**Vol 8. No. 3(2), October 2015**
**African Journal of Computing & ICT**
**ISSN 2006-1781**



# Automatic Diagnosis of Depressive Disorders using Ensemble Techniques

**[1]B. Ojeme, [2]M. Akazue & [3]E. Nwelih**
[1]Department of Computer Science, University of Cape Town, South Africa
[2]Department of Mathematics and Computer Science, Delta State University, Nigeria
[3]Department of Computer Science, University of Benin, Nigeria.
[1]ojeme@yahoo.com, [2]akazuem@gmail.com, [3]emmanuelnwelih@yahoo.com

Corresponding author: Ojeme Blessing, ojeme@yahoo.com

## ABSTRACT

Depression is widespread and often undiagnosed or misdiagnosed, globally, because of acute shortage of mental health professionals and its high comorbidity with other disorders. Though various classification algorithms have been used on depression datasets, and high classification accuracies reported in the past decade, studies have shown that more than 90% of people who suffer from depressive symptoms in the developing countries of Africa do not have diagnostic facilities and treatment. To overcome the difficulties, this paper reports the preliminary findings of a study to investigate the use of ensemble techniques for the automatic diagnosis of depression using a dataset of 580 patients (severe depression = 271, mild depression = 23, moderate depression = 272 and no-depression = 14), collected from the University of Benin Teaching Hospital-UBTH and Primary care centre in Nigeria. The performance analyses and results obtained with the machine learning algorithms, trained independently and jointly with different combinations, are discussed using varous performance metrics. The area under the receiver operating characteristics curve-AUC for ensemble classifiers shows a remarkable improvement over the individual classifers.

**Keywords**: Machine learning, esensemble techniques, depression disorders, Mental health.

African Journal of Computing & ICT Reference Format:
B. Ojeme, M. Akazue & E. Nwelih (2015): Automatic Diagnosis of Depressive Disorders using Ensemble Techniques.
Afri J Comp & ICTs Vol 8, No.3 Issue 2 Pp 31-38.

## 1 INTRODUCTION

In the past decade, various classification algorithms have been used on depression datasets and high classification accuracies have been achieved (Das & Sengur, 2010; Sacchet, Prasad, Foland-Ross, Thompson, & Gotlib, 2015; West, Mangiameli, Rampal, & West, 2005). In machine learning community, Ensemble methods are learning models that improve predictive performance by combining the opinions of multiple learning models (Daumé III, 2012). Its main advantage is the unlielihood of all the models used to make the same mistake. Ensemble methods have been used extensively for medical diagnosis (Das & Sengur, 2010; West et al., 2005). Preotiuc-Pietro (Preotiuc-Pietro, Sap, Schwartz, & Ungar, 2015) used a combination of different classifiers to determine Twitter users who self-reported having either Post-traumatic stress disorder (PTSD) or depression and achieved a high accuracy.

### 1.1 MACHINE LEARNING ALGORITHMS FOR DEPRESSION

Machine Learning (ML) is simply the training of a model from data that generalizes a decision against a performance measure. ML algorithms have been successfully applied in many fields including medical diagnosis, spam detection, credit card fraud detection, digit recognition, speech understanding, face detection, product recommendation, customer segmentation and shape detection (Witten, Frank, & Hall, 2011). Common ML problems are classification, regression, clustering and rule extraction. Some commonly used ML algorithms for solving real-world problems, such as depression diagnosis, include Bayesian networks, Artificial neural networks, Support vector machines, K-means, Decision tree and Random forest (Witten et al., 2011) are discussed briefly.

Bayesian networks: Bayesian networks is a probabilistic reasoning tool for managing imprecision of data and uncertainty of knowledge in real-world problems. Bayesian networks is constructed, either by hand (manually) or by software (from data). As a real-world problem-solving tool, Bayesian networks have been used to address problems in different areas of medicine. Curiac (Curiac, Vasile, Banias, Volosencu, & Albu, 2009) presented a Bayesian network-based analysis of four major psychiatric diseases: schizophrenia (simple and paranoid), mixed dementia (Alzheimer disease included), depressive disorder and maniac depressive psychosis.

