International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Detection of Suicidal Ideation Based on Relational Graph Attention Network with DNN Classifier

Shreekant Jere[1*], Annapurna P. Patil[2]

Submitted: 25/05/2023 **Revised**: 07/07/2023 **Accepted**: 25/07/2023

Abstract: Suicidal ideation is one of the most serious issues confronting today's youth, owing to mental illnesses such as anxiety, bipolar disorder, and depression. Suicidal ideation is difficult to categorise due to the way people use language and express themselves on Twitter or other social media platforms. Personal contextualization of such originality is difficult to accurately identify users at risk. Thus, intervening in the early stages of suicidal ideation can help to reduce the number of suicides. In this study, a Relational Graph Attention Network (RGAT) with a DNN classifier is used to detect suicidal ideation. Initially, Twitter posts are collected and preprocessed to remove unnecessary data. Bi-directional Encoder Representations from Transformers (BERT) are used for embedding the tweeter post to relate to the topic and sentimental words. The preprocessed data is then fed into Latent Dirichlet Allocation (LDA) and sentiment-based Lexicon approach to find the topic and extract sentimental words. RGAT relates the features from LDA and BERT as well as the lexicon-based sentiment approach. Both of these RGAT outcomes are concatenated and classified using the DNN algorithm. The performance metrics evaluate and compare the proposed method to existing models. The attained performance metrics like precision, accuracy, recall, and error for the proposed model are 88.22%, 88.21%, 88.19%, and 12%, respectively. The evaluated metric values of the proposed model is better than the values of the existing models. Thus the designed model using RGAT with a DNN classifier performs better and accurately detects suicidal ideation.

Keywords: *LDA, BERT, RGAT, Lexicon based sentiment approach, DNN*

1. Introduction

Suicide is one of the most serious social health issues confronting the world today [1]. According to the World Health Organisation (WHO), over one million people die by suicide each year [2]. Suicide is also the sixth most common cause of death for adults and the leading cause of death for young people due to mental disorders. Suicidal behaviour can be influenced by a variety of social and personal circumstances, including trauma or unpleasant experiences, physical or mental illness, social isolation, hopelessness, anxiety, depression, and bipolar disorder [3].

Anxiety is a common emotion or feeling that occurs during stressful situations, and the brain alerts the body to potential danger [4]. Having problems at work, making a critical choice, losing interest, or rage are all symptoms of depression. Suicide and heart attacks or strokes are more likely in people who suffer from severe depression and psychological distress [5]. Suicidal ideas or thoughts often precede the act of suicide, and many people believe that suicide is a permanent solution to temporary problems.

Social media platforms such as Twitter, Reddit, and Facebook are becoming modern networks where individuals can express their thoughts and status without worrying about social judgement [6].

People prefer to discuss suicidal ideation online rather than in person. According to research, social media posts can help people identify depression and other mental health issues. These online activities motivated them to create novel potential healthcare solutions and early suicide detection systems [7]. Because of the rapid increase in suicidal cases, identifying risk factors early can help to reduce the risk of suicide. One of the most important findings of this study is that the social stigma associated with this mental illness makes it difficult for psychiatric health professionals to advise and treat those individuals [8]. Because laboratory tests to detect suicidal ideation are not yet available, it is believed that social media could provide an opportunity to analyze a person's attitudes early on by identifying risk factors and warning signs, thereby preventing deaths.

Natural Language Processing (NLP) is utilized to assist with tasks such as sentiment analysis, detecting fake news on social media, and mental health screening [9]. It has recently begun to play a growing role in processing data from social networks. Advances in NLP technologies and deep learning models can help to address the growing interest in examining advanced techniques for diagnosing mental illness [10]. Deep learning-based approaches like Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), have drawbacks in identifying

[1]AI Labs, Accenture Solutions Pvt Ltd, VTU, Bangalore, India*
[2]MS Ramaiah Institute of Technology, VTU, Bangalore, India

suicidal disorders, such as being time-consuming, having a limited number of benchmarks for training, evaluating suicidal ideation detection performance accuracy is poor, and analyzing the linguistic traits of tweets is often insufficient for accurate suicidal intent detection. This research work proposed a mental disorder detection and suicidal ideation based on a Relational Graph Attention Network with DNN classifier to easily identify suicidal ideation person by social workers and volunteers in social service. Major contributions of the research work are provided below.

- Suicidal ideation detection using Relational Graph Attention Network and DNN classifier on Tweeter data.
- Latent Dirichlet Allocation is used to process the Twitter data and identify the topic of each tweet.
- The polarity of sentiments and emotions expressed in tweets is extracted using lexicon-based semantic analysis on Twitter data.
- BERT is used to encode the word into a vector based on the context model in order to find the relation based on sentence embedding.
- RGAT is used to relate the features from LDA and BERT encoder, as well as the features of semantic analysis and BERT encoder for the attention of risk identification and state indicator.
- The DNN technique is used to classify suicidal ideation based on the severity of risk after concatenation from the relation network.

The remaining part of the paper is structured as follows: Section 2 comprises works of literature and section 3 explains a brief about the proposed methods. Section 4 illustrates the experimental results for the designed model, and section 5 summarizes the entire research article.

2. Literature Survey

Social media has grown in popularity, and the number of active users is growing. Suicidal thoughts and expressions are now common on social media. Many studies on suicidality in social networks have been conducted in order to prevent suicide and lower the associated mortality rates. Cao, L., Zhang, H., and Feng, L. [11], developed a high-level knowledge graph with deep neural networks for the purpose of detecting suicidal thoughts on media platforms. This technique is based on a two-layer attention mechanism with explicit reason and major risk indicators for people's suicidal thoughts. This was performed to evaluate Reddit, and microblogs to demonstrate that with the personally constructed knowledge graph, the social media based on detecting suicidal thoughts, and personality, post, and experience were the top three critical factors out of the six domains of personal factors. From this domain, one's suicidal ideation may be detected based on text and image posts, anxiety level, stress period, and depressive mood.

Ghosal, S., and Jain, A. [12] modeled a fastText embedding, and XGBoost to classify the data. Earlier diagnosis of depressed content with NLP was a significant area of research since a person who was at risk required such methods. Even society and the next generation may be saved if suicide risk and depression were recognized. A novel framework was developed to differentiate between content that poses a risk of suicidal ideation and depression with the aid of the machine learning classifier XGBoost for accurate classification, TF-IDF vector for term relevance, and fastText embedding for contextual analysis.

Cusick, M., Adekkanattu, P., et al. [13] developed a weakly supervised technique to identify "current" suicidal thoughts from Electronic Health Record (EHR) systems, which were composed of unstructured clinical notes. Weakly supervised machine learning methods used imperfect labels for learning, removing the burden of manually analyzing a large dataset. After identifying a risk for suicidal ideation, a rule-based Natural Language Processing approach (NLP) is used to label the validation and training notes. The logistic classifier, Support Vector Machines (SVM), and deep learning model known as Convolutional Neural Network (CNN) were trained machine learning models to be evaluated using this large corpus of clinical notes. The use of this method for extensive document evaluation may be crucial for clinical data systems for focused suicide prevention efforts.

