ORIGINAL ARTICLE

A neuro-fuzzy approach for the diagnosis of depression

CrossMark

Subhagata Chattopadhyay *

Department of Computer Science and Engineering, National Institute of Science and Technology, Berhampur 761008, Odisha, India

Abstract   Depression is considered to be a chronic mood disorder. This paper attempts to mathematically model how psychiatrists clinically perceive the symptoms and then diagnose depression states. According to Diagnostic and Statistical Manual (DSM)-IV-TR, fourteen symptoms of adult depression have been considered. The load of each symptom and the corresponding severity of depression are measured by the psychiatrists (i.e. the domain experts). Using the Principal Component Analysis (PCA) out of fourteen symptoms (as features) seven has been extracted as latent factors. Using these features as inputs, a hybrid system consisting of Mamdani's Fuzzy logic controller (FLC) on a Feed Forward Multilayer Neural Net (FFMNN) has been developed. The output of the hybrid system was tuned by a back propagation (BPNN) algorithm. Finally, the model is validated using 302 real-world adult depression cases and 50 controls (i.e. normal population). The study concludes that the hybrid controller can diagnose and grade depression with an average accuracy of 95.50%. Finally, it is compared with the accuracies obtained by other techniques.

1. Introduction

Medical decision making as a whole is a complex process due to handling of higher dimensional, raw, and subjective clinical data. Correct decision requires an orchestration of clinician's high level perceptions and intuitions toward understanding a disease process. Correctness of the diagnosis depends on the number of times symptoms are matched with the representatives of the reference diseases (which are called as 'classical' cases). Manual diagnosis is often individualized and so as the presentation of an illness. Therefore, the appropriateness of the term classical becomes relative in clinical medicine. Applications of higher data mining technique(s) and concepts of computational intelligence have tremendous research scopes in modeling the process of clinical diagnosis due to its operational resemblance. It also invites an opportunity for the cross-disciplinary research.

1.1. The manual process of medical decision making [1]

In medical decision making there are basically two major phases. The first phase is known as the *differential diagnosis* (*DD*) and the second phase is termed as *provisional* or *final diagnosis* (*PD*). During *DD*, patients' history and sign-

* Tel.: +91 680 24924321.
E-mail address: subhagatachatterjee@yahoo.com.
Peer review under responsibility of King Saud University.

symptoms are perceived by the clinicians as the inputs. This information, in turn, is processed or analyzed according to their medical knowledge-base and experience to arrive into a diagnosis of overlapping look-alike diseases. According to the perception, medical doctors possibly assign some arbitrary weights to symptoms to define its level of representation on the overall disease load. Through multiple clinical assessments (i.e., iterations in computer science term), such weights are repeatedly evaluated and if required, updated. In clinical medicine there might be chances where different diseases present with similar patterns and vice versa. Hence, based on the inputs, doctors try to match the symptom patterns with classical case of each of the possible diseases (obtained during *DD*) by measuring the similarity. Based on the degree of similarity, they then rank the possible diseases and management strategies proposed aiming at the top most ranked disease. Results of the investigations and the preliminary treatment are closely matched with all plausible diseases (obtained during *DD*) and finally best matched disease is diagnosed. This process is referred to as *PD*. However, in reality, the process is not so trivial and straight forward. There are several iterations, needed to arrive from *DD* to *PD*. Fig. 1 shows the schematic representation of clinical diagnosis process.

1.2. Complexities in depression diagnosis

Psychiatry is the most complex domain in Medical sciences. Psychiatric diseases are not directly measurable due to vague symptomatic presentations. Results of investigations and treatment are manually correlated with the course of morbidity and such correlations could lead to biased decision-making.

The onset of depression is often unrecognizable. It may originate from several non-psychiatric illnesses, for e.g., cancers [2] or chronic joint pains [3] and so forth. It is often noticed that different diseases present with similar types of symptoms and vice versa. Also behind a set of symptoms there is a possibility of multiple diseases evident in case of depression with mania [4]. Therefore, the process of depression screening (*DD*) to its formal diagnosis (*PD*) is quite lengthy and much

complex in clinical psychiatry. Moreover, manual diagnosis might be error-prone, slow, and highly dependent on the individual skill sets.

Given these issues, the *objective* of this paper is to mathematically model the clinical decision making process in psychiatry using Soft computing and not to develop a full-fledged expert tool. It is exclusively an application-based research in mental health and cross-disciplinary in nature. Unfortunately, the number of such cross-disciplinary research is small and hence, this is an attempt to add to the existing literature of '*mental health-informatics*' research. The way existing methods are realized and implemented in a model is from the scratch and hence proposed to be new in the clinical study of depression. However, this is a future scope and may be considered as the next step toward automating the depression diagnosis.

The *contribution* of this paper lies on the mathematical modeling of the manual depression diagnosis process. Furthermore, within the model, research challenges are:

(i) Mathematically managing the overlapping grades (expressed in qualitative terms, e.g., '*mild*', '*moderate*', '*severe*' etc.) with the concept of fuzzy sets.
(ii) Technically handling large rule base in Mamdani's fuzzy logic controller, and
(iii) Making a crisp final decision with the feed-forward backpropagation part of the model.

Rest of the paper is organized as follows. Section 2 discusses various applications of fuzzy set, fuzzy logic and neural network approaches in mental health research. Section 3 details the methodology of the study. Experimental results are shown and discussed in Section 4. Finally, Section 5 discusses the key outcomes and concludes the paper.

