2014 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)
July 6-11, 2014, Beijing, China

Benefits of Fuzzy Logic in the Assessment of Intellectual Disability

Alessandro Di Nuovo
School of Computing and Mathematics
Plymouth University, UK
Faculty of Engineering and Architecture
Kore University of Enna, Italy

Santo Di Nuovo
Department of Formative Sciences
University of Catania, Italy

Serafino Buono
Psychology Operative Unit
IRCCS Oasi of Troina, Italy

Vincenzo Cutello
Department of Mathematics and Computer Science
University of Catania, Italy

***Abstract*—Among the artificial intelligence techniques that successfully support computer assisted decision making, fuzzy logic has proved to be a powerful tool in various fields. In particular it is appreciated by clinical practitioners because of their approaches to take a decision require to deal with uncertainties and vagueness in the knowledge and information. One field in which fuzzy sets theory can be applied with great benefit is psychopathology due to the high prominence of sources of uncertainty, that should be taken into account when the diagnosis of intellectual disability must be formulated. Therefore clinical psychologists have often to deal with comorbidities that make the decision process harder because they must evaluate different assessment tools for a correct diagnosis.**

In our work we investigate the application of computational intelligence methods, and in particular of approaches based on fuzzy logic and its hybridizations, in the psychological assessment by means of theoretical studies and practical experiments with data collected from patients affected by different levels of intellectual disability. In this paper we present a detailed review of the experimental application, with patients under treatment in a clinical centre, of methodologies we propose to generate fuzzy expert systems for the assessment of intellectual disability. Specifically we highlight, with numerical results, how they can be beneficial for the diagnosis and improve efficacy of the administration of psycho-diagnostic instruments and the efficiency of the assessment.

I. Introduction

Fuzzy logic based techniques are widely used to support computer assisted decision making in many fields, e.g. clinical medicine [1], modelling and control [2], computer system design [3], business games [4] and many others. The success of the fuzzy logic in intelligent decision making is due to many peculiar factors that give to fuzzy approaches an edge over other artificial intelligence methods. From our experience we believe that the key features of this success are the handling of uncertainty, the possibility to combine knowledge from experts and data to enhance the knowledge-based system and the simplicity and linguistic interpretability of the decision system. Fields like medical sciences, in which these features are crucial for the real applications, can have a major benefit from application of fuzzy logic based approaches, alone or hybridized with other computational intelligence methods. Computer assisted clinical decision making is the most directly connected with these features, as an example we recall a study that describes an approach to computer-aided medical diagnosis systems for clinical contexts using fuzzy logic [5]. In this system, knowledge from experts is compiled into fuzzy cognitive maps and logical structures to estimate a stage of disease using temporal information in symptom durations. An other example can be found in [6] where an automated brain segmentation method is proposed to build fuzzy models from magnetic resonance images and to express brain features by fuzzy membership functions for the diagnosis of Hypoxic-ischemic encephalopathy in newborns. Moreover many other studies can be found in the scientific literature that propose intelligent methodologies to help medical decision making. As an example we recall the use of fuzzy systems to process images and signals, like the fuzzy inference system presented in [7] to determine the skull and brain surface from transcranial sonography to examine skull fracture and brain disease such as hematoma, and a fuzzy estimation system of dementia severity through monitoring patient's sleep using air pressure and ultrasonic sensor systems. Meanwhile studies in data mining of medical data can contribute to extend and refine the knowledge base for the decision making, examples are [8], in which a new measure is proposed for fuzzy decision trees to help medical diagnosis by detecting and highlighting of gradual relations between attributes and the diagnostic class, and [9] in which the deep tuning of the fuzzy system generated from example data is proposed to improve performance and interpretability. Trade-off between interpretability and efficiency is a critical issue for fuzzy systems (e.g. [10]) that must primarily considered in medical applications.

Motivated by all these successful applications we study the application of fuzzy logic approaches to the psychopathological diagnosis with the aim to provide a comprehensive framework that can be beneficial for both researchers and practitioners in this field to develop and use computer software for data analysis, diagnosis and therapy decision. A correct diagnosis of a patient's level of intellectual disability, 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

978-1-4799-2072-3/14/$31.00 ©2014 IEEE

field use various diagnostic tools that are mainly designed to measure the intelligent quotient (IQ). Indeed the most widely used intelligence classification systems (i.e. [11], [12] by World Health Organization and the DSM by American Psychiatric Association [13], [14]) require IQ score to evaluate the existence and degree of intellectual disability, and to provide the patient with the right treatment. Furthermore juridical decisions about capacity to "understand and will" in order to assess legal responsibility, or to recognize invalidity, also need IQ score estimation. The main instrument to obtain IQ are the tests proposed by Wechsler ([15], [16]), that are still the most widely used instruments to measure general intelligence [17].

However, this way of evaluating a person's degree of intelligence finds limitations when used in the lower ranges of cognitive level. In some research studies or in clinical practice it is not possible to use with all subjects complex intelligence tests like the Wechslers, because they have verbal instructions to be understood and adequate linguistic and practical abilities are required in order to carry out the tasks. People with serious cognitive defects and/or intellectual disability may not have these minimum abilities, so for them it is necessary to use other tools, that must be based on simplified or non-verbal instructions that allow responses with a minimal linguistic content.

