Evolving Fuzzy C-Means: An intelligent technique for efficient diagnosis of children mental retardation level from databases with missing values

Alessandro G. Di Nuovo and Vincenzo Catania
Facoltà di Ingegneria
Dipartimento di Ingegneria, Informatica e delle Telecomunicazioni,
Università di Catania
V.le Andrea Doria, 6, 95125 Catania
adinuovo@diit.unict.it, vcatania@diit.unict.it

Santo Di Nuovo
Facoltà di Scienze della Formazione
Università di Catania
Via Ofelia 2, 95124, Catania, Italy
s.dinuovo@unict.it

Serafino Buono
Psychology Operative Unit
IRCSS Oasi di Troina
Via Conte Ruggero 73,
94018 Troina, Enna, Italy

***Abstract*— In psychopathological diagnosis, a correct classification of mental retardation level is needed to choose the better treatment for rehabilitation and to assure a quality of life suitable for the specific patient condition. In order to meet this need we studied a technique that allows to perform automatic diagnoses efficiently and reliably and at the same time is easy for psychotherapists to use. In this paper we present a new approach which integrates two well-known computational intelligence techniques: Fuzzy C-Means and Genetic Algorithms. This integration allows automatic generation of an optimized classifier for efficient recognition of mental retardation level starting from databases with missing values. The results obtained with this simple technique are encouraging: an empirical test on our database of patients showed the efficiency of the simple integrated algorithm, which also gives a great deal of useful information for diagnostic purposes.*

Keywords: Automated diagnosis, Fuzzy C-Means Clustering, Genetic algorithms, Missing data analysis, Feature selection and weighting.

1 Introduction

In psychopathological diagnosis correct classification of a patient's level of mental retardation, especially during childhood and adolescence, is of fundamental importance in order to guarantee appropriate treatment for proper rehabilitation and a quality of life suited to the patient's condition.

The methods currently adopted in this field use various diagnostic tools which require time and expertise if they are to be administered correctly. Simplifying the use of these tools would make identification of the most suitable treatment faster and more efficient. For this reason a technique that is capable of executing automatic diagnoses efficiently and reliably and at the same time is easy to use for psychological diagnosis is needed. To meet this need we investigated a new methodology, which we tested on a database of 260 previously diagnosed children, to whom we administered a psychometric intelligence scale, in a version suitable for their age, i.e. the WISC-Revised [1]. Details of this scale and of the database are given in the section 6.

The three main aims of the work were the following:

1. Automatic and simple recognition of the level of mental retardation of children by administering Wechsler intelligence scales.
2. Analysis of the set of data to discover the importance of each feature for diagnostic purposes, and consequent generation of a subset of features that would allow faster application of the scale.
3. Completion of the data. By nature, in fact, the children database featured a large amount of missing data (at least one of the subtests that make up the scale was missing for about 50% of the patterns).

Together with the limited number of patterns, this last point led us to search for a non-traditional solution, i.e. to create a simple, fast, but efficient algorithm that was capable of analyzing data sets with missing values and completing them.

Throughout our work we focused on practical application and consequently on the effect that any error of assessment might have on an already disadvantaged human being. As the aim of the tool is above all to aid clinical diagnosis, serious classification errors are inadmissible because they would lead to an incorrect diagnosis and thus make automatic diagnosis useless. We therefore chose to use a fuzzy classifier which gives indications as to the correct diagnosis even in badly classified cases, indications which can advise the psychotherapist to investigate more thoroughly, guided by the fuzzy affinities with the various groups or pathologies.

An excellent algorithm that helps to meet some of these requirements is Fuzzy C-Means (FCM), which will be discussed in greater detail in Section 3, together with techniques that can be integrated in it to complete the missing data. To achieve the aim in point 2, we decided to integrate the data mining capabilities offered by evolutionary algorithms with the classification and data completion capabilities of FCM, thus obtaining a technique which would meet all our research requirements.

The rest of the paper is organized as follows. Information about evolutionary algorithms in general, and in particular the classical genetic algorithm we chose to use, is given in Section 4. Section 5 gives details of our implementation and the solutions adopted to integrate the two algorithms. Section 6 deals with application of the solution to our database of patients. Section 7 presents our conclusions and an overview of the prospects for development our solution offers.