2. Current literature

Soft computing techniques, such as Fuzzy sets and Fuzzy logic, Artificial Neural Networks (ANN), and Genetic Algorithms (GA) are useful in handling the uncertainty and vague-
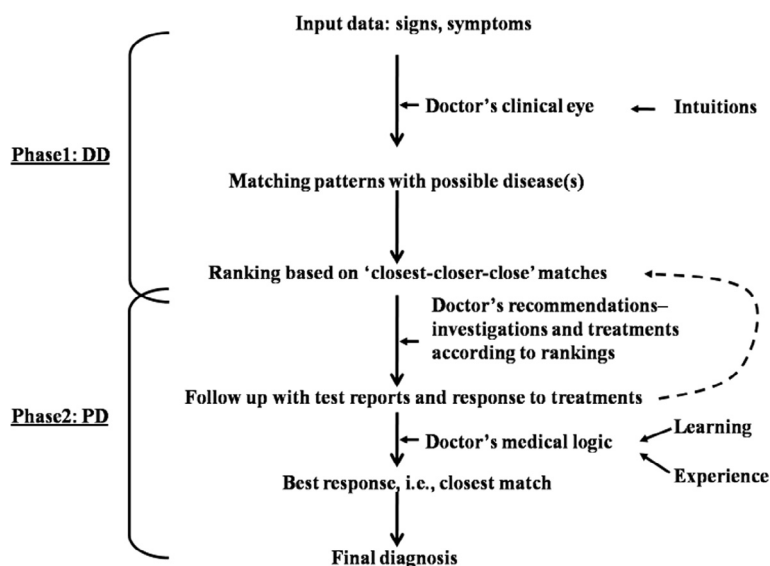

Fig. 1 The process of clinical diagnosis [1].

ness associated with the real-world data. To complement each other's merits, hybrid techniques are now being popularly used in several domains of health sciences.

In psychiatry, applications of soft computing techniques are more suitable as it can handle the issues, such as non-linearity, multidimensionality, and vagueness, compared to hard computing, such as probabilistic approaches. However, due to the lack of availability of the combined domain expertise (e.g., expertise in both psychiatry and computer science) the number of research attempts is less. Computer scientists find difficulties in understanding and modeling the disease processes, its representations, and evaluation. On the other hand, psychiatrists find it difficult to realize the merits of algorithms, its coding and necessary implementations. Despite such hindrances, however, attempts have been made by the researchers since decades in screening and diagnosing psychiatric diseases using soft computing techniques. Some recent studies have been described in the following section.

2.1. Applications of fuzzy sets and fuzzy logic

Researchers attempted to use hard computing techniques, such as heuristic and probabilistic algorithms in diagnosing psychiatric diseases [5,6]. But, they found that such algorithms are too rigid to actually evaluate the morbidity. On the other hand, the merit of fuzzy set and fuzzy logic lies on its capability to handle non-discrete qualitative inputs and analyze like human beings. Thus, these methods have been tried to screen and predict real-world adult psychiatric cases [7–11]. In these studies the authors at first statistically modeled seven psychotic diseases using the Brief psychiatric rating scale version-F2 (BPRS-F2) and Plackett–Burman design (PBD) of experiments. Then using multiple regressions, the authors identified significant predictors behind each of these illnesses. From regression equations, more experimental data had been generated. These were, in turn, clustered using Fuzzy C means (FCM) and Entropy-based fuzzy clustering (EFC) techniques (with its three possible extensions). Centroid properties of the clusters were then used for developing the corresponding FLCs (Takagi–Sugeno's type). Outputs of those FLCs were then optimized by a binary coded Genetic algorithm (GA). It was evident that FCM-based FLC predicted psychoses with more accuracy compared to EFC and its extensions.

In another study [12] a hierarchical decision-tree-based fuzzy model was proposed to screen and grade fifty cases of premenstrual syndrome (PMS), which falls under the cross-domain of gynecology and psychiatry. The authors used firing strength (FS) of fuzzy rules to evaluate the respective PMS grades. Rules with higher FS had been considered significant and had been used in designing the diagnostic rule base of PMS. The study concludes that appropriate rule base could be constructed to diagnose complex cross-domain diseases, which often lack standardized diagnostic rule sets.

Fuzzy set theory was applied in the classification of depression [13]. The authors adopted Beck Depression Inventory (BDI)-II to capture a group of young depressives. On this data, the authors attempted FCM and probabilistic based classifications (Wald's and K-means) to classify the severity of depression. After comparison, the study observed that FCM was the best technique. The authors concluded that FCM more accu-

rately realized the data structure than Wald's and K-means models.

2.2. Applications of neural network and fuzzy-neural hybrids

ANN had been used to examine the correlation of recurrent suicide attempts and the overall self-harm history of Taiwanese soldiers [14]. They identified totally ten factors which might have influence on suicidal risks. Network training with these factors was conducted using the Radial basis function (RBF). The trained model showed about 82% sensitivity and 86% specificity.

The strength of RBF and a back propagation (BPNN) algorithm had been used to develop a decision support system (DSS) in assisting rural doctors to diagnose psychiatric illnesses [15,16]. The heart of the said system is the application of BPNN as the feature extraction algorithm. The authors concluded that the proposed DSS was able to diagnose psychiatric problems with 98.75% accuracy.

