

Depressed People Detection from Bangla Social Media Status using LSTM and CNN Approach

*Tabassum Ferdous Mumu, Ishrat Jahan Munni, Amit Kumar Das**

Department of Computer Science and Engineering, East West University, Dhaka, Bangladesh

Received: January 28, 2021, Revised: February 23, 2021, Accepted: February 25, 2021, Available Online: March 06, 2021

ABSTRACT

At present, depression is the main reason for suicidal death. Depression also causes different kinds of diseases. Nowadays, people are deeply involved in social media and like to share their feelings on social media. So, it becomes easy to analyze depression through social media. In this paper, a combination of two CNN (Convolutional Neural Network) and LSTM (Long Short-Term Memory) models has been proposed to make a hybrid CNN-LSTM model, CNN has performed for the image to create a matrix, and LSTM has given the result from the given matrix. In this paper, datasets are prepared based on depression and non-depression-related status. The proposed method has been applied to that dataset. The best result has been obtained using a hybrid neural network with the word embedding technique using the Bengali Facebook status dataset. We have used the SVM (Support Vector Machine) model to predict a small dataset of Bengali Facebook status and count vectorizer to count the word in the document. Finally, this paper has built up a model that makes strength and support for deep learning architecture.

Keywords: Depression, CNN-LSTM, SVM, Word Embedding, Neural Network.



This work is licensed under a [Creative Commons Attribution-Non Commercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/).

1 Introduction

Depression is a silent killer. Depression is a serious and common medical illness, and it negatively impacts how someone feels, thinks, and behaves [1]. With the help of technology and skills, humans are now at a different level [2]. But between the rapid growth, people forget to trend their mental health. That's the world we currently live in. In a developing country like Bangladesh, things are more severe than one can look at. In Bangladesh, people do not have time to think about any type of mental illness where basic needs are lacking. This is a serious issue that takes many lives over the years. According to the CDC report (Centers for Disease Control), one of the leading causes of suicide is depression [3]. Even if the world is becoming more open-minded about human behavior and mental illness, Bangladesh is still lagging in this sector. In many countries like Bangladesh, mental illness is at a tolerable level. People with mental illness are not treated nicely. That's why depression becomes a significant issue because people cannot share their problems with others who lack trust in cognitive cases.

Nowadays, people become more and more expressive on social media than ever. It's called virtual self-disclosure. Self-disclosure can be done on various platforms of Social media. Sometimes, people think offline nobody can optically discern them; they are at liberty to express their genuine emotion. It helps some people to cope-up with the stressful life. They want to express most of the virtual life emotions that they cannot express in real life. Applications such as Twitter, Facebook, Instagram, and many other platforms write and multimedia type contents and express their emotions, feelings, and sentiments about different subjects, topics, or issues online [4]. Expressing emotion online leads to detecting a person's emotions, such as happiness, anger, depression, sadness, fear, surprises.

This work aims to find out if a person is having any kind of depression or not using a machine learning approach utilizing

social media status related to depression. The authors tried to scale the depressive status based standard scale of depression. There have two types of data set one is contained depressive status, and another is non-depressive status. This research used two main algorithms, i.e., LSTM and CNN. Differentiate between depressive and not depressive status TF-IDF (Term Frequency- Inverse Document Frequency) was used, and CNN and LSTM were used to train the model. For comparing results, logistic regression was also used to train the model. There are two different accuracies from different model training.

The study is to simply find depressive status from social media that can disclose the early stages of depression. While successful first treatment can lead to a positive result, untreated depression can lead to even life loss. The study is limited to extrovert social media users only. Because many people silently cope up with their depression. They don't express it on social media, but that does not mean that they are not depressed.

Depression is a mood disorder that causes different persistent feelings like sadness and loss of interest in life. Its effects are feeling, thinking, behavior and might result in a range of physical and emotional types of problems. Normal activities can be quite troublesome, and sometimes the range of feeling come to like life is not worth living [5]. According to WHO (World Health Organization), depression is the leading cause of disability worldwide [6]. Depression is an ongoing problem, not a passing one. It can last a week, a month, a year, or a decade. Depression is not just for a particular age of people. Every age group can have depression. CDC report says depression is twice as common among women as men. According to the American Psychological Association, around 9% of men have depression in the USA. Even children at age 3-17 age also have depression [6]. So, it is a significant problem now around the world.

