

A Study and Comparison of Prediction Algorithms for Depression Detection among Millennials: A Machine Learning Approach

Madhurima Hooda
Assistant Professor, Dept. of
IT
Amity University
Noida, India
10madhurima@gmail.com

Aashie Roy Saxena
Student, Dept. of IT
Amity University
Noida, India
saxenaashie13@gmail.com

Dr. Madhulika
Associate Professor, FET
MRIU
Faridabad, India
madhulikabhatia@gmail.com

Babita Yadav
Assistant Professor,
MVN University
Haryana, India
yadav.babita@gmail.com

Abstract— Depression is a critical problem that can affect people in diverse ways. Various treatment are there to treat people suffering from depression but challenge here is to predict those who doesn't even know that they are suffering from depression. So, to predict depression among people certain models are prepared of which overview of three major models are given in this article: a) Using Machine learning classifiers and WEKA, b) Using Imaging and Machine learning methods, c) Using the Risk factors.

Keywords— depression, machine learning, machine learning classifiers, WEKA and depressive disorder.

I. INTRODUCTION

Major depressive disorders is a disorder affecting person's way of living. Depression leads to negative affect and mood. It has such an impact on people that they started feeling ore of those leisure activities which they used to enjoy earlier [1]. Their approach is negative towards everything. This is a common disorder for every generation. But millennials report higher rates of depression than any other generation [4]. About 20% of the people suffer from depression and anxiety. They feel that stress keeps them awake all night (Health Status). Machine Learning is getting computers to program themselves. It is automating the process of automation. Machine Learning is used in computer science and other fields [2]. And so for this type of intelligent automation, machine should be learned with complex patterns as that of human brain. There are various methods used for prediction of depression such as facial images, speech tones etc. But most of the ongoing researches are based on a particular learning method. To know which prediction model is accurate a comparison between different models is necessary. This research work focuses on study of three major methods used for prediction of depression:

- Using Machine Learning Classifiers and WEKA
- Using Imaging and Machine Learning Methods
- Using Risk Factors

And to detect the most consistent and accurate of them all.

Also a software developed by University of Waikato, New Zealand called WEKA [14], is used in one of the above

prediction model. It is used for classification and depression among people. Rest of the sections are classified as: Section II gives the overview of the different prediction algorithm for depression. Section III provides the observation from the past studies Section 4 provides a basic discussion of the outcomes derived from the past studies. Section 5 concludes the research work.

II. OVERVIEW OF DIFFERENT PREDICTION ALGORITHM FOR DEPRESSION

A. Using Machine Learning Classifiers and WEKA

Machine learning is a type of artificial intelligence, a set of computational algorithms in artificial intelligence where learning is done from past experiences on a particular set of tasks or data to improvise performance. Various classifiers can be used to predict decisions in WEKA like decision trees, instance based classifiers, support vector machine, multilayer perception, logistic regression, Bayesnet etc. [14]. Supervised learning is the best for this type of research work as output label for training data set is known. Only output for test data set needs to be predicted using supervised learning. For this prediction model five major classifiers are compared to choose the best for prediction of depression:

a. Bayes Net Classifier (BN)

A Bayesian network is a probabilistic graphical model that represents a set of random variables and their dependencies via a directed acyclic graph (DAG). Learning process in Bayesian network is a two stage process : First, learn a network structure Second , learn the probability tables [3]. Structure learning can be done in WEKA using algorithms like local score metrics , global score metrics, conditional independent tests , cross validation and fixed structure.

b. Multilayer perception (MLP)

A multi layer perception (MLP) is earliest and simplest neural network developed to software effort. It maps sets of input data onto a set of appropriate outputs . MLP utilizes a supervised technique called back propagation for training the network . It is the modification of the standard linear

perceptron and can distinguish data that is not linearly separable.

c. Logistic regression

Linear regression is applicable for data with linear relations or applications for which a first order approximation is adequate [4]. But there are many applications where linear regression is not appropriate. Such case is of depression prediction for which linear regression doesn't hold good. Therefore, logistic regression is used to model this. There is a two step process of logistic regression learning methods which are capable of good performance in terms of classification accurately.