In a recent study, a group of researchers used a BPNN in classifying a set of real-world adult depression cases based on its severity/grade [17,18]. In order to do so, the authors at first collected fifteen symptoms of depression under four major constructs, such as 'motivational', 'emotional', 'cognitive', and 'vegetative and physical'. Information about symptoms was then captured by a questionnaire. Answer to each question was then quantified and the corresponding grade was assigned as 'mild'–'moderate'–'severe' under a three-point scale. The information was then fed to the network for training (incremental mode). Finally the study concluded that by the BPNN approach the network was able to classify depression cases with 89% average accuracy.

Performance of a neuro-fuzzy model in predicting weight changes in chronic schizophrenic patients exposed to atypical or typical antipsychotics for more than a couple of years had been examined [19]. The authors used FuzzyTECH 5.54 software package to generate the rule base. The model was able to predict 93% of weight changes.

To differentiate between epileptic and non-epileptic events various fuzzy arithmetic operations have been investigated [20]. A total of 244 patients were studied using the NEF-CLASS (NEuro Fuzzy CLASSification) architecture with the Artificial Neural Network and Back propagation algorithm (ANNB). Finally they were able to detect the differences with a sensitivity of 85% sensitivity, while 95.65% was achieved by ANNB.

3. Methodology

The paper aims to study a fuzzy-neural hybrid model to diagnose the severity of real-world depression cases. It focuses on the overlapping areas of diagnosis, i.e., the DD and attempts to make more discrete diagnosis, i.e., the PD. The steps of the study are as follows,

1. Depression data collection from hospitals and its processing.
2. PCA [21] has been chosen to extract the significant features (i.e. symptoms) by reducing the number of symptoms (i.e. the data dimension). This step mimics the way medical doc-

tors diagnose a disease (in this case it is depression) by extracting significant features of the illness among many such.

3. Creating input vector matrix with significant symptoms.

4. Fuzzy-neural hybrid model development using Mamdani's fuzzy logic controller [22] mapped onto a FFMNN [23], and it's tuning with a BPNN algorithm [24]. It may be noted that while training and testing the controller, the sample is divided randomly into two parts – 214 (i.e. approximately 70% of the sample) for training and remaining 88 (i.e., approximately 30%) for testing the hybrid model using a 10-fold cross validation [25]. Sugeno's controller has not been chosen because of difficulty involved in mapping of the output polynomial function with the clinical decision making.

5. Testing the performances of the proposed system on a set of real-world depression cases.

3.1. Data collection and normalization

Scores of fourteen symptoms of depression (see Table 1) have been considered in this study according to the DSM-IV-TR guidelines [26]. These symptoms denote independent factors. Corresponding grades of depression denote the dependent factor. A group of specialist doctors (mean experience of 10.4 years) consisted of five senior psychiatrists assign weights [0,1] of each of these factors/symptoms and the grades (mild, moderate, and severe) of depression. The weighted symptoms are nothing but the reflection of the load of the disease. Such 'symptom-grade' relations have been established on the basis of the clinical experience and knowledge base of the psychiatrists.

Data of total 302 adult 'psychotic depression' cases have been collected from two hospital sources in India over the period of eight months. The grades of these cases are known. For further validation of the model, a control group (no depression) has been considered. Appropriate ethical measures have been taken for preserving data privacies, i.e., the information of subjects, doctors, and hospitals. Atypical symptoms, patients aged below 19 years and above 50 years have been excluded from the study to avoid accidental inclusion of child-

hood depression, childhood schizophrenia and the early symptoms of dementia (which occurs at the older age).

The information of each case is now presented as a $N_i {}^* M_j$ matrix ('i' varies from 1 to 302 and 'j' varies from 1 to 14), which is fed to the controller as the input. Among 302 total depression cases, 214 cases (i.e., 70%) are used for training the proposed hybrid system and the remaining 30% are kept for testing its performance. It is important to note that a 10-fold cross validation has been performed to partition training and test data.

As data are raw and secondary in nature (constructed by a group of medical doctors), it has been preprocessed by checking for noise, such as data redundancy, missing values etc. Finally data have been normalized [0,1] column-wise using the max–min normalization technique. Then, data reliability has been measured by computing Cronbach's alpha (α) on the whole dataset using 95% confidence interval [27].

3.2. PCA

PCA is one useful statistical technique used to extract the hidden features from multidimensional data. Also it reduces the dimension by pruning unwanted attributes. The objective behind such dimensionality reduction of the input space is to make it lossless by preserving the original information as much as possible. Another useful merit of PCA is that, it is independent of target data and therefore learns without supervision. The steps of the PCA method are Data collection; Subtraction of *mean* of each dimension to produce a dataset having null mean; Calculation of covariance matrix; Calculation of *eigenvectors* and *eigenvalues* of the covariance matrix; Choosing components and forming a feature vector; and finally Deriving new dataset. It is important to mention that eigenvectors are sorted in descending order with the highest *eigenvalue* is the principal component (PC) of the dataset. These sorted values refer to the significant components and renders flexibility so that the components with lower *eigenvalues* could be omitted to reduce the data dimension than that of the original dataset. A new feature vector is then nothing but the matrix of the PCs that one finds significant. The final data are composed by first transposing the vector and then multiplying it on the left over mean-adjusted dataset, also transposed. PCA has been performed using MINITAB-14. In this study, symptoms with eigenvalues greater than 1 are found significant based on the cross validation. It is worth noting that α has been measured once again after the PCA to check the reliability of the data with principal components as the significant symptoms.

