Original Article

An intelligent system based on fuzzy probabilities for medical diagnosis –
a study in aphasia diagnosis ☆

*Majid Moshtagh-Khorasani[*a], Mohammad-R Akbarzadeh-T[a],*

Nader Jahangiri[b], Mehdi Khoobdel[c]

Abstract

BACKGROUND: Aphasia diagnosis is particularly challenging due to the linguistic uncertainty and vagueness, inconsistencies in the definition of aphasic syndromes, large number of measurements with imprecision, natural diversity and subjectivity in test objects as well as in opinions of experts who diagnose the disease.

METHODS: Fuzzy probability is proposed here as the basic framework for handling the uncertainties in medical diagnosis and particularly aphasia diagnosis. To efficiently construct this fuzzy probabilistic mapping, statistical analysis is performed that constructs input membership functions as well as determines an effective set of input features.

RESULTS: Considering the high sensitivity of performance measures to different distribution of testing/training sets, a statistical t-test of significance is applied to compare fuzzy approach results with NN results as well as author's earlier work using fuzzy logic. The proposed fuzzy probability estimator approach clearly provides better diagnosis for both classes of data sets. Specifically, for the first and second type of fuzzy probability classifiers, i.e. spontaneous speech and comprehensive model, P-values are 2.24E-08 and 0.0059, respectively, strongly rejecting the null hypothesis.

CONCLUSIONS: The technique is applied and compared on both comprehensive and spontaneous speech test data for diagnosis of four Aphasia types: Anomic, Broca, Global and Wernicke. Statistical analysis confirms that the proposed approach can significantly improve accuracy using fewer Aphasia features.

KEYWORDS: Aphasia, fuzzy probability, fuzzy logic, medical diagnosis, fuzzy rules.

As the information age matures, it is becoming clear that human knowledge can and should play a pivotal role in handling the complexities/uncertainties of modern decision making systems. A theory is therefore highly desirable that can systematically formulate human knowledge as well as integrate it with other information systems such as mathematical/statistical models to form more intelligent decisions. Merging these two types of information in system design is therefore essential.

To perform such combination, the main problem is how to represent human knowledge into a mathematical formula. Essentially, fuzzy systems are rule-based systems that aim to achieve this representation. The core of a fuzzy system is a knowledge base composed of the so called fuzzy if-then rules. The first step of designing a fuzzy system is consequently to obtain or define a group of linguistic fuzzy if-then rules (domain knowledge) from human experts. The next step is to merge these rules into a single system.

[a] Department of Electrical Engineering, Ferdowsi University of Mashhad, Mashhad, Iran.
[b] Department of Linguistics, Ferdowsi University of Mashhad, Mashhad, Iran.
[c] Health Research Center, Baqiyatallah University of Medical Sciences, Tehran, Iran.
* Corresponding Author
E-mail: moshtagh@kiaeee.org
☆ Part of this research was performed while author was on sabbatical as a visiting scholar at Berkeley Initiative on Soft Computing (BISC), University of California-Berkeley. E-mail: akbarzadeh@ieee.org

Various fuzzy systems apply different principles for this merging. From one perspective, fuzzy systems are nonlinear mappings that map an n-dimensional input space to an m-dimensional output space, where certain properties can be defined for this mathematical relation. But from a different perspective, fuzzy systems are knowledge-based systems extracted from human knowledge in the form of fuzzy if then rules. Converting a knowledge base into a nonlinear mapping is an outstanding privilege of the fuzzy system theory. This conversion makes the ability to use knowledge based systems (fuzzy systems) in different applications such as signal processing, communication systems and medical diagnosis/decision making systems.

Fuzzy systems is one of several ingredients of computational intelligence, or their hybrid combination, such as artificial neural networks,[1,2] fuzzy logic and fuzzy clustering,[3,4] hybrid combinations of artificial neural networks and fuzzy logic,[5] and genetic-fuzzy algorithms,[6,7] which have already been reported in the literature useful for medical diagnosis. Additionally, authors had earlier reported their preliminary success in an application of fuzzy clustering by statistical analysis.[8-10] While there is a great variation to the diagnosis approaches, the above research generally concur on the utility of fuzzy reasoning in dealing with the uncertainty and vagueness, which is typical in the type of diagnosis as well as its symptoms. Furthermore, fuzzy systems are considered attractive due to the transparency of their knowledge base. In other words, the resulting rules and membership functions can be later studied and interpreted by a medical expert for the sake of improvement or training.

In this paper, a different type of fuzzy rule-based structure is proposed for medical diagnosis that is based on fuzzy probability estimators as a key contributor to its inference engine, in contrast to our earlier work[10] that we used a regular product inference engine. The basic premise of fuzzy probabilities here is the inherent imprecision in probabilities arising out of either lack of sufficiently large data set due to prohibitive cost or inadequate measurements arising out of sensory limitations or environmental uncertainty. Fuzzy probabilities are part of a larger class of decision paradigms in which reasoning with imprecise probabilities is addressed.[11] In such a paradigm, fuzzy theory is viewed as a complement rather than a competition to probability theory in its handling of uncertainty.[12] This issue is now addressed more effectively, mainly due to the considerable advances in computing and processing information, making it reasonable to perform computations that are far more complex and less amenable to precise analysis than computations involving precise probabilities. Transition from precise probabilities to imprecise probabilities in probability theory is a form of generalization and as such it enhances the ability of probability theory to deal with real-world problems.[13]

The above paradigm of fuzzy probabilities is proposed here for medical diagnosis in general and aphasia classification in particular. After determining the corresponding fuzzy probabilities for each symptom and aphasia type, a set of rules are constructed that each corresponds to a different aphasia type. Each fuzzy rule represents a mapping between a set of features (antecedents) and a certain type of aphasia (consequent). The set of features is determined by statistical analysis of feature space. By giving a certain set of features, the output of each rule is therefore the diagnosed type of aphasia with a corresponding degree of certainty. Features are sequentially chosen based on how they contribute to improved diagnosis in training space. Unlike our earlier works,[8,9,10] features that are chosen can be still chosen if they contribute more to improved diagnosis and are not eliminated from the pool of features. Here, we also consider reducing the size of feature space by studying the effectiveness of various features in correct aphasia diagnosis. Furthermore, we consider features from the spontaneous speech as well as the comprehensive model.