2 Similar Works

A recent trend in cluster analysis is evolutionary clustering based on well known evolutionary algorithms (EAs), which have shown the potential to achieve high partitioning accuracy results. The main characteristic of EA is that they are widely applicable with success; meanwhile classical clustering approaches are often good only for certain problems.

Previous approach employed Evolutionary Strategies [2], Evolutionary Programming [3], and recently Particle Swarm Optimization [4] and Simulated Annealing [5]. Good results were also obtained through hybrid approaches with classical clustering algorithms [2, 6], especially the ones which integrates clustering and evolutionary algorithms to exploit the flexibility and adaptability of the EA together with the scalability and accuracy of classification algorithms.

Many search algorithms have been used for feature selection [7]. Among these, EAs have proven to be an effective computational method, especially in situations where the search space is uncharacterized (mathematically), not fully understood, or/and highly dimensional. There are two kinds of feature selection algorithms [8, 9]:

- Filter feature selection algorithms, which remove the irrelevant characteristics without using a learning algorithm. They are efficient processes but the feature subsets obtained may not be the best ones for a specific learning process.
- Wrapper feature selection algorithms. This kind of feature selection algorithm selects feature subsets using the precision of a classification algorithm to evaluate each candidate subset. Their problem is inefficiency, since they have to execute the classification algorithm for each evaluation.

One particular application of these methods not only selects features but also assigns them weights according to their importance for the analysis to be performed [10].

Feature selection techniques do not normally offer the possibility of classifying the sets they analyze: they are, in fact, proposed as filter techniques. There are, however, certain exceptions such as C4.5 [11], which belong to the supervised machine learning category. Another exception is the technique introduced in [12], which uses frequency-based distance measures to determine the weights to be assigned to different patterns while implementing a weighted k-nearest neighbor algorithm for classification (PEBLS).

A recent work which is technically similar to ours and uses an intelligent genetic algorithm (IGA) to design an optimal nearest neighbor (l-nn) classifier, was presented in [13]. It was shown empirically that IGA-designed classifiers outperform some existing methods, including Kuncheva and Bezdek's A-based method [14]. An improvement on the IGA-based algorithm proposed by Ho et al. is IMOEA, presented in [15]. This algorithm uses an intelligent multi-objective evolutionary algorithm to design l-nn classifiers. Both these techniques deal with data and feature reduction, but not with searching for their significance.

Discriminant analysis, given its widespread use in psychological research and also because the results it yields also include indications (F-to-Remove) as to the significance of features. For greater detail concerning this statistical technique in the pattern recognition field see [16].

2.1 Missing value analysis: A brief overview

There are various statistical methodologies for data completion in the literature (see [17] for a general discussion). The most widely used are:

- *Regression substitution (RS)*. This method uses multiple linear regression to obtain estimates of the missing values. It is applied by estimating a regression equation for each variable, using the others as predictors. Present data is then exploited to obtain the missing data.
- *EM Estimation (EME)*. This is based on the EM algorithm, which comprises two phases: "E" (Expectation), predicts an initial value for missing data using other methods (e.g. multiple linear regression); subsequently, in the "M" (Maximization) phase the missing values are calculated iteratively using the "maximum likelihood function" until the desired accuracy is reached.

In Section 6 we compare these methods with ours.

3 Fuzzy Clustering: A Brief Overview

Indicating the degree of membership (or "similarity") of a pattern to every group is called "*Fuzzy Clustering*"; this is different from "*Hard Clustering*", which simply associates them. In "*hard*" approaches, a data set is subdivided into separate partitions in which each pattern belongs to a single cluster. In the fuzzy approach, on the other hand, partitions are not created: elements are associated to each group, with a degree of membership. The result yielded by this algorithm is thus not a simple partitioning (which can, however, still be achieved by assigning each pattern to the group with which it has a higher numerical affinity, for example), but more detailed information on the relations between the patterns and groups. There are two main groups of Fuzzy Clustering techniques; the first includes algorithms that directly use fuzzy sets, while the second comprises ad-hoc algorithms such as Fuzzy C-Means. Further details of these and all other types of clustering are to be found in [18].