Many factors can play a role in the causes of depression. The main problem is that it is just the problem of living, not the real enemy. So, it is hard to fight off. There is much reason for people

to feel depressed. Even a lack of nutrition can cause depression [7].

Causes can range from biological to circumstantial. Some key reasons are horrible family history, early childhood trauma, less active brain, medical conditions, drug use, etc. Substance use problems can also cause depression. Like low esteem, mental illness, stressful events, etc. [8]. Taking about Bangladesh, the main problem is a lack of awareness about depression. According to WHO, 6 million Bangladeshi suffer from depression [9]. According to the Global School-Based Student Health Survey (2014) data, 13-17 years of children in Bangladesh, 7% attempted suicide in the last 12 months, while 8% made a suicide plan, and 5% considered trying it. The same report shows that 8% of children cannot sleep because of Depression [8].

Depression can be treated and prevented. The better option is prevention than treatment. A better understanding of depression can avoid that problem and can also help the treatment. Taking therapy, anti-depression medication is the primary treatment for depression. More income, a better relationship, and tablets are more commonly suggested treatment [9]. It is necessary to end the silence and overcome the fear of talking with someone professional so depression can avoid. Major Depression is hard to cure, but it is also preventive. In this research collection of data is based on different types of depression but just some random popular depressive status. Major Depression and minor Depression are hard to recognize solely based on status. That's why there is no level the depression only considers as a depressive status.

Depression is a significant public health concern all over the world. While successful early treatment can lead to a positive result, untreated depression can lead to even the cause of life. It is proved from many researchers that disclosure of depression, or we can say emotion can prevent the worst scenario, can release stresses. Social media is the best platform for people to disclose their emotions in this new area. Self-disclosure is in social media is a stress reliever for modern people nowadays. So, one can get the hype of their emotion through social media. We cannot guarantee that everything in social media is real. In our study, we collect some status to detect depression and at least warn their loved ones that maybe he/she is depressed.

Measurement of Depression is a depression scale that identifies if any person is alright or he/she has to undergo some treatment to cure depression. The method is to ask some questions, and the result will be according to the answer. This question is almost accurate in finding people's mental states. This scaling purpose is to recognize people at risk of developing the disorder only. The depression scale determines the severity of the depression, and the professionals will determine what type of treatment is more appropriate [2]. Depression scale can be classified by researchers and also by patients. One of the completed ranges by researchers is HAM-D (Hamilton depression) rating scaling. It is a widely used and oldest version. It takes 20-30 minutes to administer. That's why many people prefer The BDI (Beck Depression Inventory). It is a patient scale rated scale. It has high specificity and sensitivity. Currently, HAM-D and BDI are the best-validated scales [10].

For our research, the paid attention to GDS (Geriatric Depression Scale). It is designed to minimize the impact of depression symptoms. It has a yes/no format and has good sensitivity.

2 Related Works

As an Asian region, people use Facebook than twitter most of the time, so Billah and Hassan [11] locked their target on Bengali Facebook status. They tried to predict the depression from Facebook status. The collection of data is randomly on Facebook and even has suicide victim's data. Naive Bayes, Linear SVC, Logistic Regression, Multinomial Naive Bayes, and other classifiers are used. Five types of features are used in their model. Because of the lack of data model provided the best 77.96% accuracy. Traditional machine learning procedures are only used for the first Bangla attempt and small dataset. The researchers proposed that they could achieve up to 100% accuracy if they used the neural network.

Islam et. al. [4], with the help of SVM, DT (Decision Tree), KNN (K-Nearest Neighbors) performed depression analysis. They aimed to perform depression analysis on Facebook data collected from online sources. First, they proposed emotional, temporal, linguistic styles for their data collection. With their chosen classification techniques, they study each factor independently. They used the NCapture tool for data collection. Their research applied 21 types of attributes for detecting Depression from LIWC (Linguistic Inquiry and Word Count) software. They achieved accuracy between 60 to 80%.

Ramalingam et. al. [12] developed an SVM model to detect depression. They argued that men and women are classified differently according to depression. Their dataset consists of different sets of inputs like text, speech, and image—applied different classifiers to classify them correctly. The model contains several classifiers in groups. They achieved an accuracy of 82.2% in the case of males and 70.5 in the case of females.

