# Prediction of Depression among Senior Citizens using Machine Learning Classifiers

Ishita Bhakta West Bengal University of Technology Sector-1, Salt Lake, Kolkata West Bengal, India, 700064

# ABSTRACT

Depression among elderly population is an emerging problem of public health. Various socio demographic factors like age, sex, earning status, living spouse and family type etc are responsible for depression among senior people. Some co morbid conditions like visual problem, hearing difficulties, mobility problem also influence the disease. But depression can be diagnosed at earliest using predictive modeling with various influencing input variables. WEKA is a data mining tool used for prediction based on machine learning classifiers. In this paper five machine learning classifiers are compared with respect to three test options. A best method for depression prediction in aged persons also has been chosen among these five methods through comparison study.

# **General Terms**

Machine learning algorithm, Depression, Senior Citizens.

#### **Keywords**

Bayes Net classifier, Decision Table, Depression Prediction, Multi-Layer Perceptron classifier, Logistic Model, Sequential Minimal Optimization (SMO) classifier.

# 1. INTRODUCTION

In the year 1999, Government of India had declared 'senior citizen' as a person who is 60 years old or above according to National Policy on Older Persons [1]. Data from 2011 census in India indicates that 8% of Indian population is under senior citizen group. India is second most populated country in the world in terms of senior citizen [2]. It is expected that in the year 2050 the population will increase up to 20% [3]. Ageing causes internal and external changes in human body structure and functionality. Aging brain is more susceptible to depression [4]. Depression causes sadness, loneliness, sense of social separation, helplessness, ADL (Activities of Daily Living) limitation, dissatisfaction towards life, memory loss and increases suicidal tendency. The symptoms of depression are generally overlooked among this age group as they coincide with other late life problems. An episode of depression is serious if it lasts for more than two weeks. In India, one in every four elderly persons is suffering from depression. A detailed understanding of the problem of depression among the senior citizens helps the health planner and policy maker to design and implement appropriate health care delivery system. This research work will help in building a predictive model to diagnose depression among senior citizens considering various socio demographic parameters and co morbid condition. Using an appropriate predictive model overhead of manual diagnosis process can be reduced and treatment can be initiated at earliest.

Artificial intelligence introduces automation in daily life reducing manual labor using machines and their functioning software intelligently. For this intelligent automation machine Arkaprabha Sau R.G.Kar Medical College and Hospital 1 Khudiram Bose Sarani, Kolkata West Bengal, India, 700004

should be learned with pattern just like human brain. Various optimizations, statistical techniques are used to learn a machine from past experiences so that it can detect complex patterns from large and complicated data sets. All these learning procedures are collectively called machine learning algorithms. Medical condition of human being depends on various complex parameters. For analysis of these complex data machine learning techniques are very useful. Automation in medical field especially in diagnosis will be helpful for patient and medical practitioner. Ultimately this will help to provide a quality healthcare service to the patient. Now-adays machine learning techniques are used to detect benign and malignant tumor [5], predict cancerous mole, classify malignancies, measure degree of depression among adolescents [6], predict chronic kidney diseases [7,8], predict parkinson disease [9], predict heart diseases [10,11] and predict HIV [12]. In [13], authors identify depression among adolescent comparing two types of depression screening tools. Different research works are currently available to predict and detect depression based on facial images, speech tone and other features. But most of the research works concentrate on a particular learning method. There may be case where other learning methods can perform better. To find accurate prediction model comparison analysis of various learning methods is needed. There are very few works found on depression prediction among senior citizens in the literature. This study concentrates on identifying depression among senior citizens more accurately by comparing different predictive model using machine learning methodologies.

University of Waikato in New Zealand develops a software named WEKA that combines different machine learning techniques to apply them on data mining problems like data Forecasting, Classification, preprocessing, Prediction, regression etc [14]. In this research work WEKA is used for classification and prediction of depression among senior citizens. Six different machine learning classifiers are used for comparison study - Bayes Net, Logistic, Multilayer Perceptron, SMO and Decision Table. All these classifiers are used on three different test options - Using training and testing set, Cross-validation and percentage split. From comparison result best predictive model for depression prediction among senior citizens has been chosen. In current social structure depressions among senior citizens remain undiagnosed due to negligence and complicated manual diagnosis using geriatric depression scale. So if there is a model that can diagnose depression automatically from socio demographic and morbidity attributes then that will be very helpful to detect depression accurately at earliest.