d. Decision Tress

Decision trees One in which positive results are yeild that is will detect depressed people. Another in which negative results are yeild that is will detect non depressed people. Logistic function used in this model for prediction of output is given below in equation 1-

$$f(r) = 1 / 1 + e^{-r} \quad (1)$$

e. Sequential Minimal Optimisation

Sequential Minimal Optimisation is an algorithm for solving a quadratic programming problem occurs during training of support vector machines(SVM). SVMs [5] are a set of supervised are used in a world wide rane of applicative domain [11]. They are used in many feilds like artificial intelligence, engineering , data mining, etc. It works in two steps :

- First , to produce an accurate classifier.
- Then, to understand the predictive structure of the problem.

It is a form of multiple variable analysis. It is quite simple but powerful too. Further , For Prediction methodology following steps are followed:

- Load input data.
- Data is taken to preprocessing window, there filter is applied . Attribute selection filter is applied in preprocessing in WEKA.
- Choose a classifier for classification.
- Here five classifiers are and experiment are performed five times
- A test option is selected

There are three major options :

- Training and testing set: It evaluates the classifier on how well it predicts the class of data , it was trained on .
- Cross Validation: In this test datat set is automatically prepared from training data set on the basis no. Of feilds provided to limit over fitting problem.
- Percentage Split

Now for prediction output option , option for prediction is chosen.

After this the five classifiers are compared on four parameters :

- Accuracy
- ROC Area
- Precision
- RMS Error

According to which Bayes Net Classifier was the most consistent and the best classifier among the five taken for the Percentage Split method [10]. In the rest of the sections overview of two more methods used for prediction of depression is given.

B. Using Imaging and Machine Learning Methods

Major depressive disorder, is the most frequent mood disorder, highly concurrent with anxiety and stress [6]. It has an estimated 17% of lifetime prevalence. It causes major effects on quality of life, healthcare utilization etc. MRI (Magnetic Resonance Imaging) uses high frequency radio waves to produce images of a particular body organ for its detailed examination. MRI modalities to a major extent are used in depression. These are used to study brain structure from different aspects. Most common of these modalities are:

- T1-weighted imaging is used to study critical regions
- T2-weighted imaging is used to study white matter hyper intensities
- DTI (Diffusion Tensor Imaging) is used to get understanding of the brain from a microscopic level. To study the diffusion of molecules in the brain tissues.

Machine learning here is used in a unique way to test the potential of each MRI measure as a relevant biomarker of depression. Depending upon the data there are 3 tyeps of learning:

- SL – If all data is labelled
- SSL - If there is labelled data along with unlabelled
- USL - If there is unlabelled data only
- In supervised learning, training data includes desired outputs. It works only for labelled data.
- In SSL, some desired outputs are considered in the training data. It works for labelled data along with unlabelled data.

In unsupervised learning, training data doesn't include desired output. It works for unlabelled data.

The SL and SSL are divided into 2 methods –

- a. Classification Based Method – Under this method it classifies the data according to discrete categorical labels.

b. Regression Based Method - This method automatically fits data in a continuous function to work for continuous labels for the data.

Moreover learning methods are also divided into 2 more parts – Linear and Non Linear. Linear is simple but non-linear is a bit more complex. Validation Measures [7]. The name suggests that it checks the validity of the model. Moreover it assesses the model developed by learning method on how well it will be able to perform on a new unseen data. On the basis of types of learning, the validation measures differ in its type of framework. It follows three step process:

- Accuracy
- Specificity
- Sensitivity

Accuracy – Measures how accurately prediction model classifies data overall

Specificity & Sensitivity – Measures how specifically and sensitively prediction model classifies each label of the test data

For Regression Based Learning, it follows two step process –

- Correlation coefficients
- Mean squared error

It helps measure how prediction model is correlated with actual label and helps measure the level of error in the prediction model. Machine learning when used with the real world data helps in estimating an accurate prediction model. Various sources of data from all over the world are in process of being accepted to contribute in building an account model [8]. Real world data refers to the subject oriented data captured without interference or bias. Since, high bias or high variance can cause the model to be unfit or over fit. Due to these limitations of the real world data, two important methods are used that are: Cross Validation and Feature Reduction.

