# Prediction of Depression using Machine Learning Techniques: A Review of Existing Literature

Muhammad Usman
*Department of Computer Science*
*SZABIST*
Islamabad, Pakistan
dr.usman@szabist-isb.pk

Syed Haris
*Department of Computer Science*
*SZABIST*
Islamabad, Pakistan
ra_palab@szabist-isb.edu.pk

A.C.M. Fong
*Department of Computer Science*
*Western Michigan University*
Kalamazoo, MI, USA
acmfong@gmail.com

2020 IEEE 2nd International Workshop on System Biology and Biomedical Systems (SBBS) | 978-0-7381-4281-4/20/$31.00 ©2020 IEEE | DOI: 10.1109/SBBS50483.2020.9314940

***Abstract*—Depression is a serious mental disorder that negatively affects the mental health of a person, such that the person thinks, behaves, and feels in a negative way that can lead to physical and emotional problems. It can be a life-threatening problem as a person in such situation abandons hope of good and see negative aspects of situations. As per World Health Organization report of 2019, approximately 3 million people all over the world were suffering from depression between the ages of 17-25 and 40-70. Machine learning (ML) models can be trained on a training dataset so it can predict whether a person is going to develop depression or not. In the past, many researchers have worked for the prediction of depression by using ML Techniques. The purpose of this paper is to review the performance of ML techniques for early prediction of depression in older people. Different classifiers have been used, including Bayes Net, logistic regression, multilayer perceptron, sequential minimal optimization, decision table, and random forest. Our findings reveal that a maximum accuracy of 89% and maximum precision of 0.95 can be achieved from a large dataset.**

***Keywords—machine learning, prediction, classification, depression***

## I. Introduction

Depression is a disorder that badly affects physical and mental health. It can lead to feelings of sadness, anger, or loss that interfere a person's daily routine. It is also a serious medical problem and can be a leading cause of disability without proper treatment. Depression affect physical health as well for example weight gain or loss, chronic pain, heart disease, inflammation, among other illnesses. Millions of people have been found to be suffering from depression worldwide according to the world health Organization, in which youngsters and older people are affected. It is an alarming situation as many people who suffer from depression commit suicide every year because of their condition. An effective way to address the problem is that if depression can be predicted in the early stages, and then proper treatment and medication can be prescribed to patients.

There are in fact several types of depression. The most common types are:

- Major depression: It is also known as unipolar or major depressive disorder (MDD). It is a mood disorder that causes loss of interest and feeling of sadness.
- Persistent depression or bipolar disorder: It is also called dysthymia and is a continuous long term depression in which a patient feels hopeless and may lose interest in daily activities.
- Perinatal depression: It can be defined as depression in pregnancy, circularly childbirth or within the first year post-partum.
- Seasonal depression: This type of depression depends on seasons. As the name implies, seasonal depression changes according to season, for example from winter to summer or from summer to winter every year. For most people who suffer from seasonal depression, their symptoms start in the winter and continue into the fall months and making them feel moody.
- Situational depression: It is also known as adjustment disorder. It can make it tough to adjust to one's everyday life following a traumatic event. Another name for this condition is reactive depression (RD).

This paper studies the application of machine learning techniques for predicting depression. The study covers a wide range of ML classifiers, for example, Bayes Net, logistic regression, multilayer perceptron, sequential minimal optimization, decision table, and random forest. Our study reveals that LASSO with logistic regression, extreme gradient boosting, logistic regression and one of the authors using many classifiers in which the BN classifier gives better result. The review of existing literature presented in this paper aims to lay the groundwork and provide interested researchers a platform on which to build their further research upon.

The rest of the paper is organized as follows. Section II surveys the existing literature on ML classifiers for depression. It covers a wide range of ML techniques. Section III provides a critical analysis of existing research in this area and summarizes the key findings in tabular form. Finally, Section IV concludes the paper.

