How to Cite:
Bhuvaneswari, M., & Praba, V. L. (2022). An illustration of attention mechanism in LSTM algorithm for analyzing its performance. *International Journal of Health Sciences*, *6*(S2), 9010–9023. https://doi.org/10.53730/ijhs.v6nS2.7335

An illustration of attention mechanism in LSTM algorithm for analyzing its performance

M. Bhuvaneswari
Reg.No: 17224012162051, Research Scholar, Research Department of Computer Science, Rani Anna Government College for Women, Tirunelveli (Affiliated to Manonmaniam Sundaranar University, Tirunelveli)

V. Lakshmi Praba
Assistant Professor, Department of Computer Science, Rani Anna Government College for Women, Tirunelveli, India

***Abstract*---**The advent of technologies and equipments in Information Technology and Communication has increased the use of smart phones and social media by all people irrespective of age and gender. Many times people find social media posts as the outlet for draining all their emotions. Depression is a common mental health problem that ranges from low mood and loss of interest to persistent feeling of sadness and self-destruction. Mental Health Experts are interested in analyzing the physical and social media behavior of patients to diagnose depression. Attention Mechanism imitates the human cognitive of considering only the most relevant data needed for a particular analysis. LSTM is a state-of-the-art Recurrent Neural Network architecture used for depression analysis. In this paper we are further improving the LSTM Algorithm by proposing an Attention Mechanism that makes the model robust and improved. This major role of our Attention Mechanism is identifying the Prime Words and Supporting words that are relevant for the analysis. Distance is calculated between non-prime words and prime words, weights are assigned based on this calculation, higher weight being given to most relevant words. Making use of Attention helps us to focus only on the relevant data and neglect the irrelevant data. Experimental evaluation using two social media datasets showed improved performance metrics and enhanced output when the proposed Attention Mechanism is used with LSTM than the existing LSTM.

***Keywords*---**attention mechanism, neural network, depression analysis, social media analysis, LSTM

Manuscript submitted: 9 March 2022, Manuscript revised: 18 April 2022, Accepted for publication: 1 May 2022

Introduction

With the advancements of communication technologies and equipments, social media has deep rooted in our daily life with the purpose of sharing, networking, advertising etc., Users share ideas, images, videos and express their comments based on their current emotions. This creates a need for mental health experts to infer a particular user's state-of-mind by analyzing their physical behavior as well as their social media behavior. Depression Analysis of social media gathers insights in a user's posts and helps in diagnosing an early indication of depression. This facilitates initiating early remedial actions, proper treatments etc., and thus avoids devastating results. Social Media data consists of text, emoticons, abbreviations, URLs, informal language etc which makes it unstructured and challenging for analysis. This motivates the use of Attention Mechanism in analysis. Attention Mechanism is a step forward in neural networks that mimics cognitive attention of human brain. It operates by enhancing the focus on the most relevant elements of the input sequence and diminishing focus on the non-relevant parts so that the performance of the analysis is better. In this paper we devise an Attention Mechanism by assigning weights to all input elements, the highest weight being assigned to the relevant elements of the input. This work implements and tests the Attention Mechanism with the state-of-the-art LSTM (Long Short Term Memory) algorithm and proves the Attention Mechanism proposed enhances the performance of the existing LSTM.

Related Work

Gilbert, CJ Hutto Eric [1] present VADER, a simple rule-based model for general sentiment analysis, and compared its effectiveness with existing techniques. Heeyoul Choi et al. [2] proposed a 2D attention mechanism where each dimension of a context vector will receive a separate attention score which improved the translation quality of the score. Julia Ive1 et al. [3] applied a hierarchical Recurrent neural network (RNN) architecture with an attention mechanism on social media data related to mental health. Q. Cong et al. [4] proposed a attention based deep learning model for depression detection in imbalanced social media data. G. I. Winata et al. [5] proposes a LSTM with attention mechanism to classify psychological stress from self-conducted interview transcriptions. X.Ran et al. [6] proposed a LSTM-based method with attention mechanism for travel time prediction by using departure time as the aspect of the attention mechanism.

