Application of Various Machine Learning Techniques in Sentiment Analysis for Depression Detection

Soundariya R S, Nivaashini M, Tharsanee R M, Thangaraj P

Abstract*: Depression is the world's fourth leading disease and will be in the second in 2020 according to the statistics of World Health Organization. Depression affects many people irrespective of their age, geographic location, demographic or social position and more commonly affects females than males. Depression is a mental disorder which can impair many facets of human life. Though not easily detected it has intense and wide-ranging impressions. Although many researchers explored numerous techniques in predicting depression, still there is no improvement and the generations are facing higher rate of depression. It is believed that the depression detection algorithms can be more accurate and their performance can be better if they rely on artificial intelligence. On considering these factors, it is planned to perform a survey on the application of various machine learning techniques that have been used in the domain of sentimental analysis for depression detection.

Keywords–Depression detection, Machine Learning, Sentiment Analysis, SVM, Decision Trees, Depression rating mechanisms, BDI, HAMD.

I. INTRODUCTION

Major depressive disorders are a disorder affecting person'sway of living. Depression leads to negative affect and mood. Ithas such an impact on people that they started feeling ore ofthose leisure activities which they used to enjoy earlier .Their approach is negative towards everything. This is acommon disorder for every generation. But millennials reporthigher rates of depression than any other generation [4]. About20% of the people suffer from depression and anxiety. Theyfeel that stress keeps them awake all night (Health Status).

Machine Learning is getting computers to program themselves.It is systematising the method of automation. Machine Learningis utilized in computer science and other fields. And so for thistype of prediction systems, machine should be learned withmultifaceted patterns as that of human brain. There are various Techniques used for prediction of depression such as facialimages, speech tones etc. But most of the ongoing researchesare based on a particular learning method. To know whichprediction model is precise a comparison between dissimilarmodels is essential. The proposed survey focuses on study of variousmachine learning methods used for detection of depression

II. RELATED WORK

In this study, we investigate the various machine learning algorithms that can support for depression detection. Sharifa et al had worked with hybrid classifier for predicting depression based on eye movements; he had tried the combination of Gaussian Mixture Model (GMM) and Support Vector Machine (SVM) to develop an affect sensing mechanism. His research in depression detection had many perspectives like predicting positive and negative emotions, classification rate of gender based dependency modes and blinking rates. Gaussian models were generated for each subject and the generated models were fed into SVM for classification. The positions of eye movements on expressing various emotions in both genders were studied and accurate results were obtained for detecting depression [1]. Shamla et al had used gestures and finger points for depression detection. As feature extraction, functional points and action points were extracted from facial images using Active Appearance Models (AAM) and FACS. Finally, SVM was used for converting the extracted features into number of classes with the help of FERET database [2].

Truong et al designed a real-time health care system for monitoring depression based on the traditional client server architecture. User data were collected via sensors placed on terminals, the collected data was then sent to the diagnosis server for further process like feature extraction and classification. Based on the factors like convergence/divergence rate and complexity measure, new set of features like Largest Lyapunov Exponent (LLE), Sample Entrophy (SE) and Regularity Dimension (RD) were introduced. The extracted features were then passed on to different machine learning techniques like SVM, K- Neareast Neighbour (KNN) and Fuzzy C-Means (FCM). The algorithms were rated based on the performance measures like specificity and sensitivity. From the results, it was inferred that the combination of Largest Lyapunov Exponent and Regularity Dimension had shown optimised results particularly on working with SVM. The final classification results will be stored in the data centre [3]. Zhenyu et al had attempted to detect depression from speech signals collected from three categories of people like healthy, depressed and high risk. The authors had tried using filter approach and search algorithm as two stage feature selection methodologies as they had concentrated on

Revised Manuscript Received on July 08, 2019.

* Correspondence Author

Soundariya R S *, Department of Computer Science & Engg his/her department, Bannari Amman Institute of Technology, Erode, India. Email: rssoundariya@gmail.com

Nivaashini M, Department of Computer Science & Engg his/her department, Bannari Amman Institute of Technology, Erode, India. Email: nive19794@gmail.com

Tharsanee R M, Department of Computer Science & Engg his/her department, Bannari Amman Institute of Technology, Erode, India. Email: tharsh01@gmail.com

Thangaraj P, Department of Computer Science & Engg his/her department, Bannari Amman Institute of Technology, Erode, India. Email: thangarajp@bitsathy.ac.in

creating the most compressed dataset. Once the features were compressed, the reduced feature set was allowed to pass through KNN and SVM for further classification [4].

