Mathematical Statistician and Engineering Applications
ISSN: 2094-0343
2326-9865

Analysis of Major Depressive Disorder Using Ensemble Technique

Anjali Muralidharan
Department of Computer Science & Engineering
Amrita School of Engineering, Bangalore
Amrita Vishwa Vidyapeetham, India
Email: anj.3jg96@gmail.com
Dr. Kumaran.U.
Department of Computer Science & Engineering
Amrita School of Engineering, Bangalore
Amrita Vishwa Vidyapeetham, India
Email: u_kumaran@blr.amrita.edu

Article Info
Page Number: 1242-1249
Publication Issue:
Vol. 71 No. 3s (2022)

Article History
Article Received: 22 April 2022
Revised: 10 May 2022
Accepted: 15 June 2022
Publication: 19 July 2022

ABSTRACT:
Depression is a medical condition that exists on continuum of severity ranging from mild, momentary states of low mood to severe, long term symptoms that have a major impact on an individual's life. It is one of the primary causes of disability worldwide and contributes significantly to the global illness burden. The complexity for analyzing the major depressive disorder can be implemented using mental health dataset acquired from the World Health Organization. Ensemble machine learning aims to supervise on the strength and weakness of commonly used machine learning models, in order to minimize the risk in decision making. The experimental outcomes reveal that our ensemble technique used to analyze the depression brought out high acceptability in term of prediction and an accuracy of 95% was achieved.

Keywords: Depression, Mental health, Data analysis, Ensemble model, Categorization

I. INTRODUCTION

A .General System

Mental health is the most vital component of human life. Depression is a disorder that affects our mental health and which has been great thought to the society around us [1]. Depression is a situation that is mostly resulted from environmental factors and culpability. It's feeling where an individual feels totally disconnected from own self and completely difficult to connect with a deep emotional level with anyone. Certain risk factor that leads to depression is mostly to stress, illness, genetical depression and illness. Millions of people across the globe suffer from depression that accounts to 4.5 % of the total population. An extreme case of depression sometimes leads to suicide. Based on symptoms depression can be categorized into major depression disorder (MDD), bipolar disorder, postpartum depression and premenstrual dysporic disorder (PMDD). The main focus of this research is on detecting serious depressive disorder. Clinical depression is another name for major depressive illness which is a psychiatric disability where an individual is withdrawn from his daily activities and symptoms include the inability to eat, sleep, work and enjoy any daily activities that the individual used to regularly perform. The major motive of this study is to analyze how the

information is being processed by the brain for the individuals who are diagnosed with major depressive disorder.

B. Ensemble model

Ensemble machine learning model is a technology that is being used to combine multiple models and to obtain optimal performance [2]. The ensemble device modeling paradigm combines various more than one professional hypotheses to generate a single overall strategy to a problem. It reduces the opportunity of an unlucky emergence of a terrible choice from a mixed set of feasible solutions. It became initially added for class and clustering purposes and later prolonged to time collection and regression problems.

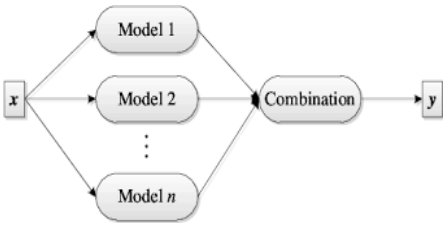

Fig 1: Ensemble model

Ensemble learning is a type of supervised learning method that implies to discrete problem solving where an individual task is being divided into different sub-tasks. An ensemble model generates single hypothesis after training and shows more flexibility while training. It's technique where we use multiple classifiers and combine together to form a better model.

C. Problem Statement

Suicide deaths due to depression are on the rise all around the world. Traditional methods applied to detect depression have detected only the symptoms in depression, but advance research aims to explore the relation between psychosocial factors such as the mental-state of an individual and quality of life style to detect major depressive disorder. We have noticed while implementing each algorithm individually it leads to either over-fitting, less accuracy and under-fitting. The presented ensemble machine-learning model that is robust in nature and performs better than all the baseline algorithms in terms of all experiments and measures.