Shun Chen and Lei Ge [7] explores the attention mechanism in LSTM network based stock price movement prediction in Hong Kong stock market. Li, W et al. [8] introduce a novel sentiment attention mechanism to select the crucial sentiment-word-relevant context words and developed an improved deep neural network to extract sequential correlation information and text local features by combining bidirectional gated recurrent units with a convolution neural network. Maupomé D et al. [9] implemented four RNN-based systems to classify the users and explore several aggregations methods to combine predictions on individual posts. Elsbeth Turcan and Kathleen McKeown [10] developed Dreaddit: A Reddit Dataset for

Stress Analysis in Social Media. This dataset is used for analysis of our proposed work. Gang Liu & Jiabao Guo [11] proposed an architecture called attention-based bidirectional long short-term memory with convolution layer (AC-BiLSTM). Yu, Xm et al. [12] proposed an attention mechanism and multi-granularity-based Bi-LSTM model (AM–Bi-LSTM) which combines the improved attention mechanism with a novel processing of multi-granularity word segmentation to handle the complex NLP in Chinese Q&A datasets. Jang, Beakcheol et al. [13] proposed an attention-based Bi-LSTM+CNN hybrid model that capitalize on the advantages of LSTM and CNN with an additional attention mechanism and tested it using IMDB Movie review dataset.

Weijiang Li et al. [14] proposed a bidirectional LSTM model with self-attention mechanism and multi-channel features (SAMF-BiLSTM) to enhance sentiment information. Kim, Y. and Choi, A. [15] proposed an LSTM network to consider changes in emotion over time and applied an attention mechanism to assign weights to the emotional states appearing at specific moments based on the peak–end rule in psychology. J. Wan et al. [16] proposed an attention based CNN-BiRNN method to recognize the individual high resolution range profile(HRRP). Su, C et al. [17] reviewed existing research on applications of Deep Learning algorithms in mental health outcome research. Baek, JW. and Chung, K [18] proposed the accident risk prediction model based on attention-mechanism LSTM using modality convergence in multi-modal. Wang, Yakun et al. [19] developed Surrounding-Aware individual Emotion Prediction model (SAEP) based on a deep encoder & decoder architecture to predict individuals future emotions. Yifeng Zhao and Deyun Chen [20] proposed the bidirectional LSTM and attention mechanism based on the expression EEG multimodal emotion recognition method. Yan ZHAO et al. [21] presents an attention- based Long Short-Term Memory (LSTM) model for depression detection to make full use of the difference between depression and non-depression between timeframes. Ghosh T. et al [22] proposed an emotion detection-based mood control framework that reorganizes the social media posts to match user's mental state based on Attention mechanism, Bidirectional Long Short Term Memory (BiLSTM), and Convolutional Neural Network (CNN).

Dataset

The datasets used for the current analysis is the publicly available depression analysis datasets – *sentiment_tweets3.csv* [23] and *Dreaddit*. *sentiment_tweets3.csv* is a twitter dataset that contains 10,314 attributes and 3 entities. The entities in the dataset are user-id, tweet (message) and label of which only the second entity, the tweet text is taken as input in this work. The depressive tweets in the dataset are labelled as 1 and the non-depressive tweets are labelled as 0. The dataset is imbalanced as it contains 3.5 times non-depressive tweets than depressive tweets. *Dreaddit* - A Reddit Dataset for Stress Analysis in Social Media consists of 1,87,444 posts from five different categories of Reddit communities in the domain – abuse, anxiety, PTSD(Post Traumatic Stress Disorder), social and others. The average length of a post in the dataset is 420 tokens with binary stress labels as 0 (non- depressive) or 1(depressive). The dataset has 112 attributes the important being – id, stress label, text, subreddit type.

Proposed Model

Attention mechanism identifies the most relevant parts of the text that contribute more to the polarity of the sentence. Figure 1 illustrates a sentence in which the highlighted words are the most important words i.e. the prime words of the sentence and the words 'very' and 'always' are the supporting words that contribute to the polarity of the prime words. Hence making use of Attention helps us to focus only on the relevant data and neglect the irrelevant data thus enhances the performance of the output achieved.

So im feeling a bit defeated; my workplace is very stressful and my marriage feels like it's crumbling because we are always arguing...

Figure 1. An Illustration of a tweet from the Dreaddit dataset

The proposed work devises and implements an Attention Mechanism to concentrate specifically on the relevant parts of the text that contribute more in depression analysis. This model assigns weight to all input words with the Prime Words being attributed the highest weights. The words that support the Prime Words in the contribution of polarity are found using the distance calculations between them. The relevant and irrelevant inputs are filtered using the Masking. Thus this Attention Mechanism ensures that words that contribute more to the polarity of the sentence are given more attention with greater weightage and less relevant words are given less importance with reduced weightage. This Attention Mechanism is implemented with the existing LSTM algorithm which enhances the performance and improves the accuracy of the results to a notable extent. The architecture of the proposed Attention Mechanism in LSTM Algorithm is given in Figure 2(a):

