



Modeling Depression Data: Feed Forward Neural Network vs. Radial Basis Function Neural Network

Subhrangsu Mukherjee¹, Kumar Ashish¹, Nirmal Baran Hui¹, Subhagata Chattopadhyay²

¹ Mechanical Engineering, NIT Durgapur, West Bengal -713209, India

² Specialty Business Unit, Nationwide, The Family of Doctors, Bangalore, India

* **Corresponding author**

Nirmal Baran Hui
Associate Professor
Mechanical Engineering
NIT Durgapur
West Bengal, 713209
India
Email: nirmalhui@gmail.com

Received: 14 July 2014; | Revised: 12 September 2014; | Accepted: 29 September 2014

Abstract

Depression is a serious affliction that affects a large fraction of the global populace. Due to the widely varying symptoms the diagnosis poses a unique problem based on uncertainty. This paper proposes an approach to tackle the aspect of uncertainty using Soft Computing techniques, which are trained using real life medical data. We have developed two forms of intelligent Neural Network models to help in obtaining a reasonably accurate diagnosis. Trials with test data have yielded nominal Mean Squared Error. Hence this could prove to be a useful tool in automated diagnosis of depression.

Keywords: Health Informatics, Depression Data, Radial Basis Function Neural Network, Back Propagation Neural Network.

1. Introduction

The usage of Artificial Intelligence in Medicine is a relatively new concept and highly relevant to the modern requirements of diagnosis [1-7]. In cases such as depression diagnosis where the uncertainty plays a major role, an intelligent algorithm can provide high levels of accuracy than the traditional methods of diagnosis. Success of such algorithms can be attributed to its ability to accept subjective data

and quantify it in order to analyze and produce an objective diagnosis.

Diagnosis of psychiatric diseases is complicated in nature due to its subjective symptoms. This is further complicated by the fact that symptoms may be progressive and early diagnosis may not be possible even by the experienced psychiatrists. Moreover there is widespread social stigma where psychiatric patients may be ostracized. The combination of all these factors makes depression a major global

health concern that affects at least one-thirds of the adult human population. The scarcity of well trained and experienced psychiatric professionals in many regions of the world makes this problem even more acute. Thus, we see the relevance and necessity of such automated intelligent diagnosis systems to help counter this menace.

Over the last two decades, a lot of work has been done to incorporate Soft Computing (SC) techniques into the handling of psychiatric disorders [8-11]. Concepts of Fuzzy-C Means (FCM) and K-means (KM) clustering techniques were used for classifying depression grades using Beck's Depression Inventory-II (BDI-II). In another study, Brief Psychiatric Rating Scale-F2 (BPRS-F2) has been used to capture the symptoms of seven look-alike psychiatric diseases and depression was one of them [12-18]. Various neural network models have also been implemented specially pertaining to suicidal tendencies stemming from depression [19-20]. Swarm intelligence techniques have also been used in the past. Hybrid SC techniques using combinations of the basic techniques have yielded good results such as neuro-fuzzy controllers [21-24]. Traditional linear regression models have also proved quite fruitful. However simple Neural Network models can be used in very low resource environments, making them tools with wide scope. Given properly obtained training data, these can yield fairly accurate diagnosis even with nominal computational and knowledge resource [25-28].

2. Developed Approaches

In this section the data collection and the construction of the two Neural Network models has been discussed.

2.1 Data Collection

'First reported' depression data were collected from the bed-tickets of three hundred and two cases taking appropriate ethical measures. Choice of a case was irrespective of gender and age. It took one year (2004-2005) to complete the task. All these cases were drug-naïve, i.e., they never took anti-depressive medications and were reporting to the hospital for

the first time. Cases presented with suicidal ideations were excluded as these cases required urgent treatment. Ten common symptoms, namely Feeling Sad (FS), Loss of Pleasure (LP), Weight Loss (WL), Insomnia (IN), Hypersomnia (H), Loss of appetite (LA) and Psychomotor Agitation (PA) were considered for this work after consulting with three senior psychiatrists (mean experience of 10.4 yrs.). The grade of each symptom and the corresponding probability of depression were unanimously assigned [0, 1] by them.

2.2. Approach 1: Back Propagation Neural Network Model Construction

An attempt has been made here to model a three-layered fully connected Feed Forward Neural Network (refer to **Fig. 1**) [29].