Adarsh, V., Kumar, P. A. et al. [14] designed an approach based on the one-shot decision to normalize the imbalance in participation across the different age groups and demographics. Additionally, an ensemble model that combines Support Vector Machines (SVM) and K-Nearest Neighbor (KNN) with intrinsic explainability in conjunction with noisy label correction frameworks was described, establishing a novel solution to the problem of distinguishing depression mood from suicidal thoughts. A label correction with the NMT further improved this strategy, and the bias variance was investigated and resolved using intrinsic explainability techniques.

Tadesse, M. M., Lin, H., Xu, B., & Yang, L. [15] designed to examine Reddit users' posts for any indicators that could expose relevant online users' attitudes towards depression. The data was trained for this purpose using machine learning methods and Natural Language Processing (NLP) techniques, and the performance was analyzed. This method helps identify a vocabulary of more dominant words in depressed reports. The outcomes demonstrated that the approach could significantly increase performance accuracy.

Chatterjee, M., Samanta, P., Kumar, P., & Sarkar, D. [16] modelled that Correct diagnosis of suicidal behaviour and proper care were two of the most effective methods of preventing suicide attempts. Researchers claimed that there was a direct connection between a person's mental state and the language that was used in their expression. The methodology aimed to examine online Twitter tweets and identify characteristics that could indicate suicidal ideation in users. Both machine learning and natural language processing techniques were applied to the data in order to train the model and evaluate its performance. The precision of the Logistic Regression classifier was achieved by extracting and combining a number of linguistic, topical, temporal sentiment, and statistical features.

Some of the drawbacks attained in reviewed articles are that they do not perform well on sparse and unstructured data, have high memory requirements [12], may require more keystrokes and are unable to adapt to the new domain [13], be more difficult to interpret and wrong selection can lead to lower predictive accuracy[14], Non-linear problems can't be solved [15]. However, the proposed model overcomes the above issues by using a Relational Graph Attention Network with a DNN classifier.

3. Proposed Methodology

Mental disorders can cause suicidal ideation due to some factors, such as depression, anxiety, and bipolar. This disorder can affect the ability to do their everyday work and activities more than a common person due to a number of factors like financial problems, life events, and various social exposure. Based on the mental disorder, the severity of suicidal ideation varies. The use of platforms such as social media to share experiences and emotions has provided new opportunities for analyzing and diagnosing mental illness and suicidal thoughts. Suicidal ideation is detected early on social media platforms, which provides an automatic evaluation for suicidal tendencies, and prevents them from hazards. To examine the severity of risk in suicidal ideation Relational Graph Attention Network and DNN classifier is proposed.

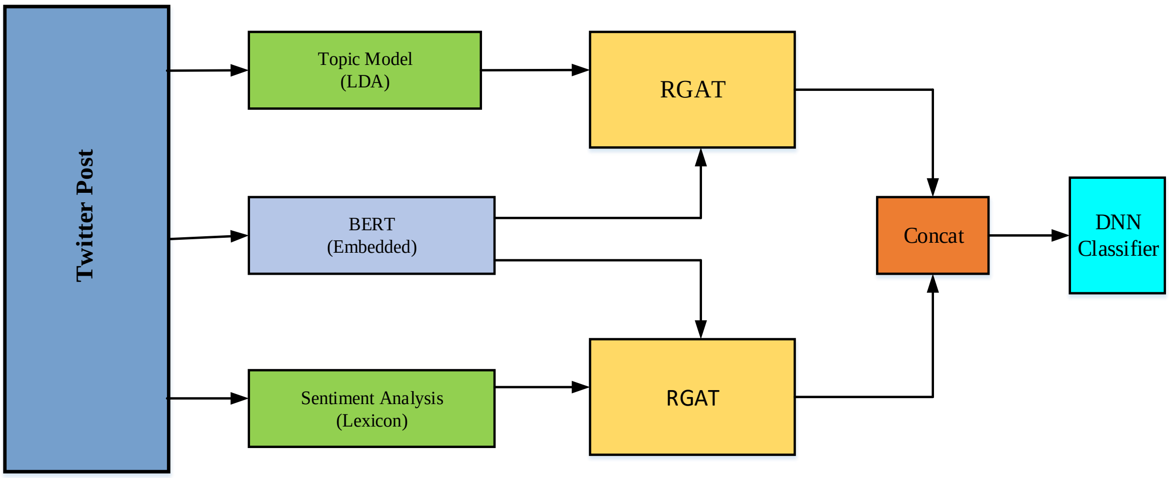

Fig 1: Flow diagram of the proposed methodology

Figure 1 illustrates the process flow of detecting mental disorders and suicidal ideation based on a Relational Graph Attention Network with a DNN classifier. The proposed method consists of two phases, are phase1 and phase 2, to extract the features based on their techniques. Initially, input is taken as a collection of Twitter posts for preprocessing. The data is given to the topic model, which is evaluated using the Latent Dirichlet Allocation (LDA) method and sentiment based on lexicon analysis to find the word of sentiments and also fed into BERT to convert the word into vector. Then data is extracted from LDA and BERT using sentence embedding output data given to the RGAT phase1. On the other hand, sentiment base lexicon analysis and BERT sentence embedding output data are given to the RGAT phase 2. Both Relational Graph Attention Networks (RGAT) are used to find the relation feature and concatenate with the DNN classifier to classify the severity of suicide.

3.1 Preprocessing of the Twitter dataset

The source set of data has more noisy and unnecessary data, which will degrade the efficiency of the topic model and sentiment analysis performance. Tweets are simple comment that contains hashtags, emojis, caption, and special symbols to convey the user's perspective. So preprocessing of data is an important step in any modelling topic and sentiment analysis model. Preprocessing data typically consists of several steps, including cleaning the data of unwanted characters (cleaning text), tokenizing the text, removing meaningless words (stopwords), and changing the word to its basic form (stemming or lemmatization). Special characters and symbols like @, #, and $ should be eliminated from the input data because they are useless for analysis by pattern removal. Then stopwords like "is," "was," "and," "or," and similar expressions have no meaning for sentiment analysis and must be eliminated from the data set. If the URL in the dataset has no subject matter and reduces the classification efficiency, it should also be eliminated. The process of tokenization involves separating a sentence into its single words or tokens. Stemming algorithm works by removing the suffix of the word according to some grammatical rules. The goal of data preprocessing is to generate clean data that is ready for analysis, resulting in more accurate and valid analysis results.

3.2 Bidirectional Encoder Representations from Transformers

One of the most effective context and word representations is BERT. The attention mechanism used by BERT is based on the transformer methodology. Attention is a method of examining the relationship between the words in a sentence. BERT considers the entire left and right context of a given the word. It should be noted that the same word can have multiple embeddings depending on the context. BERT transformer is a Google AI language pre-trained model and also learns the contextual relationship between words in a sentence [17]. Each post of users is encoded into an embedding using the pre-trained model SBERT. The pre-trained transformer model SBERT converts a sentence into a vector-length representation.