Rest of the paper is organized as – Section 2 describes predictive modeling using machine learning classifier, Section 3 describes prediction methodology, Comparison Results are discussed in Section 4 and finally Section 5 conclude the work.

# 2. PREDICTIVE MODELING USING MACHINE LEARNING CLASSIFIER

To predict valuable information from nominal and numerical dataset WEKA Classifiers are used. Machine learning is a set of computational algorithms in artificial intelligence where learning is done from experience on a particular set of tasks or data to improve performance. They are modeled using machine learning methodologies. There are three main types of machine learning - Supervised, Unsupervised and Reinforcement learning [15]. In supervised learning output label of training data set is known and hence that knowledge is applied in learning process. In unsupervised learning output label is unknown. Different classifiers can be used to train the network and predict decisions in WEKA like tree based classifier, rule based classifier, meta classifier, multi layer perceptron, support vector machine, bayes net classifiers etc. Supervised learning process is appropriate for this research work as output label for training data set is known. Only output of a test data set needs to be predicted based on supervised learning. For prediction model five classifiers are compared to choose best one for depression prediction among senior citizens. Some basic of all these classifiers are discussed in this section.

#### 2.1 BayesNet Classifier (BN)

A Bayesian network [16] is a probabilistic directed acyclic graph composed of a set of nodes and a set of edges between nodes. Node represents the random data and edge between two nodes represents conditional dependency among nodes. Learning process of Bayesian network is a two steps process – learning network, learning relationship among data. Learning can be done in WEKA using algorithms like local score metrics, conditional independence tests, global score metrics, cross-validation and fixed structure. Search algorithms that are used in these learning processes in WEKA are tabu search, simulated annealing and hill climbing. After learning the network the conditional relationship among data has been drawn using probability tables.

#### 2.2 Logistic

Logistic classification is applicable on the data set where dependent and independent attributes are dichotomous. The data set used for depression prediction has binary outcome so it cannot be modeled using linear regression. For such data logistic regression [17] is required. There are two set of experiments. One set produce positive result that is attribute set for which depressed people are recognized and another one is negative result that detects not depressed people. Logistic function (Eq. 1) is used in this model to predict the output of an experiment.

$$f(r) = \frac{1}{1+e^{-r}} \qquad \qquad \text{Eq. 1}$$

Each attribute or variable or feature has some contribution in prediction of expected outcome in a data set. Predictive capability of each attribute is measured using maximum likelihood estimation statistics. Logistic model calculate the probability of prediction of a binary outcome using input data set. This model use likelihood radio and Wald test to test statistical significance.

# 2.3 Multilayer Perceptron (MLP)

Perceptron is the earliest and simplest neural network model used for classification, pattern recognition etc consisting of a set of weighted inputs and single output [18]. Perceptron accept binary or bipolar input but output is always in bipolar form. Multilayer perceptron consists of multiple layers. Each layer contains some unit or neuron. Input sum is calculated for each unit. Activation of output unit depends on a threshold value. If input sum is greater than a threshold value then output will be activated otherwise remains deactivate. All these computations will be continued with change weight until perceptron output become equal to target output. Most challenging part of this model is determining application specific number of hidden layers and neurons in each layer.

#### 2.4 SMO

Sequential Minimal Optimization (SMO) is used to train support vector machine (SVM) [19]. A large quadratic programming (QP) optimization solution is required to train a SVM. SMO provides a simple solution of this optimization problem. It breaks the large problem into some small parts and solves each part analytically. It makes computation faster for sparse data set than other training algorithms in SVM. SMO can handle a large set of data as memory requirement is linear, not quadratic.

