Prediction and Analysis of Recurrent Depression Disorder: Deep Learning Approach

Anagha Pasalkar
Sardar Patel Institute of Technology, Andheri (w), Mumbai, India

Dhananjay Kalbande
Sardar Patel Institute of Technology, Andheri (w), Mumbai, India
Corresponding author: Anagha Pasalkar, Email: anagha.pasalkar@spit.ac.in

Mental illness, such as depression, is rampant and has been shown to affect a person's physical health. With the growth in artificial intelligence (AI) various methods are introduced to assist mental health care providers, including psychiatrists to construct proper decisions based on patient's chronicle information including sources like medical records, behavioral data, social media usage, etc. Many researchers have come up with various strategies that include various machine learning algorithms for data analysis of depression. Although there have been less attempts previously to perform the same task without making the use of pre- classified data and Word-Embedding optimization Approach. For these reasons, this study aims to identify the deep formation of the neural network among a few selected structures that will successfully complement natural language processing activities to analyze and predict depression.

Keywords: Machine Learning, Deep Learning, MDD, RCNN, RNN, CNN, DL, Word-Embedding Optimization

1 Introduction

Recurrent Depression Disorder also known as Major Depression Disorder (MDD) is a common illness characterized by persistent sadness and loss of interest in ac-

2021. In Raju Pal & Praveen K. Shukla (eds.), *SCRS Conference Proceedings on Intelligent Systems*, 35–49. Computing & Intelligent Systems, SCRS, India. https://doi.org/https://doi.org/10.52458/978-93-91842-08-6-3

tivities that one is accustomed to enjoying, coupled with a failure to perform daily tasks. Finding a mental illness can be considered a daunting task, especially because of the difficulties the nature of mental disorders. In recent years, this research center has begun to emerge. These days, researchers are given more information about human health through society and social media too. With the increasing level of difficulties in identification of mental illnesses, usage of supervised machine learning techniques has not much included because of the complexities in achieving required amounts of defined training details. One of the emerging topics and developments in technologies such as AI, ML, and Deep Learning, which converts data through layers of offline processing units provides a new paradigm for processing and retrieving huge active data from complex information. DL algorithms have shown high performance parameters depending on many rich application scenarios, including health care and medications in recent years.[2, 6]

2 Literature Survey

Recurrent Convolution Neural Networks Text Sharing [1] In this model, the authors use a repetitive structure to capture as much detail as possible when reading word presentations, which can bring much lower noise compared to traditional window-based networks. They have also used a multivariate layer that automatically determines which words play a key role in the text category in order to capture the essentials in the text. They have researched four of the most widely used data. These test results show that the proposed method exceeds the data sets of other common methods, especially in the data sets at document level.

Finding depression using a framework that integrates multiple deep modal networks with automated testing for objectives. [2] They form a framework that can detect stress with minimal human intervention: artificial intelligence test (AiME). This framework contains a summary of computerized tests using artificial intelligence, eg in-depth reading, and you can guess whether the participant is the depressed or not who is performing satisfactorily. These technologies can provide a useful tool so that mental health professionals can detect symptoms of depression, thus enabling rapid preventive interventions. In addition, reduce the challenge of observing and interpreting unusual physical symptoms and behavioral pressures by providing more meaningful assessments.

A deep learning model for detecting mental illness from user content on social media [3] In this paper the authors compiled a post for mental health communities 18 eReddit. By analyzing and learning to post user-written information,

proposed the model can accurately identify whether the user's post is psychologically specific disorders, including depression, anxiety, bipolar, borderline personality disorder, schizophrenia, and autism. The study continues to discuss the impact of these proposals model, which can serve as an additional tool for monitoring the mental health conditions of people who often use social media.

Early Detection of Depression and Treatment Response Prediction using Machine Learning: A Review [4] This paper examines the many functions available in the ML application for early diagnosis of depression. In addition, the predictive review of treatment response is also available installed. By comparison, the apparent finding of stress work has been found to be positive due to the small database of images available in depression. Authors reviewed existing stress activity using imaging bio-markers, mobility tracks, various combinations of audio, video and text.

