# Classification of Potential Multiple Sclerosis Lesions Through Automatic Knowledge Extraction by Means of Differential Evolution

Ivanoe De Falco(✉)

ICAR-CNR, Via P. Castellino 111, 80131 Naples, Italy
ivanoe.defalco@na.icar.cnr.it

**Abstract.** In this paper a classifier, designed by taking into account the user–friendliness issue, is described and is used to tackle the problem of classification of potential lesions in Multiple Sclerosis. This tool is based on the idea of making use of Differential Evolution (DE) to extract explicit knowledge from a database under the form of a set of IF–THEN rules, can use this set of rules to carry out the classification task, and can also provide clinicians with this knowledge, thus explaining the motivation for each of the proposed diagnoses. Each DE individual codes for a set of rules. The tool is compared over a database of Multiple Sclerosis potential lesions against a set of nine classification tools widely used in literature. Furthermore, the usefulness and the meaningfulness of the extracted knowledge have been assessed by comparing it against that provided by Multiple Sclerosis experts. No great differences have turned out to exist between these two forms of knowledge.

**Keywords:** Pattern Recognition · Classification · Differential Evolution · Automatic Rule Extraction · Multiple Sclerosis Diagnosis

## 1 Introduction

Multiple Sclerosis (MS) is an autoimmune disease characterized by the fact that the immune system acts harmfully on the Central Nervous System [2], causing nerve demyelination. Normally the larger part of MS lesions are small, yet they can sometimes have a diameter of some centimeters. The only way to check the development of this disease consists in clinical examination substantiated by laboratory investigations. Within these latter, magnetic resonance imaging (MRI) is very commonly used to visualize lesions [9].

MRI is currently considered as the most reliable paraclinical test with reference to the issues of diagnosis of the MS disease, evaluation of its evolution, and medical care of its effects. As a matter of fact, the use of MR images as a marker for MS requires the advice of experts and the exploitation of all their knowledge to correctly identify MS lesions.

In general, the process of finding out actual lesions for Multiple Sclerosis can be seen as a pipelining procedure composed by three tasks: the *segmentation* of


A.I. Esparcia-Alcázar et al. (Eds.): EvoApplications 2014, LNCS 8602, pp. 538–549, 2014.
DOI: 10.1007/978-3-662-45523-4_44

the MRI images into groups of homogeneous pixels/voxels representing tissues, the *labeling* of those tissues, and finally the actual *classification* step, meaning with this the assignment of each potential lesion detected to one of the possible classes, i.e. either actual lesion or non–lesion.

Yet, this process is very laborious because of the high number of MR images that must be examined and of the variability in the number of MS lesions per image, as well as in their size and spatial distribution. Furthermore, the result of the analysis of an MRI image is a set of potential lesions, some of which are actually lesions whereas others are not. Therefore, it is very important to correctly distinguish among them. This is a typical classification task, that has been up to now carried out prominently by human experts only.

In recent years Clinical Decision Support Systems (CDSSs) are becoming more and more popular in the medical domain, aiming at supporting clinicians in their whole clinical process from diagnosis and investigation to treatment and long-term care. CDSSs have been defined as 'active knowledge systems which use two or more items of patient data to generate case-specific advice' [14].

Among the many tasks that should be dealt with by clinicians, classification [7] is one of the most important and delicate, and is closely related to diagnosis. To point out its significance suffice it to say that a wrong classification leads either to false positive cases, so causing unnecessary worries and medical cares, or, even worse, to false negative diagnoses, which may cause serious illnesses to patients, and even their premature death. So it is not surprising that many classification tools have recently started to be used in the medical domain.

A desirable feature for a classifier is that it should be user–friendly as concerns both its use and the output information it can provide. Of course, this feature becomes even more important when the medical diagnostic process is considered: even if a method can correctly assign patients to diagnoses, it should not be a kind of a black–box or an oracle. Rather, it should provide clinicians with useful information on the reasons why any patient is categorized in the given way.

In this paper a tool, designed by carefully taking the above user–friendliness issue into account, is described and is used to tackle the problem of classification in Multiple Sclerosis. This system is based on Differential Evolution (DE) [11].

It is important to say here that we wish to make reference to the third above mentioned step only, i.e. the classification of potential lesions.