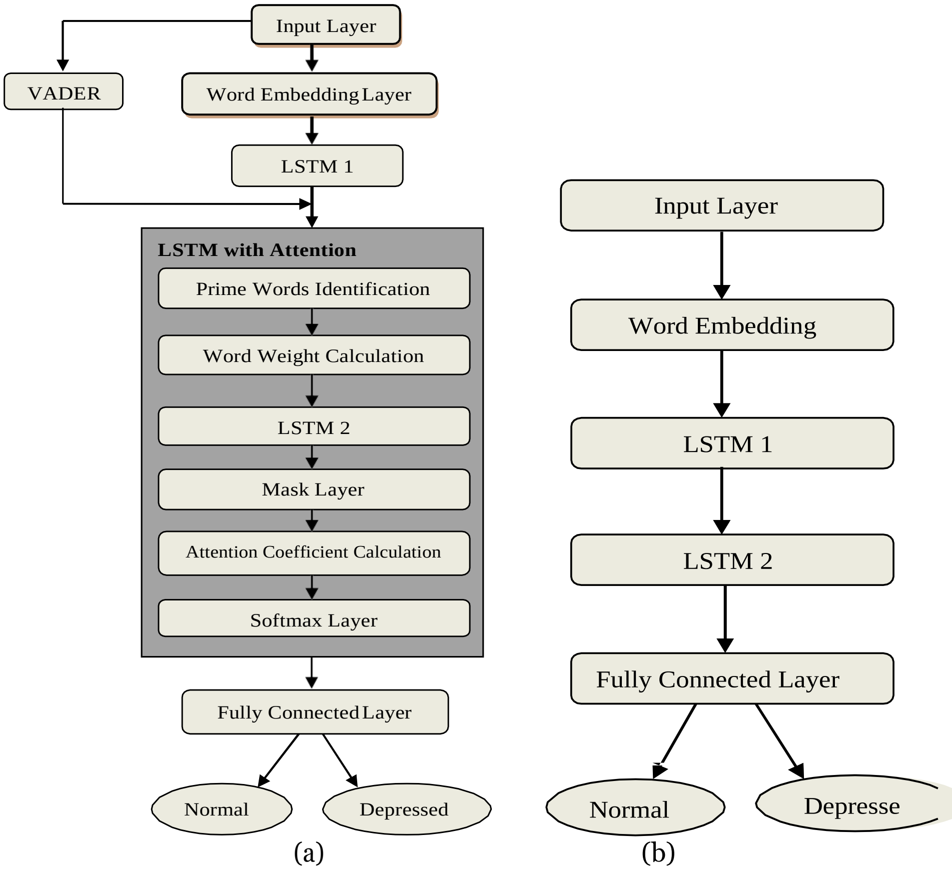

Fig.2: (a) Architecture of Attention Mechanism in LSTM (b) Architecture of LSTM

Input Layer

Pre-processing of input text from the dataset is performed. Pre-processing involves removal of hash tags, punctuation marks, stop words, URL's, expanding abbreviations etc., Pre-processed sentences are tokenized into words. Let $\boldsymbol{S}$ be the sentence with words $W_1, W_2, W_3, \ldots W_n$.

$$\boldsymbol{S} = \{\boldsymbol{W_1}, \boldsymbol{W_2}, \boldsymbol{W_3} \ldots \boldsymbol{W_n}\}$$

Embedding Layer

This layer converts input text into vectors space embeddings using the one of the popular embedding representation *GloVe*[24]. GloVe: Global Vectors for word representation is a global log-bilinear regression model for the unsupervised learning of word representations. It produces word vector space from matrix factorization of the word's co-occurrences matrix constructed from large corpora. The GloVe is a publicly available file with various vector dimensions like 50,100,200 and 300. This work makes use of the 300 dimension file '*glove.6B.300d.txt*' to get the maximum vector dimensional output. Let V be the embedding vector with individual word embedding as $V_1, V_2, V_3 \ldots V_n$.

$$\boldsymbol{V = \{V_1, V_2, V_3 \ldots V_n\}}$$

LSTM Layer 1

The embedding vectors V of the words W are fed into a LSTM layer and the output is stored as *out1*.

Prime Word Identification

This step identifies the Prime Words that contribute more to the polarity of the sentence and stores the position of these Prime Words in an array $\boldsymbol{PW}$. These words are assigned highest weightage of 1 to prioritize the words for further analysis.