This paper is organized as follows. The types of aphasia and the database are introduced in Section 2. The proposed fuzzy probability approach is explained in Section 3. Results of the proposed approach and authors' earlier work[10] in this field are then statistically compared against those of artificial neural networks in Section 4. Finally, several conclusions are expressed in Section 5.

1) 2. APHASIA DATA BASE

Aphasia is as an acquired impairment of language processes underlying receptive and expressive modalities caused by damage to certain areas of the brain, which are primarily responsible for the language function; Reasons for such brain damage can be stroke, head injury or cerebral tumors. 100,000 new cases of aphasia are caused by stroke every year in the USA alone.[14,15] Because human brain has large neural networks and various functional cerebral regions, malfunction of these networks can cause different syndromes, and different types of aphasia can be diagnosed. Aphasia diagnosis is a particularly difficult medical diagnostic job because, in addition to the typical complexities of medical diagnosis such as in natural diversity of test objects and expert opinion, there is a significant degree of added complexity by the linguistic uncertainty and vagueness in data, inconsistencies in the definition of aphasic syndromes, large number of interview questions/measurements with imprecision, and hence a natural diversity and subjectivity of opinions of experts who diagnose the disease as well. To reduce diagnostic error in the face of the problem's high complexity, Aphasic diagnosis is performed by testing for a large number of empirically co-occurring set of symptoms, rendering aphasic diagnosis a time consuming and error prone process even among multiple experts. The challenge, therefore, is to determine the most consistent and accurate estimation using fewest number of test questions.

Four major types of aphasia syndromes are listed as follows[15]:

- *Broca's Aphasia* (also called Motor or Expressive Aphasia)
- *Wernicke's Aphasia* (also called Sensory or Receptive Aphasia)
- *Global Aphasia* (also called Total Aphasia)
- *Anomic Aphasia*

While the above major types of aphasia may be clearly defined, their classic taxonomy is polytypic in the symptoms' feature space, i.e. impairments may be part of more than one syndrome.

Furthermore, definition of syndromes is probabilistic rather than crisply defined, and there is great overlap in the boundaries of resulting clusters in the feature space.[15]

Major comprehensive language tests in English speaking countries are the Western Aphasia Battery (WAB) and the Boston Diagnostic Aphasia Examination (BDAE). In German speaking countries the Aachen Aphasia Test (AAT) is the commonly used test battery. Because the AAT was used for evaluation of language function in this database, the test has been described in more detail (see table 1).[16]

The AAT-test profiles of 265 aphasic patients were gathered since 1986. The database consists of some nominal data, e.g. diagnosis of aphasia type, disease, etc., and quantitative data, i.e. the AAT scores. All patients were examined at a time when no change in the aphasia classification or in the lesion size was expected. In the first time after a brain lesion, the symptoms as well as the size of the lesion can change because parts of the disturbed tissue (penumbra) can recover or can be irreversibly destroyed.[15]

From the above 265 AAT scores, 146 patients are diagnosed with one of the four major types of aphasia (Anomic, Broca, Global, and Wernicke), while the rest of the 265 profiles are either other types of aphasia or are undecided profiles. The 146 profiles have been split in two halves here, first half as training set and the second half as testing set.

Database is available online at this address http://fuzzy.iau.dtu.dk/aphasia.nsf.

Table 1. AAT subtests[15]

Code	Test	Score Range
	Spontaneous speech	
P0 (χ_1)	Communicative behavior	0 – 5 [points]
P1 (χ_2)	Articulation and prosody	0 – 5 [points]
P2 (χ_3)	Automatized language	0 – 5 [points]
P3 (χ_4)	Semantic structure	0 – 5 [points]
P4 (χ_5)	Phonologic structure	0 – 5 [points]
P5 (χ_6)	Syntactic structure	0 – 5 [points]
T0 (χ_1)	Token test	0 – 100 [%]
T1- T5 ($\chi_8 - \chi_{12}$)	Token subtests	0 – 10 [points]
N0 (χ_{13})	Repetition	0 – 100 [%]
N1 (χ_{14})	Single phonemes	0 – 30 [points]
N2 (χ_{15})	Monosyllabic nouns	0 – 30 [points]
N3 (χ_{16})	Loan and foreign words	0 – 30 [points]
N4 (χ_{17})	Compound words	0 – 30 [points]
N5 (χ_{18})	Sentences	0 – 30 [points]
C0 (χ_{19})	Written language	0 – 100 [%]
C1 (χ_{20})	Reading aloud	0 – 30 [points]
C2 (χ_{21})	Selecting/combining on dictation	0 – 30 [points]
C3 (χ_{22})	Writing on dictation	0 – 30 [points]
B0 (χ_{23})	Confrontation naming	0 – 100 [%]
B1 (χ_{24})		0 – 30 [points]
B2 (χ_{25})		0 – 30 [points]
B3 (χ_{26})		0 – 30 [points]
B4 (χ_{27})		0 – 30 [points]
V0 (χ_{28})	Comprehension	0 – 100 [%]
V1 (χ_{29})	Auditory for words and sentences	0 – 60 [points]
V2 (χ_{30})	Reading for words and sentences	0 – 60 [points]

Two classifiers are designed:

--The first classifier uses the *spontaneous speech* subtests of the AAT, because these tests are easy to take. Such a classifier is easy to consult, because it needs only a few inputs that are relatively easy to obtain. Inputs to the classifier are selected from the six test scores P0, P1, ..., P5. Diagnosis is the output, which is one of the major aphasia types: Anomic, Broca, Global, Wernicke. The data is processed by the classifier with the highest accuracy, or better to say, with a certain degree of error. The measure of accuracy is the percentage of correct diagnoses when test data set is used as the input of the classifier.

--The second classifier may use all the subtests of the AAT, referred to as the *comprehensive model*. A major challenge is to select a small but effective set of inputs, as is explained in the following section.

The proposed fuzzy probability classifier

The proposed fuzzy probability classifier approach has three general parts:

--First, statistical parameters for each feature of each type of aphasia are separately calculated from the training set (a set of 73 input-output pairs).

--Second, the calculated statistical parameters are used to construct candidate membership functions for the fuzzy rules. A membership function is defined for each corresponding feature of each aphasia type.