Starting from DE, a classifying tool is designed that can extract explicit knowledge from a database under the form of a set of IF–THEN rules, can use this set of rules to carry out classification, can output the class assigned to each instance, and can also provide clinicians with this knowledge, thus explaining the motivation for each of the proposed diagnoses. Of course, this extracted knowledge should never be seen as a substitute of doctor's experience, rather both as a confirmation of his/her knowledge and as a set of possible suggestions to complement doctor's knowledge, to be clinically validated.

The tool described here is based on *DE* and performs *R*ule *Ex*traction, so it is referenced within this paper as *DERЕx*. The originality of the approach presented here lies in the fact that up to now DE has been used in classification tasks in combination with other tools as neural networks e.g. [10], bayes–based methods [13], fuzzy logic tools [1], nearest neighbor [12], and so on, but just seldom has it been applied on its own [8]. More importantly, in all cases in which DE has been used on its own, it has never been assigned the task of extracting by itself classification rules from databases, as it can be noted in [3], where a wide list of applications of DE is reported. Rather, DE is used for other tasks, such as optimizing parameters, optimizing membership functions, etc.

DEREx is used here for multiple sclerosis because recent experiments over other medical databases [4] showed its superiority over other classifiers.

A paramount issue when automatically extracting knowledge from databases is the investigation about whether or not the set of rules found is useful and meaningful for the clinicians. In fact, the extracted knowledge could allow achieving very good classification accuracy, yet it could be very far from the one a doctor would ever use, and it could even be just a kind of a tricky combination of values for the database attributes, without any actual medical meaning. This issue should always be addressed when using one of these rule–extracting tools.

In Section 2 our rule extractor DEREx is shortly described. Section 3 reports on the experimental results on a real Multiple Sclerosis database. The resulting explicit rules are given in Section 4. In Section 5 the extracted knowledge is compared with that provided by Multiple Sclerosis experts. Finally, in Section 6 conclusions are given and future works are outlined.

## 2 The Rule Extractor: DEREx

To face this problem of supervised classification, we have taken advantage of our DEREx tool [4] to carry out the automatic extraction of a set of explicit IF-THEN rules from the database. This tool relies on Differential Evolution.

Due to lack of space, describing DEREx with sufficient details is here impossible, and reference to [4] should be made. Just to give some necessary information, each solution in the DE population codes for a set of IF–THEN rules, each of which contains AND–connected literals on the database variables. For each class more than one rule can be contained in the individual solution, and these rules can be seen as logically connected in OR. A very important parameter is the maximum number of rules that can be contained in a set, denoted as $NR$.

During evolution, the fitness of each individual is the percentage of the cases in the training set that are correctly classified by using the set of rules encoded in that individual. Indeterminate items, i.e. the items assignable to either no class or to more than one through the set of rules encoded by the individual being evaluated, are treated as incorrectly classified during training, whereas they are assigned to exactly one class during testing by a recovery mechanism [4].

In the top pane of Fig. 1 the uppermost part says that each DE individual is in this case a vector, containing real values, representing a set of $NR$ classification rules written in sequence in the individual.

The middle part of the top pane shows that each rule is represented by a set of fields. Namely, each rule consists in a *Rule_Active* field, followed by a number of $NV$ *Literal_Representation* groups (where $NV$ is the number of variables in the database), and finally by a *Class* field. In the rule, database variables are listed sequentially, meaning that the generic $i$-th *Literal_Representation* deals with the $i$-th variable of the database. *Rule_Active* tells whether or not the rule should be considered during classification. This is decided by comparing the real value contained in this field against a real value Rule Threshold ($RT$), which is a parameter for our tool: if the former value is higher, then this rule is seen as active in the current individual and should be used in the classification process.

Each *Literal_Representation* field encodes a zero-th order literal, i.e. a literal in which only one variable is contained and is compared with one or two real values by means of relation operators. As shown in the bottom part of the top pane, on its turn this field consists of four fields, each containing a real value, as it is detailed in the following paragraphs.

The first field is the *Literal_Active* field. Similarly to the *Rule_Active* field, it determines whether or not the literal is present in the rule. Also here, a real-valued parameter Literal Threshold ($LT$) is defined, and the generic literal under account is active if and only if the value in this field is higher than $LT$.

