At the next generation \( \mathbf{v} \) replaces \( \mathbf{a} \) if and only if:
   - both vectors are feasible — namely, they satisfy all constraints in Equation (6) — and \( f(\mathbf{v}) \leq f(\mathbf{a}) \), or
   - the distinguishability measure associated to the whole fuzzy partition:

\[
\Delta(\mathbf{a}) = \frac{1}{n} \sum_{u=1}^{n} \Delta(\{a_{jk}^u\}_u) \tag{9}
\]

with \( \Delta(\{a_{jk}^u\}_u) \) computed through Equation (3) and ranging in \([0, 1]\).

Equation (7), as it can be immediately noticed, is the euclidean distance between the point \((S(\mathbf{a}, \mathbf{y}), \Delta(\mathbf{a}))\) and the point \((1, 1)\) which represents the pair of ideal values of the relative classification accuracy and the distinguishability measure achievable in the best case.

**Possibilistic constrained DE optimization algorithm**

**Initialize** the population \( P_{t=0} = \{\mathbf{a}_1, \ldots, \mathbf{a}_{NP}\} \)

**Evaluate** \( P_{t=0} \) by means of Equation (7)

Check “feasibility” of each \( \mathbf{a}_i \) through (6.ii) and (6.iii)

- \( \mathbf{a}_{best,g} \leftarrow \) best feasible individual

while \( g < G \) do

    \( g \leftarrow g + 1 \)

    for \( i = 1 \rightarrow NP \) do

        Randomly select \( k, r_1, r_2 \in (1, \ldots, NP) \), with \( k \neq r_1 \neq r_2 \)

        for \( k \rightarrow D \) do

            if \((\text{rand}_k[0,1] \leq CR) \lor (k = i)\) then

                \( v_{k,i,g} \leftarrow a_{k,best,g-1} + F \cdot (a_{k,r_1,g-1} - a_{k,r_2,g-1}) \)

            else

                \( v_{k,i,g} \leftarrow a_{k,i,g-1} \)

            end if

        end for

        Evaluate \( v_{i,g} \) and check its feasibility

        if \((v_{i,g} \text{ and } a_{i,g-1} \text{ are feasible}) \land f(v_{i,g}) \leq f(a_{i,g-1}) \lor (v_{i,g} \text{ is feasible and } a_{i,g-1} \text{ is infeasible}) \lor (v_{i,g} \text{ and } a_{i,g-1} \text{ are infeasible} \land \Delta(v_{i,g}) \leq \Delta(a_{i,g-1}))\) then

            \( a_{i,g} \leftarrow v_{i,g} \)

        else

            \( a_{i,g} \leftarrow a_{i,g-1} \)

        end if

    end for

end while

Figure 1. Pseudocode of the DE algorithm with constraint handling

- \( \mathbf{v} \) is feasible while \( \mathbf{a} \) is infeasible, or
- both vectors are infeasible, but \( \mathbf{v} \) provides a greater or equal value for the distinguishability measures used in the item ii) of Equation (6).

Thus, according to the presented selection operator: if two compared solutions are feasible, the one with lower objective function defined in Equation (7) is better; a feasible solution is always better than an infeasible one; if two compared solutions are infeasible, the less infeasible one is better.

By means of this selection operator, the DE algorithm exerts a selective pressure towards search space regions providing feasible solutions with higher objective values.

**IV. EXPERIMENTAL EVALUATION ON THE VERTEBRAL COLUMN DATASET**

The effectiveness of the proposed approach has been assessed on The Vertebral Column Data set (VCD) [27], a freely downloadable [28] recent biomedical database collected by the Group of Applied Research in Orthopaedics (GARO) of Lyon (France) for the diagnosis of several
vertebral column pathologies such as disk hernia and spondylolisthesis.

VCD contains 310 items, each one corresponding to a single patient represented by six biomechanical features (degree spondylolisthesis ($DS$), pelvic radius ($PR$), sacral slope ($SS$), lumbar lordosis angle ($LL$), pelvic tilt ($PT$) and pelvic incidence ($PI$)) derived from the shape and the orientation of the lumbar spine. To each item, one class label between “Normal” and “Abnormal” is assigned referring to the health status of the patient.

Starting from VCD we have built a FRB according to the following steps: 1) a set of crisp AND-connected rules with clauses expressing relationships between database attributes and values, has been automatically extracted through the J48 algorithm, an implementation of the C4.5 decision tree classification algorithm [29] contained in the Waikato Environment for Knowledge Analysis (WEKA) system [30]; 2) for each variable a number of linguistic terms has been identified by using the distinct variable ranges used into the extracted crisp clauses and by adopting specific medical knowledge; afterwards, the rules have been transformed into the fuzzy CNF form described in Section III.