Changye et al had tried predicting depression from internet behaviour by means of extracting features based on time – frequency. It was believed that time – frequency features can enhance the efficiency of the prediction model and hence Discrete Fourier Transform (DFT) and K-means clustering were utilized. DFT was used to extract the initial features, while K-means clustering was used to identify the features that are morecorrelated with the depressed status. Different classification techniques like Back Propagation neural networks, Logistic regression and decision tress were used for building the prediction model. From the experimental results it was found that Back Propagation neural networks outperforms the other algorithms and had shown greater accuracy in classifying healthy and unhealthy classes [5].

Huda et al had attempted to detect depression in Malay speakers using acoustic measures. Power spectral density and transition features were the two important voice featuresthat were identified as the pointers for depression. Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA) were used to attain the decisionRestrictions for the classification of Healthy and Depressed states. Resampling methods like Equal-Test-Train, Jackknife and Cross Validation were used in LDA and QDA. In future, additional acoustic features can be added to make the approach more effective [6].

Xiaowei et al had explored the use of eye movements for depression detection. For every individual, features like saccade, fixation and pupil size were extracted. Classifiers like Logistic Regression (LR),Support Vector Machine (SVM), Random Forest (RF) and Naïve Bayes (NB), were applied to a group of students into two categories: depressed and normal. The main objective of the research was to ascertain the feasibility of the utilization of eye movements for depression detection [7].

Michelle et al compared the speech and text based features for building an automatic depression detection system. The researcher had also attempted to model the speech to text automated system for detecting depression. In case of speech signals features like prosodic features, phonetics were considered and as far as text was considered, linguistic features like semantics and syntactic features were taken into account.The performance of the proposed model was evaluated based on statistical metrics like Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). The proposed method was tested using Audio/Visual Emotion Challenge (AVEC) dataset. The same dataset was used for obtaining both speech and text features. Using the AVEC dataset, transcripts were generated and from the transcripts text based features were generated. The researchers had tried adding an automatic speech recognition system into depression detection system [8].

Jian et al had proposed EEG based depression detection system along with a splitting criterion for EEG.On considering the non-linear nature of EEG signal, three non-linear features like Renyi entrophy, Correlation dimension and CO complexity were extracted. Also certain features based on frequency were extracted in order to identify the linear nature of EEG signals. Further, in order to test the differences between EEG and splitting criteria of EEG,Friedman Test and post-hoc two-tailed NemenyiTest were performed [9]. Mandar et al had concentrated on natural language processing for emotion based depression detection using Twitter API. Neutral and negative emotions were identified from Twitter data and from those emotions, depressed state was predicted. SVM and Bayesian classifiers were used for predicting various class labels using the metrics like F1-score, confusion matrix and accuracy. Among the various naïve Bayes classifiers, multinomial naïve Bayes classifier was used for classification, as it shows better performance in terms of handling large amount of data[10].Yan et al had analysed EEG signals for mild depression detection using Particle Swarm Optimization (PSO) in terms of multi-objective. The research work had also included Linear Discriminant Analysis (LDA) as part of feature extraction. The main objective of the research was to reduce the amount of misclassification rates, minimize and maximize the internal and external distance accordingly [11]. Zhenyu et al had proposed the use of ensemble methods for two class depression detection from speech signals.Different classifiers like SVM, random forest, KNN, decision trees and Naïve Bayes were used to train the learning parameters based on theconcept of same and different utterance. In general, sequential based ensemble learning methods, base learners were created either one by one or parallel and subset of the base learners were allowed for ensemble pruning, in order to surge the performance of ensembles and to minimize the level of complexity. Complementariness Pruning (CP), Margin Distance Minimization (MDM) and Ensemble Pruning via Individual Contribution ordering (EPIC) were the three different pruning methods used for research. To examine the performance of the proposed ensemble methods, four set of experiments were conducted and the results indicated that the pattern combining classifiers trained on multi-utterance achieved better results than ensemble methods joining multiple classifiers trained on the same dataset. Proper ensemble pruning may possibly expand the performance and diminish the computational cost [12].

