Journal of Harbin Engineering University
ISSN: 1006-7043

Vol 45 No. 7
July 2024

Enhanced Conditional Generative Adversarial Networks Suicidal Risk Identification from Social Networking Sites Using Text Similarity Measures

B Bhaskar Rao[#1], Chandrakant Naikodi [#2], Suresh L [#3], Sanjeevkumar Chetti[#4]
[#1]Research Scholar, Cambridge Institute of Technology, Bengaluru-560036
[#2]Professor and Chairman, Department of Studies in Computer Science, Davangere University, Davangere-577007.
[#3] Professor, and Head, Department of Information Science and Engineering, RNS Institute of Technology, Bengaluru.
[#4] Principal Director, Ministry of MSME, Government of India, Mumbai, Maharashtra, India

***Abstract*—**Publicly available social networks can be exploited for monitoring mental healthcare by employing machine learning (ML) and Artificial Intelligence (AI) methods to classify and assess the risk related with different mental health illness. By integrating text similarity measurements with Deep Learning (DL) methods, the CGANSRI model develops the accuracy of risk assessment by efficiently capturing contextual nuances. The CGANSRI can discern subtle linguistic cues that may recommend suicidal thoughts through its training on a vast corpus of social media posts. Concurrently, sophisticated measures of text similarity are employed to quantify the semantic relationships between posts, providing a comprehensive understanding of the textual content. The system efficiently investigates information provided by users to organize nuanced linguistic cues that may be associated with the risk of suicide by incorporating ATSM. This method develops the accuracy and responsiveness of organizing suicidal risk by using the combined benefits of online communications, despite the challenges posed by the dynamic and context-specific nature of such risks. Many tests and validations have been conducted using diverse datasets acquired from social networking platforms. The findings of this study designate that the CGANSRI-ATSM framework exhibits the capacity to classify suicide risk in online settings at an initial phase with enhanced accuracy and recall associated to present organizations. Our results establish that the combination of CGANSRI-ATSM Fast text yields impressive recall, precision, and accuracy measures, with caps of 85.626%, 86.524%, and 94.637%, correspondingly. This research contributes to the field of suicide prevention by using DL models and feature engineering methods to analyze social media data. By leveraging these methods, we objective to develop suicide detection and prevention efforts in the context of the widespread use of social networking media.

Keywords: Social Networking Sites, Machine Learning, Deep Learning.

1. Introduction

Worldwide, a suicide takes place once every forty seconds, as reported by the World Health Organization (WHO) [1]. Approximately 800,000 people commit suicide annually, with 20 times as many attempts as successful suicides, according to the report. However, suicide is notably underreported when compared to other causes of death. Accordingly, one million suicide deaths occur worldwide each year. The information above suggests that the leading cause of mortality for young people, especially women, is suicide. According to the well-known suicide book, there is a typical process and pattern that starts with suicidal thoughts, progresses to a suicidal attempt, and ends with an actualized suicide. Though it's not a guarantee, having suicidal thoughts increases the probability of someone trying suicide later on. Caregivers can identify the elements associated with suicidal ideation and take necessary action by closely examining the different warning indicators. The American Foundation for Suicide Prevention (AFSP) created a list of suicide risk factors and warning indicators to help those that may be in danger. In this research divide these risk factors into three categories: health variables (which include mental health and chronic pain), family history (which includes prior suicide attempts), and environmental factors (which include stress and abuse). Additional warning signs and indicators of suicide thoughts are providing by the National Institutes of Health (NIH) [2].

Social media usage has significantly increased in the past several years due to the widespread use of the Internet. Every day, billions of people use social media as a platform to voice their thoughts [3]. People share their true thoughts on social media because, in contrast to the real world, it offers a certain degree of anonymity. Simultaneously, as most social media data is accessible, it is possible to use it to detect suicide [4]. Implementing these unchanged suicide detection techniques would

prevent individual resistance and ensure that the collected data more accurately reflects real emotions [5]. Although the exact number of people who have thought about suicide is unknown, it can be inferred that a very small percentage of people have suicidal ideation. Because of this, data obtained from social media platforms frequently displays bias, as only a small percentage of users report having suicidal thoughts. Currently, deep learning or machine learning techniques are the mainstay of research on the identification of suicide using social media data. Strategies addressing this issue must be put into practice in order to preserve optimal model performance [6, 7], as directly training the classification model using imbalanced data might have a major influence. The imbalance issue in text classification is being addressed in a number of ways, including improving conventional techniques for oversampling and undersampling, modifying the model's loss function, and so on. A substantial body of research addresses issues of mental health and suicide by utilizing data from social media. In computational social science research, data from Reddit, Instagram, Tumblr, TeenLine, and Twitter have all been used frequently [8]. Since it allows users to post content reflecting mental health disorders and states of mind, such as r/Depression, r/SuicideWatch, and r/BipolarSoS, Reddit stands out as the most promising of these due to its variety, popularity as measured by the volume of content it contains, and ability to provide anonymity. Reddit content analysis can help Mental Health Professionals (MHPs) improved comprehend a patient's present condition, develop the accuracy of their diagnosis, and, if essential, recommend therapeutic options [9]. The determination of this study is to provide an overview of CGANSRI determinations, complications and possible contribution in this context of suicidal risk detection. As in this research delve deeper into the complexities of this new method, in this research will at how it overcomes the limits of existing methods while also opening up new opportunities for proactive intervention and support in mental health care. The growing field of AI applications in mental health emphasizes the significant of investigate new methods such as CGANSRI to enhance our capacities in understanding, classifying, and eventually reducing suicide risk [10]. Suicidal Risk ATSM, a method for analyzing textual content on social networking sites. This study attempts to detect small yet essential signs suggestive of suicidal ideation using advanced text similarity measures, enabling a proactive and sympathetic response to those in distress. As this research manage the difficult interplay of technology and mental health, ATSM emerges as a promising tool in the ongoing efforts to promote digital well-being and intervene when it counts the most.