3.3 Rgat Mechanism Phase 1

After preprocessing the Twitter data set, the processed data is fed into Latent Dirichlet Allocation (LDA) to find the topic of the Twitter post.

a. Latent Dirichlet allocation

An unsupervised generative probabilistic modelling technique known as LDA can be used to find hidden linguistic formations in a collection of text documents. Each topic in a document is a combination of words according to LDA, and each topic's relative importance (expressed as a variety of weights) varies depending on the document. The highest probabilities of the word will provide a deeper knowledge of the topic, which represents the probabilistic distribution of each document in latent topics by LDA's generative process. [18] All document topic distributions share the Dirichlet prior. Furthermore, each latent topic is represented as a probabilistic word distribution, and all topic word distributions have the same Dirichlet prior. A corpus D with K documents is presented, with document d containing F_d words. The generative process used by LDA to model D is as follows:

1. For each topic p ($p\epsilon\{1,\ldots,P\}$), select multinomial φ_p from Dir (β).

2. For every document m ($m\epsilon\{1,\ldots,K\}$), select multinomial θ_m from Dir (α).

3. In present document m, for each word U_n ($n\epsilon\{1,\ldots,F_d\}$)

i. Select a topic V_n from multinomial θ_m.
ii. Select a word U_n from multinomial φ_{V_n}.

The following mathematical formula represents the probability of the LDA is expressed as

$$p(D|\alpha,\beta) = \prod_{m=1}^{K}\int p(\theta_m|\alpha)\left(\prod_{n=1}^{F_d}\sum_{V_{mn}} p(V_{mn}|\theta_m)p(U_{mn}|V_{mn},\beta)\right)d\theta_m \quad (1)$$

Where the Dirichlet prior variable α for the document-topic distributionθ, the parameter β is the Dirichlet prior for the topic's word distribution φ, P is the latent topics count, and F represents the vocabulary size or how many words are in the document.V_{mn}, U_{mn} are words in each text document sampled for the current word-level variables. Latent Dirichlet Allocation (LDA) and S-BERT output are given to Relational Graph Attention Network based on the attention mechanism for further data processing.

3.4 Rgat Mechanism Phase 2

The second phase is processed for lexicon-based sentiment analysis to find the sentiment polarity of the Twitter dataset to extract the sentimental words.

i. Lexicon-based sentimental analysis

Lexicon analysis is one of the most widely used methods for analyzing sentiment words, and it also determines the sentiment based on the semantic orientation of words or phrases found in a text [19] [20]. This technique uses a dictionary of positive and negative words to rate the sentiment of each word in a tweet, either positively or negatively. Following this message representation, the dictionary assigns sentiment scores to each positive and negative word or phrase. The final estimation of the overall sentiment for the message is made using combining operations like average or sum. Using the dictionary-based sentiment analysis technique, the features from the database are retrieved, and the sentiment polarity is calculated. In order to determine the polarity of the entire sentence, SentiWordNet dictionary is used to assign polarities to individual words where the score value is compared to a predetermined threshold. If it is greater than the threshold value, the score is classified as positive otherwise, it is classified as negative. The text is assigned to express neutral sentiment when there is no difference.

The SentiWordNet database's evaluated positive and negative term scores are used to determine the sentiment orientation for each term in the sentences or tweets. Using this method, First list the term with the highest sentiment score. Based on the context or frequency of each term in the given sentence determines the score for that term. The total score of both positive (TotPosScore) and negative score (TotNegScore) are calculated based on Equation (2), where m refers to the number of terms, and tweet is represented by w.

$$TotPosScore_w = \sum_{s=1}^{m} TotPosScore + PosScore_s$$

$$TotNegScore_w = \sum_{s=1}^{m} TotNegScore + NegScore_s \quad (2)$$

To determine whether a tweet's sentiment is "positive," "negative," or "neutral," the total positive score and total negative score are added together for each tweet. The overall sentiment polarity $polarity_{swn}(w)$ for the tweet, w is provided in Equation (3).

$$polarity_{swn}(w) = \begin{cases} 1, if\ TotPosScore(w) > TotNegScore(w) \\ -1, if\ TotPosScore(w) < TotNegScore(w) \\ 0\ , otherwise \end{cases} \quad (3)$$

The sentiment score provides three metrics that quantify the sentiment of a user's tweet $polarity_{swn}(w)$ calculate for the tweet w using the SentiWordNet database. The tweet's overall sentiment is positive or 1 if the sum of its positive and negative scores is greater than zero; otherwise, it is neutral or 0; and if the sum of its negative and positive scores is lower than zero, the overall sentiment is negative or –1.

3.5 Relational graph attention network (RGAT)

A Relational Graph Attention Network (R-GAT) based on the attention mechanism is proposed by expanding the Graph Attention Network (GAT) of dependency relations encoded, and the link between opinion words and aspects is established.

i.) Graph Attention Network

A graph G with n nodes, where each node represents a word in a sentence, represents a dependency tree. The interrelation of words is represented by the edges of G and N_i is used to represent the nodes adjacent to node i. GAT considers neighbourhood node representations and updates each representation iteratively using multi-head attention.

A GAT changes the hidden state of a node at layer c+1 using multi-head attention given a node i with a hidden state h_r^i at layer r, the neighbours' node n[i], and their hidden states. The updating procedure is

$$h_{at_i}^{c+1} = \Big\|_{m=1}^{M} \sum_{j\epsilon n[i]} \alpha_{ij}^{cm} W_m^c h_j^c \quad (4)$$

$$\alpha_{ij}^{cm} = attention(i,j) \quad (5)$$

Where, ‖ represents vector concatenation. $\|_{m=1}^{M} y_i$ represent the concatenation of vectors y_1 to y_c head at layer c+, $\alpha_{ij}^{cm}$is the node i attention coefficient towards neighbour j(m indicates attention head at layer c). $W_{cm} \in R^{D/M}$ is the matrix initial states of a linear transformation, and D is the hidden states dimension. During training, an attention context vector is learned and represented as $\alpha_{cm} \in R^{\frac{2D}{M}}$ at $h_{at_i}^{c+1}$ node i attention.

ii.) Relational graph attention network

The GAT combines the dependency paths and the representations of neighbouring nodes. However, this method ignores the relationships of dependency, which might result in the loss of critical dependency data. Neighbourhood nodes should impact each other differently if their dependency relations are different. Additional relational heads are added to enhance the source GAT, and these relational heads are used to control information flow from neighbouring nodes using relation-wise gates. Particularly, compute a relational head by first converting the relations of dependency into vector representations.

$$h_{rel_i}^{c+1} = \Big\|_{k=1}^{K} \sum_{j\epsilon N_i} \beta_{ij}^{ck} W_k^c h_j^c \quad (6)$$

$$g_{ij}^{ck} = \sigma\big(relu\big(r_{ij} W_{k1} + b_{k1}\big) W_{k2} + b_{k2}\big) \quad (7)$$

$$\beta_{ij}^{cm} = \frac{e^{g_{ij}^{ck}}}{\sum_{j=1}^{N_i} e^{g_{ij}^{ck}}} \quad (8)$$

Where r_{ij} Shows nodes i and j are embedded in relation to one another. Two types of the head in R-GAT are K relational heads and M attentional heads.

$$x_i^{c+1} = h_{att_i}^{c+1} \| h_{rel_i}^{c+1} \quad (9)$$

$$h_i^{c+1} = relu(W_{c+1}\, x_i^{c+1} + b_{c+1}) \quad (10)$$

The above Equation (10) computes each node's final representation. After determining the features from the relational graph attention network concatenate the output of RGAT and gives it to a classifier.