Artificial neural network: Artificial neural network (ANN), a mathematical representation of the human neural systems is efficient in modelling and making sense of real-world clinical data in which the relationship among the variables is unknown or complex (Amato et al., 2013). ANN is quite helpful in real-world problems that do not have algorithmic solution or when there is need to pick out interesting structures from existing data. In psychiatric diagnosis, Mukherjee et al (Mukherjee, Ashish, Hui, & Chattopadhyay, 2014) used Back propagation feed forward neural network (BPFFNN) and Radial basis function neural network (RBFNN) models to detect depression. Training the models with 45 real-life medical data instances showed that the two approaches obtained the same diagnostic efficiency as clinicians.

Fuzzy logic: Fuzzy logic (FL) is a set of mathematical principles for knowledge representation that allows intermediate values to be defined between conventional binary logic like true/false, yes/no, high/low (Hasan, Sher-e-alam, & Chowdhury, 2010). Being a multivalued logic, FL imitates human reasoning sense and deals with situations when we have just one item which partly belongs to one class and partly to another. Abdullah et al (Abdullah, Zakaria, & Mohamad, 2011) proposed a design a FuzzyExpert System (FES) for the diagnosis of hypertension risk for patients aged between 20's, 30's and 40's years, divided on gender line. The proposed system, used Mamdani inference method and when tested with data collected from 10 persons with different work background was found to provide a faster, cheaper and more reliable diagnostic results compared to the traditional methods

Support Vctor Machines: A support vector machine (SVM) is a way of performing classification by finding a separating boundary (hyperplane) that separates the data into two categories (Daumé III, 2012). SVM offers a possibility to find solution to real-world problems, such as depression diagnosis, that cannot be linearly separated in the input space by making a non-linear transformation of the original input space into a high dimensional feature space, where an optimal separating hyperplane can be found. Sacchet (Sacchet et al., 2015) conducted an analysis to differentiate the depressed individuals from healthy controls using SMV in conjunction with structural global graph metrics. Data was obtained from multiple brain network properties of 32 (14 diagnosed with MDD) participants, all women aged 18-55 years, at the Stanford Center for Neurobiological Imaging. The SVM model, when tested, was able to diagnose depression with 71.88% accuracy, 71.43% sensitivity and 72.22% specificity.

K-nearest neighbor: The K-nearest neighbor (KNN) presents a simple but effective means of making classification decisions. KNN performs prediction by finding a training example V that is most similar to the test example $\tilde{V}$. Ghasemi and Khalili (Ghasemi & Khalili, 2014) conducted a research to compare the predictive strengths of multilayer perceptron (MLP) and K-nearest neighbour (KNN) for the diagnosis of bipolar disorder. With 70% of the available data used for training the models, the diagnostic results showed the superiority of the MLP model with 16% error rate to that of the KNN with 21% error rate.

## 2.0 PROBLEM STATEMENT

Depression is one of the most common psychiatric disorders globally. Depression is difficult to detect by clinicians because it shares symptoms with other physical and/or mental disorders (Chattopadhyay, 2014). The World Health Organisation (WHO, 2012) has shown that more than 90% of people who suffer from depressive symptoms in the developing countries of Africa do not have access to diagnostic facilities and treatment.

In most developing countries like Nigeria, with scarce mental health services (Ganasen et al., 2008), traditional diagnostic practice in depression services typically involves clinician-to-patient interview where judgments are made from the patient's appearance and behaviour, subjective self-reported symptoms, depression history, and current life circumstances (Baasher, Carstairs, Giel, & Hassler, 1975). The views of relatives or other third parties may be taken into account. A physical examination to check for ill health, the effects of medications or other drugs may be conducted. This intuitive model, though still in use today, is slow and leaves diagnostic decision-making entirely to the subjective clinical skills and opinion of the clinicians (Chattopadhyay, 2014).