The second field is called *Literal_Type*. It encodes the relation operator that compares the variable and the constant value(s). We have decided to take the following seven different operators into account: $<, \leq, =, \geq, >, IN, OUT$. The first five operators need one constant value, i.e. C1, whereas the latter two need two constant values C1 and C2. The operator $IN$ checks if the value of the variable contained in the literal is within the numerical range expressed by C1 and C2 in their order of appearance in the individual. The operator $OUT$, instead, checks if the value taken on by the variable in the literal is outside the range [C1 - C2], meaning that either it is lower than C1 or it is greater than C2.

The third and the fourth fields of the *Literal_Representation* field hold, respectively, the real values for the constants C1 and C2. C1 will be used for each active literal, while C2 only if the *Literal_Type* field contains IN or OUT.

Finally, the *Class* field contains the value representing the class to which all the database instances that satisfy the considered rule are assigned.

The bottom pane of Fig. 1, instead, shows an example for a database with two variables $var_1$ and $var_2$, in which we have set $NR = 3$, $RT = 0.50$, and $LT = 0.50$. Starting from the values contained in each field, it can be realized that rules 1 and 3 are active, whereas rule 2 is not. In rule 1, both literals are active, whereas in rule 3 just the literal about $var_2$ is. Therefore, the DE individual in the figure encodes the following set of rules:

$$\textit{IF } (var_1 \leq 6.14) \textit{ AND } (var_2 \geq 3.41) \textit{ THEN class=2}$$

$$\textit{IF } (var_2 \leq 3.12) \textit{ THEN class=1}$$

## 3 Experiments

DEREX has been evaluated on a dataset, opportunely anonymized, collected at the Department of Bio-Morphological and Functional Sciences of the University

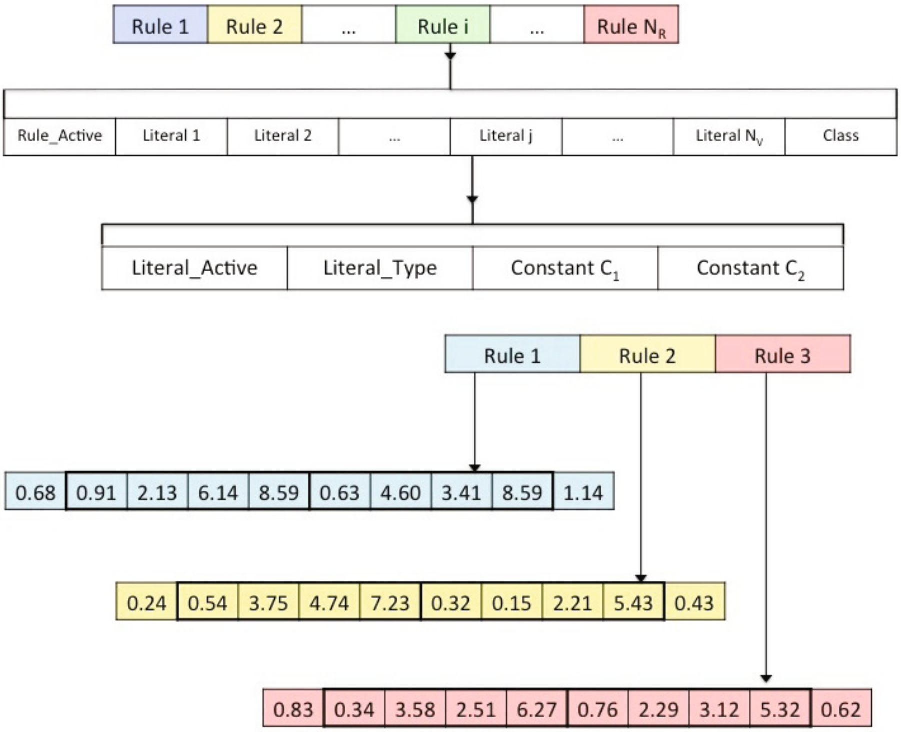


**Fig. 1.** Top: general structure of individuals in DEREx. Bottom: an example of an individual and its decoding into a set of IF–THEN rules.