This model consists of three layers, the Input Layer (IL), the Hidden Layer (HL) and the Output Layer (OL). There are ten neurons in the IL corresponding to the ten symptom inputs. The HL has variable number of neurons ranging between 3 and 15 depending upon the most optimum performance based on the training data set. The OL has only one neuron that gives the final output as degree of depression. The weight sets [V] and [W] connect the outputs of IL to inputs of HL and the outputs of HL to the inputs of OL respectively. They are optimized according to the training data as well.

2.2.1 Layer 1 (Input Layer)

The input layer accepts the ten symptoms of depression as input. The nodal output is the same as input as a linear transfer function has been used.

2.2.2 Layer 2 (Hidden Layer)

The hidden layer takes the weighted mean of outputs from the input layer as its input. The weight system used in this case is [V]. Prior to training of the system, the weight set is randomly generated using a seed based random number generator. After training the optimized weight values are used. The transfer function used is 'log-sigmoid' and thus the nodal outputs vary within 0 and 1.

$$O_{P2} = \frac{1}{1 + \exp(-a * I_{P2})}$$

where I_{P2} = Nodal Input and O_{P2} = Nodal Output for the second layer and 'a' is a constant.

2.2.3 Layer 3 (Output Layer)

The output layer takes the weighted mean of outputs from the hidden layer as its input. The weight system used in this case is [W]. Prior to training of the system, the weight set is randomly generated using a seed based random number generator. After training the optimized weight

values are used. The transfer function used is 'tan-sigmoid' and thus the nodal outputs vary within 0 and 1.

$$O_{P3} = \frac{\exp(-I_{P3}) - \exp(-I_{P3})}{\exp(-I_{P3}) + \exp(-I_{P3})}$$

where I_{P3} = Nodal Input and O_{P3} = Nodal Output for the third layer.

A batch mode of training has been adopted to optimize the performance of the controller so developed. It can be implemented using a Back-Propagation (BP) algorithm, as explained below.