Predicting Depression Severity using multi modal functions [5] This study uses a real-world, clinically valid 30-year database depressed patients and 30 health management courses, using the Support Vector Machine in stages with strategies for selecting several features. The authors state the offer the nature of the statistical features extracted, feature selection based on T-tests done better than other methods. Among the various meeting methods used here, feature integration worked very well with up to 88 percent with moderate accuracy.

Pre-Depression Disclosure: Social Network Analysis and Random Forest Strategies [6] This study used data from social media networks to find a variety of MDD diagnostic methods based on machine learning. The authors have done a good Database analysis to show the features of the articles according to the various aspects of their texts. The authors have suggested 2 different methods based on singleton machine readings twice. The first uses 2 random forest partitions (RF) with 2 borders, and the last uses 2 independent RF separators, one to find pressure studies to identify non-depressed people.

Pressure detection from social networking data using machine learning strategies [7] In this study, authors aimed to perform pressure analysis on Facebook data collected from a social media online. Investigate the effects of depression, and promote the machine learning process as an effective and awesome process.

The Stress Discovery Model Based on Emotional Analysis on Micro-blog Social [8] The authors of this study first performed a psychological analysis using words and man-made information to calculate the micro-blog's tendency toward depression. Second, the stress detection model was developed based on the proposed procedure with 10 traits of depressed users drawn from psychological research. After that 180 users and 3 types of separators are used to validate the

model, its accuracy is about 80 percent. Also, the analysis of each item is analyzed. Finally, an application was made within the proposed online mental health monitoring model.

Using data analysis and machine learning for studying and predicting depression in users on social media [9] The study suggested a system for predicting potential users frustrated, with the intrusion features of already affected users. A a combination of both tweet levels and user-level constructions used for production a robust and reliable system in which semantic embedding is trained from advanced neural networks are accepted below the level of tweet, while at the user level, i a method that uses 12 key features was used in the wide-ranging feature engineering. In addition, SVM with Word2Vec and TF-IDF below the tweet level showed 98.14 percent accuracy and 95.63 percent memory and the gradient enhances differentiation below the user level revealed 95.26 percent accuracy with a memory of 86.75 percent.

Automatic Discovery of Speech Depression Using Ensemble Convolution Neural Networks [10] This program is based on Convolution Neural Study Networks (CNNs). In the pre-processing phase, speech files are represented as a sequence of log-spectrograms and random samples to measure positive and negative samples. With the work of isolation itself, first of all, the proper construction of this work, based on the One-Dimensional Convolution Neural Networks, built. Second, several of these genres based on CNN are trained for different uses and the related concepts are compiled using the Ensemble Averaging algorithm and integrated with each speaker to determine the appropriate final decision. Proposed Combination Program gets satisfactory results from AV-EC-2016 compared to Support Vector Machines-based reference system and handicrafts, with CNN + LSTM program called DepAudionet, and one divorce case based on CNN.

3 Research Gap Identified

- It is clear that most of the subjects identified in the textbooks used are monitored separation techniques than other ML methods
- Mental health researchers can look into the possibility of using the disorder, potential real-time ML analysis data.
- Most object machine learning algorithms do not have the use of Word-Embedding.
- It is important to calculate the sense of a sentence by pointing to the left

context once the correct context of a word using word embedding

- Object Context Identification of a sentence is not done in machine learning exactly as it is classifies according to labels provided in the database
- Contrary to traditional methods, we will have to present a decision to duplicate a neural network of text separation without human-made features.
- Also, these courses are limited and focus only on a small set of mental health conditions.

4 Existing Methodology

Recursive Neural Network-Recursive Neural Network is a section of Artificial Neural Networks that can process input sequences for in-depth learning and maintain its status while processing subsequent input sequences. Traditional neural networks will process input and move on to the next regardless of their sequence. However, Recurrent NN is a biased model, where the forth-coming words are more governing than the previous ones in larger extent. Therefore, it can reduce performance when used to capture complete semantic document, because the essentials can appear anywhere in the document rather than in the end. [1]

Recurrent Neural Network- Duplicate neural networks, also known as RNN's, are a category of neural networks that allow pre-existing effects to be used as inputs while in hidden cases. RNN has the opportunity to process inputs of any computer length based on historical data. However, Recursive NN captures the meaning of a sentence semantically using its tree formation. Its effectiveness is highly dependent on the performance of the structure of the expression tree generated. Moreover, creating such a text tree indicates the running time complexity of O n^2 at least, where n is the length of the sentence, causing the model to take more amount of time for its computation when the model encounters a long text or sentence. [1] [6]

