Journal of Information Systems Engineering and Management
2025, 10(44s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Self-Adaptive Probability with Long Short-Term Memory for Early Detection of Depression in Web Document

[1]Ashritha R Murthy, [2]Anil Kumar K M

[1]Assistant Professor, Department of Computer Science and Engineering, JSS Science and Technology University ,Mysuru, Karnataka ,India
ashrithar.murthy@jssstuniv.in
[2]Professor, Department of Computer Science and Engineering, JSS Science and Technology University, Mysuru, Karnataka, India
anilkm@jssstuniv.in

ARTICLE INFO

Received: 18 Dec 2024

Revised: 10 Feb 2025

Accepted: 28 Feb 2025

ABSTRACT

The depression is increasingly prevalent across all ages groups due to fast pace of life and many people use social media to express their feelings. Consequently, social media provides valuable data for early detection of stress and depression. However, challenges arise in text feature sensitivity, as unwanted text in documents lead to overfitting, making it difficult to capture complex relationships due to affect the detection accuracy. In this research, proposed Self-Adaptive Probability – Long Short Term Memory (SD-LSTM) technique for detecting depression is efficiently capture sequence data and achieves better accuracy. The SD-LSTM ability to adjust based on input patterns helps to detect different types of text data, minimizes chances of overfitting thereby enhancing accuracy. Pre-processing involved stemming, stop word removal, lowercase, removal of non-character to eliminate unwanted text and improve clarity of sentences. The feature extraction used XLNet, capturing nuanced meaning of words in context, which is particularly essential in depression detection where word meanings change significantly depending on context. The proposed SD-LSTM achieve better accuracy of 96.45% on Dreaddit, 96.85% of accuracy on Depression Mixed and 0.92 RSDD dataset. The existing method such as Gates Recurrent Unit (GRU), Logistic Regression (LR) is evaluated the proposed method.

Keywords: Depression Detection, Gates Recurrent Unit, Self-Adaptive Probability – Long Short Term Memory, Single Gated Leak ReLU-Convolutional Neural Network and Text Feature

I. INTRODUCTION

Depression is significant risk factor for dementia and individuals suffering from dementia often experience a decline in cognitive abilities including thinking and memory.

As one of the leading causes of mental disorders, early detection and treatment of depression are crucial for enhancing chances of managing the condition and minimizing its negative impact on a person socio-economic life [1-3]. Deep Learning (DL) have been widely applied in research to gain a deeper understanding of user's mental health on social networking platforms [4]. The people often express their stress and depression on social media, providing valuable information. Analysing the linguistic patterns in depressive posts offers insights into their mental state [5]. However, there is an insurmountable issue is exploring the text word sentence in online media that detect using traditional analysing model additional it is difficult to give timely prediction outcomes of new depressed users [6]. The traditional technique is efficiently detected but limited amount of correlations maximized significantly due to growth in volume of data so in this research consider improved DL techniques and provided a better outcome in depression detection [7].

These sources offer a potential pathway to uncover insights into mental health, particularly in tasks related to stress and depression. Individuals experiencing mental health issue often struggle to express their feelings and these challenges explored through text feature analysis [8]. The collected from the Dreaddit, Depression-Mixed and RSDD datasets is general suffering from these mental aliment using text feature expressed by text feature. The text facing unwanted context is resolve the pre-processing stage then extract the feature to analysis the mid-level feature

Journal of Information Systems Engineering and Management
2025, 10(44s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

involved is consider the text, edge of the feature [9-10]. The main criticism of most of this research is about the values of extract depression scores for each of the depressive in the user timeline and provided it to DL model, enabling the consideration of temporal modelling for detecting of depression [11-12]. This research focuses on various classes related to depression and stress, aiming to detect depression and non-depression using DL technique. It identifies that depressed individuals often use words associated with rejection, negative expressions, sadness, stress, or dissatisfaction [13-14]. Therefore, DL allows computational models that are composed of multiple processing layer to learn representations of data with various depression [15]. However, challenges arise in text feature sensitivity, as unwanted text in documents lead to overfitting, making it difficult to capture complex relationships due to affect the detection accuracy. In this research, proposed Self-Adaptive Probability – Long Short Term Memory (SD-LSTM) technique for detecting depression is efficiently capture sequence data and achieves better accuracy. The SD-LSTM ability to adjust based on input patterns helps to detect different types of text data, reducing errors and minimizes the chances of overfitting thereby enhancing accuracy. The main contributions of this research are considered in below:

- Pre-processing involved stemming, stop word removal, lowercase, removal of non-character to eliminate unwanted text and improve the clarity of sentences.
- The feature extraction used XLNet, capturing the meaningful of words in context, which is particularly essential in depression detection where word meanings change significantly depending on context
- The proposed SD-LSTM technique for detecting depression is efficiently capture sequence data and achieve better accuracy. The SD-LSTM ability to adjust based on input patterns helps to detect different types of text data, reducing errors and minimizes the chances of overfitting thereby enhancing accuracy.

The rest of paper as follows: section 2 provides a related work for depression detection. Section 3 introduces the proposed method utilized by SD-LSTM Section 4 discusses result and comparative analysis. Section 5 discusses conclusion.

