EasyChair Preprint

№ 10053

Mental Health Prediction Using Data Analysis

Kush Kaushik and Rahul Kapri

May 10, 2023

MENTAL HEALTH PREDICTION USING DATA ANALYSIS

Abstract: Mental and behavioral problems exist inall societies, in all phases of life, in men and women, in the rich and poor, and in rural and urbanpopulations. Approximately 450 million individuals globally are believed to be experiencing a mental or neurological disorder., including behavioral or substance-related problems, at any given moment. Social networking sites are a common way for people to express their feelings in the modern world. These kinds of emotions are frequently examined to forecast user behavior. This study uses an ensembled deep learning network to characterize these feelings to forecast the user's mental disorder. The analysis iscarried out on the Internet social networking platform, and both convolutional and recurrent neural networks are used to create the ensembling deep-learning model. In this study, multiclass classification is used to distinguish between dementia, psychosis, and Alzheimer's disease. Themulticlass classification procedure was carried outusing the suggested prediction method.

Keywords –Mental Health, Data Analysis, Stress, Parkinson.

1. INTRODUCTION

Many people communicate their feelings and views on social media Enterprises have progressively embraced social media as a means of promoting their products or services, and social networking sites (SNS) have emerged as a pivotal element in this endeavor. By enabling users to establish connections with one another, SNS present an opportunity for enterprises to foster two-way communication with their customers [1][1], And disease like Alzheimer which are neurological illness that causes progressive and permanent memory loss (AD). The early stages of Alzheimer's disease are sometimes difficult to notice since the symptoms grow slowly and gradually. A method for diagnosing Alzheimer's disease based on feature selection has been developed[2][2]. There is no medical remedy accessible anywhere in the globe.On several social media sites or online social health forums, users routinely disclose their psychological problems or disorders under pseudonyms [3][3]. Joining an online medical group allows you to sympathize with people who have similar issues. In an effort to self-diagnose, users commonly use social media to obtain information about their symptoms. Social media has been utilized by many researchers to research mental diseases like schizophrenia, depression, and anxiety as well as people's emotional states.

Recent research has utilized different methods to study depression and anxiety among social media users. One study collected tweets from individuals claiming to have depression and used an instrument called Lexical Using an approach that combines inquiry and linguistic analysis with the LIWC tool, individuals can monitor alterations in their social engagement patterns on the Twitter platform[4][4] Another study examined postpartum depression risk among Facebook users by comparing their depression levels before and after giving birth[5][5] They used image data to predict depression in social network users by analyzing Instagram images using face recognition and colorimetric analysis. In a previous study, Velocity embedding techniques, N-gram language modeling, and user posts can be utilized together to analyze and comprehend textual data were used to predict future anxiety symptoms among individuals[6][6]

2. Literature Survey

The methods for predicting mental illness through You may divide social media into two categories: [directly getting data from users with their permission,

To collect data for predicting mental illness through social media, researchers can either obtain data directly from users by asking them to fill out questionnaires and consent to data collection, or extract data from public posts through APIs. Crowdsourcing platforms and data donation websites like Our Data Helps can be used to invite users to participate. Questionnaires like CESD and BDI are commonly used to measure depression, while SPS and SWLS can help detect suicidal ideation. Some researchers suggest using human annotators to validate regular expressions due to challenges in data collection. The utilization of pre-established data collections, like those found in CLPsych and the Personality project, is a viable option [7][7]. Two primary techniques for data acquisition include gathering data directly from users using data collection tools with their consent [8][8], or extracting data from public posts by utilizing APIs. Researchers have used various techniques such as crowdsourcing platforms,

data donation websites[9][9].Predefined datasets to collect data, often using questionnaires like CESD, BDI, SPS, and SWLS to measure depression and suicidal ideation. The collected data is often preprocessed by removing irrelevant content like stop words, tweets, hashtags, and URLs, and converting emojis into ASCII characters. Feature extraction is then performed to select only the most important features, [10][10] which helps minimize training time, improve interpretation, and avoid overfitting. Most studies in this area focus on analyzing textual content and language patterns, and some research has shown that certain linguistic patterns are associated with specific mental health conditions The LIWC study has also been useful in this regard. Overall, the goal of this research is to discover novel the classifiers and uncover binding data.