Transforming crisp rules into fuzzy ones introduces some unavoidable imprecisions in the determination of the fuzzy sets corresponding to medical concepts. Thus, one needs a fine-tuning process to suitably define the fuzzy sets associated to each implicated variable, in order to have a reliable CDSS by maintaining, at the same time, its interpretability. Because of these reasons, we have used the algorithm illustrated in Section III to tune the MF of the fuzzy sets used in the FRB.

A. Validation Approach and Parameter Setting

As in other similar approaches, we firstly perform training and then we test the generalization ability achieved.

More precisely, for the learning mechanism, a $T$-fold cross-validation has been carried out. This means that the specific database is divided into $T$ subsets, and $T$ training sessions are effected: in each session a subset is kept for testing and the algorithm runs on the items in the remaining folds. The best solution found in each training session is then evaluated on the testing set.

Finally, the average results of these $T$ instances are computed, and the best among them in terms of highest generalization ability is seen as the best solution of the tuning process and taken into account to represent the MFs of the fuzzy sets involved in the FRB.

For the experiments carried out for the VCD case of study we have chosen to set $T = 10$.

Moreover, through a preliminary series of experiments, the parameters of the DE algorithm have been set as follows: population size $NP = 30$, number of generations $G = 100$, $C/R = 0.3$ and $F = 0.8$.

B. Results for the VCD Case Study

In Table I we report the crisp AND-connected rules extracted by the J48 algorithm. Subsequently, a number of linguistic terms has been identified for each one of the six attributes involved in the database in accordance with the medical knowledge and the ranges of the variable values expressed in the extracted crisp rules. The identified linguistic terms are “low”, “medium” and “high” for the variables $DS$, $PR$ and $PT$, and “low” and “high” for the variables $SS$, $LL$ and $PI$. Finally, the rules are arranged in CNF, as shown in Table II.

The average results achieved over the $T$ folds, and those for the best fold, in terms of the greatest value of function defined in Equation (7) on the test set $Te_v$, are presented in Table III. Results are reported in terms of objective function on both the training set and the test set, respectively $f_{Tr}$ and $f_{Te}$, attained from Equation (7), accuracy on both the training set and the test set, respectively $S_{Tr}$ and $S_{Te}$ calculated by means of Equation (8), and distinguishability measure $\Delta$ of the entire FRB computed using Equation (9).

The results shown in Table III gain greater significance if they are compared with results achieved by means of J48 classification algorithm which has been used as starting point of our approach.

For J48, an accuracy value equal to 0.816 has been obtained, with 253 out of 310 VCD instances correctly classified. This value is less than both the $S_{Tr}$ and $S_{Te}$ accuracy values achieved both on average and over the best fold.

Furthermore, the relative classification accuracy value averaged over the training sets is slightly greater than that one achieved over the test sets. That seems to point out that the DE algorithm does not overfit the example data.

Finally, it is worth mentioning that the best solution, i.e.,
I, the variable results, comprehensibility of the medical rules has been measure.

The solution with the minimum objective function attained obtained for the best fold case and reported in Table IV.

Conversely, the same variable, in the “fuzzified” version values extending over overlapping ranges; this is a typical UNING ALGORITHM RESULTS FOR

Table II
FRB OBTAINED FROM VCD BY MEANS OF THE ILLUSTRATED FUZZIFICATION PROCEDURE.

<table>
<thead>
<tr>
<th>Rule</th>
<th>Membership Functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>IF (DS is low OR medium) AND PR is low AND SS is low AND PT is low THEN Abnormal</td>
<td></td>
</tr>
<tr>
<td>IF (DS is low OR medium) AND PR is medium AND SS is low AND PT is low THEN Normal</td>
<td></td>
</tr>
<tr>
<td>IF (DS is low OR medium) AND (PR is low OR medium) AND SS is low AND (PT is medium OR high) THEN Abnormal</td>
<td></td>
</tr>
<tr>
<td>IF DS is low AND (PR is low OR medium) AND SS is high AND (PT is low OR medium) THEN Normal</td>
<td></td>
</tr>
<tr>
<td>IF DS is low AND (PR is low OR medium) AND SS is high AND PT is high AND LL is low AND PI is high THEN Normal</td>
<td></td>
</tr>
<tr>
<td>IF DS is low AND (PR is low OR medium) AND SS is high AND PT is high AND LL is high THEN Abnormal</td>
<td></td>
</tr>
<tr>
<td>IF DS is medium AND (PR is low OR medium) AND SS is high THEN Abnormal</td>
<td></td>
</tr>
<tr>
<td>IF (DS is low OR medium) AND (PR is high AND SS) THEN High</td>
<td></td>
</tr>
<tr>
<td>IF DS is high THEN Abnormal</td>
<td></td>
</tr>
</tbody>
</table>