**Table 1.** The sclerosis database

| variable | unit range | description |
|---|---|---|
| surrounding white matter | 0.32 - 1.00 | Amount of White Matter enclosing a lesion |
| compactness | 0.31 - 1.98 | Degree of compactness of a lesion. For a given shape, compactness is high either if the volume is large or if the enclosing surface is small, i.e. the object is strongly compact |
| tissue contrast | 0.56 - 1.00 | Minimum color contrast to detect a WML in the multiparametric space |
| volume | 3 - 10,522 | Lesion volume in terms of the number of voxels |
| sphericity | 0.01 - 1.23 | Degree of sphericity of a lesion. The more elongated the lesion is and the more it deviates from a sphere, the lower sphericity will be |

of Naples Federico II. In particular, starting from MR brain images of 120 patients with clinically definite MS, a multiparametric segmentation procedure has been preliminarily applied to the whole data set in order to identify normal brain tissues or clusters of potentially abnormal white matter voxels, labeled as White Matter Potential Lesions (WMPLs). For each WMPL, the features described in Table 1 represent the actual input data for the DSS.

The resulting database contains 2844 items, 1905 of which represent actual lesions (class 2) and 939 showing no actual lesions (class 1).

As concerns DEREx, a *DE/rand–to–best/1/bin* mutation strategy has been used, and the values for the parameters have been set as follows: population size $N_{Pop} = 30$, number of generations $Gen = 500$, crossover ratio $CR = 0.5$, scale factor $F = 0.5$, $NR = 10$, $RT = 0.50$, and $LT = 0.50$. No preliminary tuning phase has been specifically effected over this sclerosis problem for the choice of these values, rather the same values as in [4] have been used.

In each run 10–fold cross-validation has been carried out, so that, for the generic $i$-th fold, the $i$-th 10% of the data, in their order of appearance in the database, is kept for testing, and training takes place on the remaining 90%. For the generic $i$-th fold the result is represented by the classification accuracy over the related testing set $\%C^i_{Te}$. Then, the result for the whole run is the average, over the 10 folds, of the 10 $\%C^i_{Te}$ values achieved, i.e. $Av_C = \langle \%C^i_{Te} \rangle$.

A total number of 25 runs, each of them being a 10–fold cross-validation, has been effected. The results have been collected in terms of average percentage of correct classification over the testing set over the 25 runs $Av_{CC} = \langle Av_C \rangle$.

As concerns the other classification techniques used in the comparison on this MS problem, reference has been made to the Waikato Environment for Knowledge Analysis (WEKA) system release 3.4 [6] that contains a large number of such techniques, divided into groups (Bayesian, based on functions, lazy, meta-techniques, tree-based, rule-based, other). From each such group at least one representative has been chosen. Due to lack of space, their names are shown in Tab. 2. Three rule–based classifiers have been considered, i.e. OneR, Part, and Ridor, since we are of course more interested in this kind of tools.

Similarly to what was done for DEREx, no preliminary parameter tuning has been carried out for all of the above techniques as well, so the parameter values used for each such method are those set as default in WEKA.

Furthermore, since also for these classification techniques results must be provided in terms of average results over 25 runs, for each of them either the starting seeds or some parameter values have been varied. Actually, RBF, AdaBoost, Part, and Ridor are based on a random starting seed so that the 25 runs for them have been carried out by varying this value. Some other techniques, instead, do not depend on any starting seed, so the 25 runs have been carried out as a function of a parameter typical of the technique: alpha for Bayes Net, globalBlend for KStar, bias for VFI, and minBucketSize for OneR. Finally, NBTree depends neither on an initial seed nor on any parameter, so only one run has been performed for it on the database.

Also for all of these tools 10–fold cross–validation has been carried out.

Table 2 shows the results, achieved by each technique on the database, expressed in terms of $Av_{CC}$. Namely, for each tool the average accuracy $Av_{CC}$ over the 25 runs, the highest value of $Av_C$ (*best_acc*), and the lowest one ($worst\_acc$) are shown. Also the standard deviation *std_dev* is reported for the techniques for which multiple runs are carried out. Finally, the last row of the table reports the rank for each tool based on $Av_{CC}$.

