# Study of Depression Analysis using Machine Learning Techniques

**Devakunchari Ramalingam,Vaibhav Sharma, Priyanka Zar**

***Abstract:Depression is a major health issue that imparts major impact on the stability of mind. With the extension of various social media platforms, an expansion of number of different platforms enabled people to interact and share their experiences. These provided a large dataset for identification of common traits among depressed people and identify them using various machine learning algorithms. The limit to which we can identify the depressed traits of the person is necessary to determine the level of depression. The classification plays a major role in determining the kind of help a depressed person needs and also, the person with suicidal thoughts need to be identified and helped according to his condition. This paper provides the survey about the use of machine learning techniques in the analysis of depression with their research issues.***

***Keywords:Depression, Mental Stability, Social Media, Suicidal***

## I. INTRODUCTION

Depression is quite common among different age groups and is a very serious mental illness that affects our mental state, how we perceive things and how to act on stimulus which is given to the nervous system. Everyone feels sometimes down or low in their lives but when this feeling persist for longer time instances, it affects general cognitive function of the brain. The more careful task is proper separation of the symptoms of depression from sadness or the blues.

The onset of depression first affects the mood of the person which persists for longer period when it becomes the main symptom of depression. It affects the body via low energy, sexual dysfunction, thoughts and feelings. It might lead to some serious brain functioning disabilities such as panic attacks, anxiety, fear etc. Depression is different from sadness/grief. It is often believed that depression is a result of chemical imbalance but that does not capture the complexity of the disease. Researchers suggest that depression does not occur because of chemical imbalance in brain but comes from a variety of factors such including faulty mood regulations by brain, genetic vulnerability, stressful life events, medication and mental issues. It can be several of these factors interacting with one another to bring depressed state symptoms and effects on body. Chemical imbalance involvement is present but the imbalance is not cause of depression by one chemical rather, it is by a number of chemicals which interact or are involved in this process. With this level of complexity, we can comprehend how two people having similar symptoms but having physical, social conditions different in terms of living but similar in terms of disease.

Researchers are able to identify this gene which may affect person making him vulnerable to low moods and how person responds to drug therapy for this gene. The aim of these researches is to achieve better personalized treatment for the symptoms of depression. The influence of brain plays a significant role in the onset of the symptoms relating individual to depression. Nerve cell connections, nerve cell growth and functioning of nerve cell play a part towards the low changes in mood. Still the understanding of neurological connection is not accurate but to a level is successful in determining the state of mind of the person via correct understanding of neurological balance.

According to World Health Organization (WHO), depression is the most common illness worldwide and a leading cause of mental disability. The 7.6% of people suffering from these diseases are from age group 12 which shows how early this disease can arise in an individual. The test for depression detection often are based on interaction of patient with a psychiatrist which involves physical presence of the patient. With the advancement of technology, a number of different social media platforms are able to support people to describe their feelings which were earlier suppressed. These platforms not only provide a method to analyze the mental stability according to different age groups but also, provide a platform for supporting and identification of such sets of people.

The recent report on the global burden of diseases, and in it the depression among different genders with help of data obtained from such sources predicted that unipolar depressive episodes to be 1.9% in case of men and 3.2% in case of women, and one-year prevalence have been estimated to be 5.81% in case of men and in case of women 9.54%. It is estimated that by the year 2020 if these trends continue, then the burden of depression will rise to 5.7% of total burden of diseases. There is a need of smart systems to analyze the symptoms and work on the data sets for accurate and timely prediction of emotions of a person. Also, a test based on AI algorithms under different scenarios for detection of emotional imbalance.