II. RELATED WORK

In novel times different works were being carried out in terms of machine learning techniques with different kinds of data to analyze depression. This section compares the different literature reviews and discusses the related work.

In [3], this study the authors have tried to analyze the depressive and anxiety symptoms between different age groups and performed clustering of symptom nodes. To see if symptoms grew more related with development, parametric permutation tests were employed, and network centrality was assessed to see if there were any developmental changes in the overall importance of the specific symptoms. It was indicated that highly interconnected networks where formed by the symptoms which provided a strong relation between symptoms of despair and anxiety.

In [4], this study the author attempted to do an fmri meta-analysis to see if there was any congruence between brain areas related with observing dynamic facial expressions. The

findings show that processing complex, dynamically changing features of faces requires more brain resources when there are several faces.

In [5], the author proposed an emotion categorization between healthy inviduals and depressed patients to predict depression. The study reveals that depressed invdiduals portrayed a different pattern of brain activation when compared to the healthy inviduals.

In [6], the author aims predict depression in individuals by calculating out how far it is to the nearest depression treatment centre. The data was collected using population based survey and the distance was calculated using network analysis using Geographic Information Systems.

In [7], the author presented a depression monitoring system based on artificial intelligence. Visual and vocal expressions are used to predict the Beck depression inventory II (BDI-II) scores. The feature dynamic history histogram (FDHH) is used to capture the time movement on feature space after the first facial extraction. Finally, a fusion is performed using regression approach to predict BDI-II scales using the obtained audio characteristics and FDHH.

In [8],this approach we used patient-reported data from depressed patients (n=4041, with 1949 completers) from level 1 of the Sequenced Treatment Alternatives to Relieve Depression to identify variables that are most predictive of treatment outcome, and then used these variables to train a machine-learning model to predict clinical remission. The collecting and distribution of large-scale, clinical-grade datasets is critical to the success of this method.

In [9], this approach depicts key applications of supervised and unsupervised machine learning in musculoskeletal medicine; such as diagnostic imaging, patient measurement data, and clinical decision support. Support Vector Machine is a supervised machine learning algorithm that can be used for regression as well as classification. It is, however, mostly employed to solve categorization difficulties. Unsupervised Learning is a machine learning technique in which users work with data that has not been labeled. The current literature base is being used to find areas where machine learning can perform as well as or better than humans.

In [10], this study we introduce a novel emotion-modulated learning rule to train a recurrent neural network that allows a robotic arm and a complex musculoskeletal arm to perform goal-directed tasks in this study, which is based on basic emotion modulation theory and the neural mechanisms of generating complex motor patterns. The approach performed poorly with bigger datasets, and due to the exploratory nature of the study, no correction for multiple comparisons was made.

In [11], this paper portrays a supervised Multinomial Naïve Bayes Classifier which is a form of supervised learning that classifies the market lists of anonymous places.

In [12], this paper presents how the levels of depression have been captured and features are extracted using Gabor filters and prediction is done using SVM classifier.

In [13], this work different speech samples are collected from different individuals where we collect data and depression is analyzed.

In [14], this paper, we analyze the handwritten we use ensemble machine learning model where the complexity of the model is being reduced.

In [15], this paper we study the patients with epilepsy and the severity of depression was being measured.

III. METHODOLOGY

I. Dataset

The dataset used for this entire work is Mental Health dataset where individuals with depression can be predicted over various sections which include prevalence-by-mental-and-substance abuse, depression-by-level-of-education, prevalence-of-depression-by-age, prevalence-of-depression-male, suicide rates versus prevalence. In this entry the latest estimates are being presented which are acquired from the Institute for Health Metrics and Evaluation. Mental health disorder can take up many forms and are mostly under-reported. The dataset is divided into eight parts which includes the following : Anxiety disorders, Bipolar disorders, Schizophrenia, Any mental or substance use disorder, Alcohol use disorder, Drug use disorder and Depression. In our experiment we take the Depression part of the dataset is being taken where we split it into our training set and test set. The Institute for Health Metrics and Evaluation has classified it into three major categories such as persistent depression (dysthymia) and major depressive disorder (severe).Out of the three major categories we will performing the analysis of major depressive disorder in our experimental study.