1.1 Motivation of the research

- Improved Conditional Generative Adversarial Networks, which harness the power of text similarity measurements on social networking sites to empower mental health practitioners, pave the path for precise and proactive suicidal risk identification.
- In the field of digital compassion, our research aims to transform suicide prevention by utilizing cutting-edge technology. Enhanced Conditional Generative Adversarial Networks provide a ray of hope by uncovering important insights via sophisticated text similarity measurements on social networks.
- Our innovative approach uses Enhanced Conditional Generative Adversarial Networks to decipher subtle linguistic patterns on social media, providing a transformative tool for early detection of suicidal risk and fostering a safer online community.

1.2 The main contribution of the research

- To develop a new methodology for detecting suicidal risk on social networking platforms, the proposed method combines ATSM with an enhanced version of CGANSRI.
- By integrating text similarity measurements with deep learning techniques, the CGANSRI model improves the accuracy of risk assessment by effectively capturing contextual nuances. The CGANSRI can discern subtle linguistic cues that may suggest suicidal thoughts through its training on a vast corpus of social media posts.
- Finally, this strategy enhances the accuracy and responsiveness of identifying suicidal risk by utilizing the combined benefits of online communications, despite the challenges posed by the dynamic and context-specific nature of such risks. Numerous tests and validations have been conducted using diverse datasets acquired from social networking platforms.

The remaining part of this work is structured in the following manner. In Section 2, we will discuss the literature survey of suicide detection. The proposed method is described in depth in Section 3. A discussion of the results and further analysis is included in Section 4. The conclusion and future research directions are presented in Section 5.

2. Literature survey

SNS have changed interpersonal communication and information exchange in recent years. The possibility of identifying and comprehending suicidal risk on these

platforms has grown as people disclose their thoughts, emotions, and challenges online. This survey's objective is to examine the importance of using social networking sites to identify suicide risk. The evolution of suicidal risk identification via social networking sites is one of the important topics covered in this review, which offers a thorough assessment of current models. In order to shed light on the difficulties, opportunities, and future prospects for mental health computing study and practice, this study also attempts to summarize the current state of the area.

With significance on the initial identification of suicidal ideation, Tadesse et al. [11] recognized an automated technique for distinguishing suicidal signals. Their methodcreates use of ML and DLmethods to organize content on the social networking site Reddit. They use a hybrid model called Long Short Term Memory-Convolutional Neural Network (LSTM-CNN), which combines a convolutional neural network and long short-term memory, to increase efficiency. This model's effectiveness is assessed and contrasted with other classificationapproaches. Our tests display that combining word embedding methods with neural network design yields the best organizationconsequences. These outcomes highlight the potential and effectiveness of DL architectures in creating a model that successfully predicts suicide risk in a range of text categorization tasks.

Li et al. [12] developed a Deep Hierarchical Ensemble model for Suicide Detection (DHE-SD) using a hierarchical ensemble method. A Sina Weibo dataset with over 550 thousand posts from 4521 users was used to generate the model. A validation of the algorithm's efficacy was conducted using 7329 public Weibo postings. The proposed model performs better on the recognized and public datasets. They also eliminate user postings with important suicide ideation using the sentence-level mask technique to create the model more generalizable. Even when baseline models perform poorly, the proposed method can detect social media users with suicide ideation, according to experiments.

To improve a system for distinguishing suicidal ideation, Aldhyani et al. [13] described an experimental technique. This technique, which creates use of text representation methods like TF-IDF and Word2Vec, used publicly available Reddit datasets for analysis. ML and DL methods were combined in the classification procedure. Two investigation models used CNN and Bidirectional Long Short-Term Memory (BiLSTM), in addition to the XGBoostML model. These models were employed to establish social media postings by using features from the LIWC-22 and textual content to distinguish suicidal ideation- connected posts from other kinds. Numerous criteria, with accuracy, precision, recall, and F1-scores, were used to evaluate these models. Using textual features, the CNN-BiLSTM model detected suicidal thoughts with 95% accuracy, associated to 91.5% for the model. When integrating Linguistic Inquiry and Word Count (LIWC) characteristics, XGBoost performed better than CNN-BiLSTM.

In order to organize suicidal inclinations, Bernert et al. [14] assembled research articles from Web of Science, PubMed, EMBASE, and PsycInfo that used AI, ML, or Natural Language Processing (NLP). After investigative the risk of suicide attempt, most of this research concentrated on assessing the risk of suicidal ideation, suicide fatalities, and other risk-related significances. The field is still in its primary stages of examination, but it is expanding rapidly and suggestions potential for future research, according to the authors. In their investigation, Chadha et al. [15] examined the way numerousML algorithms differentiate suicidal from non-suicidal content on Twitter. Working together, the researchers designated 112 features for the study independently from a survey of doctors and patients at a mental health hospital.

Ji et al. [16] investigated techniques for identifying suicidal thoughts using ML techniques. The researchers categorized these methods and explored their applications in diverse settings, such as suicide notes, clinical interviews, social media, questionnaires, and Electronic Health Record (EHR). They indicated that a significant portion of future suicide ideation detection would probably come from online social media content. They also included a list of publicly accessible datasets and suggested lines of inquiry for further investigation.