**Table 2.** The 10-fold classification accuracy for all the classifiers

| | DEREx | Bayes Net | RBF | Kstar | AdaBoost |
|---|---|---|---|---|---|
| $Av_{CC}$ | 87.49 | 78.07 | 82.20 | 85.60 | 82.35 |
| *best_acc* | 88.54 | 78.09 | 82.24 | 85.90 | 82.35 |
| *worst_acc* | 86.90 | 78.02 | 82.14 | 85.27 | 82.35 |
| *std_dev* | 0.46 | 0.03 | 0.04 | 0.21 | 0.00 |
| *rank* | 1 | 8 | 7 | 5 | 6 |

| | NBTree | OneR | Part | Ridor | VFI |
|---|---|---|---|---|---|
| $Av_{CC}$ | 86.81 | 72.03 | 87.20 | 86.14 | 72.40 |
| *best_acc* | — | 73.91 | 87.20 | 87.03 | 72.40 |
| *worst_acc* | — | 65.30 | 87.20 | 85.44 | 72.40 |
| *std_dev* | — | 2.89 | 0.00 | 0.59 | 0.00 |
| *rank* | 3 | 10 | 2 | 4 | 9 |

DEREx turns out to be the best tool, the runner-up being Part. It is worth noting that DEREx performs better than all the other rule–based tools.

## 4 The Advantage of DEREx: The IF-THEN Rules

The clear advantage of DEREx consists in the fact that it provides users with explicit knowledge automatically extracted from the database under the form of IF-THEN rules. In fact, it can straightforwardly express rules to perform diagnosis. Furthermore, DEREx can also perform feature extraction, since the achieved rules may contain some of the database attributes only, which can be seen as an extremely useful support for a correct diagnosis. This has turned out to be true in [4] where seven medical databases have been faced. In this way, physicians are helped with useful information. Of course, their opinion about the correctness and the usefulness of these rules is of paramount importance for medical practice. In the following the best set of rules found for the Multiple Sclerosis problem is reported. Namely, they are those with the highest percentage of correct classification on the testing set achieved on a fold in all the executions.

*IF (surrounding_white_matter* $< 0.81$*) AND (compactness IN (1.56 - 1.88)) AND (volume IN (7,224 - 10,395)) THEN lesion*

*IF (surrounding_white_matter* $> 0.40$*) AND (compactness* $\geq 0.36$*) AND (tissue_contrast* $> 0.63$*) THEN lesion*

*IF (compactness IN (0.60 - 1.62)) AND (tissue_contrast* $> 0.65$*) AND (volume* $> 8,498$*) THEN no_lesion*

*IF (surrounding_white_matter* $\geq 0.68$*) AND (compactness OUT (0.88 - 1.85)) AND (volume* $\leq 4,230$*) AND (sphericity* $\geq 0.02$*) THEN lesion*

*IF (surrounding_white_matter IN (0.41 - 0.57)) AND (tissue_contrast* $\leq 0.58$*) AND (volume* $> 1,817$*) AND (sphericity IN (0.15 - 0.19)) THEN lesion*

*IF (compactness OUT (1.39 - 1.47)) THEN no_lesion*

As it can be seen, the best individual has six active rules, four of which classify for *lesion* and two for *no_lesion*. Thus, even if the OR connector is not

**Table 3.** Statistics of the best set of rules

| | Correct Classification Rate | Sensitivity | Specificity |
|---|---|---|---|
| Training Set | 89.11% | 89.83% | 87.78% |
| Testing Set | 93.43% | 90.52% | 98.80% |

**Table 4.** Linguistic variables and terms

| variable | Terms |
|---|---|
| surrounding white matter | bit, partially, almost completely, completely |
| compactness | weak, strong |
| tissue contrast | little, great |
| volume | small, medium, large |
| sphericity | low, moderate, high |
| tissue structure | normal, abnormal |

explicitly present in DEREx, the set of rules performs implicitly an OR over each class by using as many rules as needed to achieve good classification accuracy.

This set of rules has been obtained for fold 7, and its values of correct classification rate, sensitivity, and specificity over both the training set and the testing set are reported in Tab. 3.

## 5 Comparison Between Extracted Knowledge and Experts' Knowledge

Once we have seen that DEREx is able to extract knowledge that allows classifying with a good accuracy, the question arises whether or not this automatically extracted knowledge is useful, and, even more important, if it is meaningful for experts. To investigate this, we need to compare the knowledge provided by the experts against that provided by DEREx.