Convolution Neural Network - Deep learning and training features of CNN, each input image will pass through a series of convolution layers with filters (Kernels), Pooling, fully integrated layers and use the Soft-Max function to distinguish an object with a value 0 and 1. Specifying the size of the window: the size of the small windows can lead to the loss of sensitive information, while the large windows lead to a much larger parameters which are difficult to train. [1]

5 Proposed Methodology

To address the limitations of the above model we can use Recurrent Convolution Neural Network (RCNN). [1, 2]

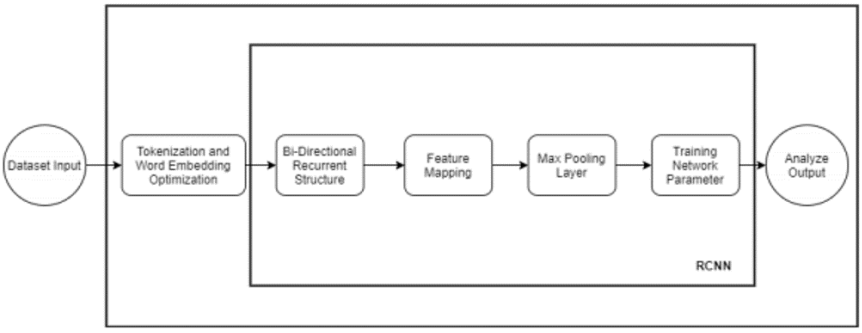

Figure 1: Block diagram of proposed system.RCNN model+Word Embedding Optimization Technique

1. Data set Exploration:

   The raw data of the proposed program was collected on various forums. For behavioral analysis and prediction, we have collected data from Twitter profiles of the users. We have collected data from Twitter as in particular the comments of the twitter user are one of the most important challenges with information on whether they can contain content that is disappointing.Postings on Twitter are especially helpful in analysis because they are often very common in day-to-day activities. This provides rich and consistent ways of capturing the moral qualities associated with one's thinking, emotions, hobbies, communication, and activities. To address this issue we use the keys to the Twitter API tokens to collect data from Twitter as our dataset. Improving data based on hashtags, retweets, and comments.

Table 1: Dataset Analysis Table

Number of User Account	200
Number of Tweets Gathered	3000
Mean of Number of Post per day	5.37
Mean of Number of Post per user	2.84

2. Data Prepossessing:

   To prepare this data for pattern processing and pattern analysis it must be passed to the data acquisition and prepossessing phase. Data processing not only improves data quality but also reduces log file size. Data prioritization involves a number of steps including data collection, data cleaning, feature set identification and setting required parameters. Feature parameters are encouraged by the stress detection process performed by domain experts with background information about depression, in which they look at the potential interactions of potential depressed users to find the right clues. For this reason, we have used four types of solid stress indicators-depressive symptoms lexicons, negative polarity(sentiments), ruminative analysis

   - Depressive Symptoms Lexicons-The first factor estimates the age of the words associated with the depressive, detailed description of the Twitter post.Posting ideas that are directly related to a particular symptom is obviously a very different behavior for people with depression. Based on this view, we suggest an embedded network for finding references related to post-depressive symptoms. To find out which sign related to depression, we use Word-Net which is a web dictionary.
   - Negative Polarity Sentiments-According to the researches on cognitive thinking, depressed people commonly show negative thinking and emotions. Therefore, depressed people on social media often expose their feeds and post as bad unity often than others. Depending on the concept, we suggest a second feature network that takes this practice into account in terms of emotions in the post.
   - Ruminative Analysis-It is known that ruminative thinking style can be displayed as repetitive thoughts and behaviors. People with depression are more likely to be depressed to express their feelings or negative experiences over and over so that the sentences in the relevant articles may appear more frequently in their posts.

3. Tokenization:

   In tokenization, the text would be separated into meaningful pieces, also known as tokens. After tokenization, word embedding is performed using Word2Vec. In this process we must apply tokenization techniques to assign random vector values to each sentence.