--Third, four fuzzy rules (one for each type of aphasia) are constructed iteratively using the above membership functions. Different combinations of atomic fuzzy propositions are iteratively examined/selected to gain the best possible fuzzy rule-base. The rule construction procedure considers the contribution of all available features one at a time in order to

avoid the time consuming alternative of an exhaustive search. The rule base uses a fuzzy probability estimator as inference engine. For calculation *MATLAB* software version 7.0 is used.

A. Calculating Statistical Parameters

For each aphasia type in the training set and for each feature, mean $\bar{x}_i^l$ and standard deviation σ_i^l are calculated as below, where $I = 1, \ldots, 30$ is index of features, $l = 1, \ldots, 4$ is index of aphasia types.

$$\bar{x}_i^l = \frac{1}{n_l} \sum_{j=1}^{n_l} x_{ij}^l, \tag{1}$$

$$\sigma_i^l = \left(\frac{1}{n_l - 1} \sum_{j=1}^{n_l} (x_{ij}^l - \bar{x}_i^l)^2 \right), \tag{2}$$

And where $j = 1, \ldots, n_1$ is index of patient record, and n_1 is the total number of patients for each aphasia type *l*.

B. Membership Function Definition

The above statistical parameter pairs $\left(\bar{x}_i^l, \sigma_i^l \right)$ are then used to define trapezoid fuzzy set A_i^l as below:

$$\mu_{A_i^l}(x_i; a_i^l, b_i^l, c_i^l, d_i^l) = \begin{cases} \frac{x_i - a_i^l}{b_i^l - a_i^l}, x_i \in \left[a_i^l, b_i^l \right) \\ 1, x_i \in \left[b_i^l, c_i^l \right] \\ \frac{x_i - d_i^l}{c_i^l - d_i^l}, x_i \in \left(c_i^l, d_i^l \right] \\ 0, x_i \in R - \left(a_i^l, d_i^l \right) \end{cases} \tag{3}$$

$$a_i^l \le b_i^l \le c_i^l \le d_i^l, a_i^l < d_i^l$$

Where $d_i^l, c_i^l, b_i^l, a_i^l$ are defined as follows:

$$\begin{aligned} a_i^l &= \bar{x}_i^l - 3\sigma_i^l \\ b_i^l &= \bar{x}_i^l - \sigma_i^l \\ c_i^l &= \bar{x}_i^l + \sigma_i^l \\ d_i^l &= \bar{x}_i^l + 3\sigma_i^l \end{aligned} \tag{4}$$

Figure 1 illustrates the histogram of a typical

feature N4 (Compound words) from feature space for patients with the four Aphasia types. As can be seen, it is difficult to determine the optimal general shape for the fuzzy probabilities due to the small available dataset. Authors implemented both trapezoid and Gaussian forms in the earlier stages of their research and confirmed that, as will be shown in the following sections, the trapezoid form provides a good performance as well as reasonable "fit" to the distribution of patient records.

C. Obtaining Fuzzy Rule-base

For each type of aphasia, a general fuzzy rule is defined as:

$Ru^{(l)}$: IF x_1 is A_1^l and … and x_m is A_m^l THEN diagnosis is (aphasia type: Anomic, Broca, Global, Wernicke) with degree u^l

Where degree of truth for each atomic fuzzy proposition $(x_i$ is $A_i^l)$ is defined by $\mu_{A_i^l}(x_i) \ge 0$. A composition of these atomic terms makes up the fuzzy rule's antecedent. The degree of certainty for lth rule, u^l, is calculated by using fuzzy probability estimator.[17] Zadeh [12] defines the probability of a fuzzy event A by

$$P(A) = \int_A f(z)dz = \int_R \mu_A(z)f(z)dz = E[\mu_A(z)], \tag{5}$$

where z is a random variable, $f(z)$ the density of z, $u_A(z)$ the membership function for the fuzzy event A, and $E[.]$ the expected value. On this basis, the probability of the fuzzy output event C is

$$P(C) = E[\mu_C(Y)] \tag{6}$$

An estimator for the probability $P(C)$ is [18]

$$\hat{P} = \frac{1}{n} \sum_{k=1}^{n} \mu_C(Y(k)) \tag{7}$$

which is interpreted as average membership degrees and is implemented as a part of the inference engine as follows:
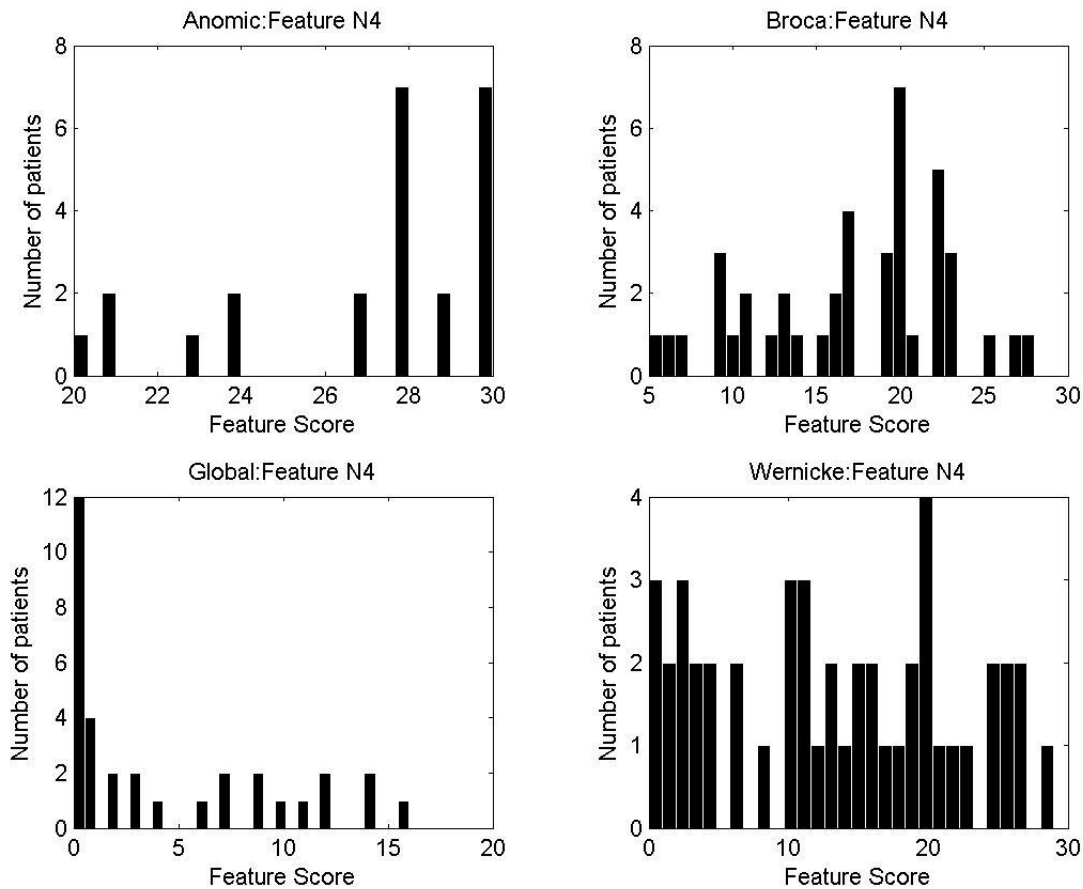
Figure 1. Histogram of a typical feature N4 (Compound words) from feature space which shows the distribution of patients scores. There are respectively 24, 42, 33 and 47 patients for Anomic, Broca, Global and Wernicke.