II. Preprocessing

At the start of this experiment, we proceed by reducing the size of the data and removing the duplicate values. Datasets might have missing values which may cause hindrance while applying the machine learning algorithms. In such scenarios it is considered as an ideal practice to perform missing data imputation where missing values are replaced in each column of the input data prior to performing the prediction task. Although a wide range of models are available to perform the imputation, the most effective is k Nearest Neighbor algorithm commonly known as nearest neighbor imputation.

KNN Imputation

The goal behind kNN approaches is to find 'k' samples in a dataset that are comparable or close in space. The 'k' samples are then used to estimate the value of the missing data points. The missing values in each sample are imputed using the mean value of the 'k'-neighbors discovered in the dataset. We use the KNN imputer class that supports KNN imputation. KNN Imputer is a data transform that is configured initially based on the approach used to estimate missing values. The default distance measure is a Nan-aware Euclidean distance measure that will not include Nan values when computing the distance between members of the training dataset. This is controlled by the metric argument. The Euclidian distance is calculated using the formula.

$$d(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

III. Feature Extraction

We use random-forest feature selection for feature selection. One of the most well-known machine learning techniques is random forests. They're useful since they have strong predictive performance, little over fitting, and are simple to comprehend. The fact that it's simple to calculate the relevance of each variable on the tree decision lends to its interpretability. As a result, calculating the contribution of each variable to the decision becomes easier. Backward Elimination: In backward elimination, we start with all of the features and remove the least significant one at a time, improving the model's performance. Repetition is done until no improvement is shown when features are removed.

Ensemble Model

Ensemble methods are a machine learning technique that combines several base models in order to produce one optimal predictive model with a higher prediction power. We use various algorithms in our ensemble machine learning model and compare the results of the algorithms when applied individually to ensemble model. The key to an optimal ensemble model is the diversity of the models used. In our study we combine the following algorithms for our prediction.

- **Multi-layer Perceptron**:

A multi-layer perceptron with a non-linear mapping between inputs and outputs was developed to overcome the limitations of a single perceptron, which cannot be applied to non-linear data. Non-linear activation functions are used by the neurons in MLP to simulate the activity of neurons in the human brain. Input and output layers, as well as one or more hidden layers with many neurons placed on top of each other, make up a Multilayer Perceptron. While neurons in a Perceptron must have a threshold-enforcing activation function, such as ReLU or sigmoid, neurons in a Multilayer Perceptron can have any activation function they want.

- **Support Vector Machine**:SVM (Support Vector Machine) is a supervised machine learning technique for solving classification and regression issues. However, it is commonly used in categorization difficulties. The value of each feature represents the value of a certain coordinate in the SVM method, and each data item is displayed as a point in n-dimensional space (where n is the number of features you have). Then, by selecting the hyper plane that best distinguishes the two classes, we may classify them.
- **Random Forest**: A classifier based on random forest. A random forest is a Meta estimator that employs averaging to increase predicted accuracy and control over-fitting by fitting a number of decision tree classifiers on different sub-samples of the dataset.

IV. EXPERIMENTAL RESULTS

For the analysis of depression, we have used the ensemble learning model which is a combination of the three algorithms ie.Multi-layer perceptron, Support Vector Machine Classifier and Random forest. Then we apply this model to our training data which is being preprocessed using the K Nearest Neighbor Imputation, where all the missing values are being handled and the test data is being fed to the trained classifier to analyze depression.

Table I shows the comparison between the performances of the algorithms when applied invidually to the dataset. The results indicate that the Support Vector machine classifier

showed very low performance when compared to other algorithms. It had a very accuracy which implied no difference to depression data before and after the training. The comparison table depicts that our ensemble machine learning model showed up a higher accuracy when compared to other models.

Classifier	Accuracy (%)
Multi-layer perceptron	80
SVM Classifier	27.98
Random-forest	81.29
Ensemble	91

Fig I AND II depicts the dataset before and after training. Here we depicts to visualization before training our model and after training the model. The factors that are being depicted are accuracy-score and training loss. From the above results it's clearly evident that our model performs well on the training data and there are no signs of under-fitting or over-fitting shown by the model.