3.1 Fuzzy C-Means

The most widely used algorithm implementing the fuzzy philosophy is "Fuzzy C-Means" (FCM), developed by Bezdek, who proposed a generalization by means of a family of objective functions [19]. Its main features are ease of implementation and computational complexity, which is linear for all parameters.

The FCM version we implemented proposes to minimize the following objective function:

$$J(U,C;X)=\sum_{j=1}^{K}\sum_{i=1}^{N}u_{ij}^{m}\cdot d(\mathbf{x}_i,\mathbf{c}_j) \tag{1}$$

where $\mathbf{c}_j$ is the prototype of the j-th cluster and $d(\bullet,\bullet)$ is a distance metric appropriately chosen from the pattern space, $\mathbf{x}_i$ is the i-th pattern, u_{ij} is the degree of truth of the i-th pattern in the j-th cluster, raised to the "fuzzyfier" m. K and N are respectively the number of clusters and the number of patterns.

m is a parameter on which the degree of fuzzyfication depends: as its value increases, so does the degree of uncertainty, until it settles at $u_{ij} = 1/K\ \forall i,j$, whereas when it gets close to 1 the result is the partitioning typical of "hard" algorithms.

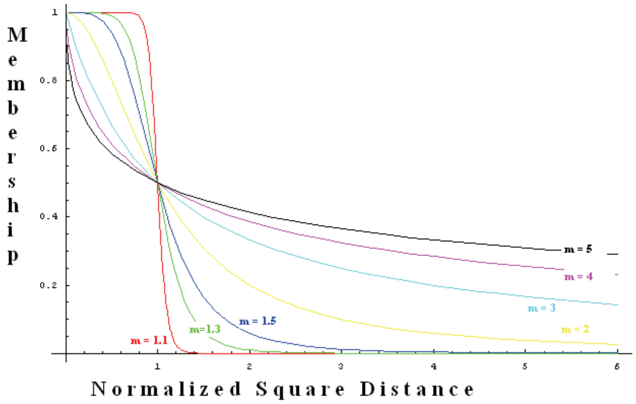

Figure 1. Variation in fuzzy membership with respect to distance with different values of the parameter m.

The accuracy by which the algorithm classifies patterns depends heavily on the value of m. Normally it is made to vary between 1.5 and 2, but sometimes better classification is achieved with higher or lower values. The optimal choice is linked to the data set being investigated and thus varies according to the application involved. There are no techniques which allow an effective value to be chosen a priori. In our implementation we use the genetic algorithm to estimate the optimal value for m.

The FCM algorithm can be summarized in the following steps:

1. Given the N patterns assume a predefined number of clusters K, where $K\in[2,N[$.
2. At step p=0, initialize the matrix $U^{(0)} = [u_{ij}]$, respecting the constraint given by

$$\sum_{j=1}^{K}u_{ij}=1 \quad \forall i \tag{2}$$

3. At step p>0, calculate the centroids $C^{(p)}$, i.e. the prototype vectors $\mathbf{c}_j$, starting from $U^{(p-1)}$:

$$\mathbf{c}_j=\frac{\sum_{i=1}^{N}u_{ij}\cdot\mathbf{x}_i}{\sum_{i=1}^{N}u_{ij}} \tag{3}$$

4. Update the matrix $U^{(p-1)}$, obtaining $U^{(p)}$:

$$u_{ij}=\frac{1}{\sum_{k=1}^{K}\left[\frac{d(\mathbf{x}_i,\mathbf{c}_j)}{d(\mathbf{x}_i,\mathbf{c}_k)}\right]^{2/(m-1)}} \tag{4}$$

5. Verify whether the STOP criterion is satisfied; if it is not, return to point 3.

The STOP criterion normally chosen is $\|U^{(p)} - U^{(p-1)}\| < \varepsilon$, with $\varepsilon\geq 0$.In order to avoid long calculation time it is preferable to choose a certain number of iterations as the STOP criterion; in this case the first condition is also applied and the algorithm is stopped when one of the two is met.