## II. RELATED WORK

Sridharan et al. presented the detection diagnostics on online social media with the help of Convolution Neural Networks (CNN) where emphasis was to get the data posted on social networking websites

**Revised Manuscript Received on May 29, 2019.**

**DevakunchariRamalingam,**Assistant Professor, Dept. of Computer Science and Engineering, SRM Institute of Science and Technology, Chennai, Tamilnadu, India. (Email: devakunchari.r@ktr.srmuniv.ac.in)

**Vaibhav Sharma,**UG student, Dept. of Computer Science and Engg., SRM Institute of Science and Technology, Chennai, Tamilnadu, India. (Email: vaibhavsharma_r@srmuniv.edu.in)

**PriyankaZar,**UG student, Dept. of Computer Science and Engg., SRM Institute of Science and Technology, Chennai, Tamilnadu, India. (Email: priyankaravindra_zar@srmuniv.edu.in)

by different users while also ensuring algorithm [1] preserves the privacy with the help of separating agents which deal with data. An agent handles users' isolation requirement by creating different communities in the system. The dominant capacity through transfer learning and feature learning of CNN is used. A different filtering mechanism is proposed for the system where the filtering of redundant data set is done so as to increase the learning of CNN and hence reducing the faulty learning of the model. Online filtering threshold works via the help of Online Setup Assistant (OSA) practice. The OSA presents the user with the set of messages preferred for each meaning; the consumer tells the classification choice to recognize or to reject messages. This method helps in formulation of a proper classification and addition of certain messages from undesired creators into the Blacklist. It also ensures that all the future block sets for unwanted messages cannot continue feeding false information to the learning system. The system make an attempt to automatically identify the potential online users with SNMDD (Social Network Mental Disorder Detection) framework that explores the various features from different data logs of OSN's and a new tensor technique for deriving latent features from multiple OSN's for SNMD (Social Network Mental Disorder) detection.

Melissa N Stolar, Margaret Lech, Shannon J Stolar, Nicholas B Allen approach [2] for the detection is very interesting and quite different from others. The model uses an optimized spectral roll off parameters for detection of depression symptoms from speech signals on the clinical dataset obtained. The adolescent speech signals were pre-processed and cleaned to remove background noise and also the cross talk of parents over the microphone. Further, the decimation of sampling frequency is done to achieve minimal amount of data and hence, achieving minimal processing time and memory requirements. A number of other pre-processing techniques such as Blind Source Separation (BSS) for reducing background noise recoded on microphone, Frame/Windowing were cleaned signal is normalized to within the amplitude range of -1 to 1 based on max amplitude, Voice Activity Detector (VAD) which works according to the production process of speech to categorize it as voiced or unvoiced. The feature analysis from the verbal inputs is done via different component algorithm for detection of feature and optimization of detected features for proper classification. Different technique employed was Spectral category (S*) and Spectral category (S). The S* category works on the principle of flux, Centroid, Entropy, Power spectral density, Formants and single roll-off whereas the spectral category S defines S* as the basic step upon which it adds optimized roll-off ranges. The classification of these features is done with the help of simple SVM classifier. In past studies gender dependence has improved depression classification either best for females, males and varied amongst features. In this study depression detection was more effective in males than females.

*Detecting depression stigma on social media*

A linguistic analysis for detecting depression was performed by Ang Li et al. [3]. A content analysis of depression-related Tweets [4] was performed by Patricia A. et al. A nationally-representative study [5] among U.S. young adults was done by Brian et al. using multiple social media platforms. Affective content analysis of Online Depression Communities was done by Thin Nguyen et al. [6]. The main aim of these systems is to efficiently design the algorithm for detection of the depression stigma. The data set is collected from websites such as Weibo and Twitter. First the collected dataset is analyzed with the help of syntax and semantics analysis which gives the sense of depression stigma among posts posted by different age groups. In this process the syntax is analyzed for finding certain key words and relevance of those key words is mode with the help of semantic analysis which finds the general emotion of the paragraph via understanding the emotion of the text also known as Emotion Detection Systems. Then the posts are classified according to the depression symptoms.