Uddin et. al. [13] research topic is mostly similar to ours. They used a special type of RNN (Recurrent Neural Network), GRU (Gated Recurrent Units), to detect Bangla language depression status. They collected 500 tweets and 210 depressive statuses using the survey. They tried to prove that questionnaires and academic interviews are not the only way to detect depression. Because their dataset only contains 1,176 data, they achieved 75.7% accuracy with GRU, but with LSTM, they gained 88.6% accuracy.

Choudhury et. al. [2] in their research measure depression of university undergraduates for the purpose to help them if they needed any treatment for their depression. Their main objective was to find the main reasons for depression and whether depression can be predicted successfully. They found that most university students are suffering from different types of depression. But in this country, depression is not a major enough issue for the university to provide a psychiatrist for students. They took a survey from university students. They used six algorithms Gradient Boost, Deep learning, Generalized Linear Model, KNN, Random Forest, and SVM, to train and test the dataset. They found that deep learning is the best method among them, with the lowest number of false negatives, followed by the Gradient Boost algorithm. From the Generalized Linear model, they get their highest accuracy, which is 74.17%.

3 Materials and Methods

People of Bangladesh face different kinds of problems. For all of these, they suffer from depression. Bangladesh is a developing country, so poverty is the main reason for depression. Nowadays, the younger generation also faces love problems. Some people even get depressed regarding their studies.

Bangladesh contains a larger population than its land, so people face job problems.

On the other hand, there are a lot of non-depression sentences. Daily people post their status, which is not under depression. This study aims to implement a combined deep learning classifier to enhance the performance of language modeling and text classification to detect depression using the dataset of Bengali Facebook status.

In the framework, datasets are transformed into word vectors using word embedding. Then the word vector is fed into the CNN layer. In the max-pooling layer, it makes a sequential layer to feed into the LSTM layer.

3.1 Data Collection

Bangla language is comparatively more complicated to recognize than English. Most of the analysis is done in the English dataset. So, finding many datasets is not possible for analysis, and experiments used two different datasets: depressive and non-depressive. But it was challenging to find depressive Bangla datasets, as the presence of depressive Bangla datasets is less frequent. From existing paper 1177 depressive and non-depressive status was collected [13]. A better performance needs more dataset. The self-developed dataset contains Bangla status from different sources like Facebook, Twitter, YouTube comment sections. There is also some translated Bangla status in the dataset. The authors collected some non-depressive data from GitHub and Kaggle. The datasets are labeled like depressive datasets as '1' and non-depressive as '0'. The primary dataset contains 5053 depressives, and the 2110 non-depressive dataset total dataset is 7163. As for deep learning, the large dataset is required authors thought about this dataset is incomplete. That's why the dataset is still in the updated phase. Sample of datasets are given in Table 1.

Table 1 Sample data set

Depressive Data	Non- Depressive
ভাল থাকুক তারা যাদের জন্য আমি আজ মৃত প্রায়।	করোনা নিয়ে সবাইকে সচেতন হতে হবে
অভিশপ্ত এই জীবন। আত্মহত্যা ছাড়া আর কোন উপায় নাই আমার কাছে।	আহারে কাজের কোন শেষ নাই
ভবিষ্যত সম্পূর্ণ অন্ধকার।	ভালো কথার ভাত নাই।।এইটাই বুঝলাম
অভিশপ্ত আমি আজ। এটাই আমার প্রাপ্য. সারা পৃথিবী পেয়ে আজ তোমাকে হারলাম।	সামান্য কিছু টাকার কাছে আজ মানবতার পরাজয়।
প্রতিকূলতার সাথে আর পেরে উঠছি না আমি। আমি এক অসুখী সূর্য। সংসার এর মায়া ত্যাগ করার সময় চলে আসছে। কিনা করলাম তোর মত স্বার্থপর এর জন্য। আর তুই কি দিলি? তোর মত স্বার্থপর এর কোনদিন ভাল হবে না।	নিরাপদে থাকুন
নিজের পাপে জর্জরিত আমি। জগতেই পাই আজ নরক যন্ত্রণা। আমার এই কদাকার রূপ আমি কাউকে দেখাবো না	পরিবার নিয়ে সুখে আছি