3.2 Missing Data Analysis with FCM

It is possible to integrate into the FCM iterations an efficient estimate of values which for various reasons have not been collected during data collection. Hathway and Bezdek [20] identified four methods for data completion that can be integrated with FCM:

- *Whole Data Strategy (WDS).* This method simply eliminates any incomplete patterns from the classification. It is only suitable if what is being sought is estimation of the centroids and not classification of all the patterns.
- *Partial Distance Strategy (PDS).* If there are missing features, this method calculates the partial distances using the features that are present and rescales the classification taking the number of missing features into account.
- *Optimal Completion Strategy (OCS).* In this approach the values of missing features (x_{ik}) are seen as further parameters to be optimized:

$$x_{ik}=\frac{\sum_{j=1}^{K}u_{ij}^{m}c_{ik}}{\sum_{j=1}^{K}u_{ij}^{m}} \tag{5}$$

and thus estimated directly by the algorithm during its execution cycle, evaluating (5) between step 4 and step 5.

- *Nearest Prototype Strategy (NPS).* This calculates the partial distances and approximates the missing values with those of the nearest prototype. The authors state that although it always yielded results in their numerical tests, the convergence of this strategy was not demonstrated.

By means of numerical tests and theoretical considerations, the authors indicate OCS as the strategy that performs best out of the four. Another advantage of this strategy is that its output is a complete data set. For these reasons, we chose to integrate this strategy in our implementation of the FCM algorithm, calling the resulting algorithm FCMOCS.

4 Genetic Algorithms

The basic idea behind these systems consists of evolving a population of potential solutions to a given problem using operators inspired by genetic variability and natural selection. The most common algorithms belonging to the evolutionary family are undoubtedly Genetic Algorithms (GAs), the classical model of which was introduced by Holland [21]. A GA operates on a population of individuals called "chromosomes", each representing a candidate solution to a given problem. An iteration of the algorithm, known as a generation, causes the current population to evolve into a new one with the aim of finding increasingly better candidates.

The evolution of a GA derives from Selection, Reproduction (Copy, Crossover, and Mutation) operators. Selection is used to determine which individuals could be used for the following Reproduction operators. According to the roulette rule, the algorithm will choose the individuals in relation to their fitness: the "fitter" a chromosome, the greater its probability of being chosen. The individuals selected for reproduction are called "parents". The first Reproduction operator is Copy. It only chooses a parent and copies it in the new generation. This technique prevents the elimination of the best individuals which could cause a regression to worse candidates. Crossover is the main reproduction operator and it exchanges portions of the chromosome of two parents to create two new individuals called "offspring". The third reproduction operator is Mutation, which allows new parts of a chromosome to be inserted into a parent. It is useful to get out of local minimum or maximum values. These operators are selected in a probabilistic way; each operator probability is defined before the algorithm starts. In our specific case, the mutation operator randomly modifies the value of a parameter chosen at random. The crossover between two configurations exchanges the value of two parameters chosen at random.

A GA has other parameters to be set before it runs, such as the number of generations, G, and the number of individuals, I, in each generation. The choice of G is generally based on experience. The same is for I that in this work we set as twice the number of parameters P to be tuned: $I = 2 \cdot P$. In this way we can express the computational complexity relating to the set being dealt with as linear, or $O(P)$. Another (linear) parameter is the number of generations, but this can be set a priori as a stop criterion, so it does not depend on the set being examined.

The classic GA we used can be summarized as follows:

1. *Set number of individuals, N.*
2. *Randomly generate initial population.*
3. *Evaluate fitness value for each individual.*
4. *Copy individual with best fitness into new population (elitism).*
5. *Select a pair of "parent" individuals from the population (using roulette wheel method).*
6. *Apply crossover and/or mutation operators to create new "offspring" to insert into the new population.*
7. *Repeat steps 5 and 6 until a new population of N individuals is generated.*
8. *Repeat steps 2 to 7 until stop criterion is satisfied.*

5 FCM-GA: An Integrated Technique for Data Analysis

Besides its fuzzy nature, the FCM algorithm was chosen due to the need to analyze a data set with up to 50% of the data missing. As seen in Section 3, there are various solutions to complete data, among which we chose OCS. The choice was due to the fact that OCS can be integrated with the FCM algorithm and also provides an effective estimate of values which were not collected during data collection. However, although the algorithm classifies and completes data efficiently, it does not allow features to be analyzed and has a parameter (the fuzzyfier m) to be optimized. We therefore opted for further integration with a classic GA, thus obtaining not only the information being sought but also a considerable improvement in classification accuracy as compared with a simple FCM. We will call the most simple version of the integrated algorithm thus obtained FCM-GA, whereas we will refer to the technique as FCMOCS-GA when the OCS data completion solution is also integrated.