In their comprehensive review of linear and nonlinear machine learning models in suicidology, Cox et al. [17] presented data. To identify suicidal thoughts and behaviours, they explored the potential incorporation of previous scientific findings into these models. Alongside data-driven research and theoretical frameworks, the authors emphasized that ML could significantly advance science. After scrutinizing past research, they concluded that advanced models might not always be the optimal choice. The lack of interpretability could outweigh any performance improvements, whether noticeable or not. The study's findings suggest that integrating predictive models with preventive measures represents the most efficient approach to focus on proactive services.

The presented method by Vioules et al. [18] evaluates suicidal warning indicators, identifies postings with

suicidal content, and automatically recognizes abrupt variations in user behavior.

The investigators created behavioral characteristics to gauge the degree of danger associated with a person's Twitter behavior. The two groups of behavioral features found were post-centric features and user-centric features. Sharma &Churi [19], focusing on the younger Indian population, examined interesting topics, including colorism, social comparison, and mental health in relation to Instagram usage. They applied structural equation modeling but were unable to find a important positive link among age and social concerns, even after controlling for characteristics such as time spent and frequency on Instagram. Moreover, study has recognized that social comparison can result in colorism and mental health difficulties.

Mbarek et al. [20] employed a variation of machine learning methods to address the challenge of forecasting suicide among users. They used a wide range of data that has proven useful in establishing suicidal users, including emotional, temporal, and account features. Suicide victims were employed to the test viability of their method, and the importance were consistent with predictions.

3. Proposed system

The proposed method organizations ATSM with an enhanced version of CGANSRI. By incorporating text similarity measurements with DL methods, the CGANSRI model develops the accuracy of risk assessment by proficiently apprehending contextual nuances. The CGANSRI can discern subtle linguistic cues that may recommend suicidal thoughts through its training on a massive corpus of social media posts. Simultaneously, sophisticated measures of text similarity are employed to quantify the semantic associations between posts, providing a comprehensive understanding of the textual content. The system effectively surveys information provided by users to organize nuanced linguistic cues that may be associated with the risk suicide by incorporating ATMS. Fig.1 shows the block diagram of the CGANSRI-ATSM technique.

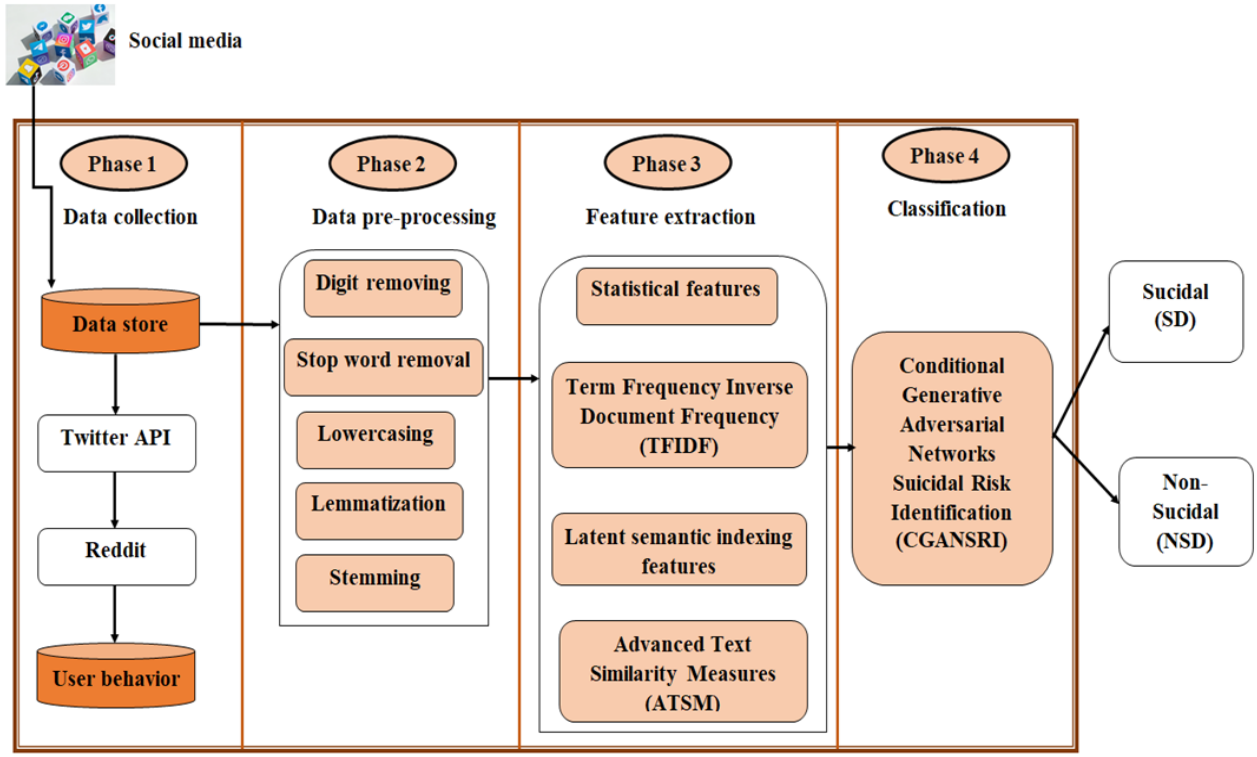

Figure 1: Proposed method ofCGANSRI-ATSM

3.1 Phase 1: Dataset and data collection

According to the suggested method, Twitter and Reddit data are used to train the algorithm for organizing symptoms of suicidal ideation [21].

Primarily, the dataset included posts from individuals seeking assistance one internet forums associated to suicide, fundamentally on subreddits dedicated to suicide discussions. These posts can be considered indicative of suicidal thoughts, given that their authors are typically individuals with such thoughts. Furthermore, normal posts from other subreddits discussing friends, family, and entertainment are also gathered. Following a manual analysis of the suicidal

posts, a set of phrases was identified and employed as Twitter search terms.