First, to minimize the effects of the keywords i.e. depressive disorder on data modeling, the posts which matched the keywords but did not reflect any stigma were categorized as non-stigma posts. Then, a certain number of posts were randomly selected from the pool of non-stigma posts to balance the number of posts in each class (i.e. stigma and non-stigma). Second, to improve the performance of the data modeling, the linguistic features which were reliable for differentiating between stigma and non-stigma posts were selected. Third, to maximize modeling performance, a series of classification models were built based on key features using four different classification algorithms, respectively, including Simple Logistic Regression (SLR), Multilayer Perceptron Neural Networks (MLPNN), Support Vector Machine (SVM), and Random Forest (RF). Each model was tested by 5-fold cross-validation. Specifically, the data was randomly divided into five subgroups with equal size. Each subgroup was then used to test the model that was built using the opposing four subgroups. After five rounds of model training, results of model training were integrated into a final model.

The results of this study confirmed that incorporating linguistic analysis methods into online detection of stigma can be useful to improve the performance of stigma reduction programs. First, social media data can be used to monitor users' stigmatizing attitudes towards mental illness. In this study, two human coders were recruited to perform content analysis on Weibo posts with keywords. The results showed that 6.09% of relevant Weibo posts reflected stigmatizing attitudes towards depression. This is inconsistent with a previous study conducted outside of China, which also employed human coders to manually process tweets and found that only 0.7% of relevant tweets reflected stigmatizing attitudes towards depression. Such inconsistency might be partly due to poor mental health literacy among Chinese people, and partly due to the role of anonymity in encouraging users to be more likely to express their stigmatizing attitudes on Sina and Weibo compared to Twitter. Besides, there were gender differences in stigmatizing attitude towards

depression, which suggests that attempts to reduce depression stigma should be gender specific.

*In-Time perception in depression*

A meta-analysis by Sven Thönesl, and Daniel Oberfeld analysis of time perception of the person is done [7]. Depressive patients frequently reported to perceive time passing as slow. Effects of depression on time passing judgments have been investigated in this model with the help of tasks such as: verbal time estimation technique, time production technique, time reproduction technique, and duration discrimination methods. Different ratings of the subjective flow of time is obtained as a result of these techniques. By using method of a classical random-effects meta-regression and robust variance estimation, the meta-analysis is aimed for evaluating the irregular results from 16 previous studies on similar topic. Furthermore, it represents a collective data of 433 depressive patients and 485 healthy subjects. As a result, it was found that the depressive patients perceive time as going slower relative to control subjects. The meta-regression also showed that there are no significant effects of interval duration. Thus, stating that the time interval analysis as not accurate method for predicting the results. However, there was a tendency of over production of short durations and underproduction of long durations in depressive patients compared to healthy controls. Depression has normal effects on the subjective flow of time and has no effect on duration based judgments.

*Filtering*

The filtering mechanism constitutes the system for preprocessing of the input where the aim is to provide the data set to the learning algorithm without any redundancies. The aim is to get proper learning dataset so as to improve the accuracy of classification of the problems by the system. The detection systems run on classification algorithms which bring out a feature in the entity that is useful for characterizing the entity accordingly. When it is fed with wrong or redundant dataset then, the learning algorithm fails to comprehend the meaning to the data set and that can lead to misclassifications of the input. The varying nature of input dataset can prove challenging to these filtering algorithms where at a time an algorithm has to filter the discrepancies in data set input in text, speech or in image. According to different sets of input the filtering mechanisms differ and their redundancies also differ. So, the filtering system must be able to overcome these all problems. For example, speech filtering systems as shown in fig. 1, here, there might be overlapping of sound in input, frequency ranges which can be non-essential for the detection of goal, and unrecognized words which might cause redundancies.

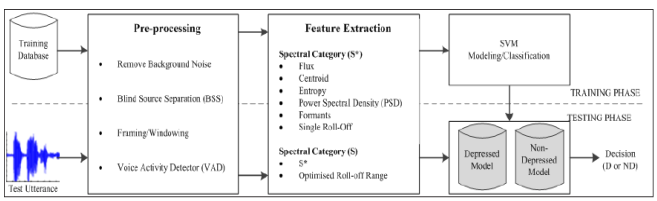


**Fig. 1. Procedural flow of acoustic depression recognition**

In case of image there are number of agents that might lead to misclassification of a feature which may be arise because of uncertainty handling in processing systems because of incomplete information, noisy images, fragmentary, imprecise, unreliable images, vague imagery, contradictory images leading to state of indecision, deficient images and overloading. However, a number of fuzzy logics and neural networks based methods have shown a great potential in determining the meaning of an image to the learning algorithm.