II. LITERATURE REVIEW

This research conduct studies on depression detection involved SD-LSTM technique. The Literature review about depression and stress detection are provided in this section, along with their advantages and limitations. Yang et al. [16] presented a DL based Gated Recurrent Units (GRUs) to explicitly method to analysis the mental states by text feature and identification of depression tendency from the Dreaddit and Depression-Mixed dataset. These was considering the knowledge-aware and Contrastive Network (KC-Net) extract mental stage knowledge combined to performed in dot product attention were used for metallization process. However, GRU model limit their ability to capture text feature because complex to analysis the long-term dependencies.

Oryngozha et al. [17] developed a machine Learning (ML) based Logistic Regression (LR) focuses on detecting and analysing streass-realted dataset like Dreaddit, ML technique using for classify text as stressed or non-stressed. These was efficient trained especially large data, does not require the input feature because it normally distributed, handle a variety of data distributions. However, LR technique was sensitive to outliers which significantly affect the model and struggle to capture complex relationship of data due to lack on the accuracy.

Ilias et al. [18] introduced a Deep Neural Network (DNN) or language models based on transformers based on Mental-Bidirectional Encoder Representation from Transformers (Mental-BERT) method consider the Depression-Mixed and Dreaddit dataset. These was considering the context from text feature, consider the meaning of word based on full context due to learn efficiently to achieve better accuracy. However, BERT models were large which makes them difficult to deploy on text with limited data such as embedded systems.

Guo et al. [19] developed a Single Gated LeakReLU-Convolutional Neural Network (SGL-CNN) technique involved the depression detection models to identify key feature from a large number of data. These was performed in activation function allows minimized feature, improve the model ability to focus on relevant feature while suppressing irrelevant leading to better accuracy. However, SGL-CNN combination of gating and leaky activation function to makes less transparent of decision making and complex to learn text feature instant were activated has challenge to analysis the depressing text.

dos Santos et al. [20] presented a depression anxiety detection by using BERT method to analysis the stress and depression disorder from self-disclosure statement with minimal annotation. These was involved pre-trained on large corpus of text enabling to analysis the text feature and allowing to effectively handle out of vocabulary word which meaning of new terms. However, BERT model read data from left context of word makes it difficult to deployed, not accuracy learn pattern of text due to large data involved in the model.

Inamdar et al. [21] presented a Natural Language Processing tools like Embeddings from Language Model for word embeddings (ELMo) with logistic regression and Support Vector Machine (SVM). These was generated the dynamic embeddings that consider the context of words in a sentence, leading to improve classification accuracy. However, ELMo embedding required large data while logistic regression and SVM were linear models that does not capture non-linear pattern.

In the overall analysis, the existing techniques have challenges arise in text feature sensitivity, as unwanted text in documents lead to overfitting, making it difficult to capture complex relationships due to affect detection accuracy. In this research, proposed SD-LSTM technique for detecting depression is efficiently capture sequence data and achieve better accuracy. The SD-LSTM ability to adjust based on input patterns helps to detect different types of text data, reducing errors and minimizes the chances of overfitting thereby enhancing accuracy.

III. PROPOSED METHODOLOGY

In this research, Proposed SDLSTM network for detecting depression is efficiently analysis the stress and allows to capture text feature, leading to better accuracy detection. Initially data obtain from the Dreaddit, Depression-Mixed and RSDD datasets. The stemming, stopwords, removing non character and lowercase this phase applied to text feature to utilized the essential words. The feature extraction using XLNet technique considers context from both directions, enabling to better sentence phrases and particularly use in depression where the meaning of words changes significantly depending on context. Figure 1 Pipeline of implemented method.

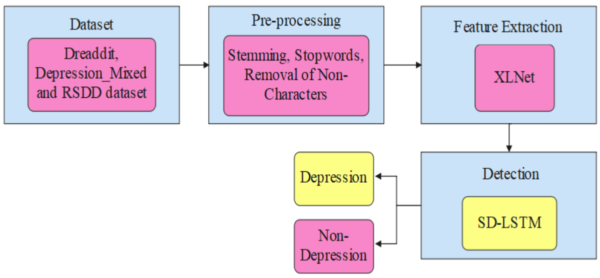

Fig. 1. Pipeline of implemented method.

A. Data Collection

In this section, proposed SD-LSTM technique is evaluated using Dreaddit, Depression_Mixed and RSDD dataset benchmark for detecting depression.

DReaddit Dataset

The Dreaddit dataset [22] is a text data collection specifically designed for detecting depression. An includes a total of 190 thousand posts from Reddit with 3,553 of them manually labelled. The dataset contains stress-indicative across 5 fields, making it a valuable resource for stress detection research.

Depression Mixed Dataset

The Depression Mixed dataset [23] is a depression detection dataset containing 2,765 posts collected from various subedits on Reddit. Each post is labeled as either depressive or non-depressive and consists of a multi-sentence monologue detailing speaker's background and current feelings. The authors specifically matched text feature with instant protocols to support accurate depression detection.