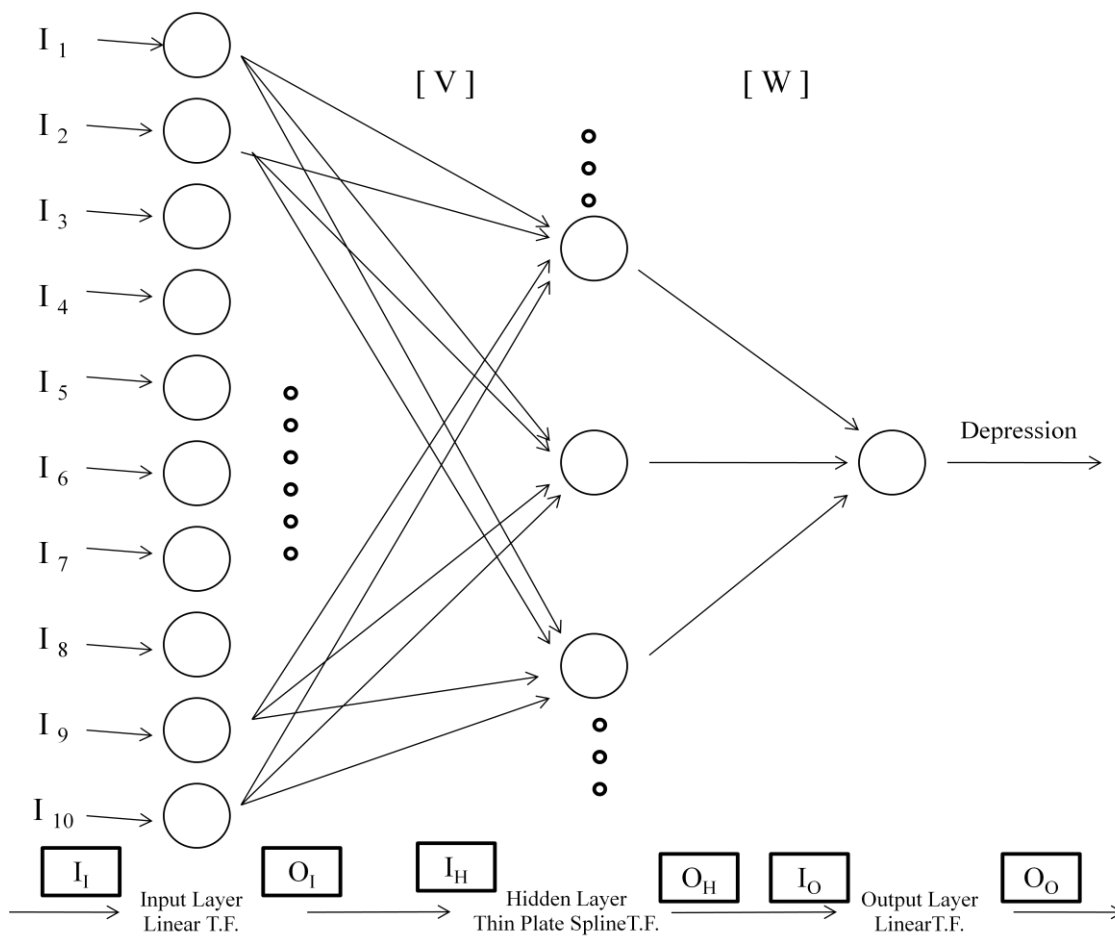


Figure 1: The multi-layered feed forward neural network model.

2.2.4 Tuning of the neural network system using a Back Propagation Algorithm

In this case the weight sets [V] and [W] are fine tuned to optimize the system. This objective

is achieved through the use of a Back Propagation (BP) algorithm.

Let us consider C training scenarios and Mean Squared Deviation (MSD) in prediction can be calculated as follows:

$$E = \frac{1}{2c} \sum_{i=1}^c (T_i - O_i)^2$$

where T_i is the target for i^{th} iteration and O_i is output of OL neuron for i^{th} iteration.

Now the change in weight value v_{ij} in weight set [V] is represented as:

$$v_{ij,updated} = v_{ij,previous} + \Delta v_{ij}$$

Where the change in v_{ij} can be calculated as:

$$\Delta v_{ij} = -\eta \frac{\partial E}{\partial v_{ij}} + \alpha \Delta v_{ij,previous}$$

where η represents the learning rate and α represents the momentum constant. Δv_{ij} represents the change in v_{ij} .

Now the value of the first term may be calculated as:

$$\frac{\partial E}{\partial v_{ij}} = \frac{\partial E}{\partial E_c} \frac{\partial E_c}{\partial O_{Oc}} \frac{\partial O_{Oc}}{\partial I_{Oc}} \frac{\partial I_{Oc}}{\partial O_{Hj}} \frac{\partial O_{Hj}}{\partial I_{Hj}} \frac{\partial I_{Hj}}{\partial v_{ij}}$$

Similarly for the weight set [W], we use the corresponding equations:

$$w_{i,updated} = w_{i,previous} + \Delta w_i$$

$$\Delta w_i = -\eta \frac{\partial E}{\partial w_i} + \alpha \Delta w_{i,previous}$$

And in this case the first term is:

$$\frac{\partial E}{\partial w_i} = \frac{\partial E}{\partial E_c} \frac{\partial E_c}{\partial O_{Oc}} \frac{\partial O_{Oc}}{\partial I_{Oc}} \frac{\partial I_{Oc}}{\partial w_i}$$

2.3. Approach 2: Radial Basis Function Neural Network

A Radial Basis Function Neural Network (RBFNN) has following salient features (i) The input and Output Layer nodes use linear transfer functions and (ii) The Hidden Layer uses a Radial Basis Function as its transfer function. **Radial Basis Functions** have their functional value dependent on the distance of the input parameter from a given mean. Examples are the Gaussian distribution function, multiquadric function, inverse quadric function and polyharmonic spline functions. In our developed system we use a special case of polyharmonic spline function – the thin plate spline function as shown below.

$$O_{P2R} = I_{P2R}^2 \log(I_{P2R})$$

where I_{P2R} = Nodal Input and O_{P2R} = Nodal Output for the second layer of RBFNN.

2.3.1 Tuning of the RBFNN System using a Back Propagation Algorithm

The tuning of this system is the same as the BPFNN system due to the similar structure. Please refer to **Section 2.2.4** for the tuning process.

3. Results and Discussions

In both the approaches, neural network model is trained using a Back Propagation algorithm off-line. A set of 45 training data sets was used. Each set consisting of ten normalized inputs I_1 to I_{10} and one target output that is the actual doctor's diagnosis. A batch mode of training was used here. First the NN model in Approach 1 is optimized taking an appropriate number of neurons in the hidden layer. The number of hidden layer neurons is varied between 3 and 15 to maintain a stable and non redundant state keeping momentum constant (α) and learning rate (η) as constant. The result is shown in **Fig. 2**. After this the number of neurons in hidden layer is kept as the one which had minimum final error, which in this case was found to be 15. Now the learning rate η is varied between 0.01 and 0.8. The results are shown in **Fig. 3**. The most optimum learning rate is found to be $\eta=0.8$ and momentum constant value α is varied between the same 0.01 and 0.8. The results can be seen in **Fig. 4**. After this system is thus optimized, the program is allowed to run for 200 epochs where the MSD is observed to stagnate at a value of 0.012187 as shown in **Fig. 5**.