The medical knowledge needed to classify WMPLs has been defined in cooperation with a team of physicians, starting from the sclerosis features contained in the faced database, and can be stated, in natural language, as follows. *The tissue composing a WMPL is abnormal if the lesion is somewhat surrounded by WM, characterized by a strong compactness and greatly contrasted in the multiparametric space. The sphericity is moderate or high in small lesions, whereas, as their volume increases, the sphericity starts decreasing progressively. Finally, as volume increases and sphericity starts lessening, a lesion can be surrounded by gradually decreasing WM and its compactness still remains high.*

As it can be seen, the experts' knowledge is based on a fuzzy view, given the use of words such as *somewhat, small, volume increases, starts decreasing, gradually decreasing*, and so on. Furthermore, as it is often the case in the medical domain, this knowledge is based on positive evidence only, i.e. it contains only sentences representing the presence of an actual lesion.

In accordance with this knowledge, the linguistic variables and the fuzzy values shown in Table 4 have been identified.

**Table 5.** IF–THEN form of experts' knowledge

| |
|---|
| 1) IF [Sphericity is (Moderate OR High)] AND [Compactness is Strong]AND [Volume is Small] AND [TissueContrast is Great] AND [SurroundingWhiteMatter is Completely] THEN [TissueStructure is Abnormal] |
| 2) IF [Sphericity is Moderate] AND [Compactness is Strong]AND [Volume is Medium] AND [TissueContrast is Great] AND [SurroundingWhiteMatter is (AlmostCompletely ORCompletely)] THEN [TissueStructure is Abnormal] |
| 3) IF [Compactness is Strong]AND [Volume is Large] AND [TissueContrast is Great] AND [SurroundingWhiteMatter is (Partially OR AlmostCompletely OR Completely)] THEN [TissueStructure is Abnormal] |
| 4) ELSE [TissueStructure is Normal] |

**Table 6.** Values for the shapes of the trapezoids

| Variables | Terms | $\alpha_1$ | $\alpha_2$ | $\alpha_3$ | $\alpha_4$ |
|---|---|---|---|---|---|
| Surrounding | Bit | 0.32 | 0.32 | 0.33 | 0.38 |
| White | Partially | 0.33 | 0.38 | 0.40 | 0.41 |
| Matter | Almost Completely | 0.40 | 0.41 | 0.46 | 0.95 |
| | Completely | 0.46 | 0.95 | 1.00 | 1.00 |
| Compactness | Weak | 0.31 | 0.31 | 0.36 | 0.74 |
| | Strong | 0.36 | 0.74 | 1.98 | 1.98 |
| Tissue | Little | 0.56 | 0.56 | 0.61 | 0.92 |
| | Great | 0.61 | 0.92 | 1.00 | 1.00 |
| Volume | Small | 3 | 3 | 3,177 | 3,529 |
| | Medium | 3,177 | 3,529 | 7,051 | 7,697 |
| | Large | 7,051 | 7,697 | 10,522 | 10,522 |
| Sphericity | Low | 0.01 | 0.01 | 0.03 | 0.10 |
| | Moderate | 0.03 | 0.10 | 1.02 | 1.03 |
| | High | 1.02 | 1.03 | 1.23 | 1.23 |

These linguistic variables and values have been used to write the three if-then rules aimed at identifying the positive cases, i.e. when a potential lesion is an actual one. Table 5 shows those rules and the default ELSE one.

It is to be noted that the knowledge provided by the experts is in a fuzzy form, whereas that extracted by DEREx contains crisp rules. To effectively compare these two sets of rules, that proposed by DEREx should be reformulated into a fuzzy form. So, it is important to transform the crisp values contained into suitable fuzzy values. To this aim, reference can be made to [5], where a DE tool was used to tune the parameters of a fuzzy system working on this database. Each fuzzy value was there represented as a trapezoid, so four real values were needed to identify the shape of each trapezoid. Those values represent the x–values for the bottom–left, top–left, top–right, and bottom–right vertices of the trapezoid, the y–values being 0 for the bottom vertices and 1 for the top ones. Table 6 shows the results of the tuning of the fuzzy values and the correspondence between the crisp values and the fuzzy ones for each database variable.

By making use of those values, the set of crisp rules found by DEREX and shown in the previous subsection can now be rewritten as shown in Tab. 7.