$$u^l = \frac{\sum_{i=1}^{m} \mu_{A_i^l}(x_i)}{m} \tag{8}$$

Where $m \leq 30$ is the total number of features that are actually used as input in the antecedent part of the rule. During testing stages, u^l is interpreted as the degree of affinity of a given patient to a particular aphasia type l during testing.

Rule generation algorithm aims to gain the highest accuracy with fewest inputs by recursively and exhaustively examining all possible first order combination of inputs. For all Aphasia classes, the same set of features is used in the rule-base. In other words, features are the same in all rules, differing only in their definition of membership functions for different aphasia types (classes). The search algorithm is stated as follows. At start, algorithm begins with only one feature, i.e. one atomic term in the antecedent. All available features and their corresponding atomic fuzzy propositions are substituted one at a time in the rule structure and the feature with the best performance is selected. The algorithm then keeps this feature's corresponding atomic proposition in the rule but doesn't omit it from the set of features; this means that a feature can be reselected in next stages. This is unlike our previous work[10] and also Huang and Gedeon's[20] in which the same search algorithm was used but the selected feature form set of features was omitted and couldn't be reselected after its first selection. In second stage, it appends another atomic term from the set of features to the

rule's antecedent, choosing the feature which produces the best accuracy in combination with previously chosen feature. This process is repeated until no further classification improvement is gained by adding a feature.

In this application, the search algorithm has been started with one feature in each rule. Following the iterative rule generation algorithm, the best accuracy is found by a combination of three features for comprehensive model (feature x_6 has been selected twice in the first and third stages of search algorithm) and two features for spontaneous speech. It is observed that increasing the number of features any further will either has no effect on it or decrease the accuracy.

For instance, below rules are obtained with comprehensive model data.

$Ru^{(1)}$: IF x_6 is A_6^1 and x_{17} is A_{17}^1 and x_6 is A_6^1 and

x_{23} is A_{23}^1 THEN diagnosis is Anomic with degree u^1

$Ru^{(2)}$: IF x_6 is A_6^2 and x_{17} is A_{17}^2 and x_6 is A_6^2 and

x_{23} is A_{23}^2 THEN diagnosis is Broca with degree u^2

$Ru^{(3)}$: IF x_6 is A_6^3 and x_{17} is A_{17}^3 and x_6 is A_6^3 and

x_{23} is A_{23}^3 THEN diagnosis is Global with degree u^3

$Ru^{(4)}$: IF x_6 is A_6^4 and x_{17} is A_{17}^4 and x_6 is A_6^4 and

x_{23} is A_{23}^4 THEN diagnosis is Wernicke with degree u^4

The Matrix $U=[u^l]$ is the output of rules Where u^l represents the degree of belonging to each lth type of aphasia, and is calculated as indicated in (8). Finally, maximum value of u^l is considered as final diagnosis for aphasia type T:

$$T = index(\max \ [u^l]) \qquad (9)$$

Figure 2 shows the fuzzy probability classifier as a diagram.

Results

The above fuzzy classifier approach is applied to two sets of data, first set is data from spontaneous speech interview, while the second is from comprehensive model data as explained in Section 2. For comparative purposes, a back-propagating neural network (NN), with network characteristics similar to those reported in[1,2] has been tested on both sets of data. Moreover, we compared the results with a hierarchical fuzzy rule based approach. The hierarchical rule based approach here is similar to Akbarzadeh-T and Moshtagh-K[10] with a difference on method of feature selection. Since
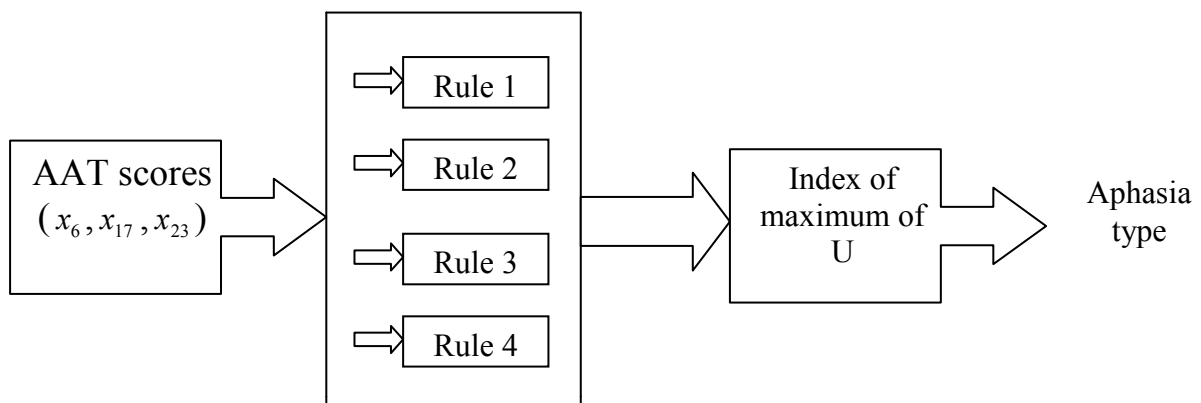

Figure 2. Fuzzy classifier diagram.

feature selection based on non-elimination yields better results, the results of the hierarchical fuzzy rule based approach are also updated in this paper. Furthermore, in order to reach reliable conclusions, a 50 fold cross validation on 50 random distributions of training sets and testing sets are considered. For the NN, the design choices are as follows:

Network topology: A multi-layer perceptron with an input layer, a hidden layer, and an output layer.