RSDD Dataset

The RSDD dataset [24] is a comprehensive resource suitable for experiments related to depression, created by annotating users from a publicly available Reddit dataset. Designed to simulate low-resource platforms, the RSDD facilitates research in early depression detection. The dataset includes 107,274 control users and 9,210 diagnosed users, covering a wide range of mental health conditions. In total, it contains data on 385,476 users, including 6,434 with bipolar disorder, 14,139 with depression, and 335,952 control users.

i. *Pre-processing*

The pre-processing [25] involved initially stemming, stop word removal, non-character eliminated and converting text to lowercase are crucial in the cleaning the text data and reduce the noise.

Stemming

In this phase, stemming is converts terms that not meaningful word into their root form, thereby focusing on the core meaning of language. This process is expressed by equation (1) shown below

$$d_i = \left\{ W_j^i, 1 \le j \le m_i \right\} \quad (1)$$

Where, m_i denoted extracted words from i^{th} documents after extracting keyword, W unique keywords are expressed by equation (2) shown below;

$$W = \left\{ b_i, 1 \le x \le k \right\} \quad (2)$$

Where k denoted total amount of words in dictionary or unique keywords from documents.

Stop word Removal

In this section, removal unwanted repeated word for example: “the”, “and” that does not contribute to the semantic meaning of the text reduce the noise and improving the focus on meaningful content.

Lowercase

To ensures that words are treated uniformly regardless of their case which helps in avoiding duplication of text feature in depression detection.

Removing of Non-Character

Removing non-characters’ elements, such as letters, special characters, HTML tags and other alphabetic characters, from the text data helps remove unnecessary noise and irrelevant data. The pre-processed data is fed to feature extraction to enhance the contextual feature from cleaned text.

ii. *Feature Extraction*

After pre-processed data, XLNet technique for feature extraction is capturing the nuanced meaning of words in context and particularly essential in depression detection where the meaning of word change significantly depending on context. The pre-training process involved using a large corpus to train on text feature, which enhances models ability to learn and introduces a permutation model. These permutation modelling objective of XLNet is expressed in equation (3) shown below;

$$\max_{\theta} \mathrm{E}_z \sim Z_T \left[\sum_{t=1}^{T} log\ p_\theta \left(x_{zt} \mid x_{Z<t} \right) \right] \quad (3)$$

Where, Z_T indicate possible set of all possible permutations of length T, sequence of index values is $\left[1,2,\ldots,T\right], z_t$ represented $t-th$ element, permutation solution in Z_T, $z \in Z_T$ respectively. The XLNet is consider 2 stream self-

Journal of Information Systems Engineering and Management
2025, 10(44s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

attention mechanism to obtain bidirectional context from recent location without mask. The text content denoted $h_\theta\left(X_{z<t}\right)$ function to similarly to standard states in transformer, encoding both context and x_{zt} respectively. The query denoted $g_\vartheta\left(x_{z<t}, z_t\right)$ has access contextual data $X_{z<t}$ and location of z_t. The every self-attention layer $m=1,..,M$ 2 streams of denoted are improved equation (4) and (5) shown below;

$$g_{z_t}^{(m)} \leftarrow Attention\left(Q=g_{z_t}^{(m-1)}, KV=h_{Z_{<t}}^{(m-1)}; \theta\right) \qquad (4)$$

$$h_{z_t}^{(m)} \leftarrow Attention\left(Q=h_{z_t}^{(m-1)}, KV=h_{Z_{<t}}^{(m-1)}; \theta\right) \qquad (5)$$

Where, Q, K, and V denoted key values of query in attention operation and convergence rate of permutation model is low. To enhance convergence rate, XLNet predicts last token in the factorization order and target subsequence conditioned by expressed in equation (6) shown below;

$$\begin{aligned} &\max_\theta \mathrm{E}_z \sim Z_T\left[\log p_\theta\left(X_{z<t} \mid X_{z\le t}\right)\right] \\ &= \mathrm{E}_z \sim Z_T\left[\sum_{t=c+1}^{|z|} \log p_\theta\left(x_{zt} \mid x_{Z<t}\right)\right] \end{aligned} \qquad (6)$$

The sequence data within recent factorization order z cause XLNet to lose some information when dealing with particularly long sequences. The 2 segments taken from a long sequence s, $\tilde{x}=s_{1:T}$, and $x=s_{T+1:2T}$ respectively. The $[1,..,T]$ denoted permutations and $[T+1,\ldots,2T]$ are processed and content $h^{(m)}$ for every self-attention layer m, is updated with memory is expressed in equation (7) shown below;

$$h_{z_t}^{(m)} \leftarrow Attn\left(Q=h_{z_t}^{(m-1)}, KV=\left[\tilde{h}^{(m-1)}, h_{Z_{<t}}^{(m-1)}\right]; \theta\right) \qquad (7)$$

Where, concatenation occurs with dimension sequence along $Attn(.)$ denoted attention operation and location of encoding is related actual location in original sequence. The input to XLNet consists of word embedding $e(x)$ and initialization matrix W, context and query are improved through several masked 2 stream attention layer.