3.2 Data Pre-processing

An input text has filtered using pre-processing to enhance a proposed model's performance by eliminating unnecessary features to process row posts before acknowledging word embedding. Pre-process of the dataset was needed. To make the dataset clean and eligible for testing, used Natural language toolkit (NLTK) [14] for pre-processing datasets. The genetic kit is a platform that is used to build Python programs that work in statistical NLP (Neuro-Linguistic Programming) with language data to apply. It includes libraries for processing text for tokenization, sorting, grouping, stemming, tagging, and semblance reasoning. We have used NLTK to remove punctuation, stop-words according to the given text document. But it did not remove all the punctuation that is used as emoticons signs. Bangla dataset contains many misspelling and incomplete words. Some spelling seems to be different from each other, but they are the same. For this type of word, stemming was done. Removed all the URLs with white-space from the dataset, and then we got our required dataset.

3.3 Proposed Neural Network

3.3.1 CNN

CNN is known as an artificial neural network that is used in recognize image and processing. It is designed especially for the processing of pixel data. Height, weight, the depth that three types of dimensions are organized in the CNN layer. CNN layer contains two components: one is feature extraction, and another is classification. In feature extraction, convolution series and pooling operation are done. And in classification, a fully connected will do their work [15]

CNN can predict using the spatial structure as its input. First, from the input data, it produces a feature map by using a filter. A convolution layer is created after padding the feature map. Then the pooling layer is performed to decrease the parameter size and execution in the network. Max polling is mostly used as a pooling layer. It reduces the size of the feature map by choosing the maximum value from every window. It does not lose any kind of important information. In a fully connected layer, flatten is performed to convert 3D or 2D data into 1D. Fully connected shows the output from the last layer of the neural network. Finally, CNN shows a sequential sentence as the result of output [16].

3.3.2 LSTM

LSTM is an artificial neural network used for the classification process and predicts the result of output in lines of sentences. For sequential series, LSTM is used. LSTM is a special type of RNN because RNN cannot maintain information for a long time, so LSTM is established. A vanishing gradient is one of the main problems of RNN that is solved by using LSTM.

LSTM mechanism contains three gates: Forget, Input, and Output gate [17]. Three gates perform differently for a sequential series. Forget gate removes cell information that is not required anymore. The input gate takes a sequential input and updates the information of the cell. Finally, the output gate gives the result of the output for the LSTM module. For each of the gates, LSTM uses an equation. All the equations are given below:

$$f_t = \sigma (W_f x_t + U_f h_{t-1} + b_f)$$

$$i_t = \sigma (W_i x_t + U_i h_{t-1} + b_i)$$

$$o_t = \sigma (W_o x_t + U_o h_{t-1} + b_o)$$

$$u_t = \tanh (W_u x_t + U_u h_{t-1} + b_u)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot u_t$$

$$h_t = o_t \odot \tanh(c_t)$$

Here, f_t represents the gate, i_t as input gate, and output gate is o_t . \odot represents the multiplication of elements step by step. The logistic sigmoid function is used for each gate. Here, u_t represents the \tanh function. The memory cell information contains in c_t and at last, h_t represents the hidden state. Three gates contain a bias vector. LSTM is key to cell state. LSTM maintains the information from the previous cell.

3.3.3 LSTM and CNN

Our model is based on the hybrid model of CNN and LSTM [18]-[28]. The workflow of our proposed model is given in Fig. 1.

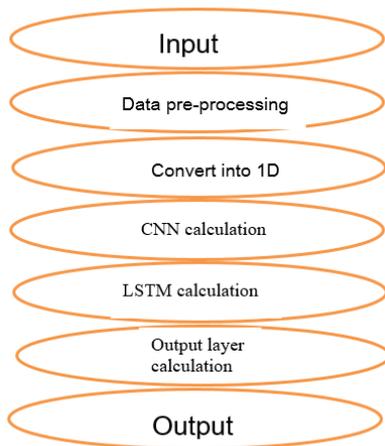


Fig. 1 Workflow of the proposed model

3.3.4 Training Model

To identify the presence of depression in social media, combined the strengths of the neural network architectures of CNN and LSTM and applied a hybrid CNN-LSTM model to classify text data. In recent years CNN was used in many text classification research, and LSTM is known for the text classification neural network. That's why the authors thought of mixing these two algorithms (see Fig. 2) to get a better result. The model takes an input and then a single number representing the probability that the tweet indicates depression. The experiment followed the Hybrid framework for detecting depression using the LSTM-CNN method.