Inputs: Test scores P1, P5, N0, and C1.

Outputs: The four classes A, B, G, and W.

Neurons: 4-5-4 neurons, read layer-wise from input layer to output layer, with sigmoid-sigmoid-linear activation functions.

Learning method: Back-propagation with momentum.

Software: *MATLAB* version 7.0

Table 2 shows the features used as inputs for the networks of both classifiers.

The detailed results of applying NN for 50 random distributions are presented in our previous paper,[10] respectively for both *spontaneous speech* and *comprehensive model* feature space.

The results are comparable to those reported earlier in some studies[1,2]; although the results of these studies only mentioned one specific distribution, i.e. the best result for NN with spontaneous speech is 87% and for comprehensive model is 92% correct diagnosis for all four major types of aphasia. The best result that was obtained in Akbarzadeh-T and Moshtagh-K[10] for classification with spontaneous speech is 90.82%, and with comprehensive model is 91.89%.

Next, the proposed fuzzy approach is applied. Table 3 shows the features used as inputs for the obtained rules in both classifiers. A 50-fold cross validation of the fuzzy approach is presented in Table 4 and Table 5, respectively, for first classifier (with *spontaneous speech*) and second classifier (with *comprehensive model*). The best result for first classifier is 96.74% and for second classifier is 94.46% correct diagnosis, as mean of correct diagnosis for the four classes.

The results of applying hierarchical fuzzy rule based approach on aphasia database as explained in Akbarzadeh-T and Moshtagh-K's work,[10] for a 50 fold cross validation of training sets and test sets, are presented in Table 6 and Table 7, respectively, for first classifier (with *spontaneous speech*) and second classifier (with *comprehensive model*). The best result for first classifier is 91.30% and for second classifier is 95.55% correct diagnosis, as mean of correct diagnoses for the four classes.

Table 2. Features for ANN classifiers[1,2]

Selected features for first classifier (spontaneous speech)		Selected features for second classifier (comprehensive model)	
Code	Description	Code	Description
P0 (X_1)	Communicative behavior	P1 (X_2)	Articulation and prosody
P1 (X_2)	Articulation and prosody	P5 (X_6)	Syntactic structure
P2 (X_3)	Automatized language	N0 (X_13)	Repetition
P3 (X_4)	Semantic structure	C1 (X_20)	Reading aloud
P4 (X_5)	Phonologic structure		
P5 (X_6)	Syntactic structure		

Table 3. Selected features for fuzzy classifiers.

Selected features for classifier with spontaneous speech features		Selected features for classifier with comprehensive model features	
Code	Description	Code	Description
P0 (X_1)	Communicative behavior	P5 (X_6)	Syntactic structure (structure of sentences, grammar)
P5 (X_6)	Syntactic structure (structure of sentences, grammar)	N4 (X_17)	
		B0 (X_23)	Compound words
			Confrontation naming

Table 4. Fuzzy probability estimator results for test sets
(first classifier (*spontaneous speech*) features: test scores P0 and P5)

Fold number	Anomic	Broca	Global	Wernicke	Mean (of all classes)	Standard deviation (of all classes)
1	91.67	85.71	100.00	86.96	91.08	6.47
2	91.67	95.24	87.50	82.61	89.25	5.44
3	100.00	95.24	100.00	82.61	94.46	8.21
4	83.33	80.95	87.50	78.26	82.51	3.92
5	91.67	90.48	93.75	73.91	87.45	9.13
6	83.33	95.24	93.75	82.61	88.73	6.69
7	100.00	100.00	100.00	86.96	96.74	6.52
8	75.00	85.71	87.50	82.61	82.71	5.52
9	83.33	90.48	87.50	82.61	85.98	3.69
10	91.67	90.48	87.50	82.61	88.06	4.04
11	91.67	100.00	93.75	82.61	92.01	7.20
12	83.33	85.71	93.75	65.22	82.00	12.05
13	75.00	90.48	81.25	82.61	82.33	6.36
14	75.00	90.48	100.00	86.96	88.11	10.33
15	83.33	95.24	100.00	82.61	90.30	8.68
16	83.33	95.24	87.50	73.91	85.00	8.88
17	75.00	95.24	81.25	78.26	82.44	8.91
18	91.67	85.71	93.75	73.91	86.26	8.91
19	83.33	95.24	87.50	82.61	87.17	5.79
20	75.00	85.71	81.25	82.61	81.14	4.50
21	83.33	95.24	93.75	73.91	86.56	9.96
22	75.00	95.24	93.75	86.96	87.74	9.22
23	83.33	85.71	93.75	82.61	86.35	5.11
24	83.33	90.48	87.50	78.26	84.89	5.30
25	83.33	95.24	87.50	82.61	87.17	5.79
26	83.33	90.48	93.75	69.57	84.28	10.73
27	75.00	80.95	93.75	69.57	79.82	10.39
28	100.00	90.48	93.75	69.57	88.45	13.19
29	91.67	90.48	87.50	78.26	86.98	6.07
30	66.67	90.48	87.50	82.61	81.81	10.61
31	75.00	90.48	93.75	82.61	85.46	8.39
32	91.67	90.48	93.75	82.61	89.63	4.87
33	83.33	95.24	87.50	69.57	83.91	10.76
34	100.00	95.24	87.50	82.61	91.34	7.77
35	66.67	95.24	93.75	82.61	84.57	13.20
36	75.00	85.71	93.75	78.26	83.18	8.35
37	91.67	85.71	81.25	91.30	87.48	4.97
38	83.33	85.71	93.75	73.91	84.18	8.17
39	83.33	85.71	81.25	82.61	83.23	1.87
40	83.33	95.24	100.00	69.57	87.03	13.59
41	91.67	90.48	87.50	86.96	89.15	2.28
42	91.67	85.71	87.50	73.91	84.70	7.61
43	83.33	80.95	93.75	73.91	82.99	8.21
44	91.67	100.00	93.75	78.26	90.92	9.15
45	83.33	85.71	93.75	91.30	88.53	4.83
46	75.00	95.24	87.50	73.91	82.91	10.27
47	83.33	85.71	81.25	78.26	82.14	3.16
48	83.33	90.48	93.75	91.30	89.72	4.48
49	91.67	100.00	87.50	78.26	89.36	9.04
50	83.33	90.48	100.00	78.26	88.02	9.43
Mean (of each column)	84.50	90.86	91.00	79.74	86.52	
standard deviation (of each column)	8.08	5.08	5.53	6.12	3.56	