The correspondence between the two kinds of knowledge can be visually understood by looking at Fig. 2. The figure contains both the sets of fuzzy rules for the presence of lesions only, namely the top pane shows the knowledge

**Table 7.** Fuzzy form of DEREX knowledge

1) IF (white matter IS (bit surrounded OR partially surrounded OR almost completely surrounded)) AND (compactness IS strong) AND (volume IS large) THEN lesion

2 ) IF (white matter IS (almost completely surrounded OR completely surrounded)) AND (compactness IS strong) AND (contrast IS great) THEN lesion

3) IF (white matter IS completely) AND (compactness IS weak) AND (volume IS small) AND (sphericity IS (moderate OR high)) THEN lesion

4) IF (white matter IS almost completely) AND (contrast IS little) AND (volume IS (medium OR large)) AND (sphericity IS moderate) THEN lesion

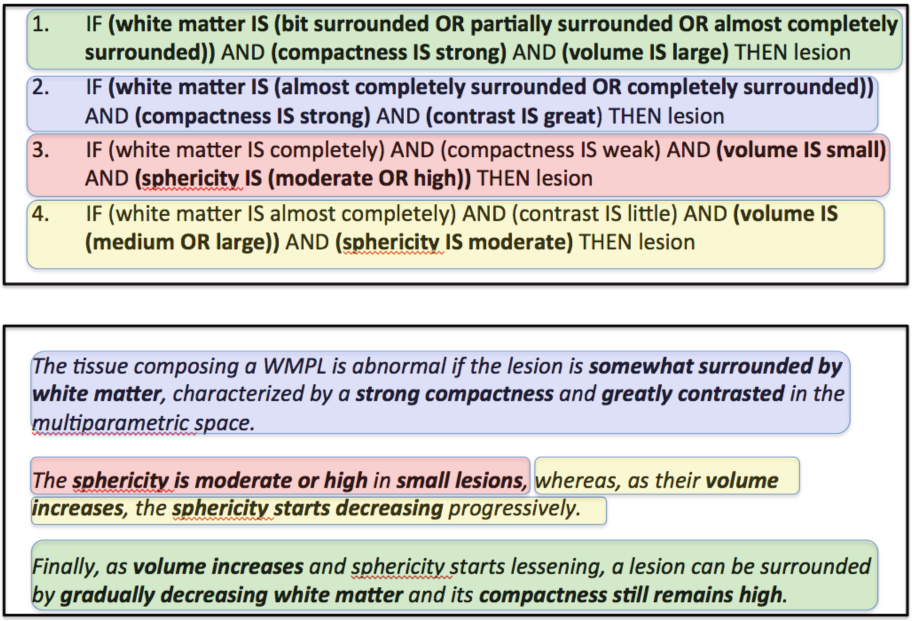


**Fig. 2.** Extracted knowledge (top pane) and experts' knowledge (bottom pane), and their correspondence

extracted, whereas the bottom one reports the experts' knowledge. In the figure similar concepts are represented by a same color in both sets.

As it can be noted, the rule 1 of DEREx almost completely corresponds to the rule 3 as stated by the experts (both are represented in green color). The variables involved are the same, apart from sphericity that is accounted for by experts whereas DEREx does not mention it, and the fuzzy values taken on by the variables are the same in both cases. Rule 2 found by our system corresponds to rule 1 (blue color represents them both). In this case the correspondence is perfect: the variables involved are the same in both rules, and so are the fuzzy values they take on. Rule 3 of DEREx corresponds to the first part of experts' rule 2 (red color represents them), and the two common variables take on exactly the same values in both rules, whereas DEREX uses two more variables. Finally, rule 4 of DEREx corresponds to the second part of rule 2 (they both are represented in yellow): also in this case DEREx uses two more variables.

In summary, from the analysis of the two kinds of knowledge it appears evident that the knowledge automatically extracted by DEREx is extremely similar to that provided by experts, hence proving that the proposed system is capable of extracting knowledge that is useful and meaningful.

The slight differences in the two sets of rules can be seen as further suggestions for the experts. For example, rule 1 of DEREx, very similar to the third rule by the experts, might suggest the doctors the question whether, in that general frame proposed by their rule, a decrease in sphericity is really important. Vice versa, rule 3 of DEREx might suggest doctors to take into account white matter and compactness too, when volume is small and sphericity is moderate or high.