On the other hand, different psycho-diagnostic tools, more suited to a subject's specific condition, could be used to indirectly derive the Wechsler IQ, but when they are used it is then necessary to reach a common metric so as to compare the reliability of the results obtained and thus ensure a homogeneous diagnosis. In this scenario the practical problem in diagnosing intellectual disability is the need to match IQ scores obtained using other tests with the Wechsler scale which, as said previously, is the most commonly used and universally recognized psycho-diagnostic tool for the diagnosis of intellectual disability.

In this paper we present a review of our efforts to assist the clinical psychologist to deal with these problems thought computer software, which implements different methodologies based on fuzzy logic and hybridizations with other soft computing techniques to extract the knowledge base from a database of examples. The software assists the diagnosis not only by means of an expert systems, but it also can deal with the common difficulties in the administration of psycho-diagnostic tools such as improving efficiency by reducing the complexity and missing data completion. So that the final approach that integrates all the three methodologies is definitely beneficial for the diagnosis and, thus, to identify the most suitable therapy for the patient.

The rest of the paper is organized as follows: in section II we present the materials and methods used in our experiments: the psychodiagnostic instruments used are introduced in section II-A, while the database of information that was collected in a clinical and research center is detailed in section II-B. Methods based on Fuzzy Logic studied to support the diagnosis are described in section II-C, then most important results of numerical experiments of their application with the database are highlighted in section III. Finally in section IV we give our final remarks and conclusion.

II. Materials and Methods

In this section we will describe the materials and methods used in our experiments, in particular in Section II-A we present the psycho-diagnostic tools administered to the database detailed in Section II-B. In the last Section II-C are also briefly presented the fuzzy techniques we propose to use in for the psychological assessment.

A. *Psychodiagnostic instruments*

This section details the psychodiagnostic instruments that were administered to build the database that is detailed in section II-B. As said previously the main instrument to obtain IQ and, thus, to assess the intellectual disability and its level are the Wechsler Intelligence Scales, that are still the most widely used instruments to measure general intelligence. On the other hand when it is not possible to administer Wechsler scales, for example subjects who have limited verbal ability can reliably do other tests; some of these tests were specially devised for intellectually retarded subjects. The problem in these cases is to to indirectly derive the Wechsler IQ, but when they are used it is then necessary to reach a common metric so as to compare the reliability of the results obtained and thus ensure a homogeneous diagnosis. In our research we addressed this problem by creating a computational models to estimate accurately and reliably the Wechsler IQ from other four instruments that are detailed in the following along with the Wechsler scales. In our experiments we used the Italian translations of the following tests: *Wechsler Intelligence Scales*. The two main versions of Wechsler scales were developed and adapted for use in several countries and languages: WAIS (Wechsler Adult Intelligence Scale) for adults (over 16) and WISC (Wechsler Intelligence Scale for Children) for children (under 17). Complete administration of a scale takes about an hour and a half. In our research we used the last Italian version of the *Wechsler Scale for Children* (WISC-III), that is the most valuable tool for intelligence assessment in subjects aged 6 to 16 years, and older if an intellectual disability is diagnosed. It comprises 12 subtests, 6 requiring verbal administration and response (Information, Similarities, Arithmetic, Vocabulary, Comprehension, Digit Span) and 6 requiring Performance (Picture Completion, Picture Arrangement, Block Design, Object Assembly, Coding, 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.
Leiter International Performance Scale (LIPS). Russell Graydon Leiter constructed it in 1927. The version we used was presented in [18], as a culture-free test of general intelligence, in the sense that it is mainly based on abstract concepts with a minimum presence of verbal components in both the test itself and in the administration procedures. The LIPS is a intelligence test for children and adolescents, with norms ranging from 2 to 20 years. For all ages, it yields an IQ and

a measure of logical ability. Leiter devised an experimental edition of the test in 1929 to assess the intelligence of those with hearing or speech impairment and with non-English speaking examinees. A remarkable feature of the Leiter is that it can be administered completely without the use of oral language, not even for instructions. Without any verbal subtests, Leiter only measures non-verbal intelligence. Because of the exclusion of language, it claims to be more accurate when testing children who for some reason have language deficits. This includes children with any of these features: Non-native speaking, autism, traumatic brain injury, speech impairment, and hearing problem. The Leiter contains 20 subtests organized into four domains: *Reasoning ; Visualization ; Memory ; Attention*.

Coloured Progressive Matrices test (CPM). It is a version for children or mentally retarded subjects developed in Great Britain by Raven [19], who had worked for several years on the factors of intelligence according to the theory of Spearman [20]. It is often referred to simply as Raven's Matrices. The test consists of incomplete pictures in which the subject has to find the missing part. This test is also considered to be culture-free, and is easy to administer, even to subjects with limited linguistic resources. Designed for younger children, the elderly, and people with moderate or severe learning difficulties, this test contains sets A and B from the standard matrices, with a further set of 12 items inserted between the two, as set Ab. Most items are presented on a coloured background to make the test visually stimulating for participants. However the very last few items in set B are presented as black-on-white, in this way, if participants exceed the tester's expectations, transition to sets C, D, and E of the standard matrices is eased.