In the similar manner, variation of error with number of neurons, momentum constant, learning rate and number of iterations for the Approach 2 is presented in Figs. 6 through 9, respectively. The best result was obtained with number of neurons = 6, $\eta=0.35$ and $\alpha=0.8$.

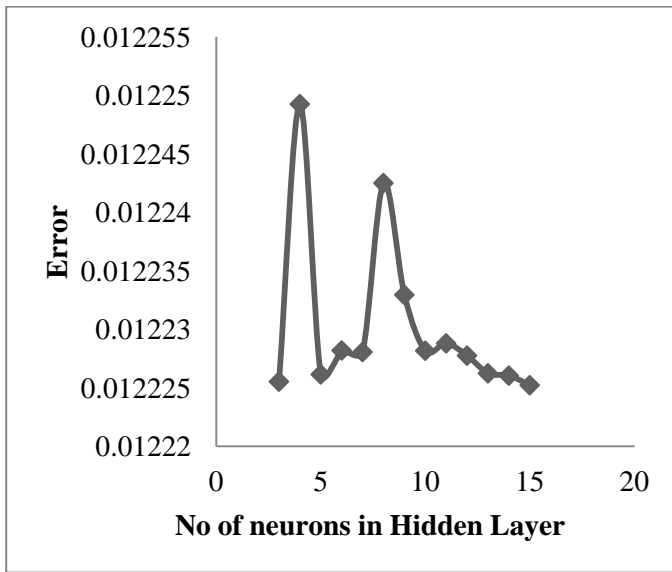


Figure 2: Variation of error with number of neurons in hidden layer in Approach 1.

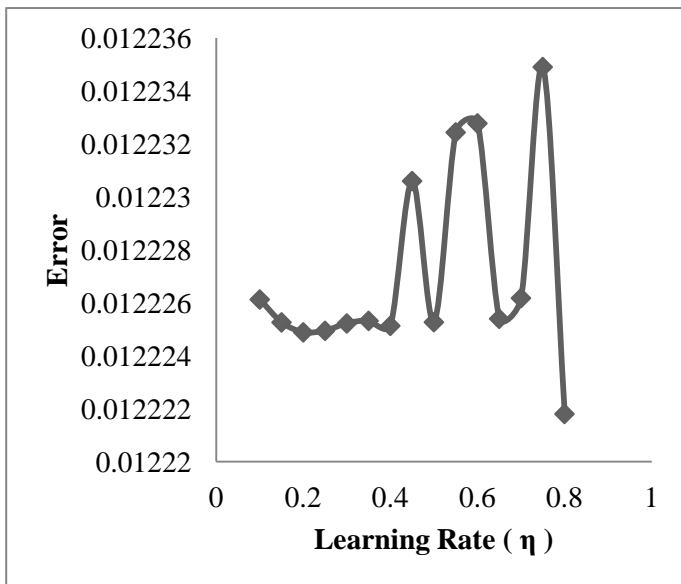


Figure 3: Variation of error with learning rate (η) in Approach 1.

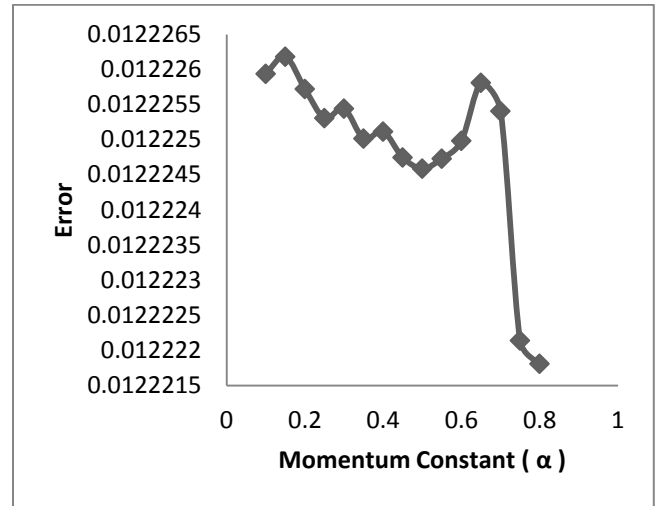


Figure 4: Variation of error with momentum constant (α) in Approach 1.