## 3 METHODOLOGY AND DATA COLLECTION

This study seeks to investigate the strenght of ensemble techniques to automatically detect depression in Nigeria and other developing countries. The steps taken to achieve the objectives are as follows:

a) Collect depression data from the mental unit of the university of Benin Teaching Hospital (UBTH) and Primary health centre in Nigeria.
b) Extract the features (symptoms of depression).
c) Build ensemble models using Weka (Waikato Environment for Knowledge Analysis), a popular, free machine learning tool (Bouckaert et al., 2013).
d) Test the performance of the built model on a set of real-world depression cases

The data used for training the machine learning algorithms consisted of 580 data instances and 23 attributes collected from the UBTH and primary health centre in Nigeria. There were 254 male and 326 female patients from 12 to 92 years old (with a mean age of 41.8 and standard deviation of 16.3). The features shown in Table 1 were identified as relevant for the screening and diagnosis of depression.

**Vol 8. No. 3(2), October 2015** **ISSN 2006-1781**

**African Journal of Computing & ICT**



**Table 1. Features of depression extracted from the dataset**

| S/N | Features | code | Data type |
|---|---|---|---|
| 1 | age | ag | integer |
| 2 | sex | se | Integer |
| 3 | marital status | ms | Integer |
| 4 | sad mood | sm | integer |
| 5 | suicidal | su | Integer |
| 6 | loss of pleasure | lp | Integer |
| 7 | insomnia | in | Integer |
| 8 | hypersomnia | hy | Integer |
| 9 | loss of appetite | la | Integer |
| 10 | psychomotor agitation | pa | Integer |
| 11 | psychomotor retardation | pa | Integer |
| 12 | loss of energy | le | Integer |
| 13 | feeling of worthlessness | fw | Integer |
| 14 | lack of thinking | lt | Integer |
| 15 | indecisiveness | id | Integer |
| 16 | recurrent thoughts of death | rt | Integer |
| 17 | impaired function | if | Integer |
| 18 | weight gain | wg | Integer |
| 19 | weight loss | wl | Integer |
| 20 | stressful life events | sl | Integer |
| 21 | financial pressure | fp | Integer |
| 22 | depression in family | df | Integer |
| 23 | employment status | es | Integer |
| 24 | depression diagnosis | | nominal |
| 25 | comorbidity | | nominal |
| 26 | treatment | | nominal |

### 3.1 Proposed Ensemble Techniques

Stratified cross validation technique (Witten et al., 2011) was used to split the dataset of 580 patients (severe depression = 271, mild depression = 23, moderate depression = 272 and no-depression = 14), into three equal folds, in which two-thirds (387) of the dataset was used for training the model while the remaining one-third (193) was used for testing. This procedure was repeated three times to ensure an even representation in training and test sets. Weka provided the platform for the data analysis, preparation, model testing and result evaluation shown in Fig. 1.

The five machine learning algorithms under study (Bayesian networks, Back-Proagation Multi-layer perceptron, Support vecor macines, K-nearest neighbor, and Fuzzy logic) were trained in the same manner, separately and then jointly.