# 2.5 Decision Table (DT)

A decision table [20] is a tree based classifier used for predictive modeling. It has two part – algorithm to design decision table and visualization part to represent graphical interface of the model. It generates a hierarchical structure of input data set depending on relation among those data. Algorithm part chooses most important attribute set from original data set and thus reduces the possibilities of over estimation. From this compact and smaller decision table output decision is taken for classification over setting decision rules among selected attributes.



Fig 1: Flow chart of predictive modeling

The general flow chart for predictive modeling is described in Fig 1. Above six classifications model are used for training, classification and prediction. All the five classifiers are applied in step "Choose Classifier Model and Predictive Model". Results are carried out for all these five classifiers and compared to choose best predictive model for depression prediction.

# 3. PREDICTION METHODOLOGY

This section describes the steps followed in this research paper for predictive modeling. Here WEKA is used as machine learning platform. Fig 2 gives a pictorial view of prediction methodology. At first the training data set is loaded

International Journal of Computer Applications (0975 – 8887) Volume 144 – No.7, June 2016

into WEKA then filters are applied on this data set for optimized result. After pre-processing, result is fed into the classifiers and corresponding test options are selected. Prediction output format is chosen to view the result in human readable format. After all these setup prediction is started and result is collected in a text file.

#### 3.1 Pre-processing

After loading input data set, filter is applied from preprocessing window. A supervised filter – attributeSelection is applied in this research work to select optimized attributes. The screenshot of WEKA preprocessing step is given in Fig 3.



Fig 2: Steps to predict the depression in WEKA



Fig 3: AttributeSelection Filter in preprocessing in WEKA

# 3.2 Classification

After applying filter, input data set is fed into classify window. After that classifiers are chosen for classification. Here five different classifiers are chosen and run the experiments five times and classification output is collected each time. Fig 4 gives an idea about how to choose classifiers in WEKA.

# 3.3 Test Option

Result of classification and prediction depends on test option selected in WEKA. There are three test options available. A brief description of each of these options is given here.

#### 3.3.1 Using Training and Testing Set

In this test option network is trained using a training set. After learning, the network is tested upon a test data set. Result is shown in classifier output window. Purpose of this test option is to verify how correctly the learned network classify and predict decision class in test data set using decision making attributes nature.

#### 3.3.2 Cross-validation

In training and testing set option both data sets are given explicitly. But in cross-validation option test data set is prepared from training set automatically using number of folds provided to limit over fitting problem. It is useful for those applications where data collection for test data set is difficult. If the number of fold is 10 then this method first partition the whole data set into two parts with no common data. Test set contains 10 samples and training set contains other 90 samples. In next iteration from these 90 samples 10 are selected as test data and former data set is included in training set. In this way variability is reduced using different partition. At the end all the results collected from different iterations are averaged and final result is shown. In other test option root mean square error is not considered for estimation of predictive model performance but cross-validation includes error for performance measurement as a result prediction result of unknown data is more accurate.

#### 3.3.3 Percentage Split

In this test option training data set is used for both testing and training. Test set contains a certain percentage of whole data set. Test set percentage is taken as user input. Suppose someone wants 30% of the training set will be used as test set. Then it should be specified in the text box. Training and testing sets are distinct. If there are 100 data then 70 data will be used for training and rest 30 data for testing.

# 3.4 Prediction

After setting all the pre-requisite of predictive model now output option for prediction is chosen from 'output prediction' box of 'more options' button. For this research work 'plaintext' is chosen as output format. Fig 5 gives an overview of how to set output format for prediction. Result of prediction is summarized in output window.