- The words are given as input into *VADER (Valence Aware Dictionary for sEntiment Reasoning)* and we get the output as class and sentiment score of words. *VADER* is a rule-based sentiment lexicon that finds the intensity of each word in the text and sums it to calculate the sentiment score of the text.
- For each word *VADER* produces two outputs – the class of the word as positive ('*pos*') or negative ('*neg*') and the polarity score of the word. If for a particular word class is '*pos*' the scores are stored in the array *PSi* and if the word belongs to class '*neg*' the scores are stored in the array *NSi* where i represents the position of words in the sentence i=1,2,...n.
- The position of prime words that contribute more to the polarity of the sentence are found and stored as Prime Word vector using the following logic:

$$\begin{array}{l}
\boldsymbol{for\ all\ i = 1\ to\ n} \\
\quad \boldsymbol{if\ class('pos')} \\
\quad\quad \boldsymbol{if\ PS_i \geq PS} \\
\quad\quad\quad \boldsymbol{PS = PS_i} \\
\quad\quad\quad \boldsymbol{PW.append(i)} \\
\quad \boldsymbol{if\ class('neg')} \\
\quad\quad \boldsymbol{if\ NS_i \geq NS} \\
\quad\quad\quad \boldsymbol{NS = NS_i} \\
\quad\quad\quad \boldsymbol{PW.append(i)}
\end{array}$$

Calculation of Word Weights

Words are assigned weight based on their importance in the sentence. Prime Words identified in the previous step are assigned a weight of 1 indicating these are most relevant for analysis. For other words the distances of the words from the Prime Words are calculated, the closest Prime Word for each word is identified and weights are assigned accordingly.

Word weights WT_i are assigned to words W_1, W_2, $W_3 \ldots W_n$ as follows:

Case 1: For all Prime Words that is $\boldsymbol{i \in PW}$

$$\boldsymbol{WT_i = 1}$$

Case 2: For all Non-Prime Words $\boldsymbol{i} \notin \boldsymbol{PW}$
The distance of the word *Wi* from all prime words in PW is found and the closest prime word is found by calculating the minimum distance and weight is assigned accordingly.

$$\boldsymbol{f(j) = dist(P_i\,, PW_j)} \;\forall\; \boldsymbol{j} \in \boldsymbol{PW}$$

$$\boldsymbol{mindist_{ij} = min(f(j))}$$

$$\boldsymbol{WT_i = [1 - (mindist_{ij} * 0.1)]} \;\forall\; \boldsymbol{i} \notin \boldsymbol{PW}$$

where P_i is the position of the word $\boldsymbol{W_i} \notin \boldsymbol{PW}$
PW_j is the Prime Word array where $j=1, 2,\ldots n$.
mindistij denotes the shortest distance from W_i to closest prime word W_j.

Here the minimum distance is multiplied by 0.1 which is the best weightage factor chosen by experimental analysis of values from 0.1 to 0.9.

LSTM Layer 2

The output of previous LSTM layer *out1* is multiplied by the calculated weight for words *WT* and fed into this second LSTM layer. The output is stored as *out2*.

$$\boldsymbol{out2 = out1 * WT}$$

Mask Layer

This layer filters out the irrelevant input from entering the next step of analysis. The mask for prime words is calculated as 1 since they are the most relevant words. For other words if the minimum distance divided by the total number of Prime Words is less than 0.3 it is considered as a Supporting Word which are less relevant words than Prime words but contribute to the polarity and are given a mask value of 1. If the minimum distance divided by the total number of prime words is greater or equal to 0.3 the words are considered irrelevant which are not needed for further analysis and hence the mask value is set to 0.

$$\boldsymbol{f(i)} = \begin{cases} \boldsymbol{Mask_i = 1\ if\ i \in PW} \\ \boldsymbol{Mask_i = 1\ if mindist_{ij}/n} < 0.3 \\ \boldsymbol{Mask_i = 0\ if\ mindist_{ij}/n\ \geq 0.3} \end{cases}$$

The previous LSTM layer output ***out2*** is masked and the new output is ***out3*** contains data with Mask = 1.

Calculation of Attention Coefficient matrix

α_{mat} the attention coefficient matrix is calculated by multiplying the masked output *out3* is multiplied with thesecond LSTM layer output *out2*.

$$\propto_{mat} = \boldsymbol{out2 * out3}$$

Softmax Layer

Softmax transforms input values into values between 0 and 1, so that they can be interpreted as probabilities. If one of the inputs is small or negative, the Softmax turns it into a small probability, and if an input is large, then it turns it into a large probability, but it will always remain between 0 and 1. $\propto_{mat}$ is fed into the Softmax layer to get the Attention Coefficient $\propto$.

$$\propto = \boldsymbol{softmax}(\propto_{mat})$$

Fully Connected Layer

Attention Coefficient α is applied to out2 and fed into the fully connected layer to classify the input into class labels as Normal(0) and Depressed(1).