Table III
TUNING ALGORITHM RESULTS FOR VERTEBRAL COLUMN DATA SET.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Best fold results</th>
<th>Average results</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f_{TV}$</td>
<td>0.219</td>
<td>0.229</td>
</tr>
<tr>
<td>$f_{TV}$</td>
<td>0.197</td>
<td>0.262</td>
</tr>
<tr>
<td>$S_{TV}$</td>
<td>0.838</td>
<td>0.825</td>
</tr>
<tr>
<td>$S_{TV}$</td>
<td>0.8709</td>
<td>0.821</td>
</tr>
<tr>
<td>$\Delta$</td>
<td>0.851</td>
<td>0.854</td>
</tr>
</tbody>
</table>

the solution with the minimum objective function attained over one of $T$ test sets, has a relative classification accuracy value much greater than those ones achieved in the other cases, although it preserves a high distinguishability measure.

For what concerns the linguistic interpretability of the results, comprehensibility of the medical rules has been improved. As instance, in the first rule represented in Table I, the variable PR is involved in two different clauses with values extending over overlapping ranges; this is a typical example of semantic redundancy in a rule base system. Conversely, the same variable, in the “fuzzified” version presented in Table II, is involved in just one conjunctive clause where the linguistic terms, low and medium, are used in place of two range values.

Additionally, the fuzzy partitioning obtained through our MF constrained-tuning algorithm permits to minimize fuzzy set overlapping; consequently it is easy to assign to each term a precise semantic sense, as it is possible to appreciate by observing Figure 2 and the MF parameters obtained for the best fold case and reported in Table IV.

V. CONCLUSION

Clinical rules can be modeled through fuzzy rule-base systems in which medical concepts are represented by linguistic variables and terms. Each FRB is even more useful as accurate in classification. Hence, it is necessary

Table IV
MF PARAMETERS CARRIED OUT FOR THE BEST FOLD CASE.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Membership Functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>DS:</td>
<td>low={−11.05, 1}, {−11.05, 1}, {−9.29, 1}, {16.78, 0}</td>
</tr>
<tr>
<td>PR:</td>
<td>medium={14.93, 0}, {163.87, 1}, {186.38, 1}, {285.84, 0}</td>
</tr>
<tr>
<td>SS:</td>
<td>high={276.71, 0}, {395.77, 1}</td>
</tr>
<tr>
<td>LL:</td>
<td>{418.54, 1}, {418.54, 1}</td>
</tr>
<tr>
<td>PT:</td>
<td>low={49.12, 1}, {129.74, 1}</td>
</tr>
<tr>
<td>PT:</td>
<td>medium={285.04, 0}, {394.43, 1}, {494.43, 1}, {494.43, 1}</td>
</tr>
<tr>
<td>PT:</td>
<td>high={160.75, 0}, {91.84, 1}, {129.83, 1}</td>
</tr>
</tbody>
</table>

Figure 2. Terms obtained from the best fold case for the variables (from top to down): DS, SS, PR, LL, PT, and PT.
to model FRB linguistic terms in order to obtain a model with a high accuracy and by preserving, at the same time, their linguistic interpretability.

In this work, this problem has been dealt with as it was optimization problem constrained by some conditions assumed for fuzzy sets. At this aim, we have imposed that any potential partition should satisfy some conditions derived from distinguishability property for linguistic terms. In particular we have used a modified DE algorithm for MF tuning with a novel evolutionary scheme designed at this scope, and we have defined a measure of distinguishability for FRBs based on Possibility Theory.

The approach has been quantitatively assessed on the VCD set, recently introduced in literature. The results, evaluated through cross-over validation method, are improved in regard to those obtained from J48, an implementation of the C4.5 decision tree classification algorithm in WEKA, in terms both of accuracy and linguistic interpretability.

In the future we plan to investigate dependence of the performance of our approach on the number of variable and fuzzy sets involved.

REFERENCES