4. Word Embedding:

   The word embedding represents the text readings where Words that have the same meaning have the same numerical representation. Embedding the Word is actually a category of strategies in which individual words are represented as vectors with real value in the previously defined vector space. Performing word embedding optimization to the vectors generated.

5. Bi-Directional Recurrent Structure:

   The Recurrent Convolution Neural Network consists of a Bi-directional Recurrent Neural Network (RNN). RNN is used for capturing the long term dependencies by considering the context of a word in relation with its neighboring words. Since a Bi-directional Recurrent Neural Network is used, it captures the meaning and the context of the sentence.

6. Feature Mapping:

   The embedded map, or activation map, is the start of a given filter release. The feature map is the result of filters applied to previous layers. The given filter is drawn through the entire previous layer, inserting one neuron at a time. Each calculation result activates a neuron and the result is collected on a feature map. In applying the function of tanh activation to the successive matrix, the elements necessary for the detection of depression are performed. Eg- To apply the function to vectors is = tanh (Wxi + by) where 'i' is a vet semantic latent, in which each semantic element will be analyzed to find the most useful representation of the text.

   - Soft-max Activation function-It is often placed as the last layer in the deep learning model.This activation function is used for general network deployment and distribution opportunities over the predicted output phases. Mathematically defined as- where, y= input vector yi=element i of the input exp(yi)=exponential function applied on yi and a normalization term applied.
   - Tan-h Activation function- Tan-h activation is a differential monotonic function. The tanh function is most commonly used for separation between two classes. The good thing is that the negative input will be mapped to the worst and zero input will be mapped near zero on the tanh graph.

6 Results and Analysis

Hyper parameters are the parameters used in DL algorithms that determine the structure of a network including the bias element, hidden layers, and variables that also determine the basic parameters essential for network training. Hyper parameters are set prior to training, pre-screening and selection. Following are the hyper-parameters related to our network structure.

- Learning Rate - The way in which the level of learning changes over time (the training of epochs) is called the learning process or the decay of the learning level. Perhaps the easiest reading schedule is to reduce the level of consecutive reading from the first large number to the smallest number. This allows for larger changes in the node weights when the learning process emerges and a small number of changes or adjustments towards the ending of the learning process.
- Batch Size - The batch size indicates the total number of samples given as an input to the training network in small groups once after which the network parameter update took place.
- Number of Epochs - The total number of epochs specifies the number of iterations during which all training data is present on the network during training.

The graph in Fig-2 and Fig-3 depicts the frequency of words observed in the dataset belonging to the depressed and non-depressed class observed in the dataset.

The Table-2 depicts the comparison table for different algorithms used in this study. Each algorithm was given training dataset as an input to train the model and analyze its accuracy. It is observed from the table that RCNN combined with word-embedding optimization has proved to give better results as compared to other algorithms.

The pie chart in Fig-4 represents different depression indicators among which tweets of negative polarity were the highest followed by depressive symptoms and ruminative analysis.

The confusion matrix is a way to summarize the performance of a division algorithm. The confusion matrix can give a better idea of whether the designed classification model is accurate and what types of errors it makes. The different Matrix levels of confusion for the proposed system are as follows:

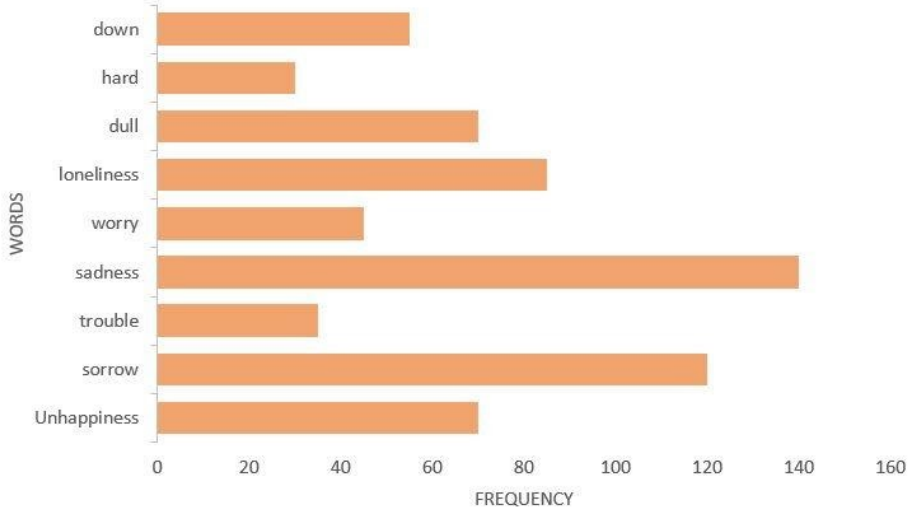

Figure 2: Depression word frequency

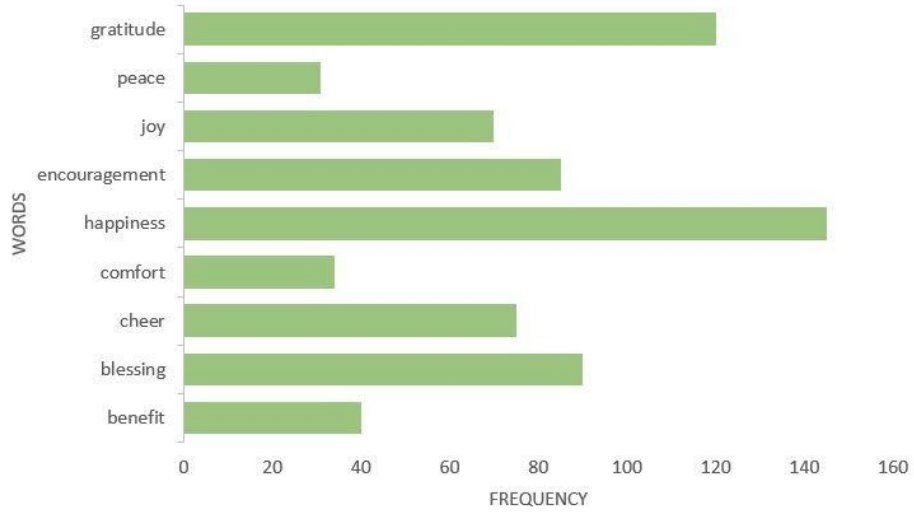

Figure 3: Non-Depression word frequency

- True Positive (TP) = The true positive value is 490 which indicates that the model was able to classify 490 depression class data from the whole given data set.

Table 2: Comparative Analysis Table

Algorithm	Batch Size	No. of Epochs	Training Accuracy
RNN	64	200	87.5322
CNN	64	200	85.611
RCNN	64	200	91.753
RCNN+Word Embedding	64	200	93.65

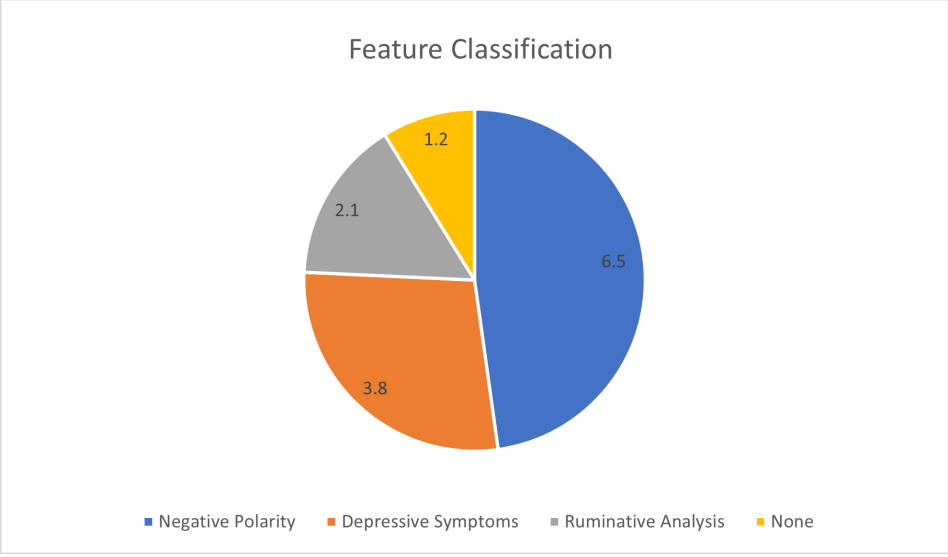

Figure 4: Depression Features Observed

- True Negative (TN) = The true negative value is 380 which indicates that the model was able to classify 380 non-depression class data from the whole given data set.
- False Positive (FP) = The false positive value is 13 which indicates that the model was able to classify 13 non-depression class data belonging to the depression class data set
- False Negative (FN) = The false positive value is 17 which indicates that the model was able to classify 17 non-depression class data belonging to the non depression class data set