Using Twitter APIs, 188,704 English-language tweets from 2,000 people were gathered that contained these search terms. Out of these 37,800 out of 188,704 tweets, 450

individual tweets were chosen for testing purposes. By manually classifying tweets as remaining, two sets of data were produced: one for individuals who expressed suicide concepts and another for those that showed no signs of suicide. The tweets that were first associated with a particular user were selected based on the frequency with which phrases suggestive of suicide were used. Two sets of tweets were subsequently manually sorted in order to categorize users. The suicide dataset contains tweets sent by people that appear to be suffering from depression. On the other hand, the typical user dataset includes users who made no mention of personal ailments, users who shared dismal articles or remarks, and users whose tweets did not appear to express any melancholy feelings.

Over the course of three months, a basic collection of data regarding each Twitter user's profile and tweets was gathered, starting with an anchor tweet. To determine the average number of hash tags, links, replies, and @mentions in each tweet, the authors used numerical data. A tweet and its associated responses and comments comprise a social media session. Rather than combining all of a user's tweets into one collection, this methodology treats each individual tweet as a separate entry for analysis. By evaluating each tweet separately, the anticipated study aims to analyze the smallest number of tweets in order to find specific indicators of depression. Moreover, integrating tweets up to a present point is made simple by working with individual tweets. Table 1 compiles the dataset's statistics.

Table 1: Dataset Information.

Definition	**Suicidal detection (SD)**	**Non- Suicidal detection (NSD)**
Number of tweets	74125	114579
Number of twitter users	445	724
Mean number of tweets per user	166.572	158.254
Mean tweet length	142.893	120.774
Mean number of emoticons per user	30.843	42.804

3.2 Phase 2: Pre-processing

To generate a word vector for classification at this stage, it is essential to eliminate noise from textual posts before applying feature extraction and embedding methods. In this process, stop words, punctuation, lowercasing, tokenization, and lemmatization are all removed. For pre-processing the dataset, we employed the Natural Language Tool Kit (NLTK) [22] and accomplished basic operations.

- Punctuation, emoji, and numerical digit removal: This method removes characters like '?,' '!,' ':,' ';,' single quotes, and emojis to improve the text's readability.
- Stop word removal: Words that don't really add anything to the model's functioning, such as "the," "a," "an," and "in," are eliminated using this method.
- Lowercasing: All words are changed to lowercase during this process.
- Tokenization: Every sentence is separated into its component words, phrases, and other pertinent information using this method.
- Lemmatization: To obtain the basic or root form of a word, it involves merging its inflected forms.
- In order for a machine learning neural network method to differentiate among posts that are suicidal and posts that are not, every text sequence in the dataset must have a consistent real-value vector. The post-padding classification technique was applied to solve this need and overcome the difficulty.

3.3 Phase 3: Feature extraction

The initial step in collecting comprehensive user data is feature extraction, which is an essential step required to achieve precise suicide ideation detection. The following are some of the features that our model made use of:

Statistical features: Our data shows that people create posts of varying lengths. Consequently, the post's length is used as a feature in the machine learning model's training.

Term Frequency Inverse Document Frequency (TFIDF): To determine a word's significance throughout the entire corpus, utilize TFIDF. Below is a definition of TFIDF [23].

$$tfidf(wf) = freq(wf) * \log \frac{N}{|\, d \in DS : wf \in d \,|} \quad (1)$$

Let WF stand for the word feature, DS for the document set, N for the total number of postings, and d for a document.

Latent semantic indexing features:

There were some issues with the features that TF-IDF developed. As the dataset expands in size, dimensionality increases. Sparsity grows when the n-gram method is applied to the dataset. The TF-IDF method was used to generate features across five tiers, with each tier consisting of 50, 100, 150, 200, and 250 new features. Singular Value Decomposition (SVD) [24] was employed to determine semantic links among the features. The most distinctive qualities were utilized to organize these five tiers.

Advanced Text Similarity Measures (ATSM):

The *Bidirectional Encoder Representations from Transformers (*BERT)/RoBERTa output is combined by *Siamese BERT (*SBERT)to produce a fixed-sized sentence embedding. We evaluate three different pooling strategies: utilizing the CLS-token output, calculating the output vectors' mean (MEAN-strategy), and determining the output vectors' maximum over time (MAX-strategy).

The provided training data determines the network structure. The following goal functions and structures are put to the test.

The main purpose is to multiply the trainable weight $W_t \in \mathbb{R}^{3n\times k}$ by the sentence embeddings u and v after concatenating them with the element-wise difference.

$$o = soft\max(W_t(a, b \mid a-b \mid)) \quad (2)$$

Optimize the cross-entropy loss, where 'n' represents the sentence embedding size, and 'k' is the number of labels. For additional details, please refer to Fig. 2, illustrating the structure of SBERT.