iii. Classification

In depression detection main goal is to categorize text data into specific classes such as depressed or non-depressed by using SD-LSTM network automatically determine whether the language used in give text indicates signs of depression. These are adapting its learning process based on probability distribution of input sequences. This allows the model to dynamically adjust the parameter to focus on probable feature or patterns related to depression leading to better accurately detect. The improvements in handling long term dependencies make SD-LSTM more capable of capturing the full context of text and reduce the change of overfitting leading to model performs.

iv. *Long Short-Term Memory*

The LSTM network, an enhance version of Recurrent Neural Network (RNN), consists of unique components knowns as memory blocks within its hidden layers. Each memory block includes cells as well as input, output and forget gates. The standard LSTM architecture introduces a forget gate, which enable network to adjust its state dynamically. The forget gate f_1 resets cell variable, causing previously stored input c_i to be forgotten while mange the input from feature vector x_t and generation of output is h_t respectively. The cell's weight information is influenced by forget gate, as previous gate cannot permanently retain information. The computations within as LSTM block operate as follows, input values are preserved in cell state only if input gate allows them, involving i_t and expected values of memory cells $\tilde{C}_t$ at each time step. These processes are evaluated in equation (8) to (13) shown below;

$$i_t = \sigma\left(W_i.\left[h_{t-1}, x_t\right] + b_i\right) \quad (8)$$

$$\tilde{C}_t = tanh(W_c\left[\left[h_{t-1}, x_t\right] + b_c\right) \quad (9)$$

$$f_1 = \sigma\left(W_f.\left[h_{t-1}, x_t\right] + b_f\right) \quad (10)$$

$$\tilde{C}_t = i_t.\tilde{C}_t + f_1.C_{t-1}^{\sim} \quad (11)$$

$$O_t = \sigma(W_x.\left[\left[h_{t-1}, x_t\right] + b_o\right) \quad (12)$$

$$h_t = O_t * tanh\left(c_t\right) \quad (13)$$

Where, $W_i.\left[h_{t-1}, x_t\right]$ and b denoted weight metrics and bias respectively. The forget gate control weight of state units of cell and forget gates, process new state of memory cell is being to updated (11). The recent memory cell, output values of gate is evaluated in (12) and final output is denoted (13) cell values. The σ, g and h are point wise nonlinear activation function and i, f, o, and c are input, forget, output gate and cell activation vector respectively. All feature of LSTM has trained using sigmoid (ϕ) and $tanh$ activation function. The proposed LSTM is avoiding issue of processing continuous input steams that are not segmented into subsequence that mean steams does not theoretically divide into small units are easily processed by network.

Proposed Self Adaptive Probabilities – Long Short-Term Memory

The SD-LSTM adapts its learning process based on probability distribution of input sequences and allows model to base on probability distribution of input sequences. It dynamically adjusts it focus on more probable feature or patterns related to depression, leading to more accurate classifications. The dropout method discards neural network unit from network temporarily according to an instant's probability during training process of depth learning network and range of neurons then set its output as 0. To support of specific characteristics of other feature thus decrease probability of overfitting in training process. The network structure generated arbitrarily is most in related application based on LSTM and evaluated probability value of selective discarding neurons in dropout to enhance self-adaptive enhance ability to detect depression across different text patterns. The mathematically expression (14) and (15) shown below;

$$\frac{N_d}{N_j} = \frac{N_q}{N} \quad (14)$$

$$N_j = N_d + N_u \quad (15)$$

The LSTM models with their complex structure to analysis the text because of small data occur in the overfitting issue reduce self-adaptive probability mechanism. The reduce overfitting issue updated in equation (16) shown below;

$$\frac{N_{qd}}{N_j} \le \frac{N_d}{N_j} \quad (16)$$

Where, formulas above N_d denoted amount of discarded nodes, N_j denoted amount of nodes for each layer, N_q denoted singular points, N indicated in single layer network. The enhance method proposed based on probability of selecting node need to remove dropout for effective prevent overfitting issue is more prominent. The self-adaptive mechanism in SD-LSTM allow model to adjust its focus based on specific input making more flexible in handling different manifestation of depression.

\

Journal of Information Systems Engineering and Management
2025, 10(44s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

IV. EXPERIMENT RESULTS

In this section, Proposed SD-LSTM technique for depression detection is efficiently achieve better accuracy of the experimental analysis is ensured using datasets namely Dreaddit, Depression_Mixed & RSDD dataset. The implementation of the proposed method is carried out using Python 3.10.12 on Windows 10 (64-bit) operating system, with an Intel Core i5 processor, and 8GB RAM. The performance of proposed method is evaluated using several performance metrics, including Accuracy, Precision, F1-score, and Recall, which are defined by equations (17) to (20) as shown below.

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{17}$$

$$Precision = \frac{TP}{(TP + FP)} \tag{18}$$

$$F1-score = 2 * \frac{(Precision * Recall)}{Precision + Recall} \tag{19}$$

$$Recall = \frac{TP}{TP + FN} \tag{20}$$

Where, TP , TN , FP , and FN illustrate True Positive, True Negative, False Positive, and False Negatives respectively.