The graph in Fig-6 depicts the training accuracy graph of RCNN algorithm

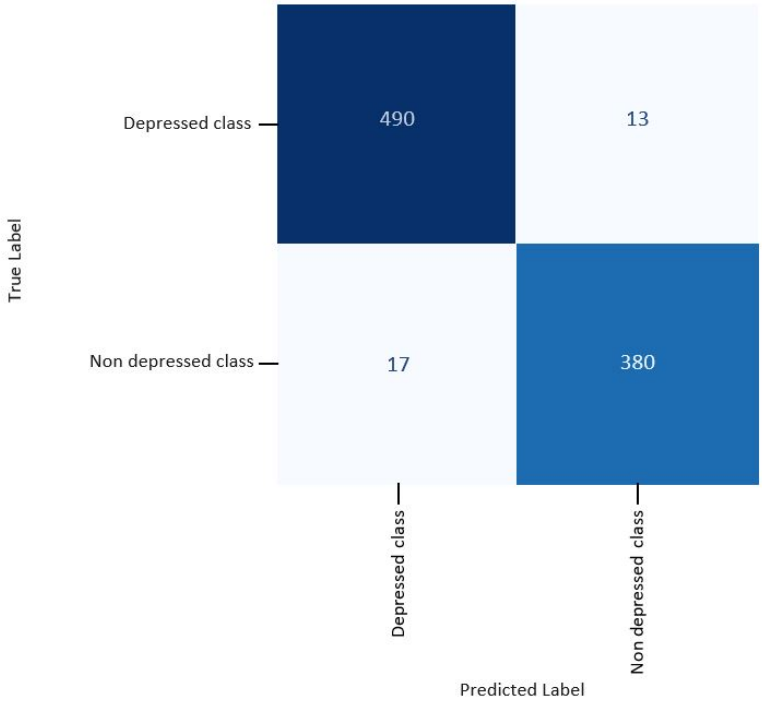

Figure 5: Confusion Matrix for the Proposed System

used in the proposed system for analyzing RDD. The system was trained to detect the depression tweets based on filtered tweets as data set. For training the system, the data set was split into a ratio of 70:30 where 70 percent of the total data set was utilized for training. The training accuracy achieved by the RCNN + Word Embedding Optimization is 93.65 percent and testing accuracy for same system is 91.37

7 Conclusion

Major depressive disorder (MDD) is the most common mental illness that has a significant impact on health quality and social and economic burden. In this study, we have introduced recurrent convolution neural network for classifying tweets as depressed.Contrary to traditional methods, we introduce a repetitive