#### Fig 4: Choose classifiers in WEKA





# 4. RESULT AND DISCUSSION

For experiment WEKA 3.8.0 is used here. Training data set is collected from a slum at Bagbazar, Kolkata which is the service area of Bagbazar Urban Health and Training Centre (UHTC) and it is also the urban field practice area of Department of Community Medicine, R. G. Kar Medical College and Hospital, Kolkata, India. All Individuals aged 60 years & above residing under the service area of Bagbazar UHTC were included in the study. Data was collected from 1st April 2016 to 30th April 2016. Total sixty senior citizens were interviewed using Geriatric Depression Scale (GDS) [21] to collect training data. Five Classifiers are compared with respect to four metrics – Accuracy, ROC area, Precision and Root Mean Square Error (RMSE). Test data set is prepared with data of ten more senior citizens with no

decision class. The result of prediction is validated manually using GDS scale. A good model should have high accuracy and high precision. ROC area indicates goodness of a predictive model. For a good predictive model ROC area should be close to one and root mean square error should be close to zero. Table 1 shows the result of five classifiers using training and test set with respect to four metrics. From the values of these metrics it is clear that SMO is more accurate and précised model for prediction of depression among senior citizens. But according to the value of ROC area and RMSE BayesNet is better than other four classification model.

	Accuracy	Precision	ROC	RMSE
Classifier	(%)		Area	
BN	91.67	0.92	0.98	0.25
Logistic	76.78	0.78	0.89	0.49
MLP	87	0.88	0.96	0.32
SMO	93.33	0.94	0.94	0.26
DT	90	0.90	0.95	0.29

Table 1. Result of prediction with training and testing set

Now if predictive model is designed with five classifiers using 10 fold cross validation then values of these four metrics are shown in Table 2. From the data of Table 2 it is clear that SMO gives high accuracy and high precision and BayesNet gives better ROC area and lower RMSE. From Table 1 and Table 2 data set it is clear that if network for depression prediction model is learned with training and test data set or cross-validation then some inconsistency of result observed. As for two metrics BayesNet classifier gives better result and for other two metrics SMO gives better result. But prediction model should be consistent with respect to all metrics.

 Table 2. Result of prediction with Cross-validation

Metrics Classifier	Accuracy (%)	Precision	ROC Area	RMSE
BN	86.67	0.87	0.96	0.32
Logistic	78.33	0.79	0.85	0.46
MLP	85 %	0.85	0.94	0.33
SMO	88.33 %	0.88	0.88	0.34
DT	80 %	0.80	0.86	0.38

Metrics Classifier	Accuracy (%)	Precision	ROC Area	RMSE
BN	95	0.95	0.99	0.22
Logistic	80	0.81	0.88	0.45
MLP	90	0.92	0.98	0.32
SMO	90	0.92	0.89	0.32
DT	95	0.95	0.96	0.25

Table 3. Result of prediction with Percentage Split

This consistency found in model learned with percentage split option. Table 3 gives the result of predictive model using five different classifiers with percentage split test option. Here 66% of the original data set is used for training and 34% for testing. It is the default split value in WEKA. As it gives optimum and best result so that percentage has been considered for experiment. In this test option among five classifiers BayesNet gives best result for all these four metrics.

So for this type of variable data set to predict depression among senior citizens percentage split is the best test option and BayesNet is the best classifier. Though BayesNet and DecisionTable provide more or less same result according to Accuracy, Precision and ROC Area, but according to root mean square error BayesNet is less error prone than Decision Table.

# 5. CONCLUSION

Depression makes Senior citizens inactive mentally and physically. It is also a common cause of mortality and morbidity among senior citizens. Prediction of depression from socio demographic and morbidity variables helps physician to detect depression as early as possible. Currently it is done manually using GDS scale. Automated predictive modeling makes the treatment procedure faster and will be helpful for medical professionals. In this research work five different learning methods – Bayes Net, Logistic, Multi Layer Perceptron, SMO and Decision Table are applied for predictive modeling. Bayes Net gave better result for percentage split testing among them.

In future other learning methods can be tried out for better accuracy. Depression is a multi factorial disease. There may be some spurious association of different factors due to confounding so optimization is needed. Different nature inspired algorithm based optimization techniques can be used for more accurate feature selection of depression prediction. Different data set from various part of the country can also be used for performance analysis of the appropriate predictive model which helps to assess generalization of the model.

#### 6. ACKNOWLEDGMENTS

Special thanks to the faculty members of the department of Community Medicine, R. G. Kar Medical Collage and Hospital, Kolkata, West Bengal, India.