After pre-processing the dataset, the dataset is ready for implementation the model. Keras was used as a neural network framework and Tensorflow as a backend library. Keras is designed for deep neural networks, as authors used deep neural algorithms for model training. For the model training, the Bangla word2 vector is used, which contains the Bangla language vocabulary. This word2 vector is used to create an embedding

matrix. As a CNN word with 2D images, the text has to convert in 2D vector array for text classification. That's why the embedding layer was used—this embedding matrix fed into the embedding layer for further execution. The embedding layer was used to minimize the cost of the loss function. Here we consider the difference between other layers with this embedding layer; then, the main difference is that the output is not like a mathematical function of the input—this pre-trained word vector was constructed for further use on the CNN model.

For this model, CNN was used as a front-end layer followed by LSTM layers and the dense output layer. CNN model was for feature mapping, and for the feature extraction, the LSTM model was used. Authors used CNN as CNN is known for taking original data to create feature maps from it. The dataset is a 1D structure. The structure was fed as input (see Fig. 2). CONV1D took the input to interpret the data into smaller square sizes. The MAXPOOLING layer then explained the result by reducing its size into an abstract dimension. The flatten layer took this output data to convert it into a vector fed into other layers for further resulting prediction. CNN was done with extracting data to transform into a vector. For the internal state to building up, LSTM was used. The sequential output was fed into the LSTM layer. LSTM is an artificial neural network architecture used in deep learning to classify processes and predict output in sentences. LSTM predicted from the previous cell, and the 3 resulting models were created.

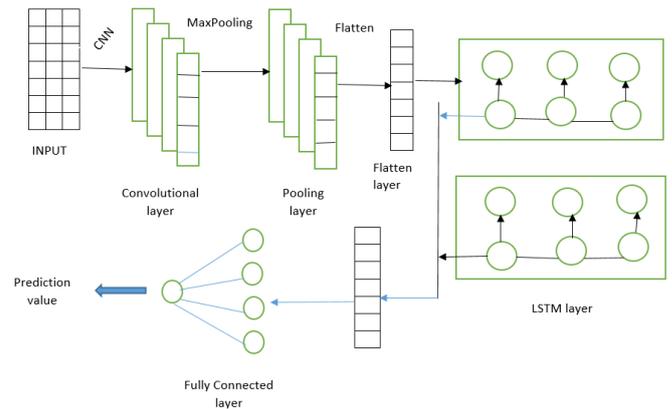


Fig. 2 Mixed mechanism of LSTM and CNN

We tried to minimize our cross-entropy error for the model. Given a training sample $x(i)$ and its true label $y(i) \in \{1, 2, \dots, k\}$ where k is the number of possible labels and the estimated probabilities $y_e(i) \in [0, 1]$ for each label $j \in \{1, 2, \dots, k\}$, the error is defined as:

$$L(x(i), y(i)) = - \sum_{j=1}^k I\{y(i) = j\} \log(y_e(i)_j) \quad [18]$$

where $I\{\text{condition}\}$ is an indicator such that $I\{\text{condition is true}\} = 1$ otherwise $I\{\text{condition is false}\} = 0$. [18]

TF-IDF vectorizer and count vectorizer of SVM used for depression prediction. TF-IDF transformer provided a matrix of the count. On the other hand, the count vectorizer transformed text into a sparse matrix of n-grams counts. Tf-Idf transformed and counted Vectorizer both perform well for prediction. That's why the authors used them both as a combined matrix. TF-IDF vectorizer showed higher accuracy than count Vectorizer, as is expected because TF-IDF balanced out with term frequency and its inverse document frequency. For the training model of TF-

IDF Vectorizer and count Vectorizer, four types of the classifier were used. Those were Naïve Bayes, SVM, Logistic Regression, and Random Forest. From the training of those classifiers, the best accuracy of TF-IDF was SVM. That's why SVM was the chosen classifier for the prediction.

SVM is a supervised machine learning algorithm for classification problems. It performs well for two classifier problems. As the authors used two types of data, depressive and non-depressive so SVM is the best selection for prediction. When a sentence is given to predict depressive or non-depressive, the count vectorizer split the word from the sentence and counts the word's value. According to the word count from the given dataset, it shows prediction. If the number of depressive words is more significant than the non-depressive word, it shows depression or vice versa.