Integrated implementation of the two algorithms caused certain problems: a pure FCM requires random initialization, on which the result depends. It is obviously not possible to operate in this way with the genetic algorithm, because the fitness value would vary from one generation to another. A way to solve this problem is to insert the initial values of the centroids among the variables of the GA. Hall *et al.* [22] studied the effects of this strategy, concluding that use of a GA caused an increase in computing time of two orders of magnitude as compared with normal execution. In normal conditions it is therefore preferable to execute the algorithm several times, starting from different initial values, which gives similar, if not identical, results.

In the case we present the analysis was performed on a data set in which the patterns had already been classified, so we chose to initialize the centroids with the mean values for the single groups. This yielded the same results obtainable with both random initialization and using the GA, but obviously meant saving a considerable amount of time.

Before initializations, the data is normalized so as to remove any numerical differences between the features and allow the algorithm to estimate their significance more efficiently. The normalization method we applied was division by the maximum number: $X_{\text{norm}} = X / max$; in our preliminary tests this method proved to perform much better than classical normalization $X_{\text{norm}} = (X - min)/(max-min)$.

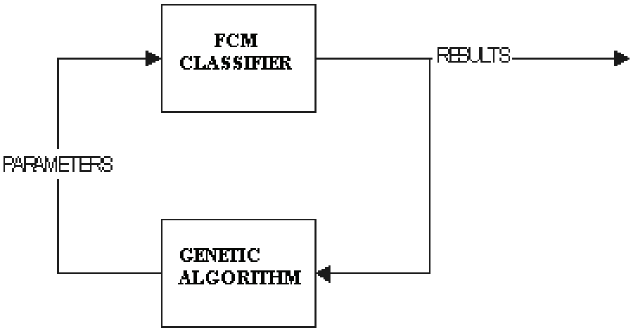

Figure 2. Systemic representation of the proposed approach.

The metric chosen was:

$$d(\mathbf{x},\mathbf{y}) = \sqrt{\sum_{k=1}^{D} w_k^2 (x_k - y_k)^2} \qquad (6)$$

where D is the size of the space of features and w_j is the weight assigned to the i-th feature, which is inserted as a parameter to be estimated by the GA.

The chromosome (Figure 3) will then be defined with as many genes as there are free parameters and each gene will be coded according to the set of values it can take. In our case study, the parameter of FCM and feature weights are mapped on a chromosome whose genes are binary coded, using a 7 bit gray code for each parameter.

FCM fuzzyfier	**The D Feature weights**		
m	w_1	…	w_D

Figure 3. Example of the structure of a chromosome: our gray binary coded implementation had 7 bits precision for each parameter.

The two integrated algorithms both have a linear complexity, FCM in the parameters N (total number of patterns), D (number of features) and K (number of clusters), which can be expressed as $O(N \bullet D \bullet K)$, while, as said, the complexity of a GA generation is $O(P)$. Hence, having chosen $P \approx 2 \bullet D$, integration of the two algorithms has a square complexity in D: $O(N \bullet K \bullet D^2)$.

The next step was to define in a simple but effective way calculation of the fitness value (F) that would exploit the execution of FCM to judge an individual's fitness to reproduce. The following function was chosen, comprising three summations, to guarantee Accuracy, Classification Consistency and Feature Selection, in that order:

$$F = \sum_{i=1}^{M} z_i + \sum_{i=1}^{M} \frac{[u_{ig_i} \cdot (1 - z_i)]}{S_1} + \sum_{j=1}^{D} \frac{(1 - w_j^2)}{S_2} \qquad (7)$$

where z_i are binary variable whose value is 1 if the predicted group is equal to the real one and 0 otherwise. g_i is the group to which the i-th element belongs, and u_{ig_i} is its affinity with the group calculated by the fuzzy clustering algorithm. S_1 e S_2 are two damping coefficients. M is the number of patterns possessing a reference diagnosis, with $K<M\leq N$.