The studies concentrate on textual elements, with only a few including picture analysis approaches. Kang et al. [11][11]

Researchers have used visual features such as color By utilizing compositions and SIFT descriptors, it is possible to extract emotional connotations from images uploaded on Twitter and Instagram.[4] Reece et al. have predicted indicators of depression in Instagram users using image color, saturation, and brightness. Guntuku et al. have also demonstrated that the VGG-Net, an image classification model, can be employed for the purpose of forecasting depression.[12][12] However, using multiple convolution neural networks for feature extraction can be time-consuming, so the suggested study[13][13]. Has used a limited number of these networks. Machine learning is becoming increasingly integrated in healthcare for disease prediction and visualization, which can assist doctors and medical analysts in making treatment and illness management decisions. Brain disease is caused by a damaged brain, not aging, and a deficiency in the brain may contribute to its growth. Moreover, this illness may cause changes in a patient's abilities, temperament, and personality[14][14]

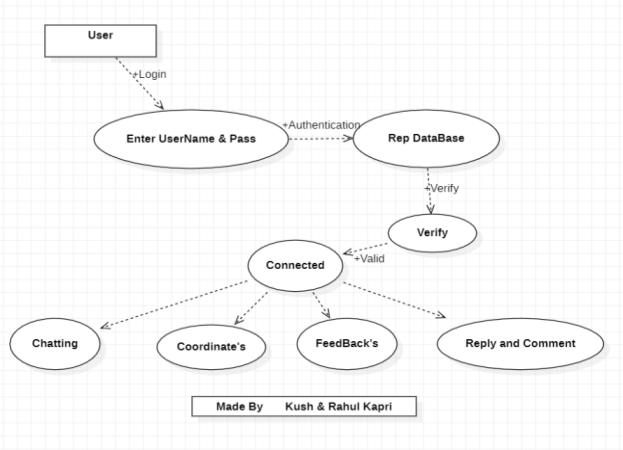
3. PROPOSED METHODOLOGY

This study categorizes mental illness into four distinct forms, namely bipolar disorder vs. non-bipolar disorder, dementia vs. non-dementia, and psychosis vs. non-psychosis. An ensemble of deep learning models is employed to build a deep learning model, which is used for classification. The novelty of this study lies in the use of an ensembled technique, combining the LSTM model and a convolutional neural network, for categorizing various forms of mental illness. Further details regarding the proposed methods will be discussed.

3.1 Dataset: The dataset, we are gathering from the user via completing a form in which they have to answer a set of questions, which are used by psychiatrists to determine the patient condition and severity.

The data was categorized Into distinct entries with labels such as anxiety versus non-anxiety, bipolar versus non-bipolar, depression versus non-depression,

Fig 3.1.0 Data Flow Diagram



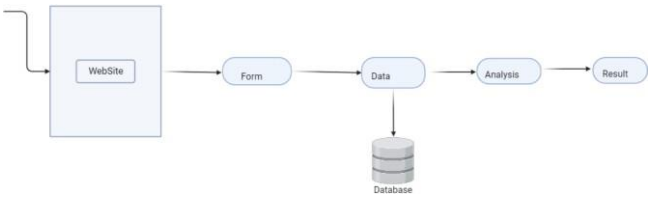
and psychotic versus non-psychotic. The dataset will undergo cleaning and lemmatization processes, among other things. It is worth noting that the most prevalent type of mental illness in china affects around 80 million people aged 18 and above. The majority of the dataset used in this study was obtained publication[15][15]

3.2 Ensemble Deep Learning Models

One tool for handling huge data is MongoDB. Genetic algorithms are used to further evaluate the data for various mental diseases, and the results are then put in MongoDB for ultimate data extraction. This strategy of data mining and information extraction decreased the overall cost of therapy. It offers the finest outcomes for clinical judgments. Using important information collected by the big data tool Mongo DB and genetic algorithm, it enables clinicians to provide more precise therapy for a variety of mental problems in less time and at

a lower cost.

Fig 3.2 Mental Health Analyzer



Researchers are attempting to forecast mental state prior to a severe mental stage using the MongoDB tool. Therefore, several gadgets established a thorough detection procedure to address the user's current situation by scrutinizing his or her everyday activities. Reasonable solutions are required to more accurately and speedily determine a patient's mental disability stage[16][16].