The Mental Development Scale (MDS). This test was devised by Ruth Griffiths and subsequently revised [21]. The scale can be used for the assessment of intelligence in children with various types of handicap. In these cases it is administered by presenting items corresponding to a much lower chronological age than that of the subject, which are assumed to be adequate for the level of development reached. The test can be administered to any subject, irrespective of its chronological age and degree of intellectual disability, and thus to subjects who cannot be assessed using other psychometric tests.

The Psycho Educational Profile (PEP). It was devised by Schopler [22], is an inventory of behaviours and abilities that gives information about a child's functional level in various areas as compared with that of children of the same age. It is designed to diagnose and assess skills and behaviours of autistic and communication-handicapped children who function between the ages of 6 months and 7 years. The resulting profile reflects individual learning and behaviour characteristics. This profile can be translated into appropriate and effective individualized educational plans and home teaching programs. It is the only tool that can be used reliably with autistic subjects or those affected by similar developmental disorders.

B. Database

To run our experiments we collected a database that comprises raw and final scores of the administration of several psycho-diagnostic instruments to 186 adults and 261 children diagnosed as mentally retarded [23]. The diagnosis was made by a team of psychiatrists and psychologists on the basis of DSM-IV clinical criteria [13]. All the subjects were tested on the Wechsler scale suitable for their age and they were diagnosed in a center specialized intellectual disability, where reference diagnoses, included in our database, were made using also other tools, such as the Vineland Adaptation Scale [24] and clinical observations and interviews. Regarding the Wechsler scales some subtests were not administered to all the subjects in the database due to specific pathologies or other reasons, so there is at least one subtest missing in 8,8% of adult cases, while this percentage rises to 50% was not administered to about 50% of the children. The occurrence of missing values is quite homogeneous, so it is not related with the intellectual disability level. We remark that 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.Further information can be found in [25]. Moreover the database comprises the results obtained by administering all the tests described above (Wechsler, LIPS, CPM, MDS and PEP) to 40 persons, 24 were Male and 16 were Female. The reason for the rather low number of people in this sub-database is that it is very difficult to administer all 5 tests to patients with intellectual disability, so we had to select the subjects from the initial greater group. Few subjects have been included that were not diagnosed as mentally retarded, because comprised in the normal range of IQ. They had been submitted to the diagnosis in the same center for relevant learning disorders, but intellectual disability failed to be demonstrated, being the disorder due mainly to motivational and/or emotional problems.

In our database we used the raw scores, i.e. scores deriving directly from the administration of each subtest, as the features of the patterns in the database. The database was chosen because it can represent the common dataset in psychological research. Our database includes patients diagnosed as affected by borderline, mild and moderate intellectual disability. In previous work [26] it was noted that only the correlation between the $WISC_{IQ}$ score and the LIPS score (which is already corrected by age as it is an IQ) is sufficiently high as to ensure the reliability of the estimate, whereas when scores are corrected by chronological age (i.e. dividing the raw scores by the patient's age expressed in months, with a saturation point at 216 months - 18 years) much higher correlations with $WISC_{IQ}$ are achieved.

We remark that, since the diagnosis is not based only on IQ scores but must also take into consideration a person's adaptive functioning, the clinical diagnosis is not made rigidly, therefore it encompasses intellectual scores, adaptive functioning scores from an adaptive behaviour rating scale based on descriptions of known abilities provided by someone familiar with the person, and also the observations of the assessment examiner who is able to find out directly from the person what he or she can understand, communicate, and like.

C. Fuzzy Approaches

1) Fuzzy C-Means with Missing data: The FCM is the most famous fuzzy clustering algorithm and it has various versions and implementations [27]. In our work we used the classical one, which proposes to minimize the following objective function with respect to fuzzy memberships $U = [u_{ij}]$ and cluster centroids $C = [\mathbf{c}_j]$:

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

where $\bar{c}_j$ is the prototype of the j-th cluster and $d(\bullet, \bullet)$ is a distance metric appropriately chosen from the pattern space, $\bar{x}_i$ 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 a hard partition (i.e. u_{ij} becomes a binary variable that identifies only if the partners belongs to the group or not). The STOP criterion normally chosen is $\left\| U^{(p)} - U^{(p-1)} \right\| < \varepsilon$, with $\varepsilon \geq 0$. In order to avoid long calculation time it is preferable to choose a certain number of iterations S 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.

The FCM implementation we used for data completion has $m = 2$, $\varepsilon = 10^{-5}$, $S = 100$, and uses the Euclidean metric as distance measure.

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. Hathaway and Bezdek [28] identified four methods for data completion that can be integrated with FCM. The authors, by means of numerical tests and theoretical considerations, indicate the 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, that is a very useful characteristic in our application scenario. For these reasons, we chose to integrate this strategy in our implementation of the FCM algorithm, calling the resulting algorithm FCM-OCS. In the Optimal Completion Strategy (OCS) approach the values of missing features (x_{ik}) are seen as further parameters to be optimized:

$$x_{ik} = \frac{\sum_{j=1}^{K} u_{ij}^m c_{jk}}{\sum_{j=1}^{K} u_{ij}^m} \qquad (2)$$

and thus estimated directly by the algorithm during its execution cycle.