3.6. DNN Classification

A DNN is a network of neurons organized in layers, with each layer receiving neuron activations from the previous layer as input and performing a basic computation. It is also known as deep learning because deep neural networks have many hidden layers, which are much larger than in other neural networks. The layer is made up of nodes, which are simply places where computation takes place. The bias and weight of each node differ for each pair of units in each two-layer succession.[21] DNN has divided into three units, namely the input unit, hidden unit, and output unit. The input layer of DNN receives the given input data. The proposed method gives different input data to the input layer. The mathematical computations on the given input data are performed in the hidden layer. There are multiple hidden layers in DNN, but the proposed method uses two hidden layers for computation. The final layer is the output layer, directly linked to the goal value the model is attempting to predict. The nodes of a deep-learning network train on a variety of characteristics based on the output of the unit before it. Nodes are able to recognize more complex features as the neural network evolves because they gather and combine data from earlier units.

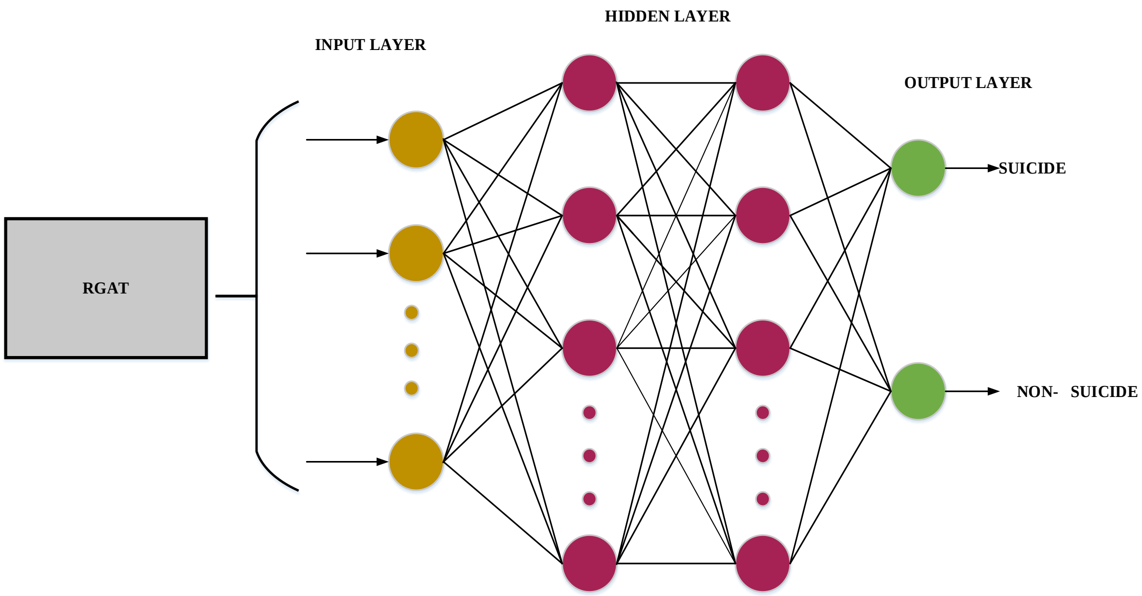

Fig 2: DNN general architecture

The general architecture of DNN is shown in figure 2, and the network calculation for three hidden layers is defined as,

$$h_a^{(1)} = \varphi^{(1)}\left(\sum_b w_{ab}^{(1)} x_b + b_b^{(1)}\right) \qquad (11)$$

$$h_a^{(2)} = \varphi^{(2)}\left(\sum_b w_{ab}^{(2)} h_b^{(1)} + b_b^{(2)}\right) \qquad (12)$$

$$h_a^{(3)} = \varphi^{(3)}\left(\sum_b w_{ab}^{(3)} h_b^{(2)} + b_b^{(3)}\right) \qquad (13)$$

$$y_a = \varphi^{(4)}\left(\sum_b w_{ab}^{(4)} h_b^{(3)} + b_b^{(4)}\right) \qquad (14)$$

Where w is the weight input, b is the bias value, x_b refer to input units, y is the output unit, $h_b^{(l)}$ and φ denotes units in the hidden layer and activation function. By using this DNN architecture, the extracted features are trained for predicting suicidal and non-suicidal ideation from Twitter posts.

4. Result and Discussion

The proposed method is implemented in order to detect mental disorders and suicidal ideation based on a relational graph attention network with a DNN classifier. This software, Python 3.8.8, is installed with an Intel(R) i5-3330S processor running at 2.70GHz, 8GB of memory (RAM), and a 64-bit operating system are used to evaluate the designed model. Twitter data is collected for preprocessing the data to remove unwanted text representation. The preprocessed data is then fed into phase 1 and phase 2 to extract the features. Identifying the topic by using Latent Dirichlet Allocation and BERT is applied to encode the tweets is performed in phase 1. Lexicon-based sentiment analysis is used to find the polarity of sentiments and encode the tweets using BERT, which is processed in phase 2. These data are given to Relational Graph Attention Network (RGAT) is used to relate the features from phase 1 and phase 2 outcomes. Then both RGAT output is concatenated and given to the DNN classifier to classify the severity of risk in suicidal ideation.

4.1 Dataset Description

The dataset contains a collection of posts from the "Suicide Watch" and "depression" subreddits of the Reddit platform, and these posts are acquired using Pushshift API. [22] This collection contains a total number of 233372 posts which are labelled as suicide and non-suicide.

4.2 Confusion metrics

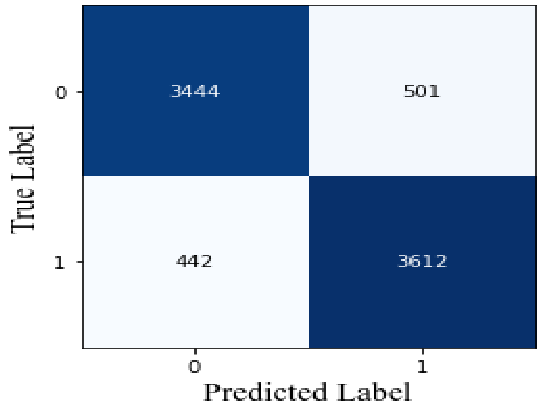

Fig 3 Confusion metrics for the proposed model

Figure 3 illustrates the confusion metrics of the proposed model. It is used to measure the effectiveness of a classification model to classify correct and incorrect predictions. Confusion metrics provide a table layout for different outcomes of the predicted results and help to visualize its outcomes. It plots a table of all a classifier's predicted and actual values. A good method has high TP and TN rates and low FP and FN rates. The column represents the predicted label, and the row represents the true label of the target variable as positive and negative. This confusion metric analyses the actual and predicted data from a given dataset. The prediction of data for 0 and 1 classes are 3444 and 3612. The total data used for testing is 7999 in that 7056 are predicted according to the actual class, and the rest of the 943 are predicted wrongly.

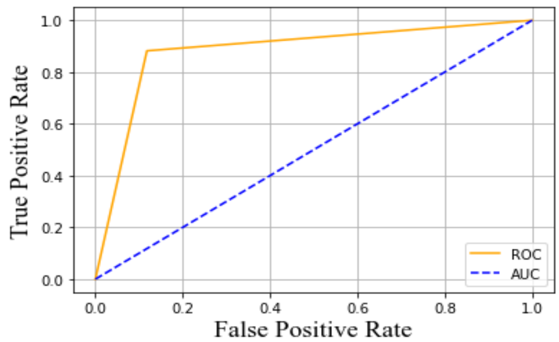

Fig 4 ROC plot for the proposed method

Figure 4 depicts the Receiver Operating Characteristic Curve (ROC) for the proposed suicidal ideation prediction. ROC (receiver operating characteristic curve) graph represents the classification model's performance across all thresholds. AUC stands for Area under the ROC Curve, which measures the total two-dimensional Area beneath the entire ROC curve. Two variables, such as the true and false positive rates, are used to plot the curve. For the best classifier, the range under ROC and AUC should be close to 1.