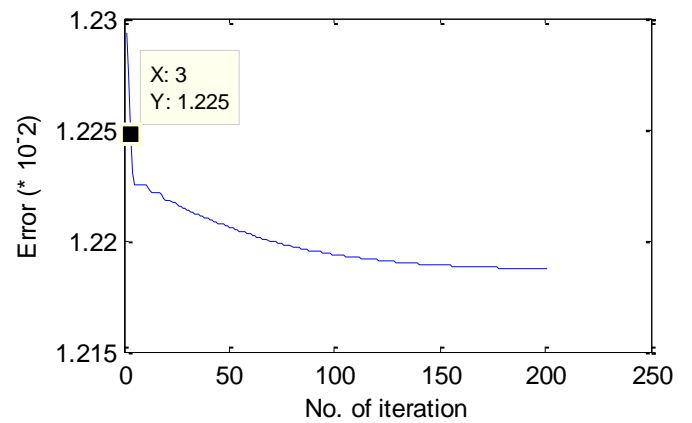


Figure 5: Error vs. no. of iterations with optimal neural network parameter in Approach 1.

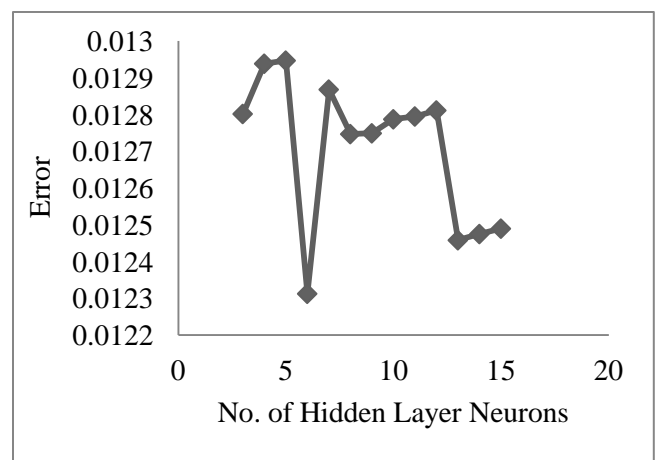


Figure 6: Variation of error with number of neurons in hidden layer in Approach 2.

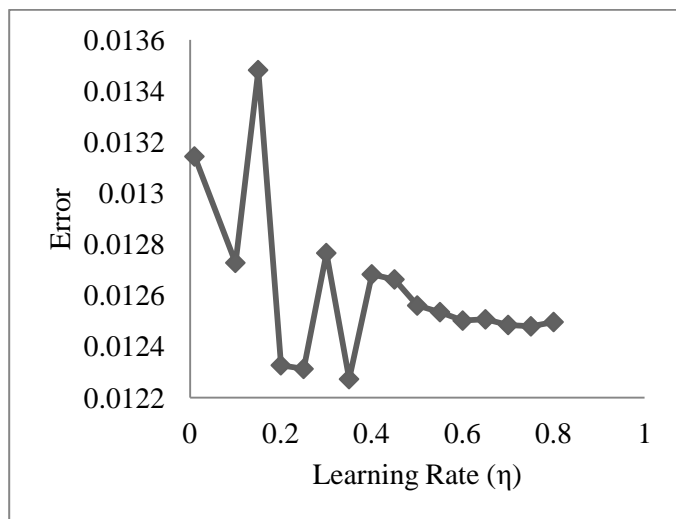


Figure 7: Variation of error with learning rate (η) in Approach 2.

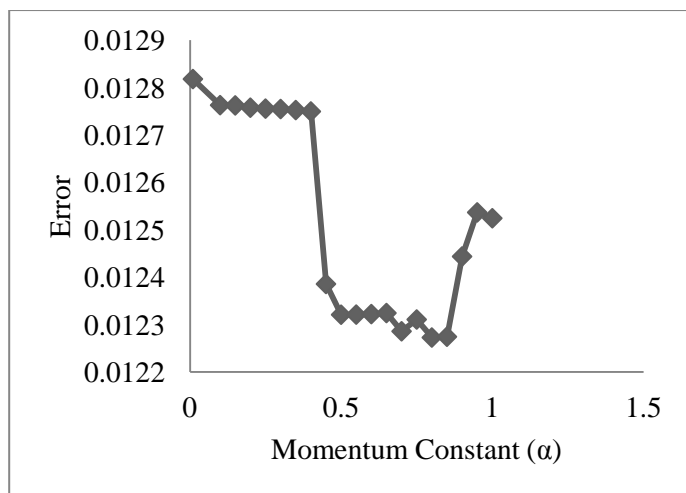


Figure 8: Variation of error with momentum constant (α) in Approach 2.