To summarize, the approach, schematized in Figure 3, is as follows:

1. *Normalize data by dividing by maximum value for each feature.*
2. *Initialize centroids and missing values.*
3. *Execute GA supplementing FCM to calculate fitness.*
4. *Select features on basis of weights assigned.*

Once the algorithm has obtained results, it is possible to build a fuzzy classifier to obtain automatically the classification for patterns with no reference, exploiting the FCM with distance (6) where the weights w_j are the optimal ones estimated by the algorithm. By setting the centroids it is possible to perform automatically new diagnoses in a single FCM iteration on the patterns to be classified alone.

The stop criterion used for the FCM was the achievement of a maximum variation (ε) lower than 0.01 or 20 iterations, whereas for the GA it was a lack of evolution in the population after 200 generations. Mutation probability used was 0.2 and crossover probability was 0.7, the population size was of 30 chromosomes. S_1 e S_2 were chosen as equal to N and D respectively. Weights range was [0,1].

6 Case study: WISC-R, a description of the method, of the database and numerical results

David Wechsler developed his intelligence scales in 1939 on the basis of his experience as a clinical psychologist at the Bellevue Psychiatric Hospital in New York, and in a revised form they are still the most widely used to measure general intelligence. They comprise various tasks, grouped into verbal and performance subtests. The two main scales were developed and adapted for use in several countries and languages: WAIS-R (Wechsler Adult Intelligence Scale – Revised) for adults (over 16) and WISC-R (Wechsler Intelligence Scale for Children – Revised) for children (under 17). The Italian version was realized by Orsini [23] for WISC-R. Complete administration of a scale takes about an hour and a half.

In psychological assessment the Wechsler scales are only one of many tools used for diagnosis. The results given below should therefore be read bearing this in mind. Further information can be found in [24]. The reference diagnosis used in the classification for our database was made using other tools, such as the Vineland Adaptation Scale [25] and clinical observations and interviews.

Unlike normal diagnosis, in which the scores obtained are weighted using a standardized adjustment, we used the raw scores, i.e. scores deriving directly from the administration of each subtest. This was due to the fact that the scale was devised and adjusted for "normal" subjects and thus has a "floor effect" when applied to mentally retarded people. By using the raw scores we allowed the algorithm to re-standardize the scales for our data set, considerably improving the degree of accuracy. To assess the intelligence level of children it is essential to relate the results to their age. We therefore weighted the raw scores by age.

WISC-R is one of the most valuable tools for intelligence assessment in subjects aged 6 to 16 years, and a valid clinical and diagnostic aid in the area of educational assessment. It comprises 12 subtests, 6 belonging to the Verbal subscale (Information, Similarities, Arithmetic, Vocabulary, Comprehension, and Digit Span) and 6 to the Performance subscale (Picture Completion, Picture Arrangement, Block Design, Object Assembly, Coding, and Mazes). The first 10 subtests are called *regular tests* and the norms of the scale are defined on the basis of these subtests. The remaining subtests (Coding and Mazes) are called *supplementary tests*. The database included 260 children who represented a quite homogeneous sample of the three retardation levels considered, as shown in Table I. The subdivision of missing values was also quite homogeneous: at least one of the subtests, in most cases Mazes, was not administered to about 50% of the subjects in each group.

TABLE I. CASES PER GROUP WITH AT LEAST ONE FEATURE MISSING.

GROUP	Borderline	Mild	Moderate
Cases with missing feature	42	43	50
Total no. of cases	84	96	81

Table II gives the results obtained both by applying our technique and using discriminant analysis on data completed with EM Estimation (EME) and Regression Substitution (RS).

TABLE II. CHILDREN DATABASE RESULTS, AVERAGES AFTER 10 TRIALS OF THE PROPOSED APPROACH.