Madhurima et al had performed a comparative analysis on various machine learning algorithms for depression detection. The analysis includes various classifiers like Bayes Net Classifier (BN), Multi-Layer Perceptron (MLP), Decision Trees, Logistic Regression, and Sequential Minimal Optimisation. In case of machine learning methods, classification and regression based method were given importance in both supervised and unsupervised algorithms. Out of all SVM shows better performance and it was known to the most widely used algorithm. The research work had also analysed the various performance metrics for both classification and regression algorithms. In case of classification, accuracy, specificity and sensitivity were the measures and in terms of regression, correlation coefficients and mean squared error were found to be the most suitable measures.The research also includes the study on feature reduction methods, cross validation methods and various risk factors for depression. Among the 5 classifiers, Bayesnet classifier had the most accurate and reliable results [13]. Arkaprabha et al had used machine learning techniques to predict the level of anxiety and depression in elderly patient. The experiments were conducted among 500 patients with ten different classifiers and tenfold cross validation methodologies. Among the classifiers like

Naïve Bayes, Bayesian Networks,J48, Multiple Layer Perceptron, Random Trees, K star, Sequential Minimal Optimisation, Random Forest outperforms the best. The performances of the classifiers were evaluated based on the parameters like precision, F1 measure and accuracy [14].

Jung et al had designed a non-intrusive sensing model to detect depression in elderly people using the data collected from various types of sensors. Foe this experimental set up, literature survey was carried out in two different perspectives: Depression Detection and Sensor based activity recognition systems.Infrared motion sensors were placed in various part of an elderly house. Contact sensors were fixed in the main door, in order to trigger the opening and closing of main door. In order to monitor the movements of elderly people inside the rom and across the room, PIR sensor were placed in each room. The feature learning module comprises of three layer algorithms to extract the features for further design and analysis of the system. Principal Component analysis was used for dimensionality reduction. Four common classification prototypes, including Support Vector Machine (SVM), Bayes Network, C4.5 decision tree, and Artificial Neural Network (ANN), were applied to identify the elderly patients suffering from the early symptoms of depression and among them Artificial Neural Network shows good results when compared to the traditional techniques like SVM, Decision trees[15].Risa et al had performed an analysis of certain features in electroencephalography (EEG) signals collected from patients with depression. The collected EEG signals were pre-processed using band pass filter. Consequently, the amplitude of the filtered signals was normalised. Once the values were normalised, frequency conversion was done using Fast Fourier Transform (FFT).Then, time series analysis was carried out for healthy patients and depressed patients. Finally two-sample t – test was performed, on analysing the changes in the t-value, different features were identified due to substantial differences [16]. Qing et al had performed depression detection for imbalanced social media data using various deep learning approaches. To deal with such data, X-A-bi-directional Long Short Term Memory (X-A-BiLSTM) was used. The approach had two segments: To handle data imbalance XGBoost was used and BiLSTM was used for efficient classification. Experiments were performed as 4 different variations such as LSTM, Attention-LSTM, Attention-BiLSTM and X-A-BiLSTM. Out of which the proposed model, X-A-BiLSTM, had greater results in terms of precision, recall and F1 Score. [17].

III. MACHINE LEARNING MODELS FOR DEPRESSION DETECTION

This Section presents the most widely used machine learning models for depression detection

A. Support Vector Machine (SVM)

SVMs, as derived from statisticallearning theory, are powerful classifiers that have been successfully applied to many pattern recognition tasks, such as classification and regression. They have shown to outperform many other machine learning methods, such as artificial neural networks and k-nearest neighbours. SVMs have attracted a great deal of attention from the machine learning community, due to their unique properties, such as good generalization performance, robustness in the presence of noise, ability to deal with high dimensional data, and fast convergence [15].

B. K-Nearest Neighbor (KNN)

KNN is widely used in text classification and pattern recognition because of its property like simplicity and operation. KNN algorithm is a classic analogy-based classification algorithm, it does not establish a classification model as SVM, it stores all the training samples until the test samples are classified. The basic process is as following: when a test sample is given, kNN algorithm searches for the n-dimensional pattern space of the training data, and finds the k training samples closest to the sample to be sorted by a certain distance measure, and finally the category is judged to be the class that has the most nearest neighbors of k-nearest neighbors. However, there are many problems: both the selection of k and distance formula and the uneven distribution of the sample set will have impacts on the classification accuracy [18].