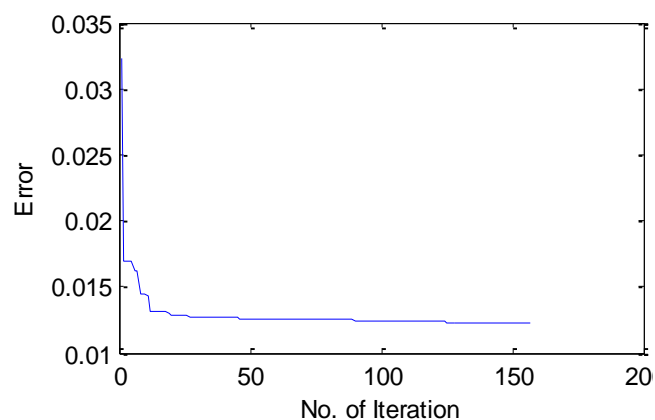


Figure 9: Error vs. no. of iterations with optimal neural network parameter in Approach 2.

Finally the optimal weight values for both the cases have been obtained and applied to find the predicted outputs of 5 test data sets. The results are presented in Table 1.

We can see that the differences between the actual target and the predicted result in both the approaches are considerably low. Moreover, the prediction obtained using the Approach 2 is more accurate than that of Approach 1 and at the same time it preserves the simplicity of algorithm of the BPFNN. The error in prediction may be attributed to: (i) Error on part of the diagnosing doctor (ii) Lack of sufficiently large training and test set and (iii) Over simplicity of the algorithm of this Neural Network model. Even though there is a clear correlation between predicted and actual data, we have attempted to increase the accuracy in our next model without compromising the simplicity and low resource requirements.

Table 1: Test Data Results

Target Depression	Predicted Depression		% error	
	Approach 1	Approach 2	Approach 1	Approach 2
0.6	0.335263	0.042501	44.12284	7.083563
0.7	0.333607	0.401875	52.34187	57.41074
0.5	0.334047	0.081333	33.19058	16.26663
0.2	0.33387	-0.23611	-66.9352	-118.053
0.4	0.334797	-0.07405	16.3008	-18.5117

4. Conclusions and Future Work

The handling of psychiatric issues still remains a widely neglected field owing to social stigmas and taboo. Furthermore the lack of experienced and trained professionals and large number of trained affected population makes the problem even more crucial. The usage of such automated intelligent systems can provide much needed relief to unattended cases in developing areas. Also such systems can prevent cases of 'false alarms' and thus save precious resources. The future scope of this work would be to obtain greater quantities of sample data, correlation and standardization of these data sets and the spreading of awareness regarding the usage of these systems. There is also a great scope for improvement of accuracy without compromising the simplicity of the system. A system should be simple and easy to handle so that even a layman can utilize and modify the system structure which best suits one's sample set. Also the simpler a system is, greater is its flexibility of usage and its scope. The requirement of minimal computational resources is also an advantage for rural and under developed areas. The rough tuning processes such as testing of various transfer functions and the flexibility of assignment of transfer functions is the subject of further work.

5. Acknowledgement

The authors gratefully acknowledge the assistance and support of psychiatrist colleagues in collecting the quantified data, developing the controller, and interpretation of the results.