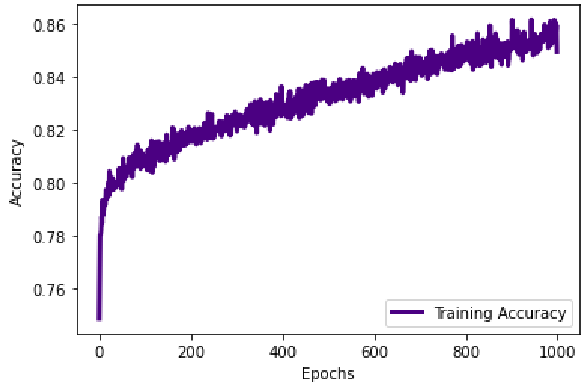

Fig 5 Training Accuracy of the proposed method

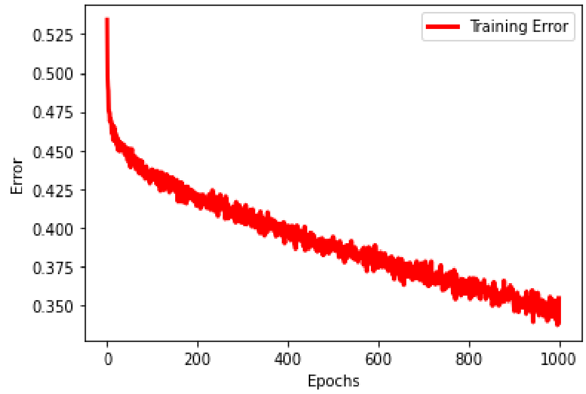

Fig 6 Training error of the proposed method

Figure 5 demonstrates the training accuracy of the proposed method. Epochs are the total number of iterations used to train the learning model with all of the training data in a single cycle. In training accuracy, the accuracy is increased for the number of an epoch. Where epoch is increased to 1000, accuracy attained to 0.86. The proposed method attains higher accuracy than the existing model. Figure 6 illustrates the training error of the proposed method. The graph represented in Training error for the proposed method is run between error and epoch. For 1000 epochs, the error that occurred in the proposed method is 0.1179. Based on the increment of the iteration, the error has been reduced.

4.3 Performance metrics for DNN classification

For the designed model, precision, accuracy, error, recall, specificity, Negative predictive value, false negative value, F1 score, false positive value, False Discovery rate, Training time, Testing time, Execution time, Kappa, and Mathew's correlation coefficient (MCC) are some of the performance metrics that are evaluated.

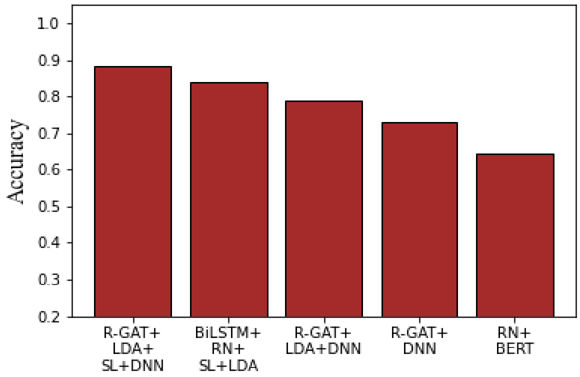

Fig 7 Accuracy comparison for proposed RGAT+LDA+SL+DNN and existing method

The proposed RGAT+LDA+SL+DNN and existing approaches are compared in terms of accuracy and are shown in figure 7. In this evaluation, an existing method such as RN+BERT, RGAT+LDA+DNN, RGAT+DNN, BiLSTM+RN+SL+LDA is compared with the proposed model RGAT+LDA+SL+DNN to determine the accuracy. The percentage of correctly classified subjects in the test dataset is referred to as accuracy. The accuracy value for RGAT+LDA+SL+DNN, BiLSTM+RN+SL+LDA, RGAT+LDA+DNN, RGAT+ DNN, and RN+BERT are 0.8821, 0.8385, 0.7901, 0.7305, and 0.6443. In comparison to previous methods, the proposed model is more accurate in detecting the type of suicidal ideation and has higher accuracy.

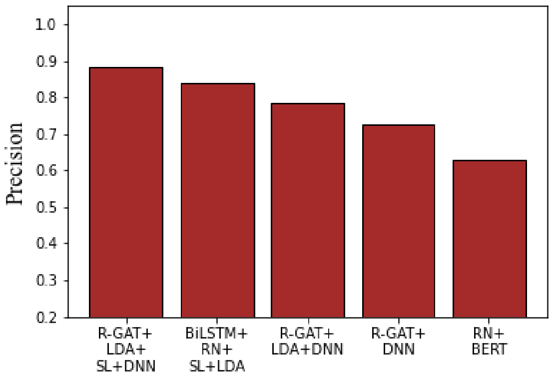

Fig 8 Comparison of proposed RGAT+LDA+SL+DNN and existing methods for precision

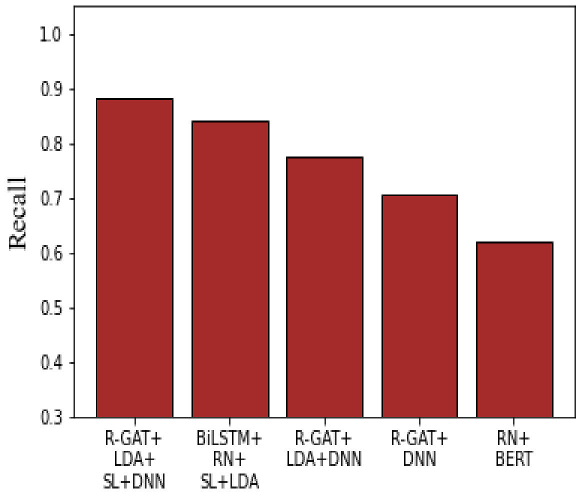

Fig 9 Recall evaluation for proposed RGAT+LDA+SL+DNN and existing techniques

The result of precision is compared between the proposed RGAT+LDA+SL+DNN, and the existing approaches are shown in Figure 8. Existing methods such as BiLSTM+RN+SL+LDA, RGAT+LDA+DNN, RGAT+DNN, and RN+BERT are compared with the proposed method. Precision depicts the relationship between true positive predicted values and full positive predicted values. The acquired precision values for the proposed RGAT+LDA+SL+DNN method and existing techniques are 0.8822, 0.8381, 0.7843, 0.7265, and 0.6303. In comparison to existing techniques, the precision rate is higher.

Figure 9 demonstrates the evaluation of recall values between the proposed RGAT+LDA+SL+DNN and the existing methods. The attained recall values for RGAT+LDA+SL+DNN, BiLSTM+RN+SL+LDA, RGAT+LDA+DNN, RGAT+ DNN, and RN+BERT are 0.8819, 0.8385, 0.7754, 0.7043 and 0.6178. This result shows the proposed method has a larger recall value than the existing model.