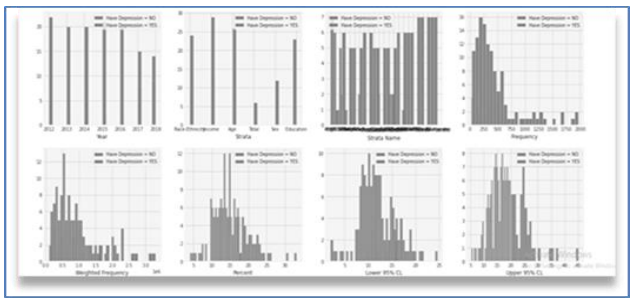

Fig I:Dataset before training

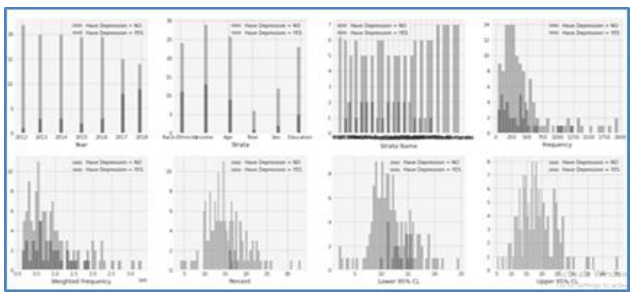

Fig II:Dataset after training

V. CONCLUSION

In this paper we proposed the ensemble machine learning technique which is a novel based approached to analyze depression. Depression is the most predominant condition that is prevalent in our society and it is considered as a crucial need to identify it and be diagnosed. The performances of the classifiers applied invidually are measured in terms of accuracy, f-score and training loss. The performance of our ensemble machine learning model is compared with other standard machine learning algorithms. From the experimental results it is proven that our ensemble machine learning model provides a high accuracy of 91%.The

key aspects of the proposed model are the accuracy, model building time and false positives. The time taken to build the model and false-positive are comparatively low when compared to multi-layer perceptron, SVM classifier and random-forest algorithms when applied individually. To summit it all up, the ensemble machine learning model applied to analyze depression is a novel based approach as it provided a higher accuracy and less model building time.

VI. REFERENCES

1. S. Jain, S. P. Narayan, R. K. Dewang, U. Bhartiya, N. Meena and V. Kumar, "A Machine Learning based Depression Analysis and Suicidal Ideation Detection System using Questionnaires and Twitter," 2019 IEEE Students Conference on Engineering and Systems (SCES), 2019, pp. 1-6, doi: 10.1109/SCES46477.2019.8977211.
2. Gupta, D. J. . (2022). A Study on Various Cloud Computing Technologies, Implementation Process, Categories and Application Use in Organisation. International Journal on Future Revolution in Computer Science &Amp; Communication Engineering, 8(1), 09–12. https://doi.org/10.17762/ijfrcsce.v8i1.2064
3. S. Ali, S. S. Tirumala and A. Sarrafzadeh, "Ensemble learning methods for decision making: Status and future prospects," 2015 International Conference on Machine Learning and Cybernetics (ICMLC), 2015, pp. 211-216, doi: 10.1109/ICMLC.2015.7340924.
4. McElroy E, Fearon P, Belsky J, Fonagy P, Patalay P. Networks of Depression and Anxiety Symptoms Across Development. J Am Acad Child Adolesc Psychiatry. 2018;57(12):964-973. doi:10.1016/j.jaac.2018.05.027
5. Kadhim, R. R., and M. Y. Kamil. “Evaluation of Machine Learning Models for Breast Cancer Diagnosis Via Histogram of Oriented Gradients Method and Histopathology Images”. International Journal on Recent and Innovation Trends in Computing and Communication, vol. 10, no. 4, Apr. 2022, pp. 36-42, doi:10.17762/ijritcc.v10i4.5532.
6. AUTHOR=Zinchenko Oksana, Yaple Zachary A., Arsalidou MarieTITLE=Brain Responses to Dynamic Facial Expressions: A Normative Meta-Analysis JOURNAL=Frontiers in Human Neuroscience VOLUME=12 YEAR=2018 PAGES=227 DOI=10.3389/fnhum.2018.00227
7. Ananthakrishnan, B., V. . Padmaja, S. . Nayagi, and V. . M. “Deep Neural Network Based Anomaly Detection for Real Time Video Surveillance”. International Journal on Recent and Innovation Trends in Computing and Communication, vol. 10, no. 4, Apr. 2022, pp. 54-64, doi:10.17762/ijritcc.v10i4.5534.
8. Groves, Samantha & Pitcher, Toni & Melzer, Tracy & Jordan, Jennifer & Carter, Janet & Malhi, Gin & Johnston, Lucy & Porter, Richard. (2017). Brain activation during processing of genuine facial emotion in depression: Preliminary findings. Journal of Affective Disorders. 225. 10.1016/j.jad.2017.07.049.
9. Roberts, T et al. “Distance to health services and treatment-seeking for depressive symptoms in rural India: a repeated cross-sectional study.” Epidemiology and psychiatric sciences vol. 29 e92. 13 Jan. 2020, doi:10.1017/S204579601900088X
10. A. Jan, H. Meng, Y. F. B. A. Gaus and F. Zhang, "Artificial Intelligent System for Automatic Depression Level Analysis Through Visual and Vocal Expressions," in IEEE Transactions on