References

- [1] Li, J. H; Land, L.; Ray P.; Chattopadhyay, S. E-Health readiness framework from electronic health records perspective, *IJIEM*, 2010, 6, 326-348. <http://aisel.aisnet.org/globdev2008/4>
- [2] Shapiro, S. D. Merging personalized medicine and biology of aging in chronic obstructive pulmonary disease, *Am J Respir. Crit. Care Med*, 2011, 184, 864-866. [DOI: 10.1164/rccm.201108-1486ED](https://doi.org/10.1164/rccm.201108-1486ED)
- [3] Tenório, J. M.; Hummel, A. D.; Cohrs, F. M.; Sdepanian, V. L.; Pisa, I. T.; Marin, H. F. Artificial intelligence techniques applied to the development of a decision-support system for diagnosing celiac disease, *IJMI*, 2011, 80, 793-802. [DOI: 10.1016/j.ijmedinf.2011.08.001](https://doi.org/10.1016/j.ijmedinf.2011.08.001)
- [4] Wang, Y.; Yan, H.; Guo, R.; LI, F.; Xia, C.; Yan, J.; Xu, Z.; Liu, G.; Xu, J. Study on intelligent syndrome differentiations in traditional Chinese medicine based on multiple information fusion methods, *IJDMB*, 2011, 5, 369-382. [DOI: 10.1016/S0254-6272\(14\)60051-0](https://doi.org/10.1016/S0254-6272(14)60051-0)
- [5] Hao, T. L.; Wu, Y. G.; Hong, B.; Zhang, L.; Chen, W. Z.; Li, G. Clinical analysis of patients with small cell prostate carcinoma, *Am. J. Biomed. Sci.* 2011, 3(1), 1-10. [DOI: 10.5099/aj110100001](https://doi.org/10.5099/aj110100001)
- [6] Shao, C.; Tian, Y.; Dong, Z.; Gao, J.; Gao, Y.; Jia, X.; Guo, G.; Wen, X.; Jiang, C.; Zhang, X. The use of principal component analysis in MALDI-TOF MS: a powerful tool for establishing a mini-optimized proteomic profile, *Am. J. Biomed. Sci.* 2012, 4(1), 85-101. [DOI: 10.5099/aj120100085](https://doi.org/10.5099/aj120100085)
- [7] Jabali, A. K. Prediction of patient's individual blood glucose levels from home monitored readings of type-I diabetics, *Am. J. Biomed. Sci.* 2013, 5(3), 188-196. [DOI: 10.5099/aj130300188](https://doi.org/10.5099/aj130300188)
- [8] Chattopadhyay, S.; Pratihari, D.K.; DeSarkar, S.C. Fuzzy logic-based screening and prediction of adult psychoses: a novel approach, *IEEE Trans. Syst. Man Cybernetics Part-A*, 2009, 39, 381-387. [DOI: 10.1109/TSMCA.2008.2010138](https://doi.org/10.1109/TSMCA.2008.2010138)
- [9] Chattopadhyay, S.; Pratihari, D.K. Towards developing intelligent autonomous systems in psychiatry: its present state and future possibilities, *Intelligent Auto. Sys. Studies in Comp. Intelligence*, 2010, 275, 143-166. [DOI: 10.1007/978-3-642-11676-6_7](https://doi.org/10.1007/978-3-642-11676-6_7)
- [10] Carson, A. The human illnesses: neuropsychiatric disorders and the nature of