a. Cross validation method is used as a means for evaluating models [13]. It checks or evaluates the accuracy of prediction model created by learning methods. Cross-Validation can be classified into various methods . But the most used methods are K-Fold Cross-Validation and Leave-One –Out Cross-Validation. The K-fold method follows a certain procedure – first it divides whole data into k equal sized sets and then takes one as test set and combining the other as training set then uses the learning method to estimate a model that describes the data by training on the training sets then tests the estimated model on the test set then figure out the appropriate validation measure to check the accuracy of the model.

b. Feature Reduction- Feature reduction method is used to reduce the number of features in the total data to number of the non-relevant features for measuring the most accurate prediction model. In supervised learning, features are reduced with the help of labels as supervised learning requires labelled data whereas in unsupervised learning, features are reduced on the basis of information available in the features. When used together, feature reduction is done on every repetition of cross validation to prevent biasing of the prediction model with information from the test set data.

C. Using Risk Factors

Major depression is a complex and is a heterogeneous disorder which not only affects a person's physical and mental health ut also disturbs his every aspect of life from his interpersonal relationship to academic functioning [12]. There are 39 major risk factors depending upon which this model is prepared , the most relevant of them all are :

- Age
- Sex
- Education
- Difficulties in paid and unpaid work
- Physical health
- Mental health
- Emotional Issues
- Discrimination
- Widespread presence of depression
- Country

These are chosen and then a follow –up is prepared upon which responses are taken from the participants, then a composite international diagnostic interview (CIDI) is conducted. Then , Using logistic regression , a model was developed in imputed data for general clustering, which works on multivariable fractional polynomial analysis . Taking the risk factors [9] into account the estimated probability of depression is of over 12 months. According to this study conducted in all European countries , major depression was found to be 13.9% in women and 8.5% in men andt he overall population suffered was 7.7%. This study aimed to identify a predictive model which can detect depression using regression based learning and these risk factors but did not address that how this risk algorithm can be best implemented in general practices to achieve good results.

II. OBSERVATIONS

Table I provided below summarizes past studies predicting depression using different methods.

Table I: Past studies predicting depression using different methods

Authors	Patient sample	Using Classifiers with percentage split method	Using Imaging	Using Cross Validation	Using Feature Reduction	Using Machine learning methods	Accuracy
Ishita Bhakta, Arkaprabha Sau, 2016 [10]	60 samples	Bayes net	-	-	-	-	95
		Logistic regression					80
		MLP					90
		SMO					90
		Decision Trees					95
Costafreda et al, 2009 [18]	74 samples	-	Smoothed grey matter voxel – based intensity values	Leave-one-out cross-validation	Voxel based morphometric And Filter method using ANOVA	Support vector machines	67.6
Fu, Cynthia HY, et al., 2008 [17]	40 samples	-	Smoothed whole brain voxel based blood oxygen level response	Leave-one-out cross-validation	-Principal component analysis	-Support vector machines (linear kernel)	68
Mwangi et al, 2012 [16]	62 samples	-	Smoothed whole brain voxel based intensity values	Leave-one-out cross-validation	-Voxel based morphometric	-Support vector machines (nonlinear Gaussian kernel) -Relevance vector machines (nonlinear Gaussian kernel)	Svm- 87.1% Rvm- 90.3%
Ma, Chaoqiong, et al. [15]	53 samples	-	Region based resting state functional connectivity (fMRI)	Leave-one-out cross-validation	-Filter method using Kendall tau rank correlation coefficient	Support vector machines (linear kernel)	94.3%

III. DISCUSSIONS

The past studies that used three major methods for prediction of depression have different outcomes:

While using machine learning classifiers and WEKA, [10] the most consistent method of them all was Bayes Net classifier for Percentage Split Method.