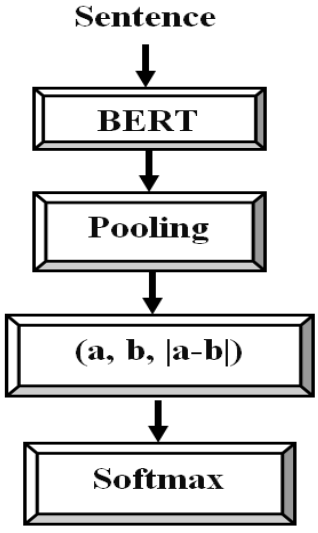

Figure 2: SBERT architecture with objective function

We used SBERT12 [25], a technique that looks for semantic similarities between two sentences, to evaluate message similarity. Document embeddings were produced using a large language model, and vector similarity was then computed using SBERT. SBERT was applied in its original form. Every pair of messages was used to calculate a similarity score for each user, and the average of these individual scores was used to establish the user's overall score. Individual message similarity scores varied from -0.214 to 1, with the first, median, and third quartiles represented by quartile values of 0.094, 0.159, and 0.232, correspondingly. The first quartile, median, and third quartile data yielded average similarity scores of 0.159, 0.178, and 0.196, respectively. These scores varied from 0.105 to 0.340

Each pair of messages was compared for comparison on three dimensions: "Topic," denoting the message's main theme; "Mindset," indicating the author's assessment of the topic as positive, neutral, or negative; and "Motivation/Intent," representing the author's underlying intentions, including needs, wants, and motivation. Three labels were available for each dimension: "Similar," "Dissimilar," and "Not enough information" (henceforth, "NEI"). These three criteria were chosen to assess more subtle markers of rumination-reflected recurrent thought patterns. To be more precise, topic similarity was used to determine whether people were concentrating on related or unrelated subjects; the former is more suggestive of ruminative thought patterns.

3.4 Phase 4: Classification

The proposed work entails using sophisticated machine learning techniques, notably cGANs, to improve the accuracy and efficiency of detecting suicidal risk from textual data on social networking sites. The use of cGANs enables conditional generation, allowing the model to produce data samples based on particular conditions [26]. The conditions in this situation would be linked to suicide risk factors obtained from text data. The addition of text similarity metrics improves the model's capacity to detect patterns and similarities in the language used by people expressing suicidal thoughts. The suggested approach aims to provide a more thorough and advanced understanding of the language linked to suicidal risk on social media by utilizing these technologies. Consequently, early detection and management to lessen possible harm are made easier. A cGAN's essential loss function seeks to optimize the maximum and minimum issues, which are comprised of discriminator and generator losses. Fig. 3 shows the architecture of cGAN.

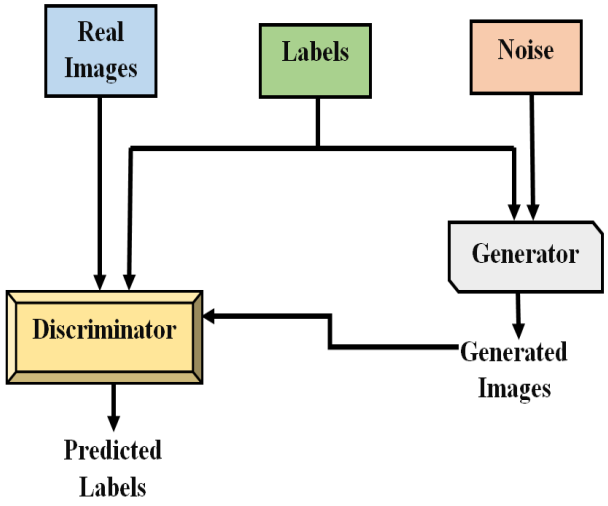

Figure 3: Architecture of CGAN method

Effectively determining if a sample's source is real or fraudulent is discriminator D's main responsibility. To ascertain this, it is necessary to assess if the expected output $D(a \mid b)$ for authentic data x is inclined toward 1, whereas the expected output $D(G(c \mid b)$ for fraudulent data x is inclined toward 0. As a result, the peak value of Eq.(3) below is reached.

$$\max_D V(D,G) = E_{a \sim p_{data}(a)} \log(D(a \mid b)) + E_{z \sim p_z(z)}[\log(1 - D(G(c \mid b)))] \quad (3)$$

The generator G's purpose is for the generated sample $G(c \mid b)$ to be recognized as actual data by the discriminator D, so that the discriminator's output $D(G(c \mid b)$ approaches 1. As a result, below Eq (4) has the minimum value

$$\min_G V(D,G) E_{z \sim p_z(z)}[\log(1 - D(G(c \mid b)))] \quad (4)$$

The discriminator D's goal is to determine whether examples are from a or $G(c \mid b)$, such that $E_{a \sim p_{data}(a)} \log(D(a \mid b))$. When this component is maximized, the discriminator D can output $D(a \mid b) = 1$ when x conforms to p_{data}. Another aspect of the problematic is that generator G wishes to fool discriminator D, resulting in $E_{z \sim p_z(z)}[\log(1 - D(G(c \mid b)$. The objective function of a CGAN is depicted in Eq (5):

$$\max_G \min_D V(D,G) = E_{a \sim p_{data}(a)} \log(D(a \mid b)) + E_{z \sim p_z(z)}[\log(1 - D(G(c \mid b)))] \quad (5)$$

Sampling from the condition variable and the noise vector at the same time is essential in the context of CGAN. Choosing an appropriate condition variable set that is in line with the generation aim is essential to improving the generator's capacity to simulate the real distribution. Directly extracting condition variables from the training data is a popular method. This makes it possible for the generator and discriminator to learn about the training set before they receive any input. As an example, using class labels as conditional variables on the adversarial network's input layer makes CGAN function as a more sophisticated unsupervised GAN, more like a weakly supervised or supervised model.