(i) Performance Analysis

In this section, proposed method involving feature extraction and detection processes is evaluated using various performance metrics, including Accuracy, Precision, F1-measure, and Recall for the Dreaddit, Depression_Mixed and RSDD dataset. The feature extraction process with datasets is represented in Figure 2 and 10. The performance of XLNet feature extraction is evaluated based on accuracy, precision, F1-measure, and recall on datasets, as described in Figure 2. Existing methods using feature extraction techniques such as BERT, VGG 16, AlexNet and ResNet are also evaluated. The XLNet method achieves a high accuracy of 96.45% on Dreaddit and 96.85% on Depression Mixed dataset. The XLNet method achieves a high accuracy of 0.92 on RSDD dataset. The feature extraction technique XLNet achieves high accuracy, reaching 96.45% and 96.85% respectively. Figure 2 to 4 for feature extraction based on 3 dataset.

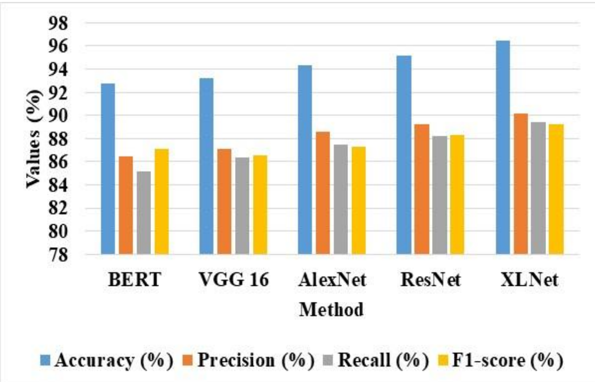

Fig. 2. Performance analysis of feature extraction on Dreaddit

Journal of Information Systems Engineering and Management
2025, 10(44s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

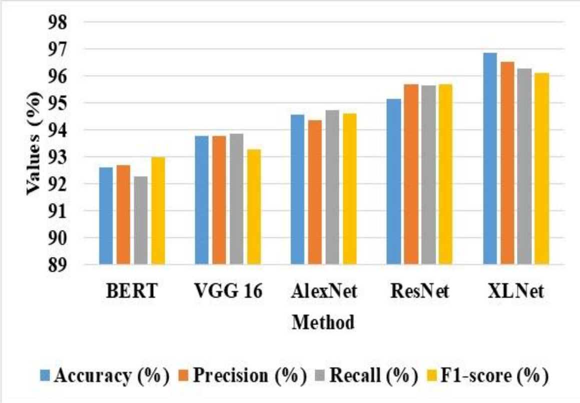

Fig. 3. Performance analysis of feature extraction on Depression_Mixed

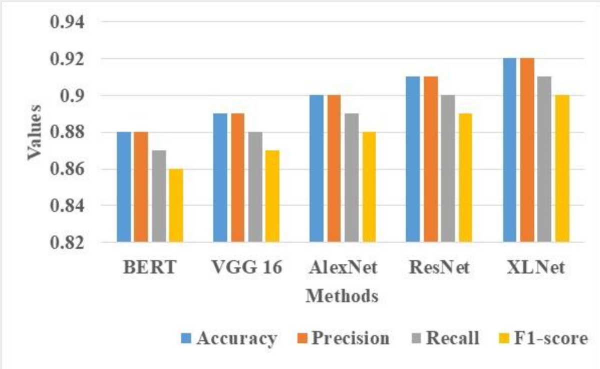

Fig. 4. Graphically represents of feature extraction on RSDD dataset

The performance of SD-LSTM for detection is evaluated based on accuracy, precision, F1-measure, and recall on datasets, as described in Table 1. Existing methods using feature extraction techniques like RNN, LSTM, Bi-LSTM and GRU are also evaluated. The SD-LSTM method achieves a high accuracy of 96.45% on Dreaddit and 96.85% of accuracy on Depression_Mixed datasets. Figure 5 to 7 for classification based on 3 dataset.

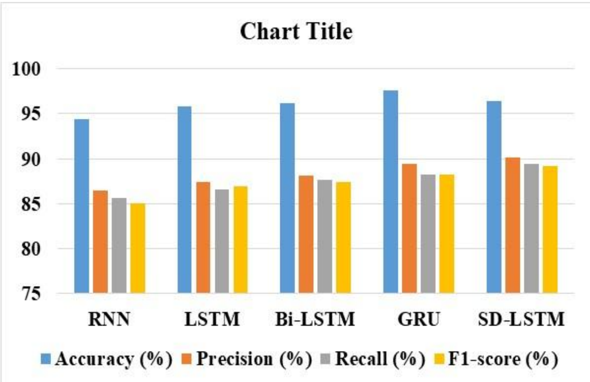

Fig. 5. Performance analysis of classification on Dreaddit

Journal of Information Systems Engineering and Management
2025, 10(44s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