2) Genetic Fuzzy C-Means for Wechsler subtests weighting and selection: In [25] we extended the FCM-OCS algorithm with its hybridization in the loop of a genetic algorithm in order to perform feature selection and weighting. The resulting algorithm was named GFCM and it is a feature selection algorithm of the wrapper type that uses GA to optimize the FCM algorithm. In our implementation information retrieved from GFCM are used to extract a fuzzy rule based classifier for the retardation level diagnosis, labeling clusters according to the data inside the cluster [27]. GFCM is also able to associate to each feature a real value (weight) that is related with its importance (i.e. it ranks selected features). Redundant and noise features were directly eliminated from GFCM during its exploration, assigning them a weight equal to zero. After the GFCM concluded its exploration an efficient classifier could be created using the optimal parameters founded by the algorithm. It is also possible to build an even smaller, but less accurate, classifier making a further feature selection on the basis of the weights assigned. To meet our requirements fitness function structure has to achieve the highest classification accuracy, but in a diagnostic system, as we stated before, also the consistency of the fuzzy partitions must be assured. These two objectives are the two faces of the same coin, in fact the first one helps diagnosis accuracy the second one its reliability. An other characteristic which can help diagnosis accuracy is the feature selection, which can eliminate redundant, unnecessary and noisy features with the advantage to simplify the diagnosis procedure. So the fitness function has to satisfy three objectives: classification accuracy, consistency and feature selection. One more than our reference algorithm. To quantify fuzzy partition consistency we chose to use a validity measure. In our implementation, we use the Xie-Beni (XB_{index}) [29] cluster validity index. Feature selection was evaluated as the number of features S_f used to classify the patterns. The Xie-Beni index should have a lower value when the data has been appropriately clustered, we expect the same is for the number of features selected. So the fitness function had to be minimized. For this reason we defined the classification accuracy as the amount of patterns misclassified, using the vector z, where the elements $\mathbf{z}_i$ are binary variable whose value is 0 if the predicted group is equal to the real one and 1 otherwise. The dimensionality of vector $\mathbf{z}$ is M, which is the number of patterns with a reference diagnosis. The three measures chosen to evaluate the objectives are not conflicting; therefore they can be seen as part of a single objective that is to build the best classifier in terms of diagnosis accuracy and reliability. So an aggregative fitness function is the best choice to guarantee maximum performances in the diagnostic scenario studied in our work. To summarize the GA had to minimize the following fitness function, comprising three sums, to guarantee the three objectives, classification accuracy, consistency and feature selection, in that order:

$$F(\mathbf{z}, XB_{index}, S_f) = \sum_{i=1}^{M} z_i + \frac{XB_{index}}{\alpha_1} + \frac{S_f}{\alpha_2} \qquad (3)$$

where α_1 and α_2 are two damping coefficients, introduced to modify the importance of consistency and feature selection. Once the algorithm ends its iterations, the data set is completed and an optimized FCM classifier is build. When new patterns are available they can be classified using the optimized FCM classifier, meanwhile, if necessary, missing data can be estimated using OCS. Once the fuzzy degree of truth matrix was obtained the diagnosis can be made associating each pattern to the group with the highest degree of truth.

3) IQ estimation from other psycho-diagnostic tools: In order to estimate the IQ score, which is necessary for diagnosis, we considered several methodologies that can create models capable of estimating a patient's Wechsler IQ on the basis of scores provided by other psycho-diagnostic tools. The fuzzy modelling methodology used here was an hybrid computation intelligence methods that trains an artificial neural network (ANN) to then derive a fuzzy system. ANNs are a computational models that try to simulate the structure and/or functional aspects of biological neural networks. Neural networks are non-linear data modelling tools. They can be used to model complex relationships between inputs and outputs or to find patterns in data. In most cases an ANN is an adaptive system that changes its structure based on external or internal information that flows through the network during the learning phase. For more information the reader can refer to [30]. More specifically, to estimate IQ from the scores obtained (x) we used an *Adaptive-Network-based Fuzzy Inference System* (ANFIS), that belongs to the class of adaptive networks that are functionally equivalent to fuzzy inference systems (FIS) originally proposed in [31]. Using a given input/output data set ANFIS constructs a FIS whose membership function parameters are tuned (adjusted) using a back-propagation algorithm alone in combination with a least squares type of method. This adjustment allows your fuzzy systems to learn from the data they are modelling. In detail for premise parameters that define membership functions, ANFIS employs gradient descent to fine-tune them. For consequent parameters that define the coefficients of each output equations, ANFIS uses the least-squares method to identify them. This approach is thus called hybrid learning method since it combines gradient descent and the least-squares method. Before the ANFIS training phase, it must be specified an initial FIS model structure of the Takagi-Sugeno type [32], that is more suited for function approximation [33]. To specify the initial model structure there are various approaches, in this work we considered two clustering approaches that are best suited for low dimensional databases. They are:

- *Subtractive Clustering* (ANFIS-SUB), that generates the initial FIS model for ANFIS training by first applying subtractive clustering on the data [34].
- *Fuzzy C-Means Clustering* (ANFIS-FCM), that generates the initial FIS using fuzzy c-means (FCM) clustering by extracting a set of rules that models the data behaviour.