## II. Literature Survey

In [1], the authors address the prediction of depression in older people using six different classifiers: Bayes Net (BN) classifier, logistic regression (LR), multilayer perceptron (MLP), sequential minimal optimization support vector machine (SMO), and Decision Table (DT). They apply WEKA Machine Learning Model to a dataset provided by the Urban Health and Training Center (UHTC) Bagbazar Kolkata. BN gives accuracy of 86.67%, precision 0.8, ROC Area 0.96, RMSE 0.32. LR gives accuracy 78.33%, Precision 0.79, ROC Area 0.85, RMSE 0.46. MLP gives accuracy 85%, precision

978-0-7381-4281-4/20/$31.00 ©2020 IEEE

0.85, ROC Area 0.94, RMSE 0.33. SMO gives accuracy 88.33%, precision 0.88, ROC Area 0.88, RMSE 0.34. DT gives accuracy 80%, Precision 0.80, ROC Area 0.86, RMSE 0.38. Therefore, among the classifiers they investigated, Bayes Net classifer gives the best result with maximum accuracy of 95% and maximum precision of 0.95 [1].

The authors of [2] address the prediction of anxiety and depression in older people using ML techniques with ten classifiers: BN, NB, Log, MLP, KS, RS, J48, RF. They apply the classifiers to a dataset of 510 old patients. The dataset is provided by kar Medical college and Hospital, Kolkata. They author also use ten-fold cross-validation. According to the authors of [2], RF gives the best result with a maximum accuracy 89 %.

In [3], the authors highlight the tool (NANA toolkit) that uses ML techniques. Wit out the practitioner, the NANA toolkit collects desired data from old people in their residential areas. The author of [3] focus on the use ML techniques applied to the data collected by the NANA toolkit from older people's home to generate an algorithm to predict the symptoms of depression in older people. In the case of logistic regression classifier, the authors use it with the LASSO method to perform regularization.

In [4], the authors highlight the use of ML techniques, integrating clinical, imaging features and predicting depression in older people. The purpose is to evaluate exact model (prediction models) for depression and proper treatment, medication prescribed to patients using ML methods with inputs of modles (multi-modal imaging,non-imaging the brain of human). 33 patients and 35 older non-depressed people were recruited individually by medicated post-recruitment. Their cognitive ability and demographics scores were recorded and brain characteristics were acquired using multi-modal magnetic resonance imaging pretreatment [4]. Using alternating decision tree (ADT) machine learning method and evaluating the prediction models for depression in older people, the authors report a maximum accuracy of 87.27% and treatment response of 89.47% [4].

The authors of [5] address the symptoms of depression in older people using ML techniques. The data are collected from 284 patient and used to predict the symptoms of depression in older people after one year. Their data are drawn from CASPER [5]. ML techniques used in their study include extreme gradient boosting (XGB) and logistic regression. The authors report that they get achieve an accuracy of 69% using logistic regression and a maximum accuracy of 74% using extreme gradient boosting [5].

In [6], the authors apply ML techniques to generate self-reported depression prediction models. The aging and health project is managed by a research center at their organization with funding from the National Institutes of Health (NIH). The purpose of the project was to understand older people's mental and physical health, how they behave, and relationship status. Data were collected from 3005 people in the years 2005-2006 and then again in 2010 and 2011 from 3377 individuals. The variables which are not related were removed. The authors apply logistic regression for classification of the relevant variables on the WEKA platform. A key finding of the paper is that of the people above than 65 years suffer from depression.

The authors of [7] apply different algorithms and methods to predict depression by recognition of human face. A combination of ML algorithms and image processing is used in a workflow summarized in Fig. 1. Recognition of features from the facial images gives the result of whether depression is detected.

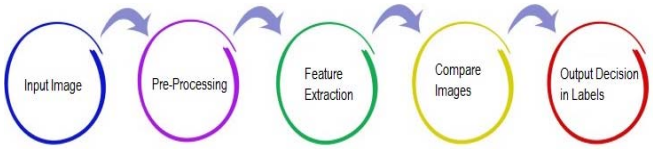


Fig. 1. Detection of depression from facial images.

In [8], the authors apply ILIOU data preprocessing method before applying ML for the prediction of depression. The performance of ILIOU data preprocessing method and principal component analysis preprocessing method was evaluated using the tenfold cross validation method assessing seven machine learning classification algorithms: nearest-neighbour classifier (IB1), C4.5 algorithm implementation (J48), random forest, multilayer perceptron (MLP), support vector machine (SMO), JRIP and fuzzy logic (FURIA). For the prediction of different types of depression, ILIOU preprocessing can be used. Maximum precision and maximum recall using SMO classifier from their original data achieved 91.2% . Using J48 classifier on PCA preprocessed data achieved maximum precision and maximum recall of 96.2%. Using MLP classifier from ILIOU Method achieved a maximum precision and maximum recall of 100%. So, ILIOU gives best result with maximum recall and maximum precision.