3.3. The neuro-fuzzy model

The study proposes and develops a fuzzy-neural hybrid approach in diagnosing the severity of depression. In doing so, Mamdani's fuzzy logic controller (FLC) has been proposed. Asymmetric triangular membership function distributions (TMFD) are considered for mapping inputs and the output, i.e., the symptoms and the grade of depression, respectively. Inputs and the output are arbitrarily assigned class labels – 'mild ('m')', 'moderate ('M')', and 'severe ('s')' with some numeric range. Its working steps are then mapped onto a FFMNN (see Fig. 2), where the *first* layer is the 'input' (I) layer; *second* layer represents hidden layer-1 (HL-1), which is

Table 1 Depression Symptoms (Dsm-Iv) And The Corresponding Abbreviations.

	Symptoms	Abbreviations
1	Feeling sad	S
2	Loss of pleasure	P
3	Weight loss	W1
4	Insomnia	I1
5	Hypersomnia	H
6	Loss of appetite	A
7	Psychomotor agitation	P1
8	Psychomotor retardation	P2
9	Loss of energy	E
10	Feeling of worthlessness	W2
11	Lack of thinking	T
12	Indecisiveness	I2
13	Recurrent thought of death	D
14	Impaired social/occupational functioning	SO

meant for the 'fuzzification' tasks i.e., calculation of membership (μ) values using TMFD (see Eq. (1)); in the *third* layer (i.e., HL-2) 'AND operations' are performed; respective 'firing strength (FS)' (refer to Eq. (2)) has been computed in the *fourth* layer (i.e., HL-3) according to the fired or activated fuzzy prules (i.e., production rules), which are obtained from the domain experts; in the *fifth* layer, Output (O) is calculated after the 'OR operation' (see Eq. (3)) using the sum of the center of the firing areas (expressed with Eq. (4)). The predicted output is then compared with the target output for each case and Mean Squared Error (MSE) is then computed. Finally, weights are updated during backpropagating with the error and the training is said to be completed when the minimum MSE value is obtained. The equations are explained as follows to understand the working principle of the controller.

$$\mu_x(A) = \Delta(x; a, b, c) \tag{1}$$

In this equation, 'μ_x' is the membership of the crisp input 'x' to the triangular fuzzy set 'A', 'ac' is the base of the triangle, and 'b' is its midpoint.

$$FS = \min(\mu_{x1}(A1), \mu_{x2}(A2), \dots \mu_{xn}(An)) \tag{2}$$

Eq. (2) expresses the method, by which firing strengths (FS) of a particular rule has been computed. In this equation, 'μ_{x1}, μ_{x2}, ..., μ_{xn}' are the membership values of the crisp inputs to the triangular fuzzy sets '$A1$, $A2$,..., An'.

$$\mu_C(f) = \max(\mu_{C_1}(f), \mu_{C_2}(f), \dots \mu_{C_n}(f)) \tag{3}$$

An example of the combined control action 'C' of the fuzzy controller (f) i.e. $\mu_C(f)$ is given by Eq. (3). Here, *OR* operation is performed.

$$O_f = \frac{\sum_{i=1}^{p} FA_j \times cn_i}{\sum_{j=1}^{p} FA_j} \tag{4}$$

Eq. (4) expresses the way defuzzification is carried out using center of area method to obtain the output of the controller 'f', denoted by 'O_f'. Here, 'FA_j' is the firing area of the 'jth' rule, 'cn_i' is the centroid of the ith area (AR) under consideration.

It is important to note that 'μ_{ij}' values and 'FS_{jk}' values resemble the connection strengths between 'I' and 'HL-1'and are denoted by 'w_{ij}' and that between 'HL-2' and 'O' layer are denoted by 'v_{jk}'. Throughout the nodal input–output processing, linear transfer functions have been considered. The developed model is then trained and tested with 214 and 88 real-world depression cases, respectively.

To measure the performance of the said hybrid controller, its sensitivity (S_n), specificity (S_p), precision (P_r), and average accuracy (A_{cc}) have been measured [28]. All values are expressed in %. Eqs. 5–8 show how these are computed. In

these equations 'p' and 'n' denote total positive and negative cases; 'tp' and 'np' are true and false positive cases and 'N' is the total number of cases.

$$S_n = \frac{tp}{p} \times 100 \tag{5}$$

$$S_p = \frac{tn}{n} \times 100 \tag{6}$$

$$P_r = \frac{tp}{tp + fp} \times 100 \tag{7}$$

$$A_{cc} = \left[S_n \times \frac{p}{N} + S_p \times \frac{N}{N + P} \right] \times 100 \tag{8}$$

4. Experimental results and discussions

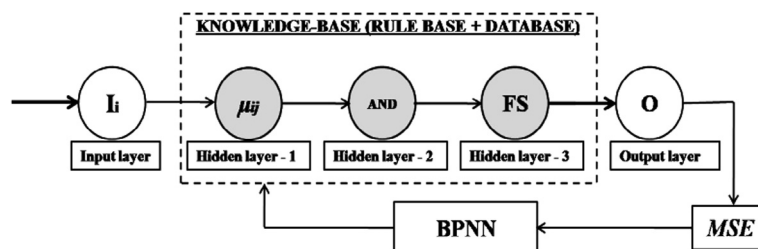This section discusses the experimental results sequentially.

4.1. Data processing

The calculated Cronbach's α is *0.87*, which is well-over 0.7, the assigned threshold [29]. Hence, the data are said to be internally consistent and reliable for PCA.