Cognitive and Developmental Systems, vol. 10, no. 3, pp. 668-680, Sept. 2018, doi: 10.1109/TCDS.2017.2721552.
11. Chekroud, Adam & Zotti, Ryan & Shehzad, Zarrar & Gueorguieva, Ralitza & Johnson, Marcia & Trivedi, Madhukar & Cannon, Tyrone & Krystal, John & Corlett, Robert. (2016). Cross-trial prediction of treatment outcome in depression: A machine learning approach. The Lancet Psychiatry. 3. 10.1016/S2215-0366(15)00471-X.
12. Nouby M. Ghazaly, M. M. A. . (2022). A Review on Engine Fault Diagnosis through Vibration Analysis . International Journal on Recent Technologies in Mechanical and Electrical Engineering, 9(2), 01–06. https://doi.org/10.17762/ijrmee.v9i2.364
13. Tack C. Artificial intelligence and machine learning | applications in musculoskeletal physiotherapy. Musculoskelet Sci Pract. 2019 Feb;39:164-169. doi: 10.1016/j.msksp.2018.11.012. Epub 2018 Nov 23. PMID: 30502096.
14. S. Pandey, Dr. Supriya M., and Shrivastava, A., "Data Classification Using Machine Learning Approach", in 3rd International Symposium on Intelligent Systems Technologies and Applications (ISTA'17), , Manipal University, Karnataka , 2017.
15. N. S. Parameswaran and Dr. Venkataraman D., "A computer vision based image processing system for depression detection among students for counseling", Indonesian Journal of Electrical Engineering and Computer Science, vol. 14, no. 1, pp. 503-512, 2019.
16. N. S. Srimadhur and Lalitha, S., "An End-to-End Model for Detection and Assessment of Depression Levels using Speech", Procedia Computer Science, vol. 171. pp. 12-21, 2020.
17. K. V. P. Nandan, Panda, M., and S. Veni, "Handwritten Digit Recognition Using Ensemble Learning", in 2020 5th International Conference on Communication and Electronics Systems (ICCES), 2020.
18. Pawan Kumar Tiwari, P. S. . (2022). Numerical Simulation of Optimized Placement of Distibuted Generators in Standard Radial Distribution System Using Improved Computations. International Journal on Recent Technologies in Mechanical and Electrical Engineering, 9(5), 10–17. https://doi.org/10.17762/ijrmee.v9i5.369
19. T. Antony, Kanmani, J., and Anjana A. P., "Study to Assess the Seizure Severity, depression and Quality of life among Patients with epilepsy at AIMS, Kochi", Indian Journal of Public Health Research & Development, vol. 9, p. 25, 2018
20. Sobha George, Francis, P. T., Paul, N., and Leelamoni, K., "Depression Among Elderly :A Comparative Study Among the Elderly Living In Old Age Homes In Ernakulam ,Kerala State,India", Internatinal Journal of Emerging Technology & Research, vol. 2, no. 4, pp. 62-66, 2016