C. Random Forest (RF)

RF is an assembly of random decision tree classifier. Predictionoutput of RF classifier is made of all prediction outputs by individualclassifier [19].

D. Decision Tress (DT)

A decision tree is a flowchart-like structure in whicheach internal node represents a “test” on an attribute (e.g.whether a coin flip comes up heads or tails), each branchrepresents the outcome of the test and each leaf node representsa class label. The paths from roots to leaves representclassification rules. It is trained by measuring decrease ofentropy [7].Decision trees are the one in which positive results are yield that is will detect depressed people. Another in which negative results are yield that is will detect non - depressed people [10].

E. Naïve Bayes (NB)

Naive Bayes Classifier is a classifier which implementsBayes theorem with a solid (naive) independence assumptions,particularly independent feature model. BayesTheorem works on conditional probability which finds out theprobability of an event given that some other event has alreadyoccurred. It predicts the conditional probability of a class giventhe set of evidences and finds the most likely class based onthe highest one [20]. A naive Bayes classifier is a famous andpopular technique because it is very fast approach and gives ahigh accuracy. It is robust to irrelevant features, but if featuresare not conditionally independent or are not Gaussian distributed,its accuracy will decrease [7].

This study aimed to identify the various machine learning classifiers/algorithms that can support to design predictive model which can detect depression and can be best implemented in general practices to achieve good results.

IV. DATASETS FOR DEPRESSION DETECTION

This Section presents the various datasets and noticeable depression rating mechanisms used by the researchers

Authors	Number of Subjects	Dataset	Type of Depression Rating Mechanism used
Shamla et al [2]	-	Face Recognition Technology (FERET) database	Hamilton Depression Rating Scale (HAMD)
Zhenyu at al [4]	300 subjects 100 - healthy controls 100 - depressed Patients 100 - high-risk people	Chinese Facial Affective Picture System (CFAPS)	Diagnostic and Statistical Manual of Mental Disorders (DSM-IV)
Xiaowei et al [5]	Female – 12 Male - 24	Chinese Facial Affective Picture System (CFAPS)	Beck Depression Inventory (BDI-II)
Huda et al [6]	Female – 17 Male - 26	-	Beck Depression Inventory (BDI-MALAY) and Beck Hopelessness Scale (BHS)
Changye et al [7]	728 Postgraduate students	-	Zung Self-rating Depression Scale, SDS
Michelle et al [8]	84 subjects	Audio/Visual Emotion Challenge (AVEC) dataset	Beck Depression Inventory (BDI-II)
Mandar et al [11]	-	Twitter API	-
Zhenyu et al [13]	184 subjects 92 - depressed patients 92 - healthy controls	Chinese Facial Affective Picture System (CFAPS)	Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) and Hamilton Rating Scale for Depression (HRSD)
Arkaprabha et al [14]	510 patients	-	Geriatric Depression Scale(GDS), Hamilton anxiety and depression scale, hospital anxiety and depression scale (HADS)
Jung et al [15]	20 elders	-	Geriatric Depression Scale (GDS)
Qing et al[17]	9000 users	-	Reddit Self-reported Depression Diagnosis (RSDD) dataset

I. Discussion and Conclusion

Depressive disorder makes people unfit not onlyphysically but mentally also. It has diverse effects on quality oflife of people. Moreover it doesn't need depression to be athigher levels to affect a person's life. Many researches aregoing on prediction of depression and from those past studiesthe three major methods were studied to determine the mostaccurate method of them all. After studying machine learning Classifiers, Feature Reduction Method, Cross ValidationMethod, the most consistent and accurate methodwas Support Vector Machines.Different data sets can be used to analyse the performance ofthe predictive model. In future

other deep learning methods can be tried forthis prediction of depression for better accuracy.

REFERENCES

[1]S. Alghowinem, R. Goecke, M. Wagner, G. Parker and M. Breakspear, "Eye movement analysis for depression detection," 2013 IEEE International Conference on Image Processing, Melbourne, VIC, 2013, pp. 4220-4224.

[2]S. Mantri, D. Patil, P. Agrawal and V. Wadhai, "Cumulative video analysis based smart framework for detection of depression disorders," 2015 International Conference on Pervasive Computing (ICPC), Pune, 2015, pp. 1-5.