Experimental Setup

System Used

The experiment is implemented in Intel Core i5 processor FHD Gaming Laptop with NVIDIA GTX 1650Ti Driver and Windows 10 Operating System. The software used for implementation of the model is *Python 3.6* in *Anaconda Navigator* user interface. The computing platform used for execution is *Cuda.*

Implementation

This work is implemented in Python with the hyper-parameter settings as given in Table 1.

TABLE 1: Hyper-parameter Settings in our model

Hyper-parameter Name	Hyper-parameter Value
Initializer	xavier_uniform
Word Embedding Dimension	300
Dropout	0.3
Number of epochs	10
Batch size	128
Hidden Layer	1
Loss function	Categorical cross-entropy
Optimizer	Adam
Learning Rate	0.001

Each input text is preprocessed and tokenized using *Tokenizer()* function. The tokens are fed into a GloVe embedding layer which gives a 300 dimension output vector. The embedding vector is passed into the first LSTM layer. The input words are also given as input into VADER to get the polarity score of the word. This is done by using the *SentimentIntensityAnalyzer* class in the *vaderSentiment* library in the *NLTK* Package. VADER sentimental analysis relies on a dictionary that map lexical features to emotion intensities known as sentiment scores. The sentiment score of a text can be obtained by summing up the intensity of each word in the text. The most contributing words of a sentence – Prime words are identified and for non-prime words their minimum distance from the prime words is calculated and weights are assigned accordingly. The weights along the embedding vector is fed into LSTM layer, masked in the Mask layer by calculating the mask and only the relevant data is passed to the next layer. Attention coefficient is calculated from the LSTM outputs, applied in Softmax layer and passed to a Fully Connected Layer for classifying into Normal or Depressed labels.

9018

Evaluation Metrics

The performance of the model is evaluated based on the four metrics – Accuracy, F1 Score, Precision and Recall.

Accuracy

Accuracy predicts how often the classifier makes correct prediction.

$$\boldsymbol{Accuracy} = \frac{\boldsymbol{TP} + \boldsymbol{TN}}{\boldsymbol{TP} + \boldsymbol{TN} + \boldsymbol{FP} + \boldsymbol{FN}}$$

Precision

Precision is the ratio of number of instances correctly classified as positive to the total number reported as positive by the model.

$$\boldsymbol{Precision} = \frac{\boldsymbol{TP}}{\boldsymbol{TP} + \boldsymbol{FP}}$$

Recall

Recall is the ratio of number of instances classified as positive to the total number of actual number of positives present in the dataset.

$$\boldsymbol{Recall} = \frac{\boldsymbol{TP}}{\boldsymbol{TP} + \boldsymbol{FN}}$$

where TP represents Number of True Positives, TN represents Number of True Negatives, FP represents Numberof False Positives and FN represents Number of False Negatives.

F1 Score

F1 Score is the weighted harmonic mean of precision and recall. These performance metrics are calculated as follows:

$$F1\,Score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

Results And Discussion

The proposed Attention Mechanism in LSTM Algorithm classified the inputs from both the datasets with improved performance metrics than the well established existing LSTM. This model classified the input data in Dreaddit dataset with approximately 60% accuracy and sentiment_tweet3.csv data with 95% accuracy. Table 2 indicates the proposed model shows a notable improvement in all the four performance metrics – Accuracy, Precision, Recall, F1 Score than the use of LSTM separately for specific subreddit types in Dreaddit dataset and all types in sentiment_tweet3 dataset. The improvement ratio percentage in Table 2 shows the proposed Attention Mechanism contributes in enhancing the performance metrics of the model.