PCA measures eigenvectors and its eigenvalues of the correlation matrix. Table 2 shows the eigenvalues of all factors. According to the concept of PCA, factors of higher eigenvalues are the principal components (PC) and are interesting to study. It may be noted that '*S' (Feeling sad)*, '*P' (Loss of pleasure)*, '*W1' (Weight loss)*, '*I1' (Insomnia)*, '*H' (Hypersom-*

Table 2 Eigen Analysis.

Symptom	Eigenvalue
S	*1.3562*
P	*1.2868*
W1	*1.1925*
I1	*1.1480*
H	*1.1275*
A	*1.0410*
P1	*1.0127*
P2	*0.9880*
E	*0.9302*
W2	*0.8979*
T	*0.8039*
I2	*0.8014*
D	*0.7243*
SO	*0.6896*

Fig. 2 The proposed neuro-fuzzy model.
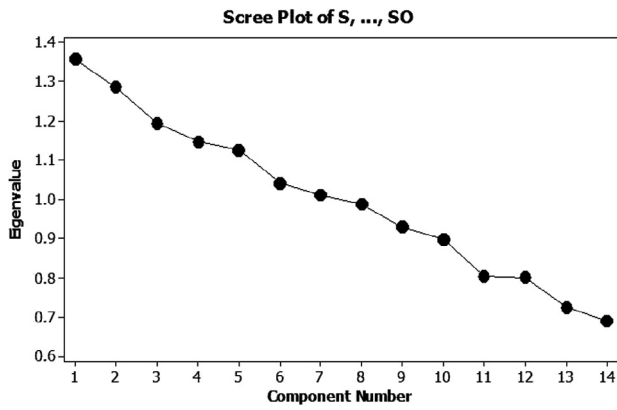
Fig. 3 Scree plots of components.

Table 3 Fuzzification Of The 1st Training Case.

I	μ(mM)	μ(Ms)	μ(s)	AND
S (0.5)	0.66	0.34	–	0.34
P (0.6)	–	1.0	–	–
W1 (0.9)	–	–	1.0	–
I1 (0.5)	0.66	0.34	–	0.34
H (0.8)	–	0.34	0.66	0.34
A (0.7)	–	0.66	0.34	0.34
P1 (0.9)	–	–	1.0	–

nia), 'A' *(Loss of appetite)*, and 'P1' *(Psychomotor agitation)* are the seven latent factors with eigenvalues higher than 1.0 (see Table 2) and hence are taken into consideration as principal components (PC). Therefore, the initial input vector space having 14-dimension initially has now been reduced to 7-dimension. It resembles the way psychiatrists extract the significant symptoms from too many at the backdrop of the illness during *DD*. Fig. 3 plots the components based on the respective eigenvalues. Because the metric of the original dataset has now been changed to 302×7, α has been checked again, which is 0.72 in this experiment and above the assigned threshold 0.7 [29].

4.2. Fuzzy-neural hybrid controller

Seven symptoms, extracted by PCA and the level of 'Depression' have been arbitrarily grouped as '*mild*'(*m*), '*moderate*'(*M*) and '*severe*' (*s*) with overlapping scales such as *mild-to-moderate (mM)*, *moderate-to-severe (Ms)* and '*very severe*' (*s*) (referring to the *DD* phase of clinical diagnosis, discussed in section I) and are represented into asymmetric TMFD (refer to Fig. 4). Such overlapping regions are the focus of this fuzzy-neural model that aims to differentiate it into 'mild', 'moderate' and 'severe' (i.e., the *PD* phase of the clinical diagnosis). A group of psychiatrists according to their experience assigned the range A Mamdani's FLC is then developed and modeled into a FFMNN structure. There are a total of seven inputs, each with two possibilities one time. Hence, there are a total
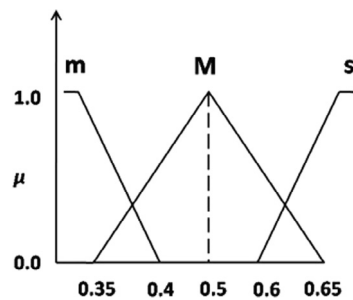
of 2^7 i.e., 128 rules that might fire among 3^7, i.e., 2187 possible rules. At first, the FLC has been trained to optimize its rule base and then tested with the optimized rule set after a 10-fold cross validation.

a. Training:

The FLC is then trained (offline) with 214 real-world depression cases obtained from 10-fold cross validation. For one training case, there are seven inputs (i.e., seven symptoms) each having three Grades of Distributions (GD). For an input, maximum two GDs can take part and hence two membership values (μ) could be computed. Membership values having absolute belongingness (i.e., 1.0) or non-belongingness (i.e., 0.0) are discarded from the study. For the first training case, 'S', 'I1', 'H', and 'A' are considerable. The rule could be stated as per the domain experts "IF 'S' is 'mM' AND 'I1' is 'mM' AND 'H' is 'Ms' AND 'A' is 'mM' THEN Depression is 'mM'" and the 'FS' of this rule is 0.34 (see Table 3).

Then the respective 'FS' was calculated. 'FS' values $\geqslant 0.5$ are considered 'strong' according to the 10-fold cross validation tests. These strong rules are used for computing the output of the controller and the rests are neglected.

It may be noted that a total of *38 rules* (which is only 30% of the total fired rules) have FS $\geqslant 0.5$. An 'OR' operation has then been performed to obtain the membership of the combined control (see Table 4). In the next step, firing areas (FA) of all fuzzy output rules for '*mM*', '*Ms*', and '*s*' have been computed (refer to fig. 5).

After defuzzification with the Centre of Area (COA) method, the final combined output of the FLC is *0.1852* i.e., it belongs to the class 'mild' (m). In addition, the *MSE* of all 214 training cases is low, i.e., *0.0008*.