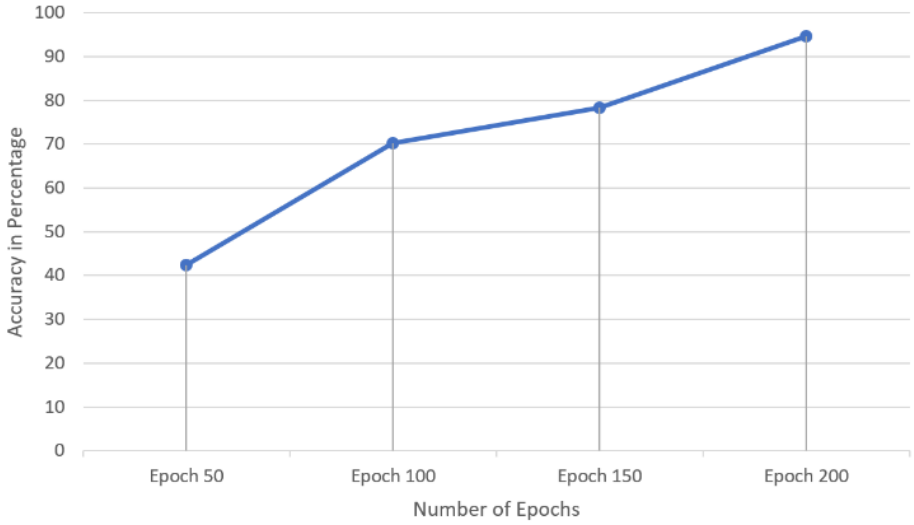

Figure 6: Accuracy graph for the Proposed System

network of neural variables that are RCNN text editing without man-made features. In our model, we use a repeating structure to enter as many details as possible when we read the presentation of words, which can produce much less noise compared to traditional window-based networks. We also use a multidisciplinary layer that automatically determines which words play important roles in the text category in order to capture the key elements of pressure analysis. RCNN captures bidirectional context eg left-sided dependence and right-hand dependence on each word in a sentence. Exploring bio signals in interpreting DL-based psychiatric disorders is the main focus of this study and, although challenging, with its promising potential to improve our understanding of psychiatry, in-depth learning algorithms are more exploratory and critical in terms of speed and accuracy compared to traditional methods.

Bibliography

[1] Recurrent Convolutional Neural Networks for Text Classification Siwei Lai, Liheng Xu, Kang Liu, Jun Zhao, Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence,2018

[2] Detecting depression using a framework combining deep multimodal neural networks with a purpose-built automated evaluation Ezekiel Victor, Zahra M Aghajan, Amy R Sewart, Ray Christian, National Library of Medicine, 2019

[3] A deep learning model for detecting mental illness from user content on social mediaJina Kim, Jieon Lee, Eunil Park Jinyoung Han IEEE Journal of Scientific Reports, 2020

[4] Early Detection of Depression and Treatment Response Prediction using Machine Learning: A Review Prajwal Kharel , Kalpana Sharma ,Sunil Dhimal , Sital Sharma Second International Conference on Advanced Computational and Communication Paradigms (ICACCP), 2019

[5] Predicting Depression Severity using multimodal functions Cornell University AI and more, 2017

[6] Early Detection of Depression: Social Network Analysis and Random Forest Techniques, Fidel Cacheda, Diego Fernandez, PhD,corresponding author Francisco J Novoa, PhD Victor Carneiro Journal of Medical Internet Research,2019

[7] Depression detection from social network data using machine learning techniques Md Rafiqul Islam Ashad Kabir, Ashir Ahmed Springer, 2018 .

[8] A Depression Detection Model Based on Sentiment Analysis in Micro-blog Social Network Xinyu Wang, Chunhong Zhang, Yang Ji, Li Sun Cambridge University Journal, 2018

2021. In Raju Pal & Praveen K. Shukla (eds.), *SCRS Conference Proceedings on Intelligent Systems*, 35–49. Computing & Intelligent Systems, SCRS, India. https://doi.org/https://doi.org/10.52458/978-93-91842-08-6-3

[9] Using data analysis and machine learning for studying and predicting depression in users on social media, Chanpreet Singh, Data Science Journal of Carleton University Ottawa, Ontario

[10] Automatic Detection of Depression in Speech Using Ensemble Convolutional Neural Networks, Adri´an V´azquez-Romero, Ascensi´on Gallardo-Antol´ın, International IEEE Journal,2020

[11] Wu, H.; Prasad, S. Convolutional Recurrent Neural Networks for hyperspectral Data Classification. Remote Sens. 2017, 9, 298.

[12] Convolution recurrent neural networks for bird audio detection in European Signal Processing Conference (EUSIPCO), 2017

[13] De Choudhury, M., Gamon, M., Counts, S., Horvitz, E. Predicting depression via social media. ICWSM. 13, 1-10 (2013)

[14] Mowery, D., Bryan, C., Conway, M. Feature studies to inform the classification of depressive symptoms from Twitter data for population health. Preprint at arXiv:1701.08229 (2017)

[15] Shen, G., Jia, J., Nie, L., Feng, F., Zhang, C., Hu, T., Chua, T.-S., Zhu, W. Depression detection via harvesting social media: A multimodal dictionary learning solution in IJCAI 3838-3844 (2017)