Table 5. Fuzzy probability estimator results for test sets
(second classifier (*comprehensive model*) features: test scores P5, N4, B0)

Fold number	Anomic	Broca	Global	Wernicke	Mean (of all classes)	Standard deviation (of all classes)
1	83.33	85.71	100.00	82.61	87.91	8.17
2	100.00	90.48	87.50	65.22	85.80	14.72
3	66.67	95.24	100.00	91.30	88.30	14.86
4	75.00	76.19	87.50	91.30	82.50	8.14
5	100.00	90.48	93.75	78.26	90.62	9.14
6	91.67	90.48	93.75	82.61	89.63	4.87
7	100.00	95.24	100.00	82.61	94.46	8.21
8	50.00	76.19	87.50	95.65	77.34	19.89
9	58.33	85.71	87.50	95.65	81.80	16.23
10	100.00	90.48	87.50	78.26	89.06	8.96
11	91.67	100.00	93.75	78.26	90.92	9.15
12	83.33	85.71	93.75	82.61	86.35	5.11
13	75.00	80.95	81.25	86.96	81.04	4.88
14	75.00	90.48	100.00	86.96	88.11	10.33
15	83.33	95.24	100.00	78.26	89.21	10.12
16	100.00	95.24	87.50	60.87	85.90	17.47
17	75.00	95.24	81.25	78.26	82.44	8.91
18	83.33	76.19	93.75	95.65	87.23	9.14
19	58.33	85.71	87.50	95.65	81.80	16.23
20	75.00	85.71	81.25	82.61	81.14	4.50
21	91.67	85.71	93.75	69.57	85.17	10.95
22	75.00	95.24	93.75	91.30	88.82	9.36
23	83.33	85.71	93.75	91.30	88.53	4.83
24	100.00	90.48	87.50	78.26	89.06	8.96
25	83.33	95.24	87.50	86.96	88.26	5.01
26	91.67	90.48	93.75	69.57	86.36	11.28
27	83.33	80.95	93.75	65.22	80.81	11.79
28	100.00	90.48	93.75	65.22	87.36	15.28
29	100.00	90.48	87.50	65.22	85.80	14.72
30	66.67	80.95	87.50	91.30	81.61	10.84
31	75.00	80.95	93.75	86.96	84.16	8.04
32	100.00	90.48	93.75	86.96	92.80	5.55
33	83.33	95.24	87.50	69.57	83.91	10.76
34	100.00	95.24	87.50	73.91	89.16	11.40
35	66.67	95.24	93.75	78.26	83.48	13.58
36	100.00	85.71	93.75	65.22	86.17	15.14
37	75.00	85.71	81.25	95.65	84.40	8.69
38	91.67	76.19	93.75	78.26	84.97	9.02
39	50.00	85.71	81.25	100.00	79.24	21.07
40	91.67	95.24	100.00	60.87	86.94	17.71
41	83.33	90.48	87.50	82.61	85.98	3.69
42	100.00	85.71	87.50	69.57	85.69	12.49
43	100.00	80.95	93.75	73.91	87.15	11.86
44	91.67	100.00	93.75	69.57	88.75	13.27
45	58.33	85.71	93.75	100.00	84.45	18.37
46	66.67	95.24	87.50	82.61	83.00	12.07
47	91.67	85.71	81.25	65.22	80.96	11.33
48	100.00	90.48	93.75	86.96	92.80	5.55
49	100.00	100.00	87.50	78.26	91.44	10.58
50	91.67	76.19	100.00	82.61	87.62	10.41
Mean (of each column)	84.33	88.57	91.00	80.61	86.13	
standard deviation (of each column)	14.64	6.60	5.53	10.77	3.70	

Table 6. Fuzzy hierarchical method results for test sets
(first classifier (*spontaneous speech*) features: test scores P0 and P5)