While using Machine learning methods, the most common used was support vector machines, it is the most popular in past

studies due to its insensitivity to high dimensional data. But there is still variability for whether a linear or nonlinear learning method was used. Therefore, to avoid complexity linear learning method should be used. While using Feature reduction method [7], it shows a high variation in their selection of feature reduction methods. Small sample sizes are used in the past studies whereas performance of feature reduction method increases with increase in sample size. And so it requires further more testing as to which feature reduction method should be optimal for prediction. While using Cross-Validation method, the most consistent method used was leave-one-out cross-validation method. It is utilized for small sample

sizes and when provided with large scale data, is highly unreliable and variant.

IV. CONCLUSION

Depressive disorder makes people unfit not only physically but mentally also. It has diverse effects on quality of life of people. Moreover it doesn't need depression to be at higher levels to affect a person's life. Many researches are going on prediction of depression and from those past studies the three major methods were studied to determine the most accurate method of them all. After studying machine learning classifiers, Feature Reduction Method, Cross Validation Method, Risk Factors the most consistent and accurate method was Bayesnet Classifier for Percentage Split testing option. Different data sets can be used to analyze the performance of the predictive model. In future other methods can be tried for this prediction of depression for better accuracy.

REFERENCES

- [1] Blanco, Joel A., and Lynn A. Barnett. "The effects of depression on leisure: varying relationships between enjoyment, sociability, participation, and desired outcomes in college students." *Leisure Sciences* 36.5 (2014): 458-478.
- [2] Domingos, Pedro. "A few useful things to know about machine learning." *Communications of the ACM* 55.10 (2012): 78-87.
- [3] Remco R. Bouckaert, "Bayesian network classifiers in WEKA", JULY 2007.
- [4] Komarek, Paul. "Logistic regression for data mining and high-dimensional classification." *Robotics Institute* (2004): 222.
- [5] Durgesh, K. SRIVASTAVA, and B. Lekha. "Data classification using support vector machine." *Journal of Theoretical and Applied Information Technology* 12.1 (2010): 1-7.
- [6] Zimmermann, Petra, et al. "Heterogeneity of DSM-IV major depressive disorder as a consequence of subthreshold bipolarity." *Archives of general psychiatry* 66.12 (2009): 1341-1352.
- [7] Patel, Meenal J., Alexander Khalaf, and Howard J. Aizenstein. "Studying depression using imaging and machine learning methods." *NeuroImage: Clinical* 10 (2016): 115-123.
- [8] Mahajan, Rajiv. "Real world data: Additional source for making clinical decisions." *International Journal of Applied and Basic Medical Research* 5.2 (2015): 82.
- [9] King, Michael, et al. "Development and validation of an international risk prediction algorithm for episodes of major depression in general practice attendees: the PredictD study." *Archives of General Psychiatry* 65.12 (2008): 1368-1376.
- [10] Bhakta, Ishita, and Arkaprabha Sau. "Prediction of Depression among Senior Citizens using Machine Learning Classifiers."
- [11] Cristina Petri, "Decision Trees", 2010.
- [12] Dobson, Keith S., and David JA Dozois, eds. *Risk factors in depression*. Academic Press, 2011.
- [13] Vanwinckelen, Gite, and Hendrik Blockeel. "On estimating model accuracy with repeated cross-validation." *BeneLearn 2012: Proceedings of the 21st Belgian-Dutch Conference on Machine Learning*. 2012.
- [14] Aksenova, Svetlana S. "Weka explorer tutorial." *School of Engineering and Computer Science California State University*.
- [15] Ma, Chaoqiong, et al. "Resting-state functional connectivity bias of middle temporal gyrus and caudate with altered gray matter volume in major depression." *PLoS One* 7.9 (2012): e45263.
- [16] Mwangi, Benson, et al. "Multi-centre diagnostic classification of individual structural neuroimaging scans from patients with major depressive disorder." *Brain* 135.5 (2012): 1508-1521.
- [17] Fu, Cynthia HY, et al. "Pattern classification of sad facial processing: toward the development of neurobiological markers in depression." *Biological psychiatry* 63.7 (2008): 656-662.
- [18] Costafreda, Sergi G., et al. "Prognostic and diagnostic potential of the structural neuroanatomy of depression." *PloS one* 4.7 (2009): e6353.