The main goal of this work is to obtain a general model, i.e. one that provides a more reliable estimate. To this purpose early stopping is a very common practice in ANN training and often produces networks that generalize well, especially when there is a relevant number of outliers like in our case. For this reason, we implement an early stopping criterion examines the reduction of mean squared error (MSE) of the training set, so the training is stopped when the reduction of MSE between training iterations becomes marginal, i.e. less than a threshold value. Note that the criterion is applied on the training set because the relatively low number of examples makes not feasible to split further the database to create a validation set as usual.

III. Experimental Results

In this section we report the most interesting results obtained in our experiments with the material and methods presented above. In particular we highlight the results of the application of the fuzzy approaches presented above in their respective scenarios. Results are presented in an order that incrementally increase the difficulty of the assessment with whom the clinical psychologist may deal to formulate a diagnosis. So we see the three proposed approaches as steps of an unified single protocol to follow during the assessment.

A. Missing Data Completion

In [35], after comparing the potential and characteristics of most common data completion techniques with the FCM-OCS, we concluded that the fuzzy based technique is the more accurate than others both for Missing Completely At Random (MCAR) and Missing Not at Random (MNAR) datasets as shown in Table I.

Then we applied this technique to the database in order to show, after estimating the missing values, how this approach can enhance the power of the verification of the hypothesis. We recall that the 'power' of experimental verification is defined as the probability of accepting the alternative hypothesis H_1 when it is in reality true. The power increases as the probability of an error of the second type decreases (error β): Power = $1 - \beta$. In other words, the power expresses the probability of not committing errors of the second type, that is, of not neglecting an experimental effect that exists in reality. For details of power analysis - which can be estimated by means of commonly used statistical software - see [36]. From our database of mentally retarded patients we extrapolated a subset of children whose attributes were the scores obtained on the Wechsler scale for children (WISC-R) and those deriving from the Vineland Adaptive Behaviour Scale (VABS) which assesses the following areas: Communication, Daily Living, Socialization and Motor Skills. The subset comprised n=60 subjects, of ages ranging from 6 years 3 months to 16 years 2 months (mean age 11.7,

Table I
Results of classification by Discriminant Analysis on the MCAR and MNAR datasets after missing data imputation with different approaches

Missing data treatment	%correctly classified	
	MCAR	MNAR
Subject Deletion	N.A.	75.8
Regression Imputation (RI)	80.8	77.4
Expectation Maximization Estimation (EME)	81.7	80.6
Fuzzy C-Means with OCS (FCM-OCS)	**82.8**	**81.7**

Table II
CORRELATION (r DI PEARSON) BETWEEN ALL SUBTESTS OF VINELAND SCALE AND THE DIGIT SPAN SUBTEST OF WISC-R, WITH THE PROBABILITY LEVEL AND THE POWER

	Subtest digit span					
	$n = 30$			$n = 60$		
Vineland	r	prob.	power	r	prob.	power
Communication	.23	.23	.20	.36	**<.01**	**.80**
Daily Living	.14	.48	.11	.24	**.06**	**.46**
Socialization	.02	.93	.03	.22	**.09**	**.40**
Motor Skills	-.10	.62	.08	.20	**.14**	**.33**

Table III
CORRELATION (r DI PEARSON) BETWEEN ALL SUBTESTS OF VINELAND SCALE AND THE MAZES SUBTEST OF WISC-R, WITH THE PROBABILITY LEVEL AND THE POWER

	Subtest Mazes of WISC-R					
	$n = 30$			$n = 60$		
Vineland	r	prob.	power	r	prob.	power
Communication	.27	.15	.30	.37	**<.01**	**.84**
Daily Living	.43	.02	.67	.45	**<.01**	**.96**
Socialization	.47	.01	.76	.50	**<.01**	**.98**
Motor Skills	.30	.11	.36	.36	**<.01**	**.80**

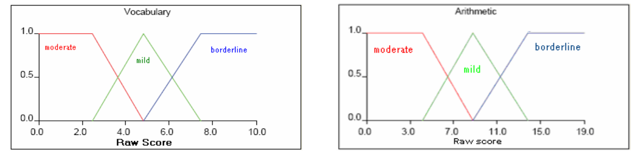
st.dev. 2.95), 21 subjects with a borderline level of mental disability, 26 moderate, and 13 mild. For 30 of these subjects the supplementary subtests of the WISC-R scale, Digit Span and Mazes, were missing. The database was completed by means of the Fuzzy C-Means with OCS technique so as to analyse the total of 60 subjects.

Tables II and III show the Pearson correlations between the Vineland and the supplementary WISC-R subtests, giving for each one the coefficient r, the probability and the power.