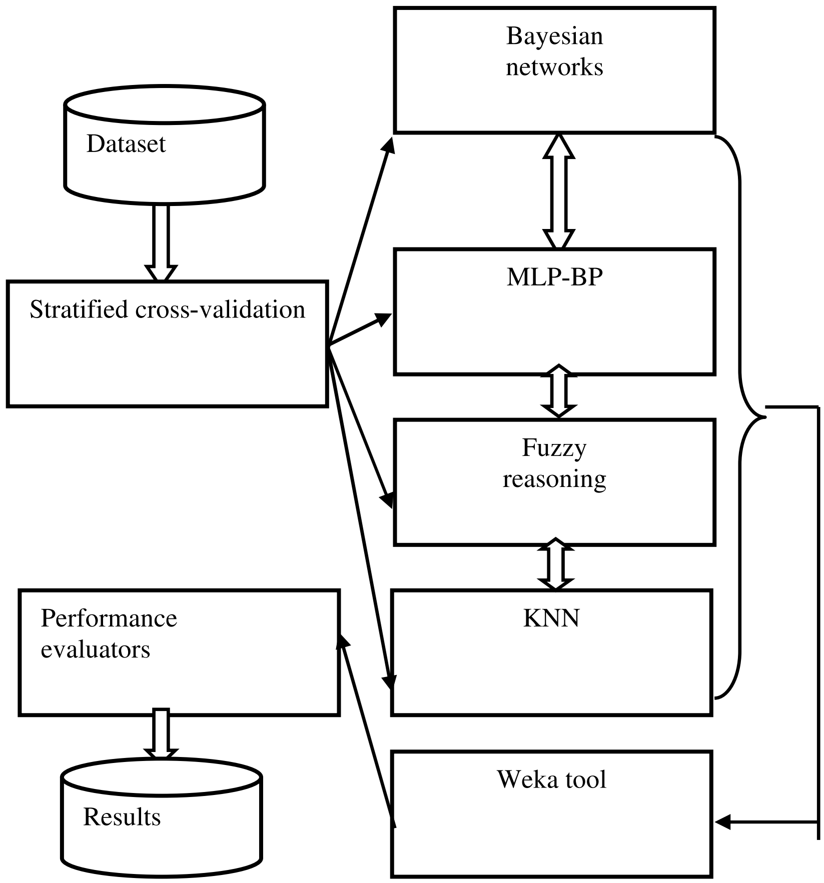


**Fig 2. Ensemble model for depression diagnosis**

## 4. RESULTS AND DISCUSSION-

### Analysis of the Performance of Proposed Techniques

The results of the machine learning techniques, trained independently, and jointly, are presented in Tables 2, 3, 4, 5 and 6. The ensemble methods, in different combinations, show minor, but consistent improvement over the scores of each individual classifier. Matthews correlation coefficient (MCC) calculated all four values (TP, FN, FP and TN) of the confusion matrix. Receiver operating characteristics (ROC) provided the area under the curve (AUC) of the plot of the true positive rate (y-axis) against the false positive rate (x-axis). An excellent classifier will have ROC area values between 0.9 and 1.0 while a poor classifier will have ROC area values between 0.6 and 0.7 (Saito & Rehmsmeier, 2015). Similar to ROC, precision provided a very powerful way of evaluating the performance of the ensemble classifiers given the imbalanced dataset used for the study. A precision of 0.876 is interpreted as 87.6% correct predictions among the positive predictions.

One perspective for future improvements is to increase the size of the dataset and modify the model to separate patients having other diseases in addition to depression. Another direction for possible improvement to the model is to reduce the dimensionality of the attributes (features) using Principal component analysis (PCA).

**Table 2.** Results of independent classifiers.

| | TPR | FPR | Prec | F-Score | MCC | ROC Area |
|---|---|---|---|---|---|---|
| BN | 0.902 | 0.084 | 0.876 | 0.885 | 0.825 | 0.975 |
| MLP | 0.938 | 0.047 | 0.921 | 0.925 | 0.895 | 0.971 |
| SVM | 0.910 | 0.079 | 0.855 | 0.881 | 0.831 | 0.916 |
| FL | 0.926 | 0.064 | 0.928 | 0.907 | 0.872 | 0.951 |
| KNN | 0.947 | 0.028 | 0.946 | 0.946 | 0.919 | 0.959 |

**Table 3. Two-classifier combination results**

| | TPR | FPR | Prec | F-Score | MCC | ROC Area |
|---|---|---|---|---|---|---|
| BN+ MLP | 0.931 | 0.056 | 0.898 | 0.912 | 0.877 | 0.982 |
| BN+ SVM | 0.910 | 0.079 | 0.855 | 0.881 | 0.831 | 0.975 |
| BN+ FL | 0.931 | 0.059 | 0.935 | 0.913 | 0.881 | 0.980 |
| BN+ KNN | 0.947 | 0.028 | 0.946 | 0.946 | 0.919 | 0.989 |
| MLP+ SVM | 0.910 | 0.079 | 0.855 | 0.881 | 0.831 | 0.973 |
| MLP+ FL | 0.931 | 0.059 | 0.935 | 0.913 | 0.881 | 0.978 |
| MLP+ KNN | 0.947 | 0.028 | 0.946 | 0.946 | 0.919 | 0.988 |
| SVM+ FL | 0.919 | 0.071 | 0.901 | 0.892 | 0.853 | 0.955 |
| SVM+ KNN | 0.924 | 0.054 | 0.901 | 0.912 | 0.869 | 0.971 |
| FL+ KNN | 0.950 | 0.028 | 0.949 | 0.949 | 0.924 | 0.980 |