		%
1	FCMOCS-GA, classification error	23.0
2	Discriminant Analysis with FCMOCS completed set, error	24.2
3	Discriminant Analysis with EME completed set, error	26.2
4	Discriminant Analysis with RS completed set, error	26.2
5	Discriminant Analysis with FCMOCS-GA selected features, error	23.9
6	Fuzzy affinity with correct group, mean	57.9
7	Fuzzy affinity of correct classified patterns, mean	67.1
8	Badly classified patterns, affinity with correct group, mean	25.2

As can be seen if we compare the results given in rows 1, 3 and 4, the technique we propose improves the accuracy, which can also be transmitted to discriminant analysis: an improvement is achieved by using the completed data set and the features selected (see rows 2 and 5). Rows 6, 7 and 8 show the mean affinity with the group associated with the diagnosis (i.e. the "correct" group) for all patterns, correctly classified patterns, and misclassified patterns. In many misclassified cases the algorithm yields a good affinity with the real group to which the subject belongs, thus providing the psychotherapist with reliable guidance as to how to use other assessment tools.

TABLE III. AVERAGE SUBTEST WEIGHTS AFTER 10 TRIALS OF THE PROPOSED APPROACH.

Subtest	Info.	Simil.	Arith.	Vocab.	Compr.	Digit Span
Weight	0.50	0.98	0.78	0.80	0.80	0.55
Subtest	Picture Compl.	Picture Arr.	Block Design	Object Asse.	Coding	Mazes
Weight	0.65	0.0	0.65	0.0	0.0	0.0

We see that 8 out of 12 tests were selected (Table III) and it is interesting to note that two of the three subtests eliminated are Coding and Mazes, i.e. the supplementary tests. This result is also affected by the fact that these are the ones with a higher number of missing data and so the algorithm "intelligently" eliminates them. Using information in Table III we were able to perform a further selection of only the 6 subtests with higher weights, obtaining a 24.2% classification error with a reduction of three quarters of an hour in the total administration time required.

7 Conclusion and Future Works

In this paper, inspired by a practical need, we have presented a new approach for the diagnosis of mental retardation levels which combines two well-known computational intelligence techniques: genetic algorithms and fuzzy c-means. The approach has a number of qualities, despite the simplicity of the two integrated algorithms. It provides four different possibilities of analysis in a single algorithm:

- Data classification. This is performed automatically by the creation of an ad hoc classifier.
- Completion of missing data.
- The assignment of weights to each feature, from which it is possible to obtain the significance of each feature.
- Feature selection on the basis of the weights assigned.

An important feature, as we have seen in comparisons with popular techniques in psychology, is that the integrated algorithm achieves an excellent degree of accuracy in both classification, the estimation of missing values, and feature selection. Along with a minimum number of parameters to be set, this considerably simplifies the work of researchers, who do not need to apply (and know) different algorithms and it thus allows it to be applied in real everyday situations occurring in psychotherapy and general medical practice. The creation of an easy-to-use automatic diagnostic tool, which at the same time tries to reduce the number of tests to be administered to a patient, is in fact extremely useful because faster (although reliable) diagnosis makes it possible to plan quickly an effective treatment. To this must be added the good scalability of the approach: for a set of N patterns and a size of D to be classified in K clusters the computational complexity is $O(N \bullet K \bullet D^2)$ for each generation executed by the GA, while it is generally in the order of hundreds. This ensures that results can be computed in a relatively short time, which makes the algorithm easier to use in practical applications even with large databases.

Despite the good results obtained, further studies are needed in order to enhance accuracy and the consistency of the results by adding further possibilities to the algorithm. It is, in fact, possible to use more advanced version of the FCM algorithm, which use a possibilistic vision for example; this would be useful from a theoretical viewpoint to make the algorithm robust to noise, and to obtain even more significant results when the set to be analyzed contains pathologies that may be simultaneously present in patients. Another extension could be the use of intelligent genetic algorithms which would allow a faster search for optimal values.

8 Acknowledgment

The authors wish to thank the IRCSS Oasi of Troina for providing the database and for supporting our research.