The field of image processing is very broad and encompasses wide variety of issues and with the limitations of dimensions removed new vast scopes provides large number of variances in the general learning of image. With 2D and 3D processing involved in a number of video and image inputs. The new findings in perceptron based networks enabled the detection of such systems.

In the case of textual inputs from different posts on social media algorithm tries to find the possible and probabilistic meanings of the text via active and passive construction of all the grammatical and general features of text and via using semantic algorithms to find out the emotion of user in case of depression analysis. The main constraint of the algorithm is to find the scenarios which are possible in the context and accordingly prediction of probabilities according to the implications of text implying to different emotions. The method of analysis includes the identification of grammatical meaning and trying to build the passive and active model for reaching probabilistic conclusion and often includes intermixing of scenarios also called as action mixing.

*Classification*

Classification is the backbone of the analysis system which aims to produce the result on the basis of the particular features of the entities constituting the dataset. Data obtained after the process of filtering is inputted to the classifier algorithms. In depression analysis model, we have different sets of inputs such as text, speech and image and according to different sets of these we use the different classifiers to properly classify them. In case of input in the form of speech, the model tries to recognize the words implied from speech and it works similar to text analysis but has added implementations of voice data which provides the information about pitch variations, intensity as well as depth of the speaker which could imply a lot towards the determination of mental state as well as emotional stability. In case of image analysis, a 2D feature vector is used to store the feature of every array of image which are divided into a number of different fragments so as to divide and study each individual part of the image. This feature vector is stored in the form of perceptron based neural networking scheme which represents the layer of information about an image vector.

A popular technique in classification is SVM (Support Vector Machines) , a classifier formally defined by decision planes which define the decision boundaries of the system.

The process of classification of an input dataset into result is step by step process and may involve the repletion of procedures to attain maximum accuracy at different levels of analysis. The model of depression analysis cannot be built upon a single group of classifiers and contains a number of classifiers in different groups. These groups are executed on a similar dataset for gaining the accuracy in the prediction. In depression analysis the first group of classification models was built to differentiate between posts with and without depression stigma. Second, to improve the performance of the data modeling, the linguistic features which were reliable for differentiating between stigma and non-stigma posts were selected. Third, to maximize modeling performance, a series of classification models were built based on key features using four different classification algorithms.

*Discussion*

A number of different researches in depression detection involve analysis of dataset to predict the abnormal behavior and suicidal thoughts among individuals. The motive of these techniques is to use the data which is available on twitter and other social media platforms such as Weibo for predicting the nature and the mind-set of individuals via analyzing their various social media posts.

The sketch of the algorithm included a Filtered wall algorithm which is able to filter unwanted messages to identify only the filters messages posted on platforms for improving the predictability for the neural network based systems to get accurate desired goals. The detection algorithm included a proper classification of the dataset systems. The results are in accordance with the input but lack the check for those users who do not express their emotions openly on the big web platforms.

The prediction of dataset needs to be accurate while predicting the person is depressed or not. The distinction algorithm analyses the clinical dataset of number of individuals for accurate prediction of mental state of individuals to tell if they are depressed or not. The model tries to strengthen the algorithm via improved classification of each individual's data according to various emotions and facial and verbal analysis techniques. For further optimization a number of models added the two stage sum feature. The system using this achieved average detection accuracy of 82.2% in case of males and 70.5 in case of females.

Time perceptron analysis provided a result based on estimation of time spent on social media platforms and relating that time to the increased anxiety levels of person. The model relates the aspect of depression symptoms which are caused by social media websites as an important source of person getting mentally stressed. The relation is that the more amount of time person spends on these websites, the more it leads to increased levels of anxiety among individuals.