Fig. 4 Triangular membership function distribution of symptoms (inputs) and the grades of depression (output).

Table 4 Output Calculations.

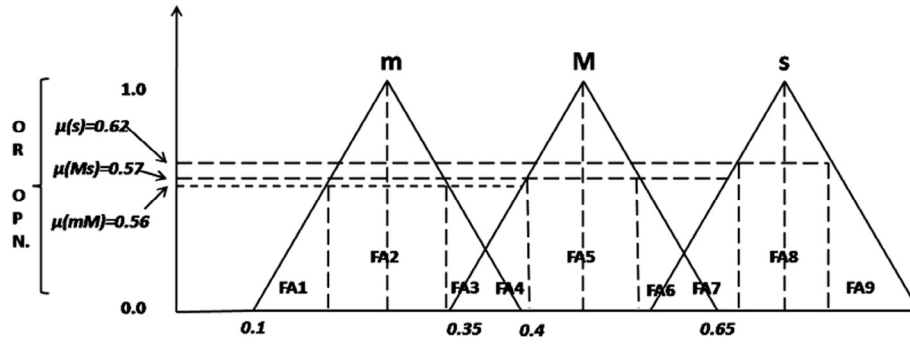| #FA	FA	Center	FA * Center
FA1	0.021	0.7911	0.016614
FA2	0.084	0.25	0.00258
FA3	0.0336	0.0767	0.219152
FA4	0.0171	0.4681	0.116432
FA5	0.1026	0.5	0.117398
FA6	0.01995	0.3412	0.074607
FA7	0.0248	0.3426	0.021
FA8	0.1116	0.75	0.0513
FA9	0.0155	0.2731	0.0837
		SUM = 3.79318	SUM = 0.702782

Fig. 5 Aggregated firing area (FA1 to FA9).

Table 5 Forward Calculations For One Test Case.

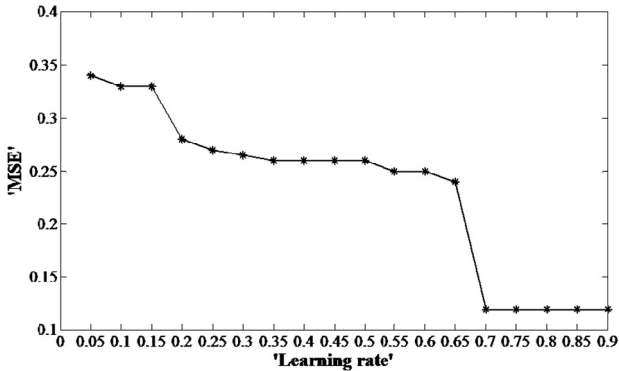| I	HL-1 (I)	HL-1(O)	HL-2 (I = O)	HL-3 (I)	HL-3 (O)	OL(I = O)
S	0.62	μ (Ms) = 0.42; μ (s) = 0.58;	0.42	IF 'S' is 'Ms' & 'P' is 'mM' … & 'P1' is 's' THEN 'Depression' is 's' (see table 5, rule # 112).	$0.52^*0.42$	0.2184
P	0.38	μ (mM) = 0.33; μ (Ms.) = 0.67;	0.33		$0.52^*0.33$	0.1716
W1	0.64	μ (Ms.) = 0.14; μ (s) = 0.86;	0.14	FS = 0.52.	$0.52^*0.14$	0.0728
I1	0.36	μ (mM) = 0.66; μ (Ms) = 0.34;	0.34		$0.52^*0.34$	0.1768
H	0.36	μ (mM) = 0.66; μ (Ms) = 0.34;	0.34		$0.52^*0.34$	0.1768
A	0.36	μ (mM) = 0.66; μ (Ms) = 0.34;	0.34		$0.52^*0.34$	0.1768
P1	0.64	μ (Ms.) = 0.14; μ (s) = 0.86;	0.14		$0.52^*0.14$	0.0728 $\sum$ = 1.066

Fig. 6 Parametric study to find the best value of the learning rate.

b. Testing

After the training, performance of the model is then tested for 88 cases with the optimized rule set. It has also been validated with 50 controls, i.e. normal population. Table 5 shows the sequence of computations for one test case as described in detail in section III-'C'. Similarly, remaining test cases are passed one-by-one though the system and the corresponding MSE has been calculated, which is found to be *0.0121*. It may be noted that the best value of η has been selected through a parametric study by varying the values of η from 0.05 to 0.95 (0.05 as the rate of increment) keeping the weights (i.e., 'w_{ij}' and 'v_{jk}' values) constant all time (see Fig. 6).

In the next step, the error needs to be minimized using a BPNN algorithm by updating the 'w_{ij}' and 'v_{jk}' values, iteratively keeping the 'η' equals to 0.70 (see Table 6). It may be noted that the 'w_{ij}' and 'v_{jk}' values are arbitrarily chosen as val-

ues between 0 and 1. These values are iteratively updated while reducing the error, so the initially considered values change accordingly.

Table 6 shows that the error could be further minimized using BPNN by iteratively updating the weight vectors. Because 1% error is negligible in medical diagnosis, especially in mental health, the weight (that resembles the information)-updating task has not been considered further.

Finally, the average sensitivity ('S_n'), specificity ('S_p'), precision ('P_r'), and accuracy ('Acc') of the hybrid controller have

Table 6 Updated weights for 1st test case.