**Table 4. Three-classifier combination results**

| | TPR | FPR | Prec | F-Score | MCC | ROC Area |
|---|---|---|---|---|---|---|
| BN+ MLP+ SVM | 0.929 | 0.062 | 0.896 | 0.909 | 0.872 | 0.982 |
| BN+ MLP+ FL | 0.934 | 0.058 | 0.900 | 0.914 | 0.881 | 0.984 |
| BN+ MLP+ KNN | 0.943 | 0.046 | 0.936 | 0.933 | 0.905 | 0.991 |
| BN+ SVM+ FL | 0.924 | 0.067 | 0.891 | 0.903 | 0.862 | 0.979 |
| BN+ FL+ KNN | 0.933 | 0.055 | 0.917 | 0.916 | 0.884 | 0.991 |
| MLP+ SVM+ FL | 0.929 | 0.062 | 0.895 | 0.906 | 0.872 | 0.977 |
| MLP+ FL+ KNN | 0.943 | 0.046 | 0.935 | 0.932 | 0.905 | 0.991 |
| SVM+ FL+ KNN | 0.929 | 0.061 | 0.932 | 0.910 | 0.878 | 0.981 |

**Table 5. Five-classifier combination results**

| | TPR | FPR | Prec | F-Score | MCC | ROC Area |
|---|---|---|---|---|---|---|
| BN+ MLP+ SVM+ FL | 0.929 | 0.062 | 0.895 | 0.908 | 0.872 | 0.984 |
| BN+ MLP+ SVM+ KNN | 0.931 | 0.059 | 0.934 | 0.913 | 0.881 | 0.990 |
| MLP+ SVM+ FL+ KNN | 0.933 | 0.058 | 0.936 | 0.914 | 0.884 | 0.990 |

**Table 6.** Four-classifier combination results

| | TPR | FPR | Prec | F-Score | MCC | ROC Area |
|---|---|---|---|---|---|---|
| BN+ MLP+ SVM+ FL+ KNN | 0.933 | 0.059 | 0.899 | 0.912 | 0.878 | 0.991 |

## 5. CONCLUSIONS AND FUTURE WORK

The diagnosis of depression still remains a major challenge because of its high comorbid factor and acute shortage of mental health professionals. Ensemble techniques, consisting of Bayesian networks, Back-Propagation MLP, SVM, Fuzzy logic and Nearest neighbour, was used to improve the needed diagnosis and prediction accuracy of depression. Though recommendations cannot be made at this stage of the research, the results from the algorithms presented offer a foundation for preliminary conclusions. It suggest that, even though the algorithms achieved high accuracy when used independently, utilizing them jointly creates a better system to support clinical decisions in predicting the level of risks of depressive disorders.

## 6. RESEARCH IMPLICATIONS AND FUTURE WORKS

There have been rampant cases of missed diagnosis of depression in Nigeria, leading to ineffective treatment and increased burden of the disorders on the sufferers. The proposed model will support clinical decisions in the diagnosis of depression.

The future scope of this work would be to modify the model to separate patients having other diseases in addition to depression.

## REFERENCES

Abdullah, A. A., Zakaria, Z., & Mohamad, N. F. 2011. Design and Development of Fuzzy Expert System for Diagnosis of Hypertension. *2011 Second International Conference on Intelligent Systems, Modelling and Simulation*, 113–117. http://doi.org/10.1109/ISMS.2011.27

Amato, F., López, A., Peña-Méndez, E. M., Vaňhara, P., Hampl, A., & Havel, J. 2013. Artificial neural networks in medical diagnosis. *Journal of Applied Biomedicine*, *11*(2), 47–58. http://doi.org/10.2478/v10136-012-0031-x

Baasher, T. A., Carstairs, G. M., Giel, R., & Hassler, F. R. 1975. Mental health services in developing countries. In *WHO seminar on the organisation of Mental Health Services*. World Health Organisation, Geneva. Retrieved from http://whqlibdoc.who.int/offset/WHO_OFFSET_22_(pt1-pt3).pdf