TABLE 2: Performance Metrics of the proposed Attention Mechanism in LSTM Algorithm model with the existing LSTM

Dataset	Subreddit type	Model	Performance Metrics			
			Accuracy	Precision	Recall	F1 score
Dreaddit	abuse	LSTM	55.56	44.39	50.26	47.00
		LSTM with Attention	55.79	55.58	50.58	53.00
		Improvement Ratio (%)	0.42	25.20	0.63	12.77
	anxiety	LSTM	59.38	29.69	50.00	37.00
		LSTM with Attention	60.83	57.67	53.64	56.00
		Improvement Ratio (%)	2.46	94.27	7.29	51.35
	PTSD	LSTM	64.29	32.14	50.00	39.00
		LSTM with Attention	64.55	63.98	53.00	58.00
		Improvement Ratio (%)	0.41	99.06	6.01	48.72
	social	LSTM	57.45	28.72	50.00	36.00
		LSTM with Attention	60.05	73.96	53.56	62.00
		Improvement Ratio (%)	4.53	157.49	7.12	72.22
sentiment_tweets3	-	LSTM	95.82	97.09	91.99	94.47
		LSTM with Attention	95.86	97.37	92.10	94.67
		Improvement Ratio (%)	0.04	0.29	0.12	0.20

Figure 3 compares the experimental results of Accuracy for the proposed Attention Mechanism in LSTM Algorithm with the state-of-the-art LSTM for the two datasets. The improvement ratio ranges from a minimum of 0.41% for PTSD sub-reddit type to 4.53% for social sub-reddit type of Dreaddit dataset. Since the accuracy of sentiment dataset is already at a maximum of 95.82% in the existing LSTM a marginal increase of 0.04 % is seen with the Attention Mechanism in LSTM Algorithm.

Figure 4 compares performance metric Precision for the proposed model and existing LSTM for the two datasets. There is a drastic improvement in this metric for Dreaddit dataset with improvement ratios of abuse, anxiety, PTSD, social as 25.20%, 94.27%, 99.06%, 157.49% respectively. Sentiment dataset shows a marginal increase of 0.29% in the proposed work than the existing LSTM. Figure 5 shows the Recall metric for the proposed work has a notable improvement than the existing LSTM. The improvement is attributed to the Attention Mechanism in the proposed work. Though abuse subreddit type shows a very little improvement of 0.63% other subreddit types show an increase ratio of 6-7%. The sentiment dataset showsan increase of 0.12%.

Fig 6 shows that proposed Attention Mechanism shows major improvement in F1 Score than the existing LSTM. Dreaddit datasets shows greater improvement ratio values as 12.77 %, 51.35%, 48.72% and 72.22% for anxiety, abuse, PTSD, social types of Dreaddit dataset respectively. The sentiment dataset shows an improvement ratio of 0.20%.