3.3 Convolutional Neural Network

In this paper, the analysis is carried out using a single CNN model. Convolution layers extract features from the input training data. Each convolution layer has a series of filters that help in feature extraction. In general, as the depth of the CNN model rises, so does the richness of features learnt by convolution layers. In a convolutional neural network, the first layer is responsible for detecting basic features, while the final layer extracts more complex features from the training data. To extract features, the convolution operation is applied to a segment of the data sample. The amount of data the filter can handle at any one time is determined by the distance covered by a single step and the value added to the edges of the input data before performing convolution may vary, such that the input data may or may not be augmented with zeros. In a study by researchers, an "Efficient K-Nearest Neighbor Classifier With Different Numbers of Nearest Neighbors" was introduced, as an example of feature extraction.[17][17].Following After passing through the ReLU activation function, the resulting output is fed into a pooling layer. A pooling layer eliminates any collected duplicate information during convolution In a research paper by M. Evanchalin et al. titled [18][18], an approach using decision Tree classifier based on Particle Swarm Optimization (PSO) has been developed on artificial intelligence was proposed in order to identify Alzheimer's illness. As a result of this layer, the data sample is smaller. Pooling is based on the assumption that neighboring pixels in a picture have substantially identical values. The average, lowest, and maximum values of four neighboring pixels are used for pooling. In the CNN model, a 29 filter is used to reduce before pooling, the input image can be reduced in size by 50 percentage. The dataset used for this study contains 430 neurological scans with age ranging from 18 to 86, out of which 109 subjects have been clinically diagnosed by dataset, 237 MRI scans with Alzheimer's disease were downloaded [19][19]. The input data may or may not undergo null padding before convolution. The convolution layer are repeated in the CNN model, typically 2-4 times for instructional purposes. The output of the convolution is processed through multilayer.

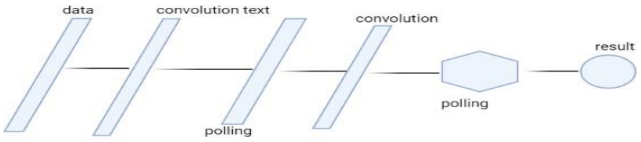


Fig 3.3 convolution Mode

3.4 Recurrent Neural Network

Our method of categorization is mainly based on an attention-based Bi-LSTM structure. Although CNNs reduce input information for prediction, every word has a distinct association with the ultimate classification. In this research, we aim to fully exploit the potential of CNN and Bi-LSTM. As, theBi-LSTM can more successfully decrypted word dependencies across extended distances. The Bi-LSTM method is depicted in Figure 1. It extractsthe last hidden layer using the characteristics of the CNN stage to produce new features. The Bi-LSTM architecture can obtain prior and subsequent contextual information[20][20], allowing for two distinct textual representations. A Random Forest technique has been proposed for regression analysis of neuroimaging data in Alzheimer's research, which is a recent and innovative approach for medical data analysis.[21][21] Nonlinear data handling is addressed in the

Popula paper. Another study employs structural features of brain MR imaging to classify Alzheimer's using WEKA and SVM algorithms. A Bi-LSTM model is used to provide a description of the sequence by using the characteristics gathered from the sequence. [22][22] The theory of early cognitive decline in the hippocampal texture has been evaluated by conducting classification training on three distinct datasets in another study. A Bi-LSTM model uses the CNN features to create a representation of the sequence, which is further improved by an attention layer that selects the most relevant features for the final categorization. In [23][23], the authors tested the theory of early cognitive losses in the hippocampal texture, and demonstrated the effectiveness of the attention mechanism can enhance the precision of predictions while simultaneously decreasing the number of trainable parameters needed. The attention mechanism assigns different weights to different parts of the input data, highlighting the most significant features for accurate prediction. For example, in the case of the statement "I'm enraged", the attention mechanism gives more importance to the word "mad" and less to "myself", allowing the model to accurately identify the statement as expressing anger despite the presence of other emotions.