It can be seen from the tables that in the complete database the first type of error probability decreases, and above all that the power of the verification, which is linked to the number of subjects in the database, greatly increases (up to 4 times). It should be remembered that the power of studies, which is unfortunately often underestimated by researchers, often determines their "success" in terms of the possibility of demonstrating an experimental hypothesis; and when - as in the case of correlation matrices - the basic analysis are used for further multivariate analysis (e.g. factorial or causal analysis) the numerousness of the database and the consequent power are essential to give sense to the whole analysis. Although there are no formal standards for power, most researchers assess the power of their tests using 0.80 as a standard for adequacy, this means in many analysis that the number of subjects must be at least 4-5 times the number of variables, and so the availability of a technique to impute missing values to complete a database is of great methodological significance for researchers.

B. *Efficient assessment of the retardation level*

In this scenario we dealt with efficiency in the administration of the Wechsler scales, that needs more than one hour and half to be completed to patient affected by intellectual disability. The main goal was to reduce the number of the subtests to be administered while maintaining a good IQ estimation performance. In this experiment the stop criterion



(a) (b)

Figure 1. Examples of Fuzzy sets for Vocabulary (a) and Arithmetic (b) subtests of WAIS-R scale.

used for the GA it was a lack of evolution in the population after 50 generations. Mutation probability used was 0.2 and crossover probability was 0.7, the population size was of 30 chromosomes. α_1 and α_2 were chosen as equal to 1 and to the total number of features (D) respectively. Weights range was [0,1].

Applying a further selection of subtests, by exploiting the significance information given by the algorithm (see Table IV), the number of subtests to be administered can be reduced to 4, while maintaining a good accuracy of 79.0% correctly classified patterns. The administration of only 4 subtests reduces the time needed for diagnosis by an hour (i.e. one third of the total time). When the process is concluded, it is also possible to use information exploited by the GFCM to build an interpretable fuzzy classifier.

In this scenario a pattern is considered as misclassified when the it is classified as belonging to a intellectual disability level different than the reference one.

Rules are formed by sets generated starting from its associated group, which represents the consequent. This way fuzzy rules extracted from our database are very transparent and readable, they could be summarized as follows:

1) `If subtests raw scores are LOW then Disability Level is MODERATE`
2) `If subtests raw scores are MEDIUM then Disability Level is MILD`
3) `If subtests raw scores are HIGH then Disability Level is BORDERLINE`

For the sake of example Figure 1 shows the fuzzy sets obtained for two subtests: Vocabulary and Arithmetic.

Throughout our work we always focused on the 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. For this reason we chose to create a classifier based on fuzzy logic, which gives indications as to the correct diagnosis even in badly classified cases. We call this characteristic of the diagnostic system "consistency" and it could be achieved by evaluating the validity of fuzzy partitions. These fuzzy indications can advise the psychotherapist to investigate more thoroughly, guided by the fuzzy degree of truth associated with the various groups (i.e. pathologies).

To explain the partition consistency Table V shows the average degree of truth with the group associated with the

Table IV
WAIS-R: SUBTEST WEIGHTS AND THEIR IMPACT ON FURTHER SELECTION.

N. of Subtests Selected	Test set (%) Classification Error	WAIS-R subtests average weights										
		Info	Com	DgtS	Simi	Arit	Voc	PicC	PicA	Blk	Obj	Digt
8	14.0	.48	0	.20	.72	.80	.30	.58	.70	0	0	.06
7	17.7	.48	0	.20	.72	.80	.30	.58	.70	0	0	0
6	17.7	.48	0	0	.72	.80	.30	.58	.70	0	0	0
5	18.9	.48	0	0	.72	.80	0	.58	.70	0	0	0
4	21.0	0	0	0	.72	.80	0	.58	.70	0	0	0

Table V
AVERAGE VALUES OF PATTERNS' FUZZY DEGREE OF TRUTH WITH REAL GROUP.

Degree of truth with real group of	Average (%)
All patterns	52.6
Correctly classified patterns	56.1
Wrong classified patterns	30.4

Table VI
CLASSIFICATION MATRIX.

	Classified as			correct
	borderline	mild	moderate	(%)
Real borderline	39	6	0	86.7
Real mild	7	73	9	82.0
Real moderate	0	4	49	92.5
Total	46	83	58	86.1

Table VII
COMPARISON OF ACCURACY AND RELIABILITY OF THE ESTIMATE WITH TEN-FOLD CROSS VALIDATION

		Best ANN	ANFIS-SUB	ANFIS-FCM
LIPS	Median	**3**	3.5	**3**
	Err>5	**22.5**	30	30
	Wrong	**10**	15	15
CPM	Median	**3.5**	4	4
	Err>5	**35**	42.5	42.5
	Wrong	**30**	**30**	32.5
MDS	Median	4	4	4
	Err>5	35	40	**32.5**
	Wrong	**20**	25	25
PEP	Median	**4.5**	**4.5**	5
	Err>5	**40**	**40**	47.5
	Wrong	30	**25**	30
LIPS	Median	2	2	2
+	Err>5	15	**12.5**	15
CPM	Wrong	**7.5**	10	12.5
PEP	Median	**3.5**	**3.5**	4
+	Err>5	**27.5**	32.5	32.5
CPM	Wrong	32.5	30	**27.5**
LIPS+	Median	**2**	2.5	**2**
CPM+	Err>5	**12.5**	20	15
MDS+	Wrong	15	25	**12.5**
ALL	Median	4	**2**	**2**
OTHERS	Err>5	22.5	22.5	**20**
	Wrong	17.5	15	**12.5**

diagnosis (i.e. the "correct" group) for all patterns, correctly classified patterns, and misclassified patterns. In many cases the algorithm yields an high degree of truth with the real group to which the subject belongs, thus providing the psychotherapist with reliable guidance as to how to use other assessment tools.