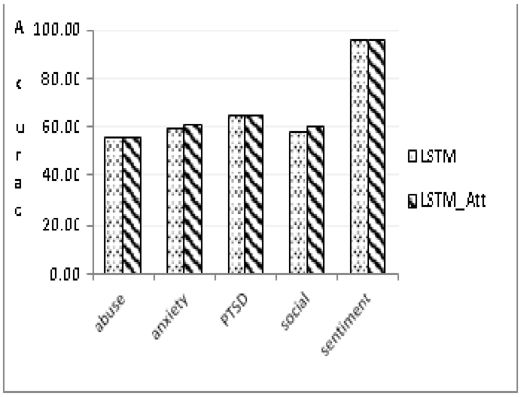

FIGURE 3. Comparison of Accuracy

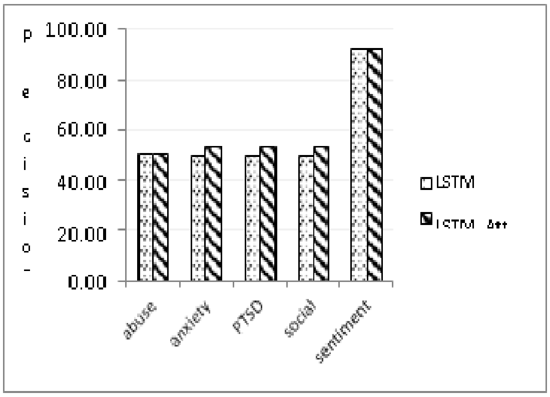

FIGURE 4. Comparison of Precision

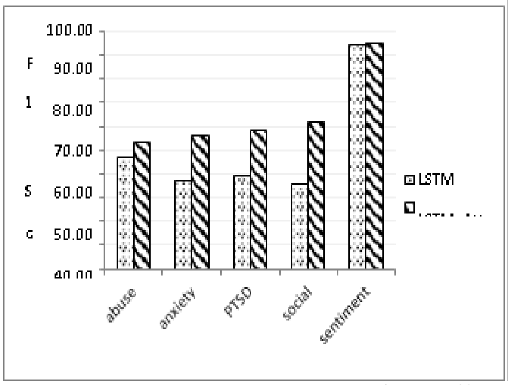

FIGURE 5 Comparison of Recall

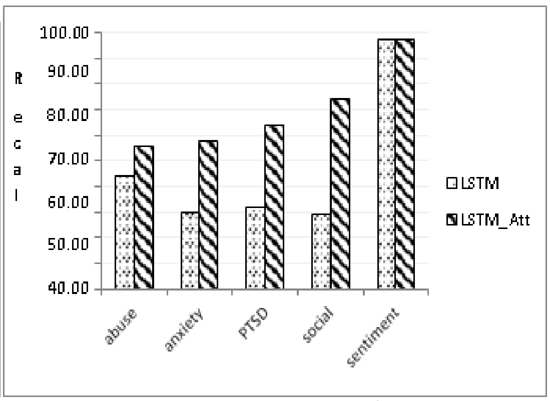

FIGURE 6 Comparison of F1 Score

In the analysis of any dataset the performance metric Precision is more important because it denotes the quality of positive predictions made by the model. When considering Dreaddit dataset the Precision metric has a percentage of improvement of minimum 25.2% to 157.49 %. Such higher improvement percentage in Precision metric indicates less false positives i.e. Normal persons are less wrongly predicted as Depressed by the proposed work.

As shown in the above discussions and results in the proposed work the Attention Mechanism plays a vital role in improving the performance metrics and output to a greater extent in Dreaddit dataset. However in sentiment dataset the existing LSTM already has a higher performance metrics so the improvement is minimum but still the Attention Mechanism enhances this performance metrics also to a certain extent.

Conclusion

Social Media forms the virtual means of interpersonal communication between people for sharing ideas and thoughts through user posts. It gives an effective insight into one's mental mood and behavior. Depression is a mental disorder that affects an individual's body, mood, behavior and daily activities. Detection of Depression at an early stage helps in rendering proper treatments and counseling as when needed and thus helps in avoiding self- destructions. The proposed Attention Mechanism makes use of Attention as the key factor to enhance the analysis process. The Attention Mechanism is then combined with the well known LSTM and tested using two social media datasets. The performance metrics and the results showed notable improvements for both the dataset. In this work we have taken only two datasets and based on the results obtain we observed Attention Mechanism plays a vital role and shows higher increase in percentage in the considered performance metrics for which the LSTM is around 50%-60%. In this we have considered only two datasets in future we can consider several datasets and analyze the performance of Attention Mechanism in LSTM.

References

1. Gilbert, CJ Hutto Eric (2014), VADER: A parsimonious rule-based model for sentiment analysis of social media text, In Eighth International Conference on Weblogs and Social Media (ICWSM-14). Available at (20/04/16), http://comp. social. gatech. edu/papers/icwsm14. VADER. hutto. pdf. 2014.
2. Heeyoul Choi, Kyunghyun Cho, Yoshua Bengio, Fine-grained attention mechanism for neural machine translation, Neurocomputing, Volume 284, 2018, Pages 171-176, ISSN 0925-2312.
3. Julia Ive1 et.al, Hierarchical neural model with attention mechanisms for the classification of social media text related to mental health, Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic, pages 69–77 New Orleans, Louisiana, June 5, 2018.
4. Q. Cong, Z. Feng, F. Li, Y. Xiang, G. Rao and C. Tao, X-A-BiLSTM: a Deep Learning Approach for Depression Detection in Imbalanced Data, 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2018, pp. 1624-1627, doi: 10.1109/BIBM.2018.8621230.