Bouckaert, R. R., Frank, E., Hall, M., Kirkby, R., Reutemann, P., Seewald, A., & Scuse, D. 2014. *WEKA Manual for Version 3-7-12*. University of Waikato, Hamilton, New Zealand. Retrieved from papers3://publication/uuid/24E005A2-AA1B-4614-BAF5-4D92C4F37413

Chattopadhyay, S. 2014. A neuro-fuzzy approach for the diagnosis of depression. *Elsevier: Applied Computing and Informatics*, *In Press*(February), 19. http://doi.org/10.1016/j.aci.2014.01.001

Curiac, D.-I., Vasile, G., Banias, O., Volosencu, C., & Albu, a. 2009. Bayesian network model for diagnosis of psychiatric diseases. *Proceedings of the ITI 2009 31st International Conference on Information Technology Interfaces*, 61–66. http://doi.org/10.1109/ITI.2009.5196055

Das, R., & Sengur, A. 2010. Evaluation of ensemble methods for diagnosing of valvular heart disease. *Expert Systems with Applications*, *37*(7), 5110–5115. http://doi.org/10.1016/j.eswa.2009.12.085

Daumé III, H. (2012). *A Course in Machine Learning*. Department of computer science, University of maryland.

Ganasen, K. A., Parker, S., Hugo, C. J., Stein, D. J., Emsley, R. A., & Seedat, S. 2008. Mental health literacy: focus on developing countries. *African Journal of Psychiatry*, *11*(1), 23–28. http://doi.org/10.4314/ajpsy.v11i1.30251

Ghasemi, M. M., & Khalili, M. 2014. A hybrid KNN-MLP algorithm to diagnose bipolar disorder. *Journal of Advanced Computer Science & Technology*, *4*(1), 1–5. http://doi.org/10.14419/jacst.v4i1.3922

Hasan, M. A., Sher-e-alam, K., & Chowdhury, A. R. 2010. Human Disease Diagnosis Using a Fuzzy Expert System. *Journal of Computing*, *2*(6), 66–70.

Mukherjee, S., Ashish, K., Hui, N. B., & Chattopadhyay, S. 2014. Modeling Depression Data : Feed Forward Neural Network vs . Radial Basis Function Neural

Network. *American Journal of Biomedical Sciences*, *6*(3), 166–174.

Preotiuc-Pietro, D., Sap, M., Schwartz, H. A., & Ungar, L. 2015. Mental Illness Detection at the World Well-Being Project for the CLPsych 2015 Shared Task. In *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality, Denver, Colorado, USA, June. North American Chapter of the Association for Computational Linguistics.*

Sacchet, M. D., Prasad, G., Foland-Ross, L. C., Thompson, P. M., & Gotlib, I. H. 2015. Support Vector Machine Classification of Major Depressive Disorder Using Diffusion-Weighted Neuroimaging and Graph Theory. *Frontiers in Psychiatry*, *6*(February), 1–10. http://doi.org/10.3389/fpsyt.2015.00021

Saito, T., & Rehmsmeier, M. 2015. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PloS One*, *10*(3), e0118432. http://doi.org/10.1371/journal.pone.0118432

West, D., Mangiameli, P., Rampal, R., & West, V. 2005. Ensemble strategies for a medical diagnostic decision support system: A breast cancer diagnosis application. *European Journal of Operational Research*, *162*(2), 532–551. http://doi.org/10.1016/j.ejor.2003.10.013

WHO. 2012. Depression: A global public health concern. Available at http://www.who.int/mental_health/management/depression/who_paper_depression_wfmh_2012.pdf. *WHO Department of Mental Health and Substance Abuse*, 1–8. Retrieved from http://www.who.int/mental_health/management/depression/who_paper_depression_wfmh_2012.pdf

Witten, I., Frank, E., & Hall, M. 2011. *Data Mining : Practical Machine Learning Tools and Techniques* (Third Edit). Morgan KaufmannPublishers. http://doi.org/10.1002/1521-3773(20010316)40:6<9823::AID-ANIE9823>3.3.CO;2-C