4. RESULT ANALYSIS:

Various regions across the globe are using big data in mental health research for different purposes. The process of machine learning, also known as "training data" or "initial assessment data," involves establishing a baseline level of expertise in the program to effectively utilize and access the information. This approach aims to improve the necessary skills required for the program to work efficiently[24][24]

Diagnosis	Percentage	Number
Depression	25.3	10
Social	31.2	12
Anxiety Disease	10.2	24
Personality	34.2	45

Fig 4.0.0 Data Monitor

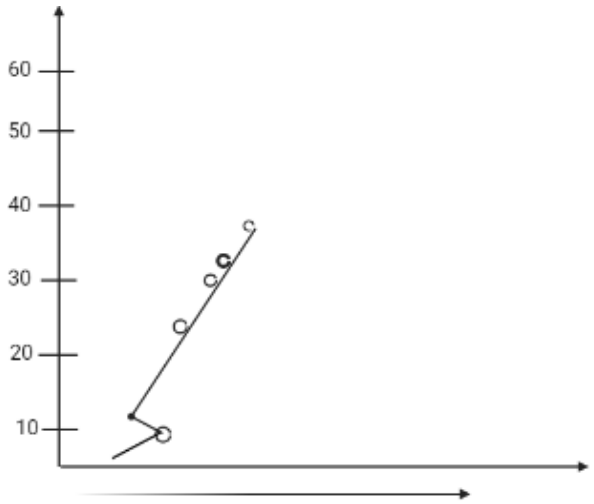


Fig 4.0.1 : Health Out Analysis

Primary	N	%	Sleep
Mood	799	9.7	1.7
Anxiety	523	25.6	1.98
Personality	678	31.1	2.1
Other	445	13	2.33

Fig 4.0.2 : Health OutCome

5.Conclusion

The deep learning model is used in this article to classify various forms of mental disorders. For or agnizing mental disorders such as anxiety, or disorder, dementia, and psychosis, the emblem deeplearning model employs while CNNs are primarily used for image classification, RNNs are more commonly used for natural language processing and speech recognition tasks. Based on measures such as accuracy, and precision, the ensemble deep learning model outperformed the current models. Different forms of Capsule networks, which belong to the family of convolutional neural networks, have the ability to incorporate a range of features used in future studies to identify mental illness using photographs.We reviewed many sorts of mental diseases as well as realistic, economical, and potential solutions to improve mental healthcare facilities. The digital mental health sector is advancing at a faster pace than scientific assessment, Therefore, it is imperative that researchers thoroughly investigate and analyze data using a diverse range of machine learning algorithms to determine the most accurate algorithm among them[25][25]. It is clear that the clinical community needs to keep up. To reduce the mortality rate of mental patients and prevent them from engaging in any criminal activities through early detection, numerous intelligent healthcare systems and tools have been developed.

6. References

[1] A. Wibowo, S.-C. Chen, U. Wiangin, Y. Ma, and A. Ruangkanjanases, "Customer behavior as an outcome of social mediamarketing: the role of social media marketing activity and customerexperience," *Sustainability*, vol. 13, no. 1, p. 189, 2021

[2] Alzheimer's Association,"2016 Alzheimer's disease facts and figures," Alzheimer's & Dementia, vol. 12(4), pp. 459-509, 2016

[3] C. S. Lee, G. N. Paul, J. W. Sallie, E. N. David, "Cognitive and system factors contributing to diagnostic errors in radiology," American Journal of Roentgenology, vol. 201(3), pp. 611-617, 2013

[4] J. Smith-Merry, G. Goggin, A. Campbell, K. McKenzie, B. Ridout, and C. Baylosis, "Social connection and online engagement:insights from interviews with users of amental health online forum," JMIR Mental Health, vol. 6, no. 3, Article ID e11084,2019

[5] D. P. Dudău and F. A. Sava, "Performing multilingual analysis with linguistic inquiryand word Count 2015 (LIWC2015). An equivalence study of four languages," Frontiers in Psychology, vol. 12, 2021.

[6] A. G. Reece and C. M. Danforth, "Instagram photos reveal predictivemarkers of depression," *EPJ Data Science*,vol. 6, no. 1, 2017.