Fold number	Anomic	Broca	Global	Wernicke	Mean (of all classes)	Standard deviation (of all classes)
1	100.00	85.71	100.00	73.91	89.91	12.61
2	91.67	95.24	87.50	52.17	81.64	19.90
3	100.00	85.71	100.00	56.52	85.56	20.50
4	91.67	80.95	87.50	69.57	82.42	9.64
5	100.00	90.48	93.75	52.17	84.10	21.65
6	83.33	95.24	93.75	69.57	85.47	11.85
7	100.00	100.00	100.00	65.22	91.30	17.39
8	75.00	85.71	87.50	82.61	82.71	5.52
9	83.33	90.48	87.50	78.26	84.89	5.30
10	91.67	90.48	87.50	78.26	86.98	6.07
11	91.67	100.00	93.75	56.52	85.48	19.63
12	83.33	85.71	93.75	69.57	83.09	10.06
13	75.00	90.48	81.25	82.61	82.33	6.36
14	75.00	90.48	100.00	86.96	88.11	10.33
15	91.67	95.24	100.00	56.52	85.86	19.85
16	75.00	95.24	87.50	73.91	82.91	10.27
17	75.00	95.24	81.25	78.26	82.44	8.91
18	100.00	85.71	93.75	73.91	88.34	11.26
19	83.33	95.24	87.50	78.26	86.08	7.18
20	75.00	85.71	81.25	82.61	81.14	4.50
21	83.33	95.24	93.75	65.22	84.38	13.83
22	75.00	95.24	93.75	86.96	87.74	9.22
23	83.33	85.71	93.75	82.61	86.35	5.11
24	83.33	90.48	87.50	78.26	84.89	5.30
25	91.67	95.24	87.50	78.26	88.17	7.32
26	83.33	90.48	93.75	65.22	83.19	12.75
27	83.33	80.95	93.75	56.52	78.64	15.76
28	100.00	90.48	93.75	60.87	86.27	17.39
29	91.67	90.48	87.50	69.57	84.80	10.31
30	66.67	90.48	87.50	86.96	82.90	10.93
31	75.00	90.48	93.75	82.61	85.46	8.39
32	91.67	90.48	93.75	82.61	89.63	4.87
33	83.33	95.24	87.50	69.57	83.91	10.76
34	100.00	95.24	87.50	73.91	89.16	11.40
35	66.67	95.24	93.75	82.61	84.57	13.20
36	83.33	71.43	100.00	65.22	79.99	15.31
37	91.67	85.71	81.25	82.61	85.31	4.63
38	83.33	71.43	100.00	69.57	81.08	14.01
39	91.67	85.71	81.25	78.26	84.22	5.83
40	83.33	80.95	100.00	69.57	83.46	12.56
41	100.00	90.48	87.50	78.26	89.06	8.96
42	91.67	85.71	87.50	65.22	82.52	11.80
43	83.33	66.67	100.00	69.57	79.89	15.25
44	91.67	100.00	93.75	73.91	89.83	11.19
45	83.33	85.71	93.75	91.30	88.53	4.83
46	75.00	95.24	87.50	60.87	79.65	15.04
47	91.67	85.71	81.25	65.22	80.96	11.33
48	91.67	90.48	93.75	82.61	89.63	4.87
49	100.00	100.00	87.50	73.91	90.35	12.44
50	91.67	71.43	100.00	69.57	83.17	15.04
Mean (of each column)	86.67	89.14	91.38	72.70	84.97	
standard deviation (of each column)	9.22	7.52	5.91	9.78	3.16	

Table 7. Fuzzy hierarchical method results for test sets
(second classifier (*comprehensive model*) features: test scores P5, N4, B0)

Fold number	Anomic	Broca	Global	Wernicke	Mean (of all classes)	Standard deviation (of all classes)
1	66.67	85.71	100.00	82.61	83.75	13.68
2	91.67	90.48	87.50	69.57	84.80	10.31
3	75.00	90.48	100.00	91.30	89.20	10.40
4	83.33	76.19	87.50	91.30	84.58	6.47
5	91.67	80.95	93.75	78.26	86.16	7.69
6	75.00	90.48	93.75	91.30	87.63	8.54
7	100.00	95.24	100.00	86.96	95.55	6.15
8	50.00	85.71	87.50	91.30	78.63	19.23
9	58.33	95.24	87.50	91.30	83.09	16.81
10	100.00	95.24	87.50	82.61	91.34	7.77
11	66.67	95.24	93.75	82.61	84.57	13.20
12	83.33	80.95	93.75	82.61	85.16	5.81
13	75.00	90.48	75.00	86.96	81.86	8.05
14	75.00	90.48	100.00	86.96	88.11	10.33
15	83.33	85.71	100.00	82.61	87.91	8.17
16	66.67	95.24	87.50	73.91	80.83	12.92
17	75.00	95.24	81.25	82.61	83.52	8.48
18	83.33	71.43	93.75	91.30	84.95	10.05
19	66.67	90.48	87.50	95.65	85.07	12.73
20	50.00	85.71	75.00	95.65	76.59	19.63
21	100.00	80.95	93.75	69.57	86.07	13.56
22	66.67	95.24	93.75	91.30	86.74	13.48
23	83.33	80.95	93.75	95.65	88.42	7.36
24	83.33	90.48	87.50	82.61	85.98	3.69
25	66.67	85.71	87.50	95.65	83.88	12.27
26	83.33	90.48	93.75	69.57	84.28	10.73
27	83.33	61.90	93.75	65.22	76.05	15.10
28	66.67	85.71	93.75	82.61	82.18	11.36
29	66.67	90.48	87.50	82.61	81.81	10.61
30	66.67	90.48	87.50	82.61	81.81	10.61
31	75.00	90.48	93.75	86.96	86.55	8.18
32	91.67	85.71	93.75	86.96	89.52	3.81
33	58.33	100.00	87.50	69.57	78.85	18.52
34	83.33	95.24	87.50	82.61	87.17	5.79
35	66.67	95.24	93.75	82.61	84.57	13.20
36	75.00	80.95	100.00	82.61	84.64	10.75
37	75.00	80.95	81.25	91.30	82.13	6.76
38	83.33	80.95	100.00	82.61	86.72	8.91
39	50.00	85.71	68.75	95.65	75.03	20.04
40	83.33	80.95	100.00	69.57	83.46	12.56
41	83.33	90.48	87.50	82.61	85.98	3.69
42	100.00	76.19	87.50	82.61	86.57	10.08
43	83.33	76.19	93.75	78.26	82.88	7.84
44	58.33	100.00	93.75	82.61	83.67	18.36
45	66.67	76.19	93.75	100.00	84.15	15.41
46	75.00	95.24	87.50	82.61	85.09	8.50
47	91.67	85.71	75.00	73.91	81.57	8.58
48	91.67	90.48	93.75	95.65	92.89	2.29
49	100.00	100.00	87.50	82.61	92.53	8.86
50	100.00	71.43	100.00	86.96	89.60	13.58
Mean (of each column)	77.50	87.14	90.63	84.26	84.88	
Standard deviation (of each column)	13.70	8.12	7.17	8.14	4.06	

Finally, we compare the proposed fuzzy probability estimator approach here with neural networks and our previous fuzzy hierarchical approach as explained in Akbarzadeh-T and Moshtagh-K's work.[10] We should indicate that, for a fair comparison, we applied the search algorithm which is explained in section 3 to our previous fuzzy hierarchical approach as well.[10] Considering the high sensitivity of performance measures to different distribution of testing/training sets, a statistical t-test of significance is applied here to compare both

fuzzy approaches results with NN results. The proposed fuzzy probability estimator approach clearly provides better diagnosis for both classes of data sets. Specifically, for the first and second type of fuzzy probability classifiers, i.e. *spontaneous speech and comprehensive model*, p values are 2.24E-08 and 0.0059, respectively, strongly rejecting the null hypothesis. While the change in feature selection improved performance of our previous fuzzy hierarchical approach,[10] the significance of statistical results did not change. It can also be deduced that both hierarchical fuzzy and fuzzy probabilities perform comparatively better when there is more uncertainty, i.e. fewer test data as in *spontaneous speech model*. It should be also noted that the fuzzy probability estimator approach

is able to reach this level of performance using significantly fewer number of features than NN, but the number of measurements are the same in both fuzzy approaches (see table 8). It should also be mentioned that the fuzzy probability estimator calculates quicker than neural networks and fuzzy hierarchical method. The total execution time for one training/testing simulation is 12.53 seconds for the neural networks and 2.60 seconds for the fuzzy probability estimator approach, also in our previous work it was 4.35 seconds for hierarchical fuzzy rule based method. Fuzzy probability estimator approach is about five times quicker than neural networks and about 1.7 times quicker than our previous work, the hierarchical fuzzy rule based method.