Iteration	'w_{ij}'	'v_{jk}'	MSE
0	0.42, 0.33, 0.14, 0.34, 0.34, 0.34, 0.14	0.52	0.0121
1	0.4219, 0.3315, 0.1406, 0.3416, 0.3416, 0.3416, 0.1406	0.5153	0.0120

Table 7 Prediction Analysis Of The Hybrid Controller On Test Data.

Prediction	'mM'	'Ms'	's'
tp	27	28	29
fp	00	00	00
tn	00	00	00
fn	01	01	02
Total	28	29	31
S_n (%)	96.42	96.55	93.54
S_p (%)	0.0	0.0	0.0
P_r (%)	100%	100%	100%
Acc (%)	92.03 (avg.)		

Table 8 Comparative studies of various soft computing techniques in automating depression.

Author(s)	Year	Technique(s)	Sensitivity	Specificity	Accuracy (avg.)
Tai & Chiu	2007	RBFN [14]	82%	86%	–
Lan et al.	2008	Fuzzy-Tech 5.54 software [19]	–	–	93%
de Carvalho et al.	2008	Neuro-fuzzy NEFCLASS and ANNB techniques [20]	–	–	95.65%
Suhasini et al.	2010	BPNN [15]	–	–	98.75%
Suhasini et al.	2011	BPNN [16]	–	–	-do-
Chattopadhyay et al.	2011	BPNN [17]	–	–	89%
Chattopadhyay et al.	2011	BPNN + ANFIS (neuro-fuzzy with Sugeno's technique) [18]	–	–	97.33%
Chattopadhyay et al. (current study)	*2012*	*Neuro-fuzzy model with Mamdani's technique*	*–*	*–*	*95.50%*

been measured by comparing with the target depression grades (see Table 7). In case of 50 control data (i.e. normal population), none of the cases are misdiagnosed as 'depression'. Hence, it is not shown as a table. The average accuracy of the hybrid controller is 95.50% (see Table 7), which is comparable to the other controllers shown in Table 8.

4.2.1. Comparison

Table 8 shows the comparison of the outcome of this study with the previous works, where several soft computing techniques have been used in the diagnosis of depression.

5. Discussions

This paper proposes a neuro-fuzzy model consisting of Mamdani's technique mapped onto a feedforward backpropagation neural network. The aim is to model the manual process of clinical diagnosis of depression. On a sample of 302 real-world depression cases, the model has been trained and tested. The observations are as follows.

- The obtained depression data are reliable both before and after PCA (Cronbach's α value are over 0.7). While diagnosing an illness, medical doctors also search for reliability of the information given to them by the patients and their relatives or the colleagues.
- PCA could reduce data dimension by extracting the significant latent factors, such as 'S' (*Feeling sad*), 'P' (*Loss of pleasure*), 'W1' (*Weight loss*), 'I1' (*Insomnia*), 'H' (*Hypersomnia*), 'A' (*Loss of appetite*), and 'P1' (*Psychomotor agitation*) to qualify for the depression state. It mimics the process by which medical doctors eventually reduce the amount of information based on the importance of the information pertinent to the illness.
- Out of total 3^7 (i.e., 2187) rules, 2^7 (i.e. 128) rules are fired. Now, with 10-fold cross validation it is noted that the rules having firing strength > 0.5 may be considered as 'strong' rules, which are ultimately required to develop the final rule base of the controller. With this technique, 98.27% rules, which are not so important, are pruned. Such pruning is important to reduce the complexity of the controller. Medical doctors also possess their own rule bases depending on their levels of experiences. This technique mimics the way medical doctors actually prune the number of unimportant

rules to reduce the diagnostic complexity and converge into a diagnosis. Hence, the paper argues that, it is the rule quality and quantity that are important for decision-making.
- With the learning rate (η), the developed hybrid model is able to automate the diagnosis process, especially from *DD* (i.e. overlapped diagnoses, which are handled by the principle of fuzzy sets) to *PD* (i.e. specific diagnoses, which are the output of combined control process) with 100% average precision and 95.50% accuracy. It is therefore a comparable result with respect to several others works (see Table 8). Medical doctors also learn either from observations or by examples and iteratively converge to the diagnosis. The qualitative information is modeled by the concept of fuzzy sets and the interpretation is tuned iteratively by adjusting the connector weights (i.e., their state of knowledge base).

Future work: Expert tool development is the next step of any controller design. However, it is a time-taking process to achieve the required amount of standardization and user satisfaction. Keeping all these in mind, we propose that such a hybrid model could be implemented as a full-fledged expert system and given to the psychiatrists for its' initial use. The authors are currently working in the Graphic User Interface (GUI) development and its standardization.

References

[1] S. Chattopadhyay, Fuzzy Logic Based Screening and Prediction of Adult Psychoses (Ph.D. thesis), Indian Institute of Technology, Kharagpur, India, 2007.
[2] M. Gannon, M. Qasesem, V. Snow, Q. Snooks, Pain management and the primary care encounter opportunities for quality improvement, J. Prim. Care Community Health 2 (2011) 37–44.
[3] D.M. Clarke, Psychological adaptation, demoralization and depression in people with cancer, in: D.W. Kissane, M. Maj, N. Sartorius (Eds.), Depression and Cancer, John Wiley & Sons Ltd, Chichester, UK, 2010 (ch. 2).
[4] I. Pacchiarotti, L. Mazzarini, G.D. Kotzalidis, M. Valentí, A.M. A. Nivoli, G. Sani, C. Torrent, A. Murru, J. Sanchez-Moreno, B. Patrizi, P. Girardi, E. Vieta, F. Colom, Mania and depression. Mixed, not stirred, J. Affect. Disord. 133 (1) (2011) 105–113.
[5] S. Chattopadhyay, Psyconsultant I: a DSM-IV-based screening tool for adult psychiatric disorders in Indian rural health center, Internet J. Med. Inform. 3 (2006) (Serial Online).