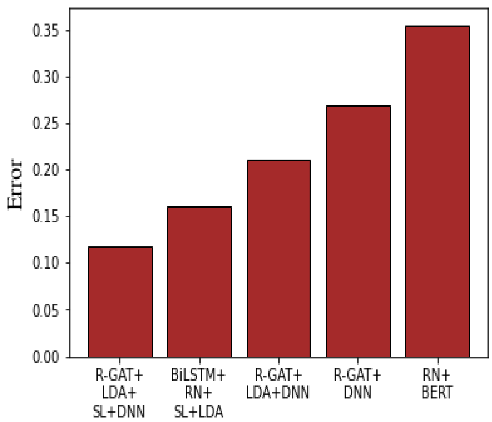

Fig 10 Error comparison between the proposed RGAT+LDA+SL+DNN method and the existing method

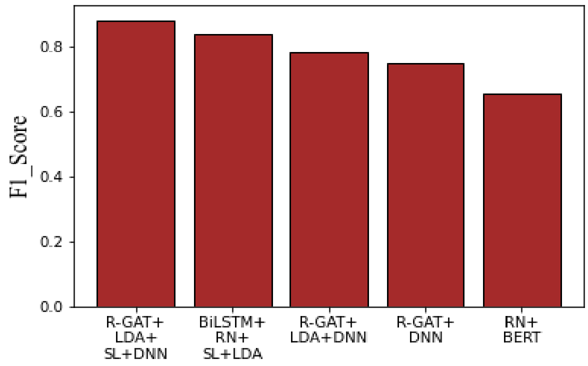

Figure 11 F1 score comparison for proposed RGAT+LDA+SL+DNN and existing method

Figure 10 depicts the evaluation error metrics for proposed and existing methods. The error values obtained for the proposed and modified method are 1-0.8821, 1-0.8385, 1-0.7901, 1-0.7305, and 1-0.6443. As a result, the proposed RGAT+LDA+SL+DNN method has less error than the existing model. F1 score is compared with the proposed RGAT+LDA+SL+DNN and existing model, which are illustrated in figure 11. The F1 score is defined as a combination of Precision and Recall. The ratio of precision and recall metrics multiplied by their sum is called the F1 score, which is twice that ratio. The attained F1 score values of RGAT+LDA+SL+DNN, BiLSTM+RN+SL+LDA, RGAT+LDA+DNN, RGAT+DNN, and RN+BERT are 0.8821, 0.8377, 0.7832, 0.7467 and 0.6522. As a result, the proposed method has a higher F1 score than the existing model.

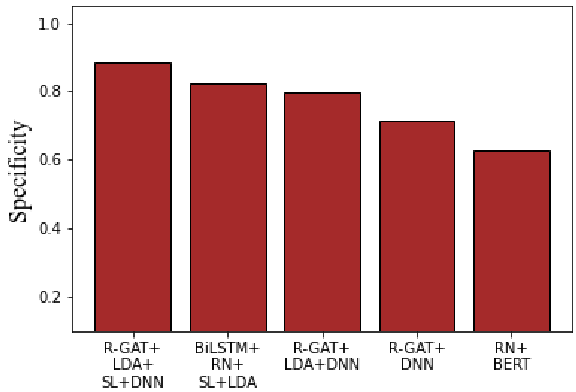

Fig 12 Specificity comparison of proposed RGAT+LDA+SL+DNN and existing techniques

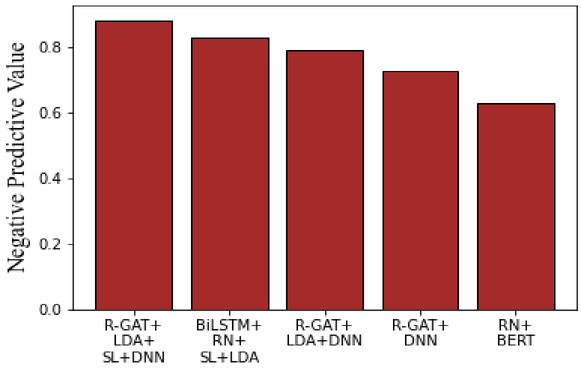

Figure 13 Comparison of Negative Predictive Value for proposed RGAT+LDA+SL+DNN and existing method

The estimation of specificity metrics for the proposed and existing models is shown in Figure 12. The proportion of true negatives correctly identified by the model is referred to as specificity. This implies that a percentage of actual negatives will be predicted as positive, which could be referred to as false positives. This significant proportion is also referred to as the True Negative Rate. The attained specificity value for the proposed RGAT+LDA+SL+DNN and existing method is 0.8828, 0.8256, 0.7987, 0.7143, and 0.6265. Compared to the existing methods, the proposed RGAT+LDA+SL+DNN model achieves a higher level of specificity. Figure 13 shows the negative predictive value comparison between the proposed RGAT+LDA+SL+DNN and the existing techniques. The negative predictive value (NPV) for RGAT+LDA+SL+DNN, BiLSTM+RN+SL+LDA, RGAT+LDA+DNN, RGAT+DNN, and RN+BERT are 0.8822, 0.8302, 0.7893, 0.7286 and 0.6301. As a result, the NPV value of the proposed method is more than the value of the existing techniques.

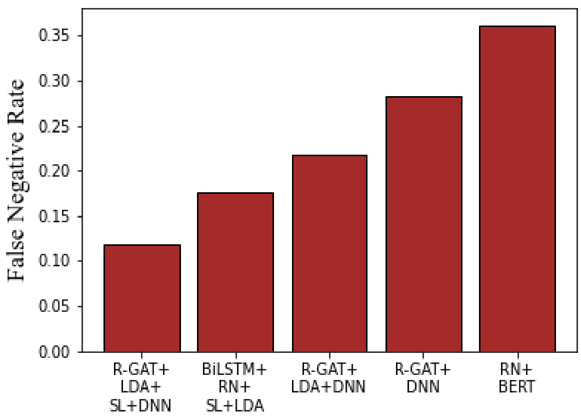

Fig 14 Evaluation of False Negative Rate (FNR) for proposed RGAT+LDA+SL+DNN and existing techniques

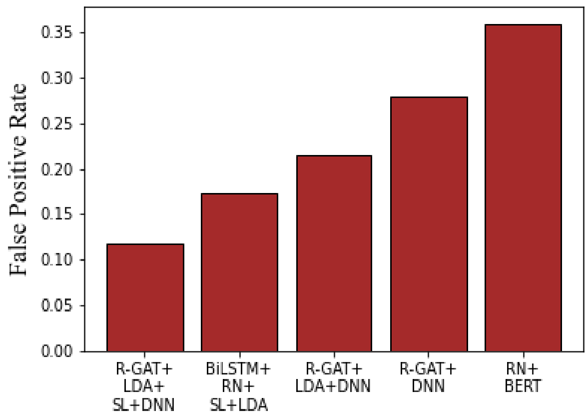

Fig 15 Evaluation of proposed RGAT+LDA+SL+DNN and existing techniques for False Positive Rate (FPR)

Figure 14 shows the comparison of the proposed RGAT+LDA+SL+DNN and the existing techniques for false negative rates (FNR). The False Negative rate values are 0.118, 0.175, 0.218, 0.283, and 0.362. Figure 15 depicts a comparison of the proposed method and existing approaches in terms of false positive rate (FPR). The false positive values for RGAT+ LDA+SL+DNN, BiLSTM+RN+SL+LDA, RGAT+ LDA+DNN, RGAT+DNN, and RN+BERT are 0.1180, 0.173, 0.215, 0.280 and 0.360. These obtained FPR and FNR values indicate that the proposed model performs better than the existing techniques.