5. G. I. Winata, O. P. Kampman and P. Fung, "Attention-Based LSTM for Psychological Stress Detection from Spoken Language Using Distant Supervision," *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 6204-6208, doi: 10.1109/ICASSP.2018.8461990
6. Ran, X.; Shan, Z.; Fang, Y.; Lin, C. An LSTM-Based Method with Attention Mechanism for Travel Time Prediction. *Sensors* 2019, *19*, 861. https://doi.org/10.3390/s19040861
7. Shun Chen and Lei Ge, Exploring the attention mechanism in LSTM-based Hong Kong stock price movement prediction, Quantitative Finance, vol.19 (9),pages 1507-1515, 2019, https://doi.org/10.1080/14697688.2019.1622287
8. Li, W.; Liu, P.; Zhang, Q.; Liu, W. An Improved Approach for Text Sentiment Classification Based on a Deep Neural Network via a Sentiment Attention Mechanism. *Future Internet* 2019, *11*, 96. https://doi.org/10.3390/fi11040096
9. Maupomé D., Queudot M., Meurs MJ. (2019) Inter and Intra Document Attention for Depression Risk Assessment. In: Meurs MJ., Rudzicz F. (eds) Advances in Artificial Intelligence. Canadian AI 2019. Lecture Notes in Computer Science, vol 11489. Springer, Cham. https://doi.org/10.1007/978-3-030-18305-9_27
10. Elsbeth Turcan and Kathleen McKeown, Dreaddit: A Reddit Dataset for Stress Analysis in Social Media, 2019,arXiv, eprint - 1911.00133
11. Gang Liu & Jiabao Guo, Bidirectional LSTM with attention mechanism and convolutional layer for text Classification, Neurocomputing, Volume 337, 14 April 2019, Pages 325-338
12. Yu, Xm., Feng, Wz., Wang, H. et al. An attention mechanism and multi-granularity-based Bi-LSTM model for Chinese Q&A system. Soft Comput 24, 5831–5845 (2020). https://doi.org/10.1007/s00500-019-04367-8
13. Jang, Beakcheol and Kim, Myeonghwi and Harerimana, Gaspard and Kang, Sang-ug and Kim, Jong Wook, Bi-LSTM Model to Increase Accuracy in Text Classification: Combining Word2vec CNN and Attention Mechanism, Applied Sciences, Volume 10, 2020 (17), Article Number - 5841, https://doi.org/10.3390/app10175841, ISSN 2076-3417
14. Weijiang Li, Fang Qi, Ming Tang, Zhengtao Yu, Bidirectional LSTM with self-attention mechanism and multi- channel features for sentiment classification, Neurocomputing, Volume 387, 2020, Pages 63-77, ISSN 0925- 2312, https://doi.org/10.1016/j.neucom.2020.01.006
15. Kim, Y., Choi, A. EEG-Based Emotion Classification Using Long Short-Term Memory Network with Attention Mechanism. *Sensors* 2020, *20*, 6727. https://doi.org/10.3390/s20236727
16. J. Wan, B. Chen, Y. Liu, Y. Yuan, H. Liu and L. Jin, Recognizing the HRRP by Combining CNN and BiRNN With Attention Mechanism, in IEEE Access, vol. 8, pp. 20828-20837, 2020, doi: 10.1109/ACCESS.2020.2969450.
17. Su, C., Xu, Z., Pathak, J. *et al.* Deep learning in mental health outcome research: a scoping review. *Transl Psychiatry* 10, 116 (2020). https://doi.org/10.1038/s41398-020-0780-3
18. Baek, JW., Chung, K. Accident risk prediction model based on attention-mechanism LSTM using modality convergence in multimodal. Pers Ubiquit Comput (2021). https://doi.org/10.1007/s00779-021-01552-1

19. Wang, Yakun and Du, Yajun and Hu, Jinrong and Li, Xianyong and Chen, Xiaoliang, SAEP: A Surrounding- Aware Individual Emotion Prediction Model Combined with T-LSTM and Memory Attention Mechanism, Applied Sciences, Volume 11, 2021, No. 23, Article-Number 11111, https://www.doi.org/10.3390/app112311111, ISSN 2076-3417
20. Yifeng Zhao, Deyun Chen, "Expression EEG Multimodal Emotion Recognition Method Based on the Bidirectional LSTM and Attention Mechanism", Computational and Mathematical Methods in Medicine, vol. 2021, Article ID 9967592, https://doi.org/10.1155/2021/9967592
21. Yan ZHAO and Yue XIE and Ruiyu LIANG and Li ZHANG and Li ZHAO and Chengyu LIU, Detecting Depression from Speech through an Attentive LSTM Network, IEICE Transactions on Information and Systems, volume E104.D, No.11, pages 2019-2023, year 2021, https://doi.org/10.1587/transinf.2020EDL8132
22. Ghosh T. et al. (2021) An Attention-Based Mood Controlling Framework for Social Media Users. In: Mahmud M., Kaiser M.S., Vassanelli S., Dai Q., Zhong N. (eds) Brain Informatics. BI 2021. Lecture Notes in Computer Science, vol 12960. Springer, Cham. https://doi.org/10.1007/978-3-030-86993-9_23.
23. https://github.com/viritaromero/Detecting-Depression-in-Tweets/blob/master sentiment_tweets3.csv.
24. Jeffrey Pennington, Richard Socher, Christopher D. Manning, GloVe: Global Vectors for Word Representation, Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 1532–1543, October 25-29, 2014, Doha, Qatar.

.