9 References

[1] D. Wechsler, *WISC-R Wechsler Intelligence Scale for Child Revised – Manual*. San Antonio, TX.: The Psychological Corporation, 1974.

[2] R. Babuska, *Fuzzy Modeling for Control*. Dordrecht: Kluwer Academic Press, 1998.

[3] D. B. Fogel and P. K. Simpson, "Evolving fuzzy clusters," presented at International Conference on Neural Networks, 1993.

[4] D. W. van der Merwe and A. P. Engelbrecht, "Data clustering using particle swarm optimization," presented at Congress on Evolutionary Computation, Camberra, Australia, 2003.

[5] S. Bandyopadhyay, "Simulated Annealing Using Reversible Jump Markov Chain Monte Carlo Algorithm for Fuzzy Clustering'," *IEEE Transactions on Knowledge and Data Engineering* vol. 17, pp. 479-490, 2005.

[6] B. Yuan, G. J. Klir, and J. F. Swan-Stone, "Evolutionary fuzzy c-means clustering algorithm," presented at FUZZ-IEEE '95, 1995.

[7] A. Jain and D. Zongker, "Feature Selection: Evaluation, Application, and Small Sample Performance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, pp. 153-158, 1997.

[8] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artificial Intelligence*, vol. 97, pp. 273-324, 1997.

[9] H. Liu and H. Motoda, *Feature selection for knowledge discovery and data mining*: Kluwer Academic Publishers, 1998.

[10] F. Hussein, R. K. Ward, and N. N. Kharma, "Genetic Algorithms for Feature Selection and Weighting, A Review and Study," presented at ICDAR 2001, 2001.

[11] J. R. Quinlan, *C4.5: Programs for Machine Learning*: Morgan Kauffman, 1993.

[12] S. Cost and S. Salzberg, "A Weighted Nearest Algorithm with Symbolic Features," *Machine Learning*, vol. 10, pp. 57-78, 1993.

[13] S.-Y. Ho, C.-C. Liu, and S. Liu, "Design of an optimal nearest neighbor classifier using an intelligent genetic algorithm," *Pattern Recognition Letters*, vol. 23, pp. 1495-1503, 2002.

[14] L. I. Kuncheva and J. C. Bezdek, "Nearest prototype classification: Clustering, genetic algorithms or random search?," *IEEE Trans. on System Man and Cybernetics, part C*, vol. 28, pp. 160-164, 1998.

[15] J.-H. Chen, H.-M. Chen, and H. S.-Y., "Design of Nearest Neighbor Classifiers Using an Intelligent Multi-Objective Evolutionary Algorithm," presented at PRICAI 2004, 2004.

[16] G. J. McLachlan, *Discriminant analysis and statistical pattern recognition*. New York: Wiley, 1992.

[17] R. J. A. Little and D. B. Rubin, *Statistical Analysis with Missing Data*, 2nd ed. New York: Wiley, 2002.

[18] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data Clustering: A review," *ACM Computing Surveys*, vol. 31, 1999.

[19] J. C. Bezdek, *Pattern Recognition with fuzzy objective function algorithms*. New York: Plenum Press, 1981.

[20] R. J. Hathaway and J. C. Bezdek, "Fuzzy c-Means Clustering of Incomplete Data," *IEEE Transactions on Systems, Man and Cybernetics – Part B: Cybernetics*, vol. 31, 2001.

[21] H. J. Holland, *Adaptation in Natural and Artificial Systems*. MI: University of Michigan Press, 1975.

[22] L. O. Hall, I. B. Ozyurt, and J. C. Bezdek, "Clustering with a genetically optimized approach," *IEEE Transactions on Evolutionary Computation*, vol. 3, pp. 103-112, 1999.

[23] A. Orsini, *WISC-R Contributo alla Taratura Italiana*. Milano: O.S. Organizzazioni Speciali, 1997.

[24] S. Di Nuovo and S. Buono, *Strumenti Psicodiagnostici per il ritardo mentale*. Milano: Franco Angeli, 2002.

[25] S. S. Sparrow, D. A. Balla, and D. V. Cicchetti, *The Vineland Adaptive Behavior Scales*. Circle Pines, MN: America Guidance Service, 1984.