In [9], the authors address depression and anxiety among seafarers using ML technology. Anxiety and depression are mental disorder that affects all age groups from youngsters to the elderly and include men and women. The authors use ML to detect depression and anxiety in seafarers or older people. More than 400 seafarers were interviewed at Haldia Dock Complex (HDC) in India and the data were collected from 470 people. Different machine learning classifiers are used to detect depression and anxiety, including CatBoost, random forest, support vector machine, logistic regression, and naïve Bayes. Among the classifiers evaluated, Catboost gives the best result with maximum accuracy and maximum precision of 82.6% and 84.1%. A final data set was prepared with 14 features from January to July in the year of 2016. 10-fold cross validation is also used on training set.

The authors of [10] address EEG-based functional connectivity to detect major depressive disorder (MMD) using ML techniques. In their paper, the authors experiment with the following classifier: SVM, LR, and NB. The dataset is provided by Hospital Universiti Sains Malaysia (HUSM). These 3 classifiers are selected because they are easily trained on small datasets. These classifiers give different accuracy, sensitivity, specificity and f-m measures. SVM gives accuracy of 98%, sensitivity of 99.9%, specificity of 95%, and f-m of 0.97, LR gives accuracy of 91.7%, sensitivity of 86.66%, specificity of 96.6%, and f-m of 0.90. NB gives accuracy of 93.6%, sensitivity of 100%, specificity of 87.9%, and f-m of 0.95. From the results, the authors conclude that SVM gives the best result with maximum accuracy.

Another paper that combines machine learning, image analysis, and EEG for depression prediction is [11]. In the paper, the authors propose a novel method for depression recognition of Electroencephalography. They apply an ensemble machine learning model that consists of deep forest

and SVM. The results of their ensemble is compared with three traditional ML modes, namely SVM, RF, and k-nearest neighbor (KNN). The authors furthermore add spatial information of EEG caps by image conversion method, and they perform processing and analysis of EEG signals on multiple frequency bands.

## III. Summary of Research

Table I summarizes the ten papers reviewed in Section II. The researchers have used machine learning techniques and different classifiers, such as Bayes Net, multilayer perceptron, SMO and decision table, LASSO with logistic regression, extreme gradient boosting in order to obtain maximum accuracy. Different algorithms are used on different datasets and have different accuracy measured. The maximum accuracy achieved is 93% from Bayes Net classifier.

TABLE I. Summary of Findings

| Ref | Max Accuracy | Max Precision | Strengths / Limitations |
|---|---|---|---|
| 1 | 93.3 | 94 | Few classifiers. Results not well explained. Small dataset. |
| 2 | 89 | 89.1 | Many classifiers. Results well explained. Large dataset. Also considered anxiety. |
| 3 | - | - | Lacks details on the dataset. Did not explain results well. |
| 4 | 74 | - | Lacks details on tools and techniques. Small dataset. |
| 5 | 87.27 | - | Few classifiers. Small dataset. |
| 6 | - | - | Few classifiers. Large dataset. Results well explained with graphical aid. |
| 7 | - | - | Results not well explained. Small dataset. Incorporates image processing. |
| 8 | 91.2 | 91.2 | Many classifiers. Small dataset. Results well explained. |
| 9 | 82.6 | 84.1 | Few classifiers. Small dataset. Results well explained with graphical aid. |
| 10 | 98 | - | Few classifiers. Small dataset. |
| 11 | 89.02 (ensemble) 82.06 (baseline SVM) | | Ensemble model comprising deep forest and SVM, compared with three baselines models: SVM, RF, KNN. Small dataset. Results well explained with graphical aid. |

In table I, the column Ref refers to the cited papers from [1] through [10]. Maximum accuracy and maximum precision are the best reported results among the classifiers evaluated. Missing entries mean the values were not reported. A most frequent reason for missing values is that the authors used other performance measures. The strengths and weaknesses column evaluates each paper in terms of factors, such as the number of classifiers considered, size of dataset(s), and whether the observed results are well explained by the respective authors.