Table 8. Comparing fuzzy approaches with neural networks

	First classifier (using spontaneous speech data)			Second classifier (using comprehensive data)		
	Fuzzy probability estimator approach	Fuzzy hierarchical rule based approach	Neural networks	Fuzzy probability estimator approach	Fuzzy hierarchical rule based approach	Neural networks
Maximum gained accuracy	96.74	91.30	90.82	94.46	95.55	91.89
Average for 50 fold cross validation	86.52	84.97	80.23	86.13	84.88	83.47
Number of used features	2	2	6	3	3	4
P values	2.24E-08	1.88E-06	---	0.0059	0.1257	---

Conclusion

Fuzzy probability estimator can be considered as a type of perception based probability theory. In perception based probability theory everything and especially truth and possibility is or is allowed to be a matter of degree. This is in contrast to standard probability theory where only likelihood is a matter of degree. In this paper, we proposed the use of fuzzy probabilities for better medical classification and implemented it on Aphasia diagnosis. While Aphasia presents a particularly challenging medical diagnosis due to various complexities such as in linguistic uncertainty and vagueness in measurement data as well as inconsistencies in the definition of aphasic syndromes them-

selves, this paper concludes that fuzzy probabilities are an appropriate decision making paradigm to reach good accuracy in its prediction and classification.

The proposed fuzzy probability approach begins with statistical analysis of the input feature space for defining the membership functions and reducing the size of the feature space. The fuzzy system performs the diagnosis by detecting the affinity of a new data set to the previously trained clusters of data. The clusters are defined by membership functions that are chosen after statistical analysis of the feature space. The proposed method is then applied to the aphasia database at AAT and results are compared with those of previously

reported neural networks. Statistical analysis reveals that the proposed fuzzy approach has a better performance for accuracy while also using fewer features as compared with artificial neural networks. In fact, due to the high level of conflict and vagueness in the data set, it is observed that using more inputs will not necessarily produce better accuracy, while the choice of features will significantly influence classification performance. It is also observed that proposed fuzzy probability estimator has a better performance than our previously proposed hierarchical fuzzy rule based approach.

Conflict of Interest
Authors have no conflicts of interest.

Authors' Contributions
MMKh carried out the design and coordinated the study, participated in all of different parts of this study and prepared the manuscript.

MRAT provide assistance in the design of the study and participated in manuscript preparation.

NJ and MKh provides assistance for all other non engineering parts of the study.

All authors have read and approved the content of the manuscript.

References
1. Axer H, Jantzen J, Keyserlingk DGV, Berks G. Aphasia Classification Using Neural Networks. European Symposium on Intelligent Techniques Aachen 2000;111-5.
2. Axer H, Jantzen J, Graf VK. An aphasia database on the internet: a model for computer-assisted analysis in aphasiology. Brain Lang 2000; 75(3):390-8.
3. Berks G, Keyserlingk DGV, Jantzen J, Dotoli M, Axer H. Fuzzy clustering - A versatile mean to explore medical databases. ESIT 2000;14-5.
4. Castellano G, Fanelli AM, Mencar C. A fuzzy clustering approach for mining diagnostic rules. IEEE 2003.
5. Axer H, Jantzen J, Berks G, Keyserlingk DGV. Diagnosis of aphasia using neural and fuzzy techniques. Symposium on Computational Intelligence and Learning 2000.
6. Tsakonas A, Dounias G, Jantzen J, Axer H, Bjerregaard B, von Keyserlingk DG. Evolving rule-based systems in two medical domains using genetic programming. Artif Intell Med 2004; 32(3):195-216.
7. Tsakonas A. A comparison of classification accuracy of four genetic programming-evolved intelligent structures. Information Sciences 2006; 176(6):691-724.
8. Moshtagh-K M, Akbarzadeh MR. Diagnosis and classification of aphasia using fuzzy techniques. Ferdowsi university of Mashhad; 2009.
9. Investigating different features of aphasia and using statistical analysis to obtain fuzzy rules to diagnose different types of aphasia. Tehran: University of Khajeh Nasir Tousi; 2005.
10. Akbarzadeh T, Moshtagh-Khorasani M. A hierarchical fuzzy rule-based approach to aphasia diagnosis. J Biomed Inform 2007; 40(5):465-75.
11. Walley P. Statistical reasoning with imprecise probabilities (monographs on statistics and applied probability). London: Chapman & Hall; 1990.
12. Zadeh LA. Probability Measures of Fuzzy Events. Journal of Mathematical Analysis and Applications 1968; 23(2):421-7.
13. Zadeh LA. Toward a perception-based theory of probabilistic reasoning with imprecise probabilities. Journal of statistical planning and inference 2002; 105(1):233-64.
14. Axer H, Jantzen J, Berks G, Südfeld G, Keyserlingk DGV. The aphasia database on the web: description of a model for problems of classification in medicine. European Symposium on Intelligent Techniques 2000;104-10.

15. Huber W, Poeck K, Willmes K. The aachen aphasia test. Adv Neurol 1984; 42:291-303.
16. Angelika K , Heike T. Data-based fuzzy rule test for fuzzy modelling. Fuzzy Sets and Systems 2001; 123(3):343-58.
17. Bandemer H, Gottwald S. Einfuehrung in fuzzy-methoden. 3th ed. Wiley-VCH; 1993.
18. Dubois DJ. Fuzzy sets and systems: theory and applications. 1st ed. Academic Press; 1980.