- human brain, *BJPsych.*, 2012, 200, 85. DOI: [10.1192/bjp.bp.111.094722](https://doi.org/10.1192/bjp.bp.111.094722)
- [11] Bramesfeld, A.; Grobe, T.; Schwartz, F.W. Prevalence of depression diagnosis and prescription of antidepressants in East and West Germany: an analysis of health insurance data, *Social Psychiatry and Psychiatric Epidemiology*, 2010, 45, 329-335. DOI: [10.1164/rccm.201108-1486ED](https://doi.org/10.1164/rccm.201108-1486ED)
- [12] WHO: Mental Health and Substance Abuse. URL: <http://www.searo.who.int/en/Section1174/Section1199/Section1567/Section18268101.html> [last accessed on 09/09/2014]
- [13] Robinson, O. J.; Overstreet, C.; Letkiewicz, A.; Grillon, C.; Depressed mood enhances anxiety to unpredictable threat, *Psychological Medicine*, 2011, 1-11. DOI: [10.1017/S0033291711002583](https://doi.org/10.1017/S0033291711002583)
- [14] Eloul, L.; Ambusaidi, A.; Al-Adwai, S. Silent epidemic of depression in women in the Middle East and North Africa region, *Sultan Qaboos Univ. Med J*, 2009, 9, 5-15. <http://www.ncbi.nlm.nih.gov/pubmed/21509269>
- [15] Surbey, A.K. Adaptive significance of low levels cooperation in depression, *Evolution and Human Behavior*, 2011, 32, 29-40. DOI: [10.1016/j.evolhumbehav.2010.08.009](https://doi.org/10.1016/j.evolhumbehav.2010.08.009)
- [16] Titmarsh S.; Goodyer, I. Psychiatric diagnosis needs a more scientific approach, *Progress in Neurology and Psychiatry*, 2011, 15(4), 21-22. DOI: [10.1002/pnp.206](https://doi.org/10.1002/pnp.206)
- [17] Kwan, B.M.; Dimidjian, S.; Rizvi, S.L. Treatment preference, engagement, and clinical improvement in pharmacotherapy versus psychotherapy for depression, *Behaviour Research and Therapy*, 2010, 48, 799-804. DOI: [10.1016/j.brat.2010.04.003](https://doi.org/10.1016/j.brat.2010.04.003)
- [18] Yu, S-C.; Lin U. H. Applications of fuzzy theory on health care: an example of depression disorder classification based on FCM, *WSEAS Trans of Information Science and Applications*, 2008, 5(1), 31-36. <http://www.worldses.org/journals/information/information-2008.htm>
- [19] Chattopadhyay, S.; Pratihar, D.K.; DeSarkar, S.C. Some studies on fuzzy clustering of psychoses data, *Intl. J. of Business Intelligence and Data mining*, 2007, 2, 143-159. DOI: [10.1504/IJBIDM.2007.013934](https://doi.org/10.1504/IJBIDM.2007.013934)
- [20] Chattopadhyay, S.; Kaur, P.; Rabhi, F.; Acharya, U. R. Neural network approaches to grade adult depression, *J. of Medical Systems*, 2011, 36(5), 2803-2815. DOI: [10.1007/s10916-011-9759-1](https://doi.org/10.1007/s10916-011-9759-1)
- [21] Tai, Y-M.; Chiu, H-W. Artificial neural network analysis on suicide and self-harm history of Taiwanese soldiers, in *Second Intl. Conf. on Innovative Comp., Inform. and Control, 2007 (ICICIC'07)*, 363 – 363. <http://doi.ieeecomputersociety.org/10.1109/ICICIC.2007.186>
- [22] Chattopadhyay S. Neurofuzzy models to automate the grading of old-age depression, *Expert Sys.: J. of Knowledge Engg.*, 2014, 31(1), 48-55. DOI: [10.1111/exsy.12000](https://doi.org/10.1111/exsy.12000)
- [23] Chattopadhyay S. A Prototype Depression Screening Tool for Rural Healthcare: A Step towards e-Health Informatics, *J. Med. Imaging and Health Inf.*, 2012, 2, 244-249. <http://dx.doi.org/10.1166/jmihi.2012.1088>
- [24] Chattopadhyay S.; Banerjee S.; Rabhi F.A.; Acharya R. U. A Case-based Reasoning System for Complex Medical Diagnoses, *Expert Systems: the Journal of Knowledge Engineering*, 2013, 30(1), 12-20. DOI: [10.1111/j.1468-0394.2012.00618.x](https://doi.org/10.1111/j.1468-0394.2012.00618.x)
- [25] Chattopadhyay S.; Acharya U. R. A Novel Mathematical Approach to Diagnose Premenstrual Syndrome, *Journal of Medical Systems*, 2012, 36(4), 2177-2186. DOI: [10.1007/s10916-011-9683-4](https://doi.org/10.1007/s10916-011-9683-4)
- [26] Smit, F.; Van Hout, H.; Van Oppen, P.; Van Der Horst, H.; Beekman, A.; Van Marwijk, H.; Cost-effectiveness of a stepped care intervention to prevent depression and anxiety in late life: randomized trial, *BJPsych*, 2010, 196, 319-325. DOI: [10.1192/bjp.bp.109.069617](https://doi.org/10.1192/bjp.bp.109.069617)
- [27] NIMH : The Numbers Count : Mental Disorders in America URL : <http://www.nimh.nih.gov/health/publication/s/the-numbers-count-mental-disorders-in-america/index.shtml> (last accessed on 12/05/2014).

- [28] WHO : Depression URL : <http://www.who.int/mediacentre/factsheets/fs369/en> (last accessed on 12/05/2014).
- [29] Hui, N. B.; Mahendar, V.; Pratihar, D. K. Time-optimal collision-free navigation of a car-like robot using a neuro-fuzzy approach,

Fuzzy Sets & Systems, 2006, 157, 2171-2204. DOI: [10.1016/j.fss.2006.04.004](https://doi.org/10.1016/j.fss.2006.04.004)