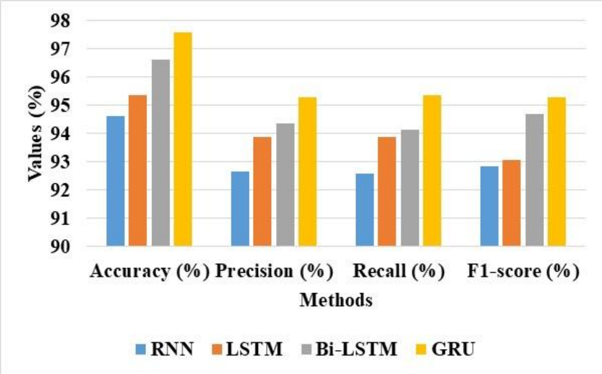

Fig. 6. Performance analysis of classification on Depression_Mixed

TABLE II

PERFORMANCE OF CLASSIFICATION ON DATASETS

The performance of SD-LSTM for classification is evaluated based on accuracy, precision, F1-measure, and recall on datasets, as described in figure 2. Existing methods using feature extraction techniques such as RNN, LSTM, Bi-LSTM and GRU are also evaluated. The SD-LSTM method achieves a high accuracy of 0.92 on RSDD dataset. The classification SD-LSTM technique achieves high accuracy of 0.92, respectively.

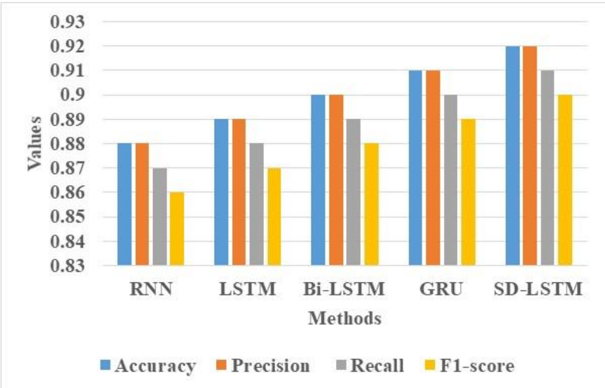

Fig. 7. Graphically represents of classification on RSDD dataset

Figure 8 to 10 for k-fold values based on 3 dataset, presents detection results using different K-fold values for the proposed method on the Dreaddit and Depression Mixed dataset. Compared to values of 2, 4, 7 and 9, using 5-fold cross-validation ensures that model is trained text feature, while also validating on a reasonable subset.

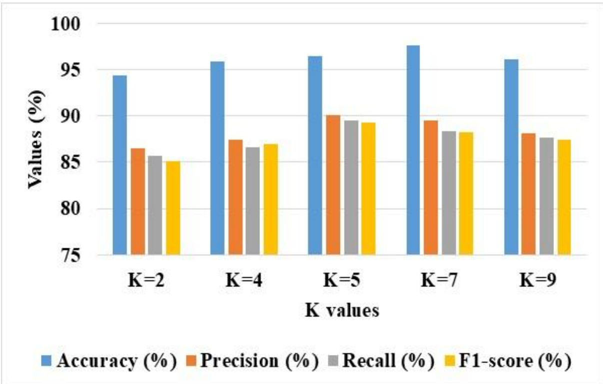

Fig. 8. Graphically represented the K fold values on Dreaddit dataset

Journal of Information Systems Engineering and Management
2025, 10(44s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

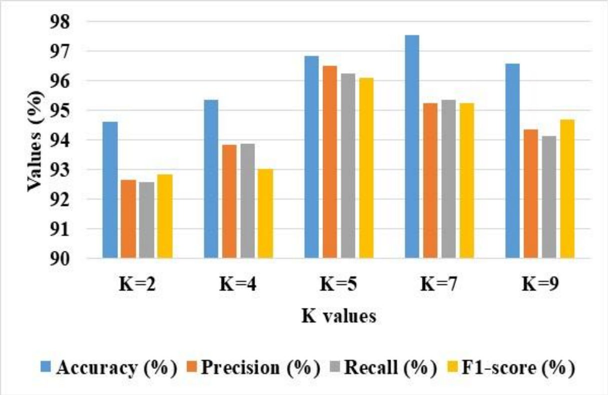

Fig. 9. Graphically represented the K fold values on Depression_Mixed dataset

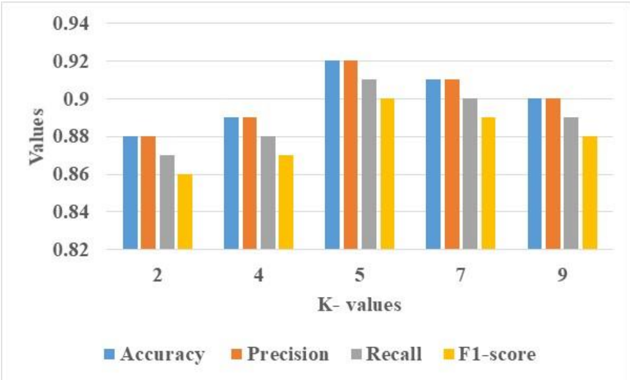

Fig. 10. Graphically represented the K fold values on RSDD dataset

(ii) Comparative Analysis

The performance of proposed SD-LSTM technique is compared to existing methods, including GRU [16], LR [17], Mental BERT [18], SGL-CNN [19], ELMO (Logistic Regression and SVM) [21] and BERT [20]. In this research, proposed SD-LSTM method achieves high accuracy, reaching 93.79% on Dreaddit dataset, 96.85% on Depression Mixed dataset and 0.92 RSDD dataset. Table 1 and 2 describes a comparative analysis of proposed method using Dreaddit, depression mixed and RSDD datasets. The SD-LSTM method achieves better accuracy by effectively capturing sequence data through its self-adaptive probability mechanism and reduce overfitting. This adaptability helps in accurately detecting different types of text data, making the model more resilient to variations in the data.