[6] S. Chattopadhyay, A computerized tool for screening of adult psychiatric illnesses: a third-world perspective, J. Clin. Inform. Telemed. 3 (2006) 1–5.

[7] S. Chattopadhyay, D.K. Pratihar, S.C. De Sarkar, Some studies on fuzzy clustering of psychosis data, Int. J. Bus. Intell. Data Min. 2 (2007) 143–159.

[8] S. Chattopadhyay, P. Ray, M.B. Lee, H.S. Chen, Towards the design of an e-health system for suicide prevention, in: Proceedings of the Eleventh IASTED International Conference on Artificial Intelligence, Palma de Mallorca, Spain, 2010, pp. 191–196.

[9] S. Chattopadhyay, D.K. Pratihar, S.C. De Sarkar, Developing fuzzy classifiers to predict the chance of occurrence of adult psychoses, Knowledge Based Syst. 20 (2008) 479–497.

[10] S. Chattopadhyay, D.K. Pratihar, S.C. De Sarkar, Fuzzy logic-based screening and prediction of adult psychoses: a novel approach, IEEE Trans. Syst. Man Cybern. A Syst. Hum. 39 (2009) 381–387.

[11] S. Chattopadhyay, D.K. Pratihar, S.C. De Sarkar, Statistical modelling of psychoses data, Comput. Methods Programs Biomed. 100 (2010) 222–236.

[12] S. Chattopadhyay, U.R. Acharya, A novel mathematical approach to diagnose premenstrual syndrome, J. Med. Syst. 36 (4) (2012) 2177–2186.

[13] S.-C. Yu, Y.-H. Lin, Applications of fuzzy theory on health care: an example of depression disorder classification based on FCM, WSEAS Trans. Info. Sci. Appl. 5 (2008) 31–36.

[14] Y.-M. Tai, H.-W. Chiu, Artificial neural network analysis on suicide and self-harm history of Taiwanese soldiers, in: Proceedings of the Second International Conference on Innovative Computing, Information and Control (ICICIC), Kumamoto, Japan, 2007, p. 363.

[15] A. Suhasini, S. Palanivel, V. Ramalingam, Multi decision support model for psychiatry problem, Int. J. Comput. Appl. 1 (2010) 61–69.

[16] A. Suhasini, S. Palanivel, V. Ramalingam, Multimodel decision support system for psychiatry problem, Expert Syst. Appl. 38 (2011) 4990–4997.

[17] S. Chattopadhyay, P. Kaur, F. Rabhi, U.R. Acharya, An automated system to diagnose the severity of adult depression, in: D. Jana, P., Pal (Eds.), Proceedings of Second International Conference on Emerging Applications of Information Technology (CSI EAIT-2011), IEEE Computer Society and Conference Publishing Services, Kolkata, India, 2011, pp. 121–124.

[18] S. Chattopadhyay, P. Kaur, F. Rabhi, U.R. Acharya, Neural network approaches to grade adult depression, J. Med. Syst. 36 (5) (2012) 2803–2815.

[19] T.H. Lan, E.W. Loh, M.S. Wu, P. Chou, T.Y. Lan, H.-J. Chiu, Performance of a neuro-fuzzy model in predicting weight changes of chronic schizophrenic patients exposed to antipsychotics NFM performance on chronic schizophrenic patients, Mol. Psychiatry 13 (2008) 1129–1137.

[20] L.M.F. de Carvalho, S.M. Nassar, F.M. de Azevedo, H.J.T. de Carvalho, L.L. Monteiro, C.M.Z. Rech, A neurofuzzy system to support in the diagnostic of epileptic events and non-epileptic events using different arithmetical operations, Arq. Neuropsiquiatr. 66 (2a) (2008) 179–183.

[21] K. Pearson, On Lines and Planes of Closest Fit to Systems of Points in Space, Philosophical Magazine, 1901. <http://stat.smmu.edu.cn/history/pearson1901.PDf>, 6, pp. 559–572.

[22] E.H. Mamdani, Application of fuzzy algorithms for the control of a simple dynamic plant, in: Proc. IEEE, 1974, pp. 121–159.

[23] F. Rosenblatt, The perceptron: a probabilistic model for information storage and organization in the brain, Psychol. Rev. 65 (1958) 386–408.

[24] D.E. Rumelhart, G.E. Hilton, R.J. Williams, Learning internal representations by error propagation, Parallel Distributed Processing, vol. 1, MIT Press, Cambridge, MA, 1986, pp. 318–362.

[25] P.A. Devijver, J. Kittler, Pattern Recognition: A Statistical Approach, Prentice-Hall, London, UK, 1982.

[26] Diagnostic and Statistical Manual (DSM-IV) of Mental Disorders, fourth ed., American Psychiatry Association, 1994.

[27] L.J. Cronbach, Coefficient alpha and the internal structure of tests, Psychometrika 16 (1951) 297–333.

[28] J. Han, M. Kamber, Data Mining: Concepts and Techniques, second ed., 2006, p. 361.

[29] J. Nunnaly, Psychometric Theory, McGraw-Hill, New York, 1978.