4 Experimental Results

4.1 Performance Evaluation

With the small dataset in hand, the LSTM and CNN model only got 81.05% accuracy. For comparison, the model we also trained the model with Logistic Regression and got 81.49% accuracy. To get this accuracy, we have done multiple tuning, and after that, we have selected the tuning value. The parameter setting for the tuning is given in Table 2.

Table 2 The list of values that are used in work

Hyper-parameter	Value
Embedding Dim	300
Hidden Units of LSTM	128
The activation function of LSTM	Sigmoid
Batch Size	10
No of CNN Layers	1
No of Convolution filter	32
Activation function	RELU
Max pool size	2
Regularization	Dropout operation
Dropout rate	0.2
Learning rate	0.1

TF-IDF and count vector models with different classifiers got us different accuracy for TF-IDF Vectorizer and counted Vectorizer as shown in Fig. 3. The best efficiency of the TF-IDF vectorizer was 67.62%, and the count vectorizer was 75.05%. That's why this TF-IDF and count vector model is unable to predict correctly. Sometimes it gets the wrong result.

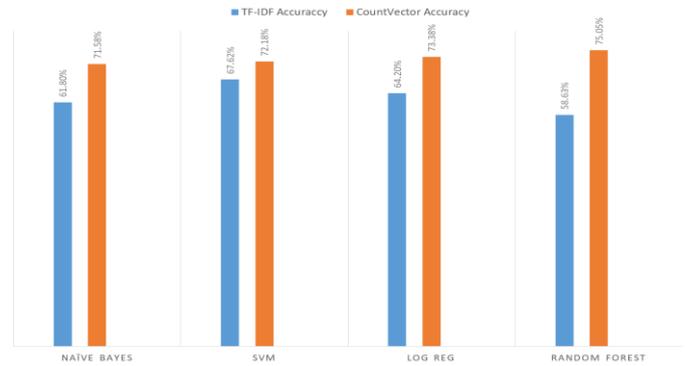


Fig. 3 Accuracy graph of TF-IDF Vectorizer and Count Vectorizer

4.2 Comparison of Proposed Model

Comparing with the international journal [11] helped determine the model's efficiency. The ground of comparison was different algorithms and accuracy. They have chosen Naive Bayes, Multinomial Naive Bayes, Logistic Regression, and Linear SVC as their main algorithms with 1000 Bangla dataset. While we achieved 81.05% with LSTM and CNN, this proposed research is different from this research. Still, the authors tried to compare the deep learning model training accuracy with ML model training accuracy. That's the main reason for the selection of this particular research paper.

That comparison research showed that only 53.93% could be achieved with unigram and emoticon while unigram, emoticon, and depressive word lists showed different accuracy for different algorithms. With our TF-IDF vectorizer, and count vector accuracy of TF-IDF is much better than this proposed method as shown in Fig. 4. Different batch sizes, feature sizes, and different classifiers are responsible for different accuracy.

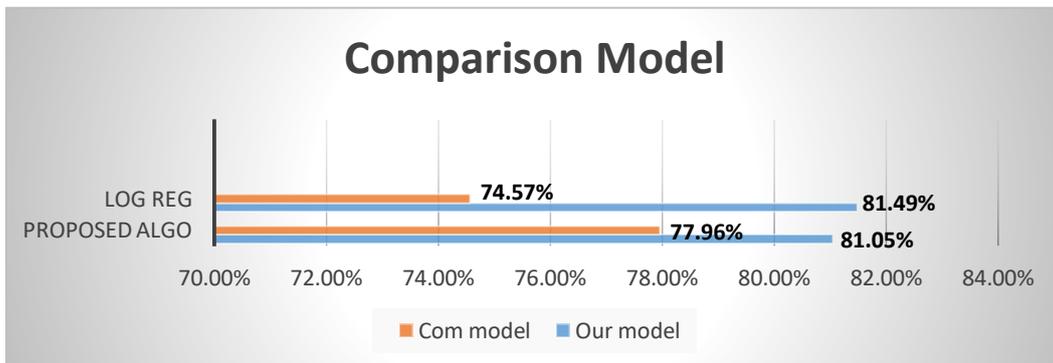


Fig. 4 Comparison model (Series 1: LSTM-CNN, Series 2: Comparison research)

5 Discussion

The authors used the CNN-LSTM hybrid model to execute their system. As far [29], a paper has already been completed with this hybrid model using English Dataset. But the new thing in this paper is that the authors used the same hybrid model using the Bangla dataset. But as the Bangla dataset is not available so much, the authors have created a small dataset for execution. After executing the hybrid model using the Bangla dataset, it performed well in detecting depression. In this modern era, most people are found in depression because of work stress and some other problem.