TABLE I

COMPARATIVE ANALYSIS OF PROPOSED METHOD ON DREADDIT AND DEPRESSION MIXED DATASET

Methods	Datasets	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
GRU [16]	Dreaddit	NA	84.1	83.3	83.5
LR [17]		NA	77.78	NA	NA
Mental BERT [18]		NA	NA	80.28	80.04
ELMO (Logistic Regressio		76	74	70	74

Journal of Information Systems Engineering and Management
2025, 10(44s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

n and SVM) [21]					
Proposed SD-LSTM method		96.45	90.12	89.45	89.25
GRU [16]	Depression_Mixed	NA	95.5	95.3	95.4
Mental BERT [18]		91.15	89.57	93.14	91.17
Proposed SD-LSTM method		96.85	96.52	96.25	96.11

TABLE II

COMPARATIVE ANALYSIS OF PROPOSED METHOD ON RSDD DATASET

Methods	Datasets	Precision	Recall	F1-score
SGL-CNN [19]	RSDD	0.86	0.84	0.85
BERT [20]		0.85	0.77	0.63
Proposed SD-LSTM method		0.92	0.91	0.90

(iii) Discussion

The advantage of proposed method and limitation of exiting method like GRU [16] model limit their ability to capture text feature because complex to analysis the long-term dependencies. The LR [17] technique was sensitive to outliers which significantly affect the model and struggle to capture complex relationship of data due to lack on the accuracy. BERT [18] models were large which makes them difficult to deploy on text with limited data such as embedded systems. SGL-CNN [19] combination of gating and leaky activation function to makes less transparent of decision making and complex to learn text feature instant were activated has challenge to analysis the depressing text. BERT [20] model read data from left context of word makes it difficult to deployed, not accuracy learn pattern of text due to large data involved in the model. The improvements in handling long term dependencies make SD-LSTM more capable of capturing the full context of text and reduce the change of overfitting leading to model performs.

V. CONCLUSION

In this research, proposed SD-LSTM technique for detecting depression is efficiently capture sequence data and achieve better accuracy. The SD-LSTM ability to adjust based on input patterns helps to detect different types of text data, reducing errors and minimizes the chances of overfitting thereby enhancing accuracy. Initially data obtained from Dreaddit, Depression Mixed and RSDD dataset and pre-processing involved stemming, stop word removal, lowercase, removal of non-character to eliminate unwanted text and improve the clarity of sentences. The XLNet used feature extraction, capturing the nuanced meaning of words in context, which is particularly essential in depression detection where word meanings change significantly depending on context. The proposed SD-LSTM achieve better accuracy of 96.45% on Dreaddit, 96.85% of accuracy on Depression Mixed and 0.92 of accuracy on RSDD datasets. The existing method such as GRU, SGL-CNN are evaluated the proposed method. Future work will consider the hybrid technique to analysis the various text features efficiently

ACKNOWLEDGEMENTS

The heading of the Acknowledgments section and the References section must not be numbered. Causal Productions wishes to acknowledge Michael Shell and other contributors for developing and maintaining the IEEE

Journal of Information Systems Engineering and Management
2025, 10(44s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

LaTeX style files which have been used in the preparation of this template. To see the list of contributors, please refer to top of file IEEETran.cls in the IEEE LaTeX distribution.

REFERENCES

[1] A. Pérez, J. Parapar, and Á. Barreiro, "Automatic depression score estimation with word embedding models," *Artif. Intell. Med.*, vol. 132, p. 102380, Oct. 2022.

[2] M. Z. Uddin, K. K. Dysthe, A. Følstad, and P. B. Brandtzaeg, "Deep learning for prediction of depressive symptoms in a large textual dataset," *Neural Comput. Appl.*, vol. 34, no. 1, pp. 721–744, Jan. 2022.

[3] L. Ansari, S. Ji, Q. Chen, and E. Cambria, "Ensemble Hybrid Learning Methods for Automated Depression Detection," *IEEE Trans. Comput. Social Syst.*, vol. 10, no. 1, pp. 211–219, Feb. 2023.

[4] T. Sharma, R. Panchendrarajan, and A. Saxena, "Characterisation of Mental Health Conditions in Social Media Using Deep Learning Techniques," in *Deep Learning for Social Media Data Analytics*, *Studies in Big Data*, 1st ed. vol. 113, T. P. Hong, L. Serrano-Estrada, A. Saxena, A. Biswas, Eds. Springer: Cham, 2022, pp. 157–176.

[5] L. Ilias, S. Mouzakitis, and D. Askounis, "Calibration of Transformer-Based Models for Identifying Stress and Depression in Social Media," *IEEE Trans. Comput. Social Syst.*, vol. 11, no. 2, pp. 1979–1990, Apr. 2024.