[3]T. C. Thang, H. T. Tran, V. D. Trong, D. V. Nguyen, H. T. Le and T. D. Pham, "Design and implementation of an e-Health system for depression detection," 2015 IEEE International Conference on Consumer Electronics (ICCE), Las Vegas, NV, 2015, pp. 74-75.

[4]Z. Liu et al., "Detection of depression in speech," 2015 International Conference on Affective Computing and Intelligent Interaction (ACII), Xi'an, 2015, pp. 743-747.

[5]X. Li, T. Cao, S. Sun, B. Hu and M. Ratcliffe, "Classification study on eye movement data: Towards a new approach in depression detection," 2016 IEEE Congress on Evolutionary Computation (CEC), Vancouver, BC, 2016, pp. 1227-1232.

[6]H. Azam et al., "Classifications of clinical depression detection using acoustic measures in Malay speakers," 2016 IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES), Kuala Lumpur, 2016, pp. 606-610.

[7]C. Zhu, B. Li, A. Li and T. Zhu, "Predicting Depression from Internet Behaviors by Time-Frequency Features," 2016 IEEE/WIC/ACM International Conference on Web Intelligence (WI), Omaha, NE, 2016, pp. 383-390.

[8]M. R. Morales and R. Levitan, "Speech vs. text: A comparative analysis of features for depression detection systems," 2016 IEEE Spoken Language Technology Workshop (SLT), San Diego, CA, 2016, pp. 136-143.

[9]A novel depression detection method based on pevasice ECG and ECG splitting criteria

[10] A Study and Comparison of Prediction Algorithms for depression detection among millenials_ML approach

[11] M. Deshpande and V. Rao, "Depression detection using emotion artificial intelligence," 2017 International Conference on Intelligent Sustainable Systems (ICISS), Palladam, 2017, pp. 858-862.

[12] Y. Guo, H. Zhang and C. Pang, "EEG-based mild depression detection using multi-objective particle swarm optimization," 2017 29th Chinese Control And Decision Conference (CCDC), Chongqing, 2017, pp. 4980-4984.

[13] Z. Liu, C. Li, X. Gao, G. Wang and J. Yang, "Ensemble-based depression detection in speech," 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Kansas City, MO, 2017, pp. 975-980.

[14] Predicting anxiety and depression in elderly patients using ML

[15] J. Kim, N. Liu, H. Tan and C. Chu, "Unobtrusive Monitoring to Detect Depression for Elderly With Chronic Illnesses," in IEEE Sensors Journal, vol. 17, no. 17, pp. 5694-5704, 1 Sept.1, 2017.

[16] R. Nakamura and Y. Mitsukura, "Feature Analysis of Electroencephalography in Patients with Depression," 2018 IEEE Life Sciences Conference (LSC), Montreal, QC, 2018, pp. 53-56.

[17] X_A_BiLSTM a Deep Learning Approach for DD in imbalanced data

[18] J. Huang, Y. Wei, J. Yi and M. Liu, "An Improved kNN Based on Class Contribution and Feature Weighting," 2018 10th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA), Changsha, 2018, pp. 313-316.

[19] Beura S., Majhi B., Dash R., ET AL.: 'Classification of mammogramusing two-dimensional discrete orthonormal S-transform forbreast cancer detection', Healthc. Technol. Lett., 2015, 2, (2),pp. 46–51

[20] A. N Hasan, B. Twala, and T. Marwala,Moving Towards AccurateMonitoring and Prediction of Gold Mine Underground Dam Levels, IEEEIJCNN (WCCI) proceedings, Beijing, China, 2014.

AUTHORS PROFILE

Soundariya R S *, Department of Computer Science & Engg his/her department, Bannari Amman Institute of Technology, Erode, India. Email: rssoundariya@gmail.com

Nivaashini M , Department of Computer Science & Engg his/her department, Bannari Amman Institute of Technology, Erode, India. Email: nive19794@gmail.com

Tharsanee R M Department of Computer Science & Engg his/her department, Bannari Amman Institute of Technology, Erode, India. Email: tharsh01@gmail.com

Thangaraj P , Department of Computer Science & Engg his/her department, Bannari Amman Institute of Technology, Erode, India. Email: thangarajp@bitsathy.ac.in