## IV. Conclusion

Depression makes people, especially the elderly, physically and mentally affected. Early prediction of depression can make the procedure of treatment faster and will likely be helpful for patients and medical practitioners. For this purpose, researchers have used machine learning techniques to perform prediction of people likely to develop depression. This paper has provided a review of different classifiers that have been applied to predict depression in different settings. The classifiers studied include Bayes Net, multilayer perceptron, SMO, MLP, KS, RS, J48, RF, and RT, LASSO, logistic regression and extreme gradient boosting models. Researchers have applied subsets of these classifiers on different datasets and reported their findings, such as the maximum accuracy achieved in their study. Some of the authors have used NANA and WEKA tools. The review presented in this paper serves as a starting point for other researchers who are interested to conduct further research in this emerging area.

## References


[1] I. Bhakta and A. Sau " Prediction of depression among senior citizens using machine learning classifiers", International Journal of Computer Applications, vol. 144, no. 7, pp. 11-16, June 2016. DOI: 10.5120/ijca2016910429

[2] A. Sau and I. Bhakta, "Predicting anxiety and depression in elderly patients using machine learning technology," Healthcare Technology Letters, vol. 4, no. 6, pp. 238-243, 12 2017, doi: 10.1049/htl.2016.0096.

[3] J. A. Andrews, R. F. Harrison, L. Brown, L.M. MacLean, F. Hwang, T. Smith, E.A. Williams, C. Timon, T. Adlam, H. Khadra, and A.J. Astell, "Using the NANA toolkit at home to predict older adults' future depression", Journal of Affective Disorders, vol. 213, pp. 187–190, 2017. https://doi.org/10.1016/j.jad.2017.02.019.

[4] M. J. Patel, C. Andreescu, J.C. Price, K.L. Edelman, C.F. Reynolds III, and H. J. Aizenstein, (2015). Machine learning approaches for integrating clinical and imaging features in late - life depression classification and response prediction. International journal of geriatric psychiatry, 30(10), 1056-1067

[5] C.M. Hatton, L.W. Paton, D. McMillan, J. Cussens, S. Gilbody, and P.A. Tiffin, "Predicting persistent depressive symptoms in older adults: A machine learning approach to personalised mental healthcare", Journal of Affective Disorders, vol. 246, pp. 857-860, 2019. doi: 10.1016/j.jad.2018.12.095.

[6] J. Choi, J. Choi, and H.T. Jung, "Applying machine-learning techniques to build self-reported depression prediction models" , Computers, Informatics, Nursing, Kluwer Health, vol. 36, no. 7, pp. 317-321, 2018. DOI: 10.1097/CIN.0000000000000463

[7] A. Pawar, S. Bandal, R. Borate, P. Jadhav, and S. Punjabi, "Depression analysis using image processing and Machine learning" International Journal of Scientific Research and Engineering Development, vol. 2, no. 5, pp. 470-473, September 2019

[8] T. Iliou, G. Konstantopoulou, M. Ntekouli, C. Lymperopoulou, K. Assimakopoulos, D. Galiatsatos, and G. Anastassopoulos, "ILIOU machine learning preprocessing method for depression type prediction", Evolving Systems, vol. 10, pp. 29–39, 2019. https://doi.org/10.1007/s12530-017-9205-9

[9] A. Sau, and I. Bhakta, Ishita. "Screening of anxiety and depression among seafarers using machine learning technology", Informatics in Medicine Unlocked, vol. 16, 2019. 100228. 10.1016/j.imu.2019.100228.

[10] W. Mumtaz, S. S. A. Ali, M.A.M. Yasin, A.S. Malik, ″A machine learning framework involving EEG-based functional connectivity to diagnose major depressive disorder (MDD), Medical & Biological Engineering & Computing, vol. 56, 2017. 10.1007/s11517-017-1685-z

[11] X. Li, X.. Zhang, J. Zhu, W. Mao, S. Sun, Z. Wang, Z., C. Xia, and B. Hu, "Depression recognition using machine learning methods with different feature generation strategies", Artificial Intelligence in Medicine, 99, 101696, 2019.
https://doi.org/10.1016/j.artmed.2019.07.004