On the other hand, social media is a common way to express their emotion to decrease their sorrow or expand their happiness. So, it is easy to find out someone's emotional status or daily status through social media. This kind of data is most famous for the proposed system. Sometimes treatment can be given through social media. The authors have executed a system that can predict depressive and non-depressive status. After knowing that the person is in extreme level depression, some sort of treatment can be done through social media. This procedure can reduce the cause of death due to depression.

6 Conclusion

This paper tried to implement a hybrid algorithm to recognize depression in social media using Bangla datasets. Natural Toolkit was used for pre-processing the dataset. The dataset contains both depressive and non-depressive Facebook Bengali status. Word vector was used, which is full of the vocabulary of all Bangla words. This word vector helps to create an embedding matrix. In the embedding layer, an embedding matrix creates to feed into the CNN model. After getting a sequential result that is fed into the LSTM model, Support Vector Machine was used for prediction with a short number of datasets at the end of all processes. Count vectorizer showed the best performance than TF-IDF tokenizer. TF-IDF is unable to give precise results because of the small data set. Proposed better algorithms than LSTM and better datasets could achieve more efficiency. We used LSTM and CNN for our model, but in the future, the work would build the model using better RNN. After detecting depression, then the program will suggest, according to the result, if they need any counseling or not.

References

- [1] "What Is Depression?" <https://www.psychiatry.org/patients-families/depression/what-is-depression> (accessed Jun. 06, 2020).
- [2] Choudhury, A.A., Khan, M.R.H., Nahim, N.Z., Tulon, S.R., Islam, S. and Chakrabarty, A., 2019, June. Predicting depression in Bangladeshi undergraduates using machine learning. In 2019 IEEE Region 10 Symposium (TENSYPMP) (pp. 789-794). IEEE. DOI: 10.1109/TENSYPMP46218.2019.8971369.
- [3] "Can Depression Really Kill You?" <https://www.verywellmind.com/can-depression-kill-you-1067514> (accessed Jun. 06, 2020).
- [4] Islam, M.R., Kabir, M.A., Ahmed, A., Kamal, A.R.M., Wang, H. and Ulhaq, A., 2018. Depression detection from social network data using machine learning techniques. *Health Information Science and Systems*, 6(1), pp.1-12. DOI: 10.1007/s13755-018-0046-0.
- [5] "Depression (major depressive disorder) - Symptoms and causes - Mayo Clinic." <https://www.mayoclinic.org/diseases-conditions/depression/symptoms-causes/syc-20356007> (accessed Jun. 06, 2020).
- [6] "What is mental illness? - MindFreedom International (MFI)." <https://mindfreedom.org/kb/voices-for-choices/voices-for-choices-what-is-mental-illness/> (accessed Jun. 06, 2020).
- [7] "Depression: What it is, symptoms, causes, treatment, types, and more." <https://www.medicalnewstoday.com/articles/8933> (accessed Jun. 06, 2020).
- [8] "World Health Organization, Depression: Let's talk." <http://www.searo.who.int/bangladesh/enbanwhd2017/en/> (accessed Jun. 06, 2020).
- [9] Selim, N., 2010. Cultural dimensions of depression in Bangladesh: a qualitative study in two villages of Matlab. *Journal of Health, Population, and Nutrition*, 28(1), p.95.
- [10] Anderson, J.E., Michalak, E.E. and Lam, R.W., 2002. Depression in primary care: Tools for screening, diagnosis, and measuring response to treatment. *British Columbia Medical Journal*, 44(8), pp.415-419.
- [11] Billah, M. and Hassan, E., Depression Detection from Bangla Facebook Status using Machine Learning Approach. *International Journal of Computer Applications*, 975, p.8887. DOI: 10.5120/ijca2019919314.
- [12] Ramalingam, D., Sharma, V. and Zar, P., 2019. Study of depression analysis using machine learning techniques. *Int. J. Innov. Technol. Explor. Eng.* 8(7C2), pp.187-191.
- [13] Uddin, A.H., Bapery, D. and Arif, A.S.M., 2019, July. Depression Analysis from Social Media Data in Bangla Language using Long Short Term Memory (LSTM) Recurrent Neural Network Technique. In 2019 International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering (IC4ME2) (pp. 1-4). IEEE. DOI: 10.1109/IC4ME247184.2019.9036528.
- [14] "NLTK Book." <https://www.nltk.org/book/> (accessed Jun. 05, 2020).
- [15] "Understanding of Convolutional Neural Network (CNN) — Deep Learning." <https://medium.com/@RaghavPrabhu/understanding-of-convolutional-neural-network-cnn-deep-learning-99760835f148> (accessed Jun. 14, 2020).
- [16] "An intuitive guide to Convolutional Neural Networks." <https://www.freecodecamp.org/news/an-intuitive-guide-to-convolutional-neural-networks-260c2de0a050/> (accessed Jun. 14, 2020).
- [17] "Long Short Term Memory | Architecture Of LSTM. " <https://www.analyticsvidhya.com/blog/2017/12/fundamentals-of-deep-learning-introduction-to-lstm/> (accessed Jun. 14, 2020).
- [18] Zhou, C., Sun, C., Liu, Z. and Lau, F., 2015. A C-LSTM neural network for text classification. arXiv preprint arXiv:1511.08630. Available: <http://arxiv.org/abs/1511.08630>.
- [19] Rakib, O.F., Akter, S., Khan, M.A., Das, A.K. and Habibullah, K.M., 2019, December. Bangla word prediction and sentence completion using GRU: an extended version of RNN on N-gram language model. In