Algorithm 1: Enhanced Conditional Generative Adversarial Networks Suicidal Risk Identification

Step 1: Input-Posts from Tweets &Reddit, C_name
Step 2: Output-Sucidal (SD) and Non Sucidal (NSD)
Step 3: for i=0 from 1 to n do // n= number of post
T[i]= S_{input}[i]
Txt.csv=T[i]
End for
Step 4: T1=tokens(Txt.csv)
Step 5: T1=tokens_lowercase(T1)
Step 6: T1=tokens_remove(T1)
Step 7: T1=tokens_Stemming(T1)
Step 8: Txt2.csv=T1
Step 9: PC= tm_map ((lowercase (Txt), remove (Txt), remove (punctuation (Txt), removenum (Txt). Stemming (Txt).
Step 10: for i from 1 to n do
S_len[i] =nchar (Txt2 [i])
End for
Step 11: Token.dfm=dfm (tokens_ngram (Txt2, n=1:4))
Step 12: trim=dfm_trm (Token.dfm, min_dfreq, min_tfreq)
Step 12: TFI_feat= trim.Tfidf
Step 13: LSA_feat=SVD (TFI_feat)
Step 14: Training. Simi = BERT (LSA_Feat)
Step 15: for I from 1 to n do
ATSM[i] = mean (Training.simi[i, SD])
End for
Step 16: Classify (C_name, ATSM)
Step 17: End

4. Result and discussion

This section covers the complexities of modern techniques, evaluation criteria, and experimental design. Our recommended model is developed using the PyTorch framework, trained simultaneously on a single GPU (Nvidia Tesla V100), and prevented from overfitting by using the early halting strategy. Additionally, each model's batch size is fixed at 64, and the dropout percentage is 0.3. ReLU is the activation function used. The number of epochs for the proposed model is 5. RMSprop and Adam optimizers are used with this model. Support vector machines (SVM), long short-term

memory (LSTM), convolutional neural networks (CNNs), and random forests (RF) are some of the existing systems used in this section.

4.1 Evaluation metrics

The study utilized standard assessment criteria to ascertain how the CGANSRI-ATSM models differentiated between suicidal and non-suicidal post content. The primary focus was on analysing the results of metrics pertaining to false-positive and false-negative classifications. The evaluation considered several metrics, including F1-score, Precision, Accuracy, and Recall. The following elucidates the method by which they were determined:

False Positive (FP), True Negative (TN),True Positive (TP), and False Negative (FN) are the four values displayed by the metrics. When the classifier successfully predicts the negative class, it is a true negative; when the classification model correctly predicts the positive class, it is a true positive. On the other hand, a false positive happens when the classifier predicts the positive class incorrectly. Both true positive and true negative are desired results for a classifier; however, a false negative occurs when the classifier erroneously guesses the negative class. All values would be focused on the main diagonal in an ideal classifier, indicating that there are neither false positives (FP = 0) nor false negatives (FN = 0), leading to flawless classification.

Accuracy: Calculating the ratio of correctly predicted instances to total predictions is a commonly used metric for assessing machine learning classifier performance.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

Although accuracy is a useful metric, it does not reveal additional information about the performance of the classifier in cases where the dataset is unbalanced.

Precision: By calculating the ratio of true positives to all anticipated positives, this statistic yields the True Positive Rate (TPR).

$$\Pr ecision = \frac{TP}{TPR} \tag{7}$$

Recall: This measure establishes the proportion of true positives to total actual positives (TAP). When a high cost is linked to false-negative results, recall is desirable in order to choose the optimal model, for example. Misclassifying an at-risk individual can have serious effects on prediction, similar to suicidal risk identification.

$$\mathrm{Re}call = \frac{TP}{TAP} \tag{8}$$

F1-score: This measure is used to assess how well a model or classifier performs when it has to strike a balance between recall and precision. It is especially helpful in situations where there is a significant percentage of real negative values in the dataset and the dataset is unbalanced. Typically, false positives and false negatives play a significant part in learning models. The goal of the F1 score is to minimize the impact of actual negative values by giving these values more weight.

$$F1 - Score = 2 * \frac{\Pr ecision * \mathrm{Re}call}{\Pr ecision + \mathrm{Re}call} \tag{9}$$

4.1.1 Accuracy Analysis

Table 2: Accuracy Analysis for CGANSRI-ATSM method

Methods	**Accuracy**
SVM	82.5
CNN	86.6
LSTM	87.7
LR	87
CGANSRI-ATSM	93.6

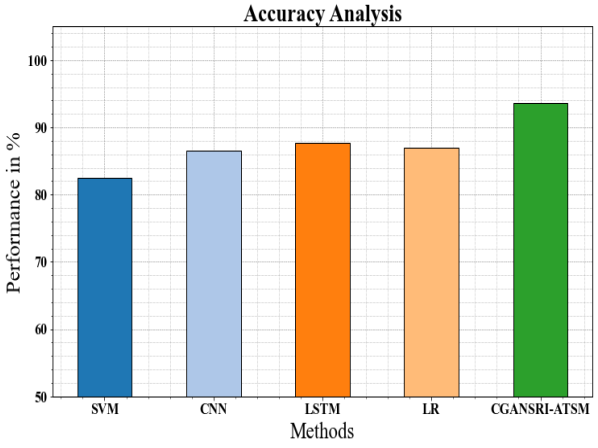

Figure 4: Accuracy Analysis for CGANSRI-ATSM method

Fig. 4 and Tab. 2 present an accuracy comparison of the CGANSRI-ATSM strategy with other current methods. The graph shows how the DL method maintains accuracy while operating more efficiently. The accuracy scores of the SVM, CNN, LSTM, and LR models are 82.5%, 86.6%, 87.7%, and 87%, correspondingly, whereas the CGANSRI-ATSM model has a 93.6% accuracy.