## III. CONCLUSION

There are various methods which are used for detection of the depression among people of various ages. The method employed by these systems uses the method of detection via analyzing the posts on social media, doing syntax and semantic analysis for emotion detection of the person, so as to predict the depression levels among various age groups. The algorithms are designed to analyze the tweet for emotion detection as well as for detection of suicidal thoughts among people on social media. The mechanism does analysis of the tweets for prediction of depression without checking the validity of tweets. Social media is an open platform where many people refrain from telling their true emotions that might relate to depression they are facing, and so the model analysis here are mostly based on the prediction from posts using various machine learning algorithms. The main requirement of model is to be perfectly able to predict the result as there are a number of implementations that require verification of data before predicting the thoughts or posts of the person as suicidal or non-suicidal.

## REFERENCES

1. S.Sridharan, AkilaBanu, M. Bakkiyalakshmi, A. Buvana P; "Detection and Diagnosis on online Social network Mental Disorders using conventional Neural Networks"; International Journal of Engineering Science and Computing; 2018.
2. Melissa N Stolar, Margaret Lech, Shannon J Stolar, Nicholas B Allen; "Detection of Adolescent Depression from Speech Using Optimised Spectral Roll-Off Parameters"; Biomedical Journal of Scientific & Technical Research; 2018.
3. Ang Li, Dongdong Jiao, Tingshao Zhu, "Detecting depression stigma on social media: A linguistic analysis"; Journal of Affective Disorders; Volume 232; pp. 358-362; 2018.
4. Patricia A. Cavazos-Rehg, Melissa J. Krauss, ShainaSowles, Sarah Connolly, Carlos Rosas,MeghanaBharadwaj, and Laura J. Bierut; "A content analysis of depression-related Tweets"; Computers in Human Behavior; Volume 54; pp. 351-357; 2016.
5. Brian A. Primack, ArielShensa, César G.Escobar-Viera, Erica L. Barrett, Jaime E.Sidani, Jason B. Colditz, A. EveretteJames, "Use of multiple social media platforms and symptoms of depression and anxiety: A nationally-representative study among U.S. young adults"; Depression and Anxiety; Volume 33; pp. 323-331; 2016.
6. Thin Nguyen, Dinh Phung, Bo Dao, SvethaVenkatesh, MichaelBerk; "Affective and Content Analysis of Online Depression Communities"; IEEE Transactions on Affective Computing; Volume 5; pp. 217-226; 2014.
7. Sven Thönes, Daniel Oberfeld; "Time perception in depression: a meta-analysis"; Journal of Affective Disorders; Volume 175; pp. 359-372; 2015.
8. Tan Tze Ern Shannon, Dai Jingwen Annie, See Swee Lan, "Speech analysis and depression"; Asia-Pacific Signal and Information Processing Association Annual Summit and Conference; pp. 1-4; 2016.
9. Amir HosseinYazdavar; Mohammad Saied Mahdavinejad; Goonmeet Bajaj; KrishnaprasadThirunara, "Mental Health Analysis Via Social Media Data"; IEEE International Conference on Healthcare Informatics; pp. 459-460; 2018.
10. Quan Hu ; Ang Li ; Fei Heng ; Jianpeng Li; Tingshao Zhu; "Predicting Depression of Social Media User on Different Observation Windows"; IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology; pp. 361-364; 2015.

11. Hugo D. Caledron-Vilca, William I.Wun-Rafael, Roberto Miranda-Loarte; "Simulation of Sucidial Tendency by using Machine Learning"; 36th International Conference of the Chilean Computer Science Society; pp. 1-6; 2017.
12. Mandar Deshpande,Vignesh Rao; "Depression Detection using Emotional Artificial Intelligence"; International Conference on Intelligent Sustainable Systems; pp. 858-862; 2017.
13. Maryam Mohammed Aldarwish; Hafiz Farooq Ahmad; "Predicting Depression Levels Using Social Media Posts"; IEEE 13th International Symposium on Autonomous Decentralized System; pp. 277-280; 2017.
14. Shweta Oak; "Depression Detection and analysis by using speech or text as the input"; The AAAI 2017 Spring Symposium; 2017.