#### 7. REFERENCES

- National Policy on Older Persons. New Delhi: Ministry of Social Justice and Empowerment; Government of India; 1999.
- [2] Santosh A, Kumar A, Rao BV, Patil RS. Magnitude of depression among geriatric population and factors associated with it in the urban slum, Bashanagar, field practice area of SSIMS and RC Davangere-a cross sectional study. International Journal of Medical and Pharmaceutical Sciences. 2014; 4(7):20-6.
- [3] Ingle GK, Nath A. Geriatric health in India: Concerns and solutions. Indian Journal of community medicine. 2008; 33(4):214.
- [4] Srivastava S, Kumar A, Khurana H, Tiwari SC, Akbar S. Short-term course and outcome of late-life depression. Journal of Geriatric Mental Health. 2015;2(2):96.
- [5] Cruz, J. A., & Wishart, D. S. Applications of machine learning in cancer prediction and prognosis. Cancer informatics. 2006; 2.
- [6] Ooi, K. E. B., Low, L. S. A., Lech, M., & Allen, N. Prediction of clinical depression in adolescents using facial image analysis. In WIAMIS 2011: 12th International Workshop on Image Analysis for Multimedia Interactive Services, Delft, The Netherlands, April 13-15, 2011. TU Delft; EWI; MM; PRB.
- [7] Jena, L., & Kamila, N. K. Distributed Data Mining Classification Algorithms for Prediction of Chronic-Kidney-Disease. in International Journal of Emerging Research in Management & Technology, v 4.11.
- [8] Baby, P. S., & Vital, T. P. Statistical Analysis and Predicting Kidney Diseases using Machine Learning

Algorithms. International Journal of Engineering Research and Technology. 2015; 4(7). ESRSA Publications.

- [9] Shubham Bind et al. A Survey of Machine Learning Based Approaches for Parkinson Disease Prediction. International Journal of Computer Science and Information Technologies. 2015; 6(2):1648-1655.
- [10] Chitra, R., & Seenivasagam, V. Heart disease prediction system using supervised learning classifier. Bonfring International Journal of Software Engineering and Soft Computing, 2013; 3(1).
- [11] Jabbar, M. A., Deekshatulu, B. L., & Chandra, P. Heart disease prediction system using associative classification and genetic algorithm. arXiv preprint arXiv. 2013;1303.5919.
- [12] Sibanda, W., & Pretorius, P. Novel application of Multi-Layer Perceptrons (MLP) neural networks to model HIV in South Africa using Seroprevalence data from antenatal clinics. International Journal of Computer Applications. 2011; 35(5).
- [13] R. E. Roberts, et al. Screening for Adolescent Depression: A Comparison of Depression Scales. Journal of the American Academy of Child & Adolescent Psychiatry. 2011; 30:58-66.

- [14] Aksenova, S. S. WEKA Explorer Tutorial.
- [15] Mitchell, T. M. 1997. Machine learning. Machine Learning.
- [16] Bouckaert, R. R. Bayesian network classifiers in weka. Department of Computer Science, University of Waikato. 2004.
- [17] Komarek, P. 2004. Logistic regression for data mining and high-dimensional classification. Robotics Institute, 222.
- [18] Mahajan, Manish, and Rajdev Tiwari. Introduction to Soft Computing. New Delhi: Acme Learning, 2010. Print.
- [19] Platt, John. Fast Training of Support Vector Machines using Sequential Minimal Optimization. Advances in Kernel Methods – Support Vector Learning, B. cholkopf, C. Burges, A. Smola, eds., 1998. MIT Press.
- [20] Jena, L., & Kamila, N. K. 2015. Distributed Data Mining Classification Algorithms for Prediction of Chronic-Kidney-Disease.
- [21] Yesavage, J. A., Brink, T. L., & Rose, T. L. 2000. Geriatric depression scale (GDS). Handbook of psychiatric measures. Washington DC: American Psychiatric Association, 544-6.