## 6 Conclusions and Future Work

In this paper, an approach based on Differential Evolution for the automatic classification of potential lesions in a Multiple Sclerosis database has been followed. Namely, a tool called DEREx has been used, which automatically extracts explicit knowledge from the database under the form of IF–THEN rules containing AND–connected literals on the database variables.

Firstly, DEREx has been run and the most effective set of rules in terms of highest classification accuracy in a ten–fold cross–validation has been found. Secondly, the tool has been compared over the same database against a set of nine classification tools widely used in literature.

The results have proven the viability and the effectiveness of the proposed approach, since this turns out to provide the highest classification accuracy.

The advantage of DEREx consists in providing users with explicit knowledge automatically extracted from the database, since it can straightforwardly express IF-THEN rules to perform diagnosis, differently from many of its competitors.

Attention has been paid to the usefulness of the extracted knowledge, by comparing it against that provided by Multiple Sclerosis experts. Results have shown that the two different kinds of knowledge are actually quite similar, thus proving the quality of the approach followed, at least for this problem.

Future work will involve investigation about the influence of the parameters $NR$, $RT$, and $LT$ on solution quality.

Another issue is that DEREx is based on the basic version of DE. Yet, more recently, several enhanced DE versions have appeared that aim at softening the main problem DE suffers from, i.e. that of a limited amount of search moves. So, we aim to use some new versions to further improve DEREx performance.

Moreover, closer cooperation with physicians will be set. This will involve receiving other real databases from them, also with reference to different diseases.

Finally, since physicians make often reference to fuzzy concepts as small, high, etc., the tool will be improved to automatically extract fuzzy rules too.

## References


1. Aliev, R.A., Pedrycz, W., Guirimov, B.G., Aliev, R.R., Ilhan, U., Babagil, M., Mammadli, S.: Type-2 fuzzy neural networks with fuzzy clustering and differential evolution optimization. Information Sciences **181**(9), 1591–1608 (2011)
2. Bobholz, J.A., Gremley, S.: Multiple sclerosis and other demyelinating disorders. In: Schoenberg, M.R., Scott, J.G. (eds.) The Little Black Book of Neuropsychology: A Syndrome-Based Approach, pp. 647–662. Springer (2011)
3. Das, S., Suganthan, P.N.: Differential evolution: A survey of the state-of-the-art. IEEE Transactions on Evolutionary Computation **15**(1), 4–31 (2011)
4. De Falco, I.: Differential evolution for automatic rule extraction from medical databases. Applied Soft Computing **13**(2), 1265–1283 (2013)
5. Esposito, M., De Falco, I., De Pietro, G.: An evolutionary-fuzzy dss for assessing health status in multiple sclerosis disease. International Journal of Medical Informatics **80**(12), e245–e254 (2011)
6. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The weka data mining software: An update. SIGKDD Explorations **11**(1), 10–18 (2009)
7. Han, J., Kamber, M.: Data mining: concept and techniques. Morgan Kaufmann (2001)
8. Maulik, U., Saha, I.: Automatic fuzzy clustering using modified differential evolution for image classification. IEEE Transactions on Geoscience and Remote Sensing **48**(9), 3503–3510 (2010)
9. Miller, D., Grossman, R., Reingold, S., McFarland, H.F.: The role of magnetic resonance techniques in understanding and managing multiple sclerosis. Brain **121** 3–24 (1998)
10. Özbakir, L., Baykasoğlu, A., Kulluka, S.: A soft computing-based approach for integrated training and rule extraction from artificial neural networks: Difaconn-miner. Applied Soft Computing **10**(1), 304–317 (2010)
11. Price, K., Storn, R.: Differential evolution: Numerical optimization made easy. Dr. Dobb's Journal, 18–24 (1997)
12. Triguero, I., García, S., Herrera, F.: Differential evolution for optimizing the positioning of prototypes in nearest neighbor classification. Pattern Recognition **44**, 901–916 (2011)
13. Wu, J., Cai, Z.: Attribute weighting via differential evolution algorithm for attribute weighted naive bayes (wnb). Journal of Computational Information Systems **7**(5), 1672–1679 (2011)
14. Wyatt, J.C., Spiegelhalter, D.J.: Field trials of medical decision-aids: potential problems and solutions. In: Proceedings of the Annual Symposium on Computer Application in Medical Care, pp. 3–7 (1991)