4.1.2 Precision Analysis

Table 3: Precision Analysis for CGANSRI-ATSM method

Methods	Precision
SVM	81.1
CNN	87.8
LSTM	90.8
LR	83
CGANSRI-ATSM	92.9

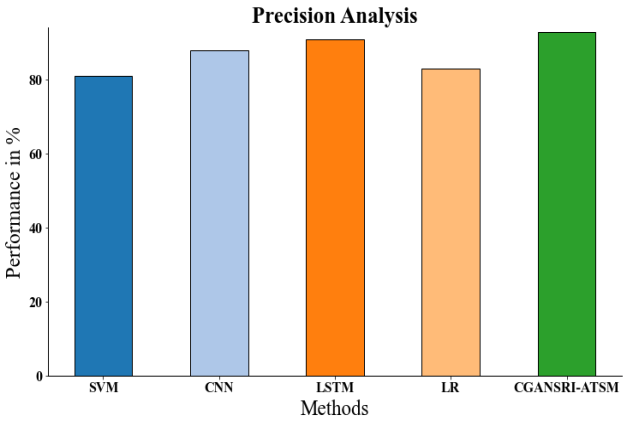

Figure 5: Precision Analysis for CGANSRI-ATSM method

Fig. 5 and Tab. 3 associate the precision of the CGANSRI-ATSM approach to other present systems. The graph displays how the DL technique maintains precision while operating more efficiently. The CGANSRI-ATSM model has a precision of 92.9%, compared to the SVM, CNN, LSTM, and LR models' precision scores of 81.1%, 87.8%, 90.8%, and 83% respectively.

4.1.3 Recall Analysis

Table 4: Recall Analysis for CGANSRI-ATSM method

Methods	Recall
SVM	84.5
CNN	89.8
LSTM	86.5
LR	82
CGANSRI-ATSM	90.3

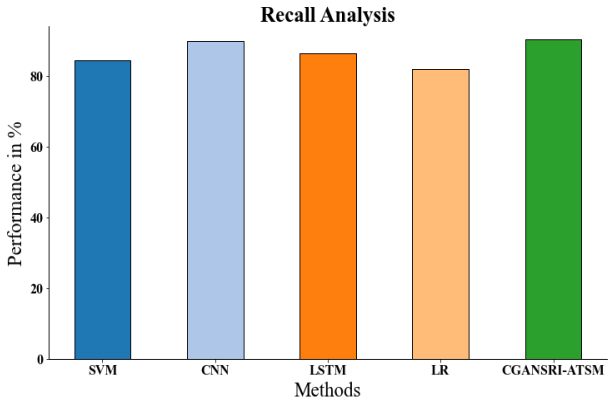

Figure 6: Recall Analysis for CGANSRI-ATSM method

Fig.6 and Tab.4 associate the recall of the CGANSRI-ATSM technique to that of numerous other present approaches. The graph shows how the DL technique maintains recall while operating more efficiently.The CGANSRI-ATSM model accomplishes a recall of 90.3%, surpassing the recall values of 84.5%, 89.8%, 86.5%, and 82% for the SVM, CNN, LSTM, and LR models, respectively.

4.1.4 F-Score Analysis

Table 5: F-Score Analysis for CGANSRI-ATSM method

Methods	F-Score
SVM	82.8
CNN	88.8
LSTM	88.6
LR	81
CGANSRI-ATSM	91.7

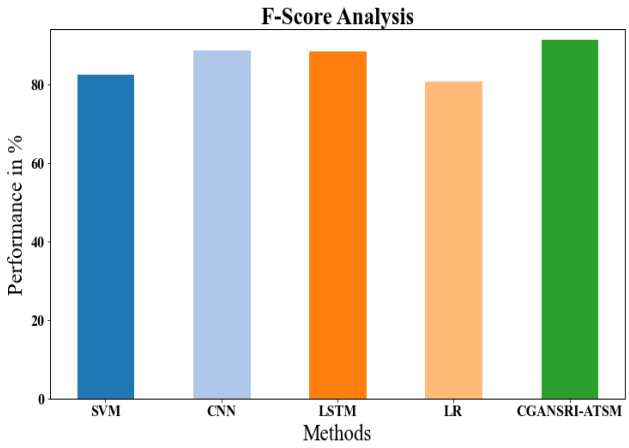

Figure 7: F-Score Analysis for CGANSRI-ATSM method.

Fig. 7 and Tab. 5 associate the F-score of the CGANSRI-ATSM technique to many other strategies currently in use. The graph shows how the DL method maintains F-score while operating more efficiently. The CGANSRI-

ATSM model has an F-score of 91.7%, which is higher than the F-score values of 82.8%, 88.8%, 88.6%, and 81% for the SVM, CNN, LSTM, and LR models respectively.

4.1.5 Execution time Analysis

Table 6: Execution time Analysis for CGANSRI-ATSM method

Methods	Execution time
SVM	179.65
CNN	213.52
LSTM	168.75
RF	245.62
CGANSRI-ATSM	136.92

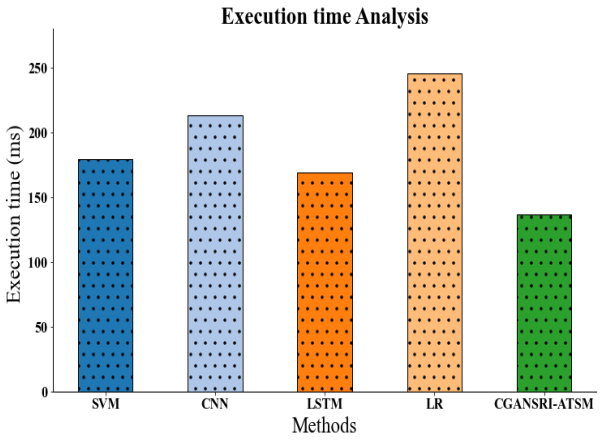

Figure 8: Execution time Analysis for CGANSRI-ATSM method

The execution times of the suggested CGANSRI-ATSM methodology and existing methods are contrasted in Tab. 6 and Fig. 8, with the CGANSRI-ATSM technique outperforming all of the other methods. For instance, the proposed CGANSRI-ATSM technique executed in a mere 136.92 ms. On the other hand, the execution times of other popular techniques, including SVM, CNN, LSTM, and RF, are 179.65 ms, 213.52 ms, 168.75 ms, and 245.62 ms, respectively.