[6] Y. Wang, Z. Wang, C. Li, Y. Zhang, and H. Wang, "Online social network individual depression detection using a multitask heterogenous modality fusion approach," *Inf. Sci.*, vol. 609, pp. 727–749, Sep. 2022.

[7] H. Kour and M. K. Gupta, "An hybrid deep learning approach for depression prediction from user tweets using feature-rich CNN and bi-directional LSTM," *Multimedia Tools Appl.*, vol. 81, no. 17, pp. 23649–23685, Jul. 2022.

[8] H. Zogan, I. Razzak, X. Wang, S. Jameel, and G. Xu, "Explainable depression detection with multi-aspect features using a hybrid deep learning model on social media," *World Wide Web*, vol. 25, no. 1, pp. 281–304, Jan. 2022.

[9] N. Farruque, R. Goebel, S. Sivapalan, and O. Zaïane, "Deep temporal modelling of clinical depression through social media text," *Nat. Lang. Process. J.*, vol. 6, p. 100052, Mar. 2024.

[10] T. Nijhawan, G. Attigeri, and T. Ananthakrishna, "Stress detection using natural language processing and machine learning over social interactions," *J. Big Data*, vol. 9, p. 33, Mar. 2022.

[11] M. A. Wani, M. A. ELAffendi, K. A. Shakil, A. S. Imran, and A. A. A. El-Latif, "Depression Screening in Humans With AI and Deep Learning Techniques," *IEEE Trans. Comput. Social Syst.*, vol. 10, no. 4, pp. 2074–2089, Aug. 2023.

[12] U. Ahmed, G. Srivastava, U. Yun, and J. C.-W. Lin, "EANDC: An explainable attention network based deep adaptive clustering model for mental health treatment," *Future Gener. Comput. Syst.*, vol. 130, pp. 106–113, May 2022.

[13] Vandana, N. Marriwala, and D. Chaudhary, "A hybrid model for depression detection using deep learning," *Meas.: Sens.*, vol. 25, p. 100587, Feb. 2023.

[14] T. Liu, J. Meyerhoff, J. C. Eichstaedt, C. J. Karr, S. M. Kaiser, K. P. Kording, D. C. Mohr, and L. H. Ungar, "The relationship between text message sentiment and self-reported depression," *Journal of Affective Disorders*, vol. 302, pp. 7–14, Apr. 2022.

[15] Y. Wu, Z. Liu, J. Yuan, B. Chen, H. Cai, L. Liu, Y. Zhao, H. Mei, J. Deng, Y. Bao, and B. Hu, "PIE: A Personalized Information Embedded model for text-based depression detection," *Inf. Process. Manage.*, vol. 61, no. 6, pp. 103830–103830, Nov. 2024.

[16] K. Yang, T. Zhang, and S. Ananiadou, "A mental state Knowledge–aware and Contrastive Network for early stress and depression detection on social media," *Inf. Process. Manage.*, vol. 59, no. 4, p. 102961, Jul. 2022.

[17] N. Oryngozha, P. Shamoi, and A. Igali, "Detection and Analysis of Stress-Related Posts in Reddit's Acamedic Communities," *IEEE Access*, vol. 12, pp. 14932–14948, Jan. 2024.

[18] L. Ilias, S. Mouzakitis and D. Askounis, "Calibration of Transformer-Based Models for Identifying Stress and Depression in Social Media," IEEE Trans. Comput. Social Syst., vol. 11, no. 2, pp. 1979–1990, Apr. 2024.

[19] Y. Guo, Z. Zhang, and X. Xu, "Research on the detection model of mental illness of online forum users based on convolutional network," BMC psychology, vol. 11, no. 1, p. 424, Dec. 2023.

[20] W. R. dos Santos, R. L. de Oliveira, and I. Paraboni, "SetembroBR: a social media corpus for depression and anxiety disorder prediction," Lang. Resour. Eval., vol. 58, no. 1, pp. 273–300, Mar. 2024.

Journal of Information Systems Engineering and Management
2025, 10(44s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

[21] S. Inamdar, R. Chapekar, S. Gite, and B. Pradhan, “Machine Learning Driven Mental Stress Detection on Reddit Posts Using Natural Language Processing,” Hum.-Centric Intell. Syst., vol. 3, no.2, pp. 80–91, Jun. 2023.
[22] Dreaddit Dataset. Available: https://paperswithcode.com/dataset/dreaddit. Accessed: Aug 2024.
[23] Depression Mixed dataset. Available: https://huggingface.co/migueladarlo/distilbert-depression-mixed. Accessed: Aug 2024.
[24] RSDD dataset. Available: https://paperswithcode.com/dataset/rsdd-time. Accessed: Aug 2024.
[25] M. Yarlagadda, K. G. Rao, and A. Srikrishna, “Frequent itemset-based feature selection and Rider Moth Search Algorithm for document clustering,” J. King Saud Univ. Comput. Inf. Sci., vol. 34, no. 4, pp. 1098–1109, Apr. 2022.