[7] B. Roark, M. Saraclar, and M. Collins, "Discriminative n-gram language modeling," *Computer Speech & Language*, vol. 21, no. 2, pp. 373–392, 2007. View at: Publisher Site | Google Scholar

[8] D. Howard, M. M. Maslej, J. Lee, J. Ritchie, G. Woollard, and L. French, "Transfer learning for risk classification of social media posts: model evaluation study," *Journal of Medical Internet Research*, vol. 22, no. 5, 2020.View at: Publisher Site | Google Scholar

[9] R. Masood, F. Ramiandrisoa, and A. Aker, "UDE at Erisk 2019: early risk prediction on theinternet," *CEUR Workshop Proceedings*, vol.2380, 2019. View at: Google Scholar

[10] U. Naseem, I. Razzak, M. Khushi, P. W. Eklund,and J. Kim, "COVIDSenti: a large-scalebenchmark twitter data set for COVID-19 sentiment analysis," *IEEE Transactions onComputational Social Systems*, vol. 8, no. 4, 2021.View at: Publisher Site | Google Scholar

[11] G. Coppersmith, C. Harman, and M. Dredze, "Measuring post Traumatic Stress Disorder in Twitter," in *Proceedings of the EighthInternational AAAI Conference on Weblogs and Social Media*, Balitmore, MD, USA, 2014.Viewat: Google Scholar

[12] K. Kang, C. Yoon, and E. Y. Kim, "Identifying Depressive Users in Twitter Using Multimodal Analysis," in *Proceedings of the 2016 International Conference on Big Data and SmartComputing (BigComp)*, Hong Kong, China,January 2016. View at: Publisher Site | Google Scholar

[13] S. C. Guntuku, W. Lin, J. Carpenter, W. K. Ng, L. H. Ungar, and D. Preotiuc-Pietro, "Studying Personality through the Content of Posted and Liked Images on Twitter," in *Proceedings of the 2017 ACM on Web Science Conference*, New York, NY, USA, June 2017.View at: Publisher Site | Google Scholar

[14] R. Nair, S. Vishwakarma, M. Soni, T. Patel, and S. Joshi, "Detection of COVID-19 cases through X-ray images using hybrid deep neural network," *World Journal of Engineering*, vol. 19, 2021.View at: Publisher Site | Google Scholar

[15] C. S. Lee, G. N. Paul, J. W. Sallie, E. N. David, "Cognitive and system factorscontributing to diagnostic errors in radiology," American Journal of Roentgenology, vol. 201(3), pp. 611-617,2013.

[16] J. Kim, J. Lee, E. Park, and J. Han, "A deeplearning model for detecting mental illnessfrom user content on social

[17] Shichao Zhang, Senior Member, IEEE, XuelongLi, Fellow, IEEE, Ming Zong, Xiaofeng Zhu, and RuiliWang. ―The paperEfficient kNN Classification WithDifferent Numbers of Nearest Neighbors‖. (Shichao Zhang, 2017)

[18] M.Evanchalin,Sweety,G.Wiselin Jiji"Detection of Alzheimer Disease in Brain Images Using PSO and Decision Tree Approach", 2014 IEEE International Conference on Advanced Communication Control and Computing Technologies (ICACCCT).

[19] Open Access Series of Imaging Studies (OASIS), online: http://www.oasis- brains.org

[20] K.Venkatachalam, A.Devipriya,J.Maniraj, M.Sivaram, A.Ambikapathy, Iraj SAmiri, "A Novel Method of motor imagery classification using eeg signal", Journal Artificial Intelligence in Medicine Elsevier, Volume 103, March 2020, 101787

[21] Yasoda, K., Ponmagal, R.S., Bhuvaneshwari, K.S. K

Venkatachalam, “ Automatic detection and classification of EEGartifacts using fuzzy kernel SVM and wavelet
ICA (WICA)” Soft Computing Journal (2020).

[22] V.R. Balaji, Maheswaran S, M. RajeshBabu, M. Kowsigan, Prabhu E., Venkatachalam K,Combining statistical models using modified spectral subtractionmethod.

[23] Malar, A.C.J., Kowsigan, M., Krishnamoorthy, N. S. Karthick, E. Prabhu &
K. Venkatachalam (2020). Multi constraints applied energy efficient routing technique based on ant colony optimization used fordisaster resilient location detection in mobile ad-hoc network. Journal of Ambient Intelligence and Humanized Computing, 01767-9. IF 4.5

[24] Alzheimer's Association, "2016 Alzheimer's disease facts and figures," Alzheimer's & Dementia, vol. 12(4), pp. 459-509, 2016

[25]W. Bosl, A. Tierney, H. Tager-Flusberg, and C. Nelson, “EEGcomplexity as a biomarker for autism spectrum disorder risk,”BMC Medicine, vol. 9, no. 1, 2019