From Table VI it can be seen that the algorithm classifies the three intellectual disability levels with greater accuracy at the moderate level: this can be accounted for by the fact that these subjects are those that have more marked pathological features. It should be noted that the algorithm is not penalized by the greater number of mild cases in the data set. In no cases there was a shift of two levels. This confirms the consistency of the results obtained.

C. IQ estimation from different psychometric instruments

When complex tests as the Wechsler scales, which are the most commonly used and universally recognized parameter for the diagnosis of degrees of retardation, are not applicable, it is necessary to use other psycho-diagnostic tools more suited for the subject's specific condition. But to ensure a homogeneous diagnosis it is necessary to reach a common metric, thus, the aim of our work is to build models able to estimate accurately and reliably the Wechsler IQ, starting from different psycho-diagnostic tools. In [26] we reviewed several approaches, based on artificial neural networks, with the aim to find the best estimator of the Wechsler IQ from other alternative psycho-diagnostic tools presented above. Table VII presents a summary of the numerical results with a comparison between the best results obtained by Artificial Neural Networks (ANN) and the ANFIS approach. In Table VII we indicate the accuracy of the estimate by absolute error measures, absolute error is defined as the absolute value of the difference between the estimated and real values. The absolute error (err) is then calculated using the following formula: $err = |y - \hat{y}|$, where y is the real score and $\hat{y}$ is the estimated score rounded (up or down) to the closest integer; in this way a lower value corresponds to greater accuracy. After an extensive series of tests with all ANNs methods presented in this section, the threshold for the early stopping was chosen for all methods as less than 25% of MSE reduction. In the meanwhile a maximum duration limit was thus chosen as 20 epochs, this because in our particular case the estimated score is rounded (up or down) to the closest integer, so small changes in MSE will not modify the final performance of the ANN. Finally it should be noted that, thanks to early stopping, different number of epochs were found for the different methods and also for each fold in cross-validation, but rarely the learning phase reached the maximum duration limit of 20 epochs.

From Table VII we see that ANFIS models can be considered as effective as the best ANN when a single test is used as input, while when many test are used as inputs ANFIS models achieve a better estimation performance.

IV. Conclusion

In this paper we reviewed the experimental application of three approaches based on fuzzy logic to deal with common problems that the clinical psychologist must address in the assessment of intellectual disability. We reported numerical results of their application on a database of patients to highlight the benefits of the use of fuzzy logic in the IQ estimation and in the diagnosis of intellectual disability. Therefore algorithms that replicate the approaches showed in this work are implemented in a computer software that is used by clinicians to convert the scores obtained in the alternative tests to Wechsler IQ, and an empirical trial confirmed the validity of this automatized assessment for the clinical diagnosis of intellectual disability [23].

Acknowledgment

The work of Alessandro Di Nuovo was partially supported by the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement n. 288899 (Robot-Era).

References

[1] C. Helgason, T. Jobe, O. De Leon, and D. Mazumdar, "A structural representation of anticipatory thought process using the example of clinical medicine and the physician," in *Fuzzy Systems (FUZZ), 2013 IEEE International Conference on*, 2013, pp. 1–4.

[2] J. M. Sousa and U. Kaymak, *Fuzzy decision making in modeling and control*. World Scientific Singapore, 2002, vol. 27.

[3] A. G. Di Nuovo, M. Palesi, D. Patti, G. Ascia, and V. Catania, "Fuzzy decision making in embedded system design," in *Proceedings of the 4th international conference on hardware/software codesign and system synthesis*. ACM, 2006, pp. 223–228.

[4] F. O. Oderanti and P. D. Wilde, "Dynamics of business games with management of fuzzy rules for decision making," *International Journal of Production Economics*, vol. 128, no. 1, pp. 96 – 109, 2010.

[5] P. R. Innocent and R. I. John, "Computer aided fuzzy medical diagnosis," *Information Sciences*, vol. 162, no. 2, pp. 81–104, 2004.

[6] S. Kobashi, K. Kuramoto, Y. Wakata, K. Ando, R. Ishikura, T. Ishikawa, S. Hirota, and Y. Hata, "Newborn brain mr image segmentation using deformable surface model based on fuzzy knowledge models," *International Journal of Intelligent Computing in Medical Sciences & Image Processing*, vol. 5, no. 2, pp. 115–124, 2013.

[7] Y. Hata, S. Kobashi, K. Kondo, Y. Kitamura, and T. Yanagida, "Transcranial ultrasonography system for visualizing skull and brain surface aided by fuzzy expert system," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 35, no. 6, pp. 1360–1373, 2005.