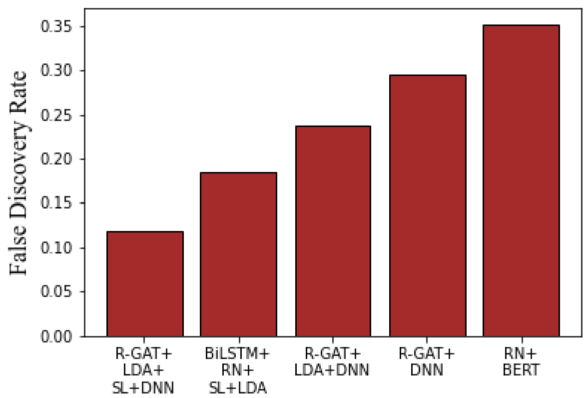

Fig 16 False Discovery Rate evaluation for proposed RGAT+LDA+SL+DNN and existing techniques

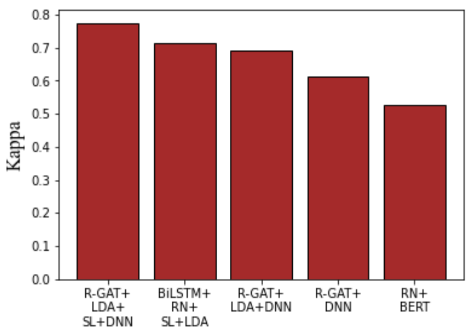

Fig 17 Kappa evaluation of proposed RGAT+LDA+SL+DNN and existing techniques

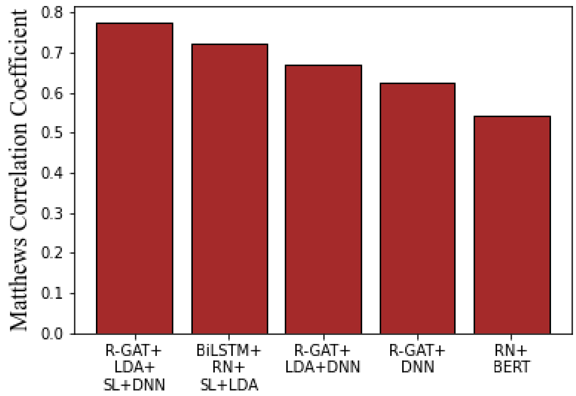

Fig 18 Matthews Correlation Coefficient (MCC) evaluation for proposed RGAT+LDA+SL+DNN and existing method

Figure 16 depicts the comparison of the false discovery rate between the proposed method and the existing techniques. The False Discovery rate values are 0.1177, 0.1843, 0.2376, 0.2954, and 0.3521. The proposed method achieved a lower false recovery rate than the existing models. The comparison of kappa metrics and Matthews Correlation Coefficient (MCC) for the proposed method and existing model are illustrated in figure 17 and figure 18. MCC is used in binary classification statistical analysis. It is a measure of the accuracy of the test. MCC provides a number between 1 and -1, with 1 indicating perfect predictions and -1 indicating the worst predictions. The attained kappa values and MCC of RGAT+LDA+SL+DNN, BiLSTM+RN+SL+LDA, RGAT+LDA+DNN, RGAT+DNN, and RN+BERT are 0.7757, 0.7143, 0.6898, 0.6144, 0.5265 and 0.7767, 0.7221, 0.6702, 0.6234, 0.5421, respectively. Thus the Kappa and MCC value of the proposed method is increased than the existing techniques.

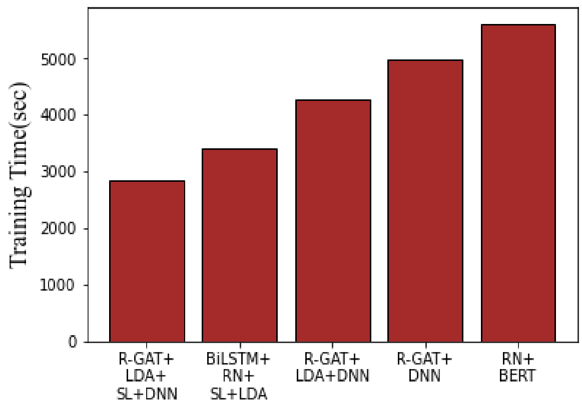

Fig 19 Training time evaluation for proposed RGAT+LDA+SL+DNN and existing techniques

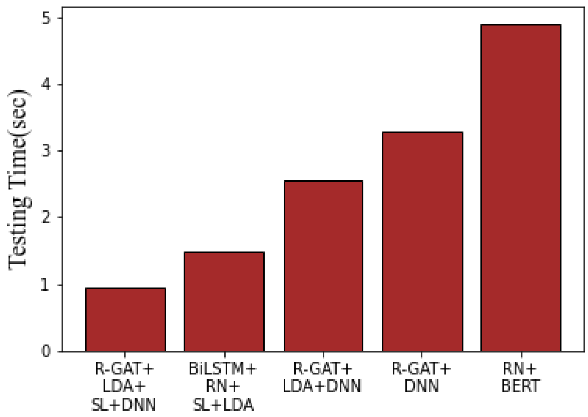

Fig 20 Testing time evaluation for proposed RGAT+LDA+SL+DNN and existing techniques

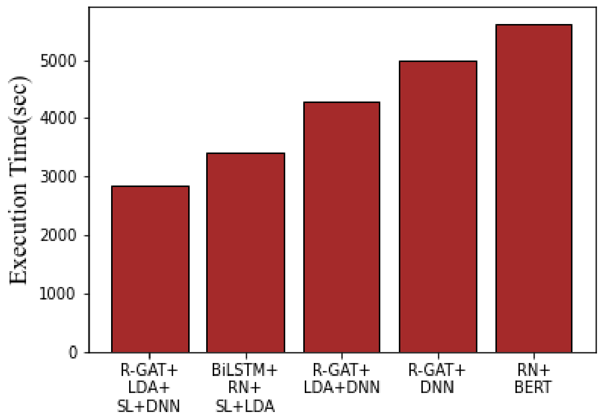

Fig 21 Evaluation of execution time for proposed RGAT+LDA+SL+DNN and existing techniques

Figure 19 illustrates the evaluation of training time for proposed RGAT+LDA+SL+DNN and existing techniques. The training time for RGAT+LDA+SL+DNN, BiLSTM+RN+SL+LDA, RGAT+LDA+DNN, RGAT+DNN, and RN+BERT are 2849.744, 3421.34, 4287.90, 4982.32 and 5623.78 in seconds. As a result, the proposed approach requires less training time than the existing model. The comparison of testing time between the proposed RGAT+LDA+SL+DNN and the existing approaches is shown in Figure 20. The running value of testing time for the proposed and existing methods are 0.931425, 1.4765, 2.5463, 3.2843, and 4.9210. The figure explains the testing time is less than the existing model. Figure 21 illustrates the execution times of the proposed and existing methods. The attained value of execution time for RGAT+LDA+SL+DNN, BiLSTM+RN+SL+LDA, RGAT+LDA+DNN, RGAT+DNN, and RN+BERT is 2850.6754, 3422.8165, 4290.4463, 4985.6043 and 5628.701in seconds. As a result, execution time is less for computing the proposed method.