- 2019 International Conference on Sustainable Technologies for Industry 4.0 (STI) (pp. 1-6). IEEE.
- [20] Emon, E.A., Rahman, S., Banarjee, J., Das, A.K. and Mitra, T., 2019, June. A deep learning approach to detect abusive bengali text. In 2019 7th International Conference on Smart Computing & Communications (ICSCC) (pp. 1-5). IEEE.
- [21] Hossain, M.M., Labib, M.F., Rifat, A.S., Das, A.K. and Mukta, M., 2019, June. Auto-correction of English to Bengali Transliteration System using Levenshtein Distance. In 2019 7th International Conference on Smart Computing & Communications (ICSCC) (pp. 1-5). IEEE.
- [22] Drovo, M.D., Chowdhury, M., Uday, S.I. and Das, A.K., 2019, June. Named Entity Recognition in Bengali Text Using Merged Hidden Markov Model and Rule Base Approach. In 2019 7th International Conference on Smart Computing & Communications (ICSCC) (pp. 1-5). IEEE.
- [23] Biswas, E. and Das, A.K., 2019, June. Symptom-Based Disease Detection System In Bengali Using Convolution Neural Network. In 2019 7th International Conference on Smart Computing & Communications (ICSCC) (pp. 1-5). IEEE.
- [24] Das, A.K., Ashrafi, A. and Ahmmad, M., 2019, February. Joint Cognition of Both Human and Machine for Predicting Criminal Punishment in Judicial System. In 2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS) (pp. 36-40). IEEE.
- [25] Tuhin, R.A., Paul, B.K., Nawrine, F., Akter, M. and Das, A.K., 2019, February. An automated system of sentiment analysis from Bangla text using supervised learning techniques. In 2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS) (pp. 360-364). IEEE.
- [26] Islam, J., Mubassira, M., Islam, M.R. and Das, A.K., 2019, February. A speech recognition system for Bengali language using recurrent Neural network. In 2019 IEEE 4th international conference on computer and communication systems (ICCCS) (pp. 73-76). IEEE.
- [27] Bhuiyan, M., Rahman, A., Ullah, M. and Das, A.K., 2019. iHealthcare: Predictive model analysis concerning big data applications for interactive healthcare systems. *Applied Sciences*, 9(16), p.3365.
- [28] Labib, M.F., Rifat, A.S., Hossain, M.M., Das, A.K. and Nawrine, F., 2019, June. Road accident analysis and prediction of accident severity by using machine learning in Bangladesh. In 2019 7th International Conference on Smart Computing & Communications (ICSCC) (pp. 1-5). IEEE.
- [29] Tadesse, M.M., Lin, H., Xu, B. and Yang, L., 2020. Detection of suicide ideation in social media forums using deep learning. *Algorithms*, 13(1), p.7. DOI: 10.3390/a13010007.