[8] C. Marsala, "Gradual fuzzy decision trees to help medical diagnosis," in *FUZZ-IEEE*, 2012, pp. 1–6.

[9] A. G. Di Nuovo and V. Catania, "Genetic tuning of fuzzy rule deep structures for efficient knowledge extraction from medical data," in *Systems, Man and Cybernetics, 2006. SMC '06. IEEE International Conference on*, vol. 6, 2006, pp. 5053–5058.

[10] A. Di Nuovo and G. Ascia, "A fuzzy system index to preserve interpretability in deep tuning of fuzzy rule based classifiers," *Journal of Intelligent and Fuzzy Systems*, vol. 25, no. 2, pp. 493–504, 2013.

[11] W.H.O., *The ICD-10 Classification of mental and behavioural disorders: clinical description and diagnostic guidelines*. Geneva: World Health Organization, 1992.

[12] ——, *ICF: International Classification of Functioning, Disability and Health*. Geneva: World Health Organization, 2001.

[13] A.P.A., *Diagnostic Statistic Manual, text revised (DSM-IV-TR)*. Washington: American Psychiatric Association, 2000.

[14] ——, *Diagnostic and Statistical Manual of Mental Disorders (DSM-5)*. Washington: American Psychiatric Association, 2013.

[15] D. Wechsler, *Wechsler Adult Intelligence Scale-III (WAIS-III)*. San Antonio, TX: The Psychological Corporation, 1997.

[16] ——, *Wechsler Intelligence Scale for Children - Third Edition*. San Antonio, TX: The Psychological Corporation, 1991.

[17] R. Colom, "Intelligence assessment," in *Encyclopedia of Applied Psychology*, C. Spielberger, Ed. New York: Elsevier, 2004, pp. 307 – 314.

[18] R. Leiter, *Leiter International Performance Scale*. Stoelting, Chicago, 1979.

[19] J. Raven, *Guide to using the Coloured Progressive Matrices. Sets A, Ab B*. H.K. Lewis, London, 1949.

[20] C. Spearman, ""general intelligence" objectively determined and measured," *American Journal of Psychology*, vol. 15, pp. 201–ï£¡293, 1904.

[21] R. Griffithis, *The abilities of young children, revised*. The Test Agency, London, 1984.

[22] E. Schopler, R. Rechler, A. Bashford, M. Lansing, and L. Marcus, *PsychoEducational Profile Revised (PEP-R)*, ser. Individualized assessment and treatment for autistic and developmental disabled children. Pro-ed, Austin, Texas, 1990, vol. I.

[23] S. D. Nuovo and S. Buono, "Diagnosis of intellectual disability: comparison between clinical criteria and automatized procedures," *Life Span and Disability*, vol. 12, no. 2, pp. 151–159, 2009.

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

[25] A. Di Nuovo, V. Catania, S. Di Nuovo, and S. Buono, "Psychology with soft computing: An integrated approach and its applications," *Applied Soft Computing*, vol. 8, pp. 829–837, 2008.

[26] A. G. Di Nuovo, S. Di Nuovo, and S. Buono, "Intelligent quotient estimation of mental retarded people from different psychometric instruments using artificial neural networks," *Artificial Intelligence in Medicine*, 2012.

[27] J. C. Bezdek, J. Keller, R. Krishnapuram, and N. R. Pal, *Fuzzy Models and algorithms for pattern recognition and image processing*, ser. The handbooks of fuzzy sets series. Springer, 1999.

[28] R. Hathaway and J. Bezdek, "Fuzzy c-means clustering of incomplete data," *IEEE Transactions on Systems, Man and Cybernetics - Part B: Cybernetics*, vol. 31, no. 5, pp. 735–744, 2001.

[29] X. L. Xie and G. Beni, "A validity measure for fuzzy clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 8, pp. 841–847, 1991.

[30] S. Haykin, *Neural Networks: A Comprehensive Foundation*. Prentice Hall, 1999.

[31] J.-S. R. Jang, "ANFIS: Adaptive-network-based fuzzy inference systems," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 23, no. 3, pp. 665–685, May 1993.

[32] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its application to modeling and control," *IEEE Transactions on System, Man and Cybernetics*, vol. 15, pp. 116–132, 1985.

[33] K. Zeng, N.-Y. Zhang, and W.-L. Xu, "A comparative study on sufficient conditions for takagi-sugeno fuzzy systems as universal approximators," *IEEE Transactions on Fuzzy Systems*, vol. 8, no. 6, pp. 773–778, Dec. 2000.

[34] S. Chiu, "Fuzzy model identification based on cluster estimation," *Journal of Intelligent & Fuzzy Systems*, vol. 2, no. 3, pp. 267–278, 1994.

[35] A. G. Di Nuovo, "Missing data analysis with fuzzy c-means: A study of its application in a psychological scenario," *Expert Systems with Applications*, vol. 38, no. 6, pp. 6793–6797, 2011.

[36] K. R. Murphy and B. Myors, *Statistical Power Analysis: A Simple and General Model for Traditional and Modern Hypothesis Tests*, 2nd ed. Taylor & Francis, 2003.