5. Conclusion

Suicidal ideation can now be analyzed and detected in new ways with the help of social media platforms to express one's thoughts and feelings. A relational graph attention mechanism and DNN Classifier are designed for detecting suicidal ideation. The Twitter dataset is collected and preprocessed to remove unnecessary data. The first phase of the solution pipeline relates the features of Latent Dirichlet Allocation and sentence embedding in BERT, and the second phase relates lexicon-based semantic analysis and BERT. The results of the first and second phases are then concatenated to classify suicidal ideation. Finally, the DNN classifier is used to detect suicidal ideation by training and testing the data from the concatenated RGAT. Performance metrics like accuracy, precision, recall, and error are examined for this proposed model in comparison to existing models. The attained performance metric values for the proposed method are 88.21%, 88.22%, 88.19%, and 12%. These metric values of the proposed model outperform the existing models. The proposed model's performance will be improved in the future by analyzing mental disorders and suicidal ideation with different datasets to easily identify suicidal thoughts and types of mental disorders.

Reference

[1] Kumar, A., Trueman, T. E., & Abinesh, A. K. (2021). Suicidal risk identification in social media. Procedia Computer Science, 189, 368-373.

[2] Ji, S., Pan, S., Li, X., Cambria, E., Long, G., & Huang, Z. (2020). Suicidal ideation detection: A review of machine learning methods and applications. IEEE Transactions on Computational Social Systems, 8(1), 214-226.

[3] Heckler, W. F., de Carvalho, J. V., & Barbosa, J. L. V. (2022). Machine learning for suicidal ideation identification: A systematic literature review. Computers in Human Behavior, 128, 107095.

[4] Renjith, S., Abraham, A., Jyothi, S. B., Chandran, L., & Thomson, J. (2022). An ensemble deep learning technique for detecting suicidal ideation from posts in social media platforms. Journal of King Saud University-Computer and Information Sciences, 34(10), 9564-9575.

[5] Cao, L., Zhang, H., & Feng, L. (2020). Building and using personal knowledge graph to improve suicidal ideation detection on social media. IEEE Transactions on Multimedia, 24, 87-102.

[6] Lin, G. M., Nagamine, M., Yang, S. N., Tai, Y. M., Lin, C., & Sato, H. (2020). Machine learning based suicide ideation prediction for military personnel. IEEE journal of biomedical and health informatics, 24(7), 1907-1916.

[7] Bhardwaj, T., Gupta, P., Goyal, A., Nagpal, A., & Jha, V. (2022, June). A Review on Suicidal Ideation Detection Based on Machine Learning and Deep Learning Techniques. In 2022 IEEE World AI IoT Congress (AIIoT) (pp. 027-031). IEEE.

[8] Sarsam, S. M., Al-Samarraie, H., Alzahrani, A. I., Alnumay, W., & Smith, A. P. (2021). A lexicon-based approach to detecting suicide-related messages on Twitter. Biomedical Signal Processing and Control, 65, 102355.

[9] Swain, D., Khandelwal, A., Joshi, C., Gawas, A., Roy, P., & Zad, V. (2021). A Suicide Prediction System Based on Twitter Tweets Using Sentiment Analysis and Machine Learning. In Machine Learning and Information Processing: Proceedings of ICMLIP 2020 (pp. 45-58). Springer Singapore.

[10] Rabani, S. T., Khan, Q. R., & Khanday, A. (2021). A Novel Approach to Predict the Level of Suicidal Ideation on Social Networks Using Machine and Ensemble Learning. ICTACT J. SOFT Comput, 11(2), 2288-2293.

[11] Cao, L., Zhang, H., & Feng, L. (2020). Building and using personal knowledge graph to improve suicidal ideation detection on social media. IEEE Transactions on Multimedia, 24, 87-102.

[12] Ghosal, S., & Jain, A. (2023). Depression and Suicide Risk Detection on Social Media using fastText Embedding and XGBoost Classifier . Procedia Computer Science, 218, 1631-1639

[13] Cusick, M., Adekkanattu, P., Campion Jr, T. R., Sholle, E. T., Myers, A., Banerjee, S., & Pathak, J. (2021). Using weak supervision and deep learning to classify clinical notes for identification of current suicidal ideation. Journal of psychiatric research, 136, 95-102..

[14] Adarsh, V., Kumar, P. A., Lavanya, V., & Gangadharan, G. R. (2023). Fair and Explainable Depression Detection in Social Media . Information Processing & Management , 60(1), 103168.

[15] Tadesse, M. M., Lin, H., Xu, B., & Yang, L. (2019). Detection of depression-related posts in reddit social media forum. IEEE Access, 7, 44883-44893.

[16] Chatterjee, M., Kumar, P., Samanta, P., & Sarkar, D. (2022). Suicide ideation detection from online social media: A multi-modal feature based technique. International Journal of Information Management Data Insights, 2(2), 100103.

[17] Pota, M., Ventura, M., Fujita, H., & Esposito, M. (2021). Multilingual evaluation of preprocessing for BERT-based sentiment analysis of tweets. Expert Systems with Applications, 181, 115119.

[18] Uthirapathy, S. E., & Sandanam, D. (2023). Topic Modelling and Opinion Analysis On Climate Change Twitter Data Using LDA And BERT Model. Procedia Computer Science, 218, 908-917.

[19] Ojeda-Hernández, M., López-Rodríguez, D., & Mora, Á. (2023). Lexicon-based sentiment analysis in texts using Formal Concept Analysis. International Journal of Approximate Reasoning.

[20] Thangavel, P., & Lourdusamy, R. (2023). A lexicon-based approach for sentiment analysis of multimodal content in tweets. Multimedia Tools and Applications, 1-24.

[21] Wang, T., Liu, L., Liu, N., Zhang, H., Zhang, L., & Feng, S. (2020). A multi-label text classification method via dynamic semantic representation model and deep neural network. Applied Intelligence, 50, 2339-2351.

[22] N. Komati. (2021) Suicide and depression detection. [Online]. Available: https://www. kaggle.com/ datasets/nikhileswarkomati/suicide-watch

[23] Emma Smith, Deep Learning for Gesture Recognition and Human-Computer Interaction , Machine Learning Applications Conference Proceedings, Vol 3 2023.

[24] Paul Garcia, Ian Martin, Laura López, Sigurðsson Ólafur, Matti Virtanen. Deep Learning Models for Intelligent Tutoring Systems. Kuwait Journal of Machine Learning, 2(1). Retrieved from http://kuwaitjournals.com/index.php/kjml/article/view/167

[25] Veeraiah, V., Anand, R., Mishra, K. N., Dhabliya, D., Ajagekar, S. S., & Kanse, R. (2022). Investigating scope of energy efficient routing in adhoc network. Paper presented at the PDGC 2022 - 2022 7th International Conference on Parallel, Distributed and Grid Computing, 681-686. doi:10.1109/PDGC56933.2022.10053344 Retrieved from www.scopus.com