4.2 Discussion

The enhanced CGANSRI employed in this study demonstrated promising results in the context of suicidal risk identification from social networking sites. Leveraging advanced text similarity measures, the model exhibited improved performance in accurately discerning and classifying textual patterns associated with suicidal tendencies. The findings suggest that integrating enhanced CGANs with text similarity metrics holds potential for developing effective tools to identify and address suicidal risks within online social platforms. The experimental consequences display that the suggested work achieves robustness in comparison to current approaches, with an accuracy of 94.637%, precision of 86.524%, recall of 85.626%, f-score of 92.736%, and execution time of 156.92ms. The limitation of the research is the possible partiality and imprecision in classifying suicidal risk, given the important dependence of the algorithm on text similarity metrics, which might not encompass the intricacy and milieu of personal statements on social media platforms.

5. Conclusion

Individuals have become increasingly comfortable revealing their personal thoughts on SNS due to the improved ubiquity of these platforms and the connected societal stigma. In conclusion, the obtainable technique recommendations a viable way for detecting suicidal risk on social networking sites. This novel technology uses deep learning and sophisticated text analysis to improve the accuracy and efficiency of suicide risk detection. This technique displays great promise in assisting mental health providers and support systems in understanding the essentials of language and context. For those who may be at risk, this capacity allows for faster support and action. The combination of GANs and ATSM not only creates the improvement of technology in tackling important societal challenges, but it also highlights the value of interdisciplinary cooperation in promoting mental health investigation and interventions. In the future, it will be significant to establish people who are suicidally inclined precisely, taking into account emotions such as anger and disgust. A sophisticated classifier that can distinguish these emotions will be significant in directing numerous kinds of suicidal individuals to the right services for intervention.

References

[1] Suicide,https://www.who.int/health-topics/suicide#tab=tab_1, (Accessed on 09/11/2022).

[2] Risk factors, protective factors, and warning signs — afsp, https://afsp.org/riskfactors-protective -factors-and- warning-signs/, (Accessed on 09/11/2022).

[3] Zogan, H., Razzak, I., Jameel, S., & Xu, G. (2021). Depressionnet: A novel summarization boosted deep framework for depression detection on social media. *arXiv preprint arXiv:2105.10878*.

[4] Shing, H. C., Resnik, P., &Oard, D. W. (2020, July). A prioritization model for suicidality risk assessment. In *Proceedings of the 58th annual meeting of the association for computational linguistics* (pp. 8124-

8137).https://doi.org/10.18653/v1/2020.acl-main.723

[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.https://doi.org/10.1109/TMM.2020.3046867

[6] Tang, T., Tang, X., & Yuan, T. (2020). Fine-tuning BERT for multi-label sentiment analysis in unbalanced code-switching text. *IEEE Access*, *8*, 193248-193256.https://doi.org/10.1109/ACCESS.2020.3030468

[7] Zhang, H., Wang, Y., Zhang, Z., Guan, F., Zhang, H., &Guo, Z. (2021). Artificial intelligence, social media, and suicide prevention: Principle of beneficence besides respect for autonomy. *The American Journal of Bioethics*, *21*(7), 43-45.https://doi.org/10.1080/15265161.2021.1928793

[8] Cavazos-Rehg, P. A., Krauss, M. J., Sowles, S. J., Connolly, S., Rosas, C., Bharadwaj, M., ... &Bierut, L. J. (2016). An analysis of depression, self-harm, and suicidal ideation content on Tumblr. *Crisis*.https://doi.org/10.1027/0227-5910/a000409

[9] Sowles, S. J., Krauss, M. J., Gebremedhn, L., & Cavazos-Rehg, P. A. (2017). "I feel like I've hit the bottom and have no idea what to do": Supportive social networking on Reddit for individuals with a desire to quit cannabis use. *Substance abuse*, *38*(4), 477-482.https://doi.org/10.1080/08897077.2017.1354956

[10] Chiong, R., Budhi, G. S., Dhakal, S., &Chiong, F. (2021). A textual-based featuring approach for depression detection using machine learning classifiers and social media texts. *Computers in Biology and Medicine*, *135*, 104499.https://doi.org/10.1016/j.compbiomed.2021.104499

[11] Tadesse, M. M., Lin, H., Xu, B., & Yang, L. (2019). Detection of suicide ideation in social media forums using deep learning. *Algorithms*, *13*(1), 7.https://doi.org/10.3390/a13010007

[12] Li, Z., Zhou, J., An, Z., Cheng, W., & Hu, B. (2022). Deep hierarchical ensemble model for suicide detection on imbalanced social media data. *Entropy*, *24*(4), 442.https://doi.org/10.3390/e24040442

[13] Aldhyani, T. H., Alsubari, S. N., Alshebami, A. S., Alkahtani, H., & Ahmed, Z. A. (2022). Detecting and analyzing suicidal ideation on social media using deep learning and machine learning models. *International journal of environmental research and public health*, *19*(19), 12635.https://doi.org/10.3390/ijerph191912635

[14] Bernert, R. A., Hilberg, A. M., Melia, R., Kim, J. P., Shah, N. H., &Abnousi, F. (2020). Artificial intelligence and suicide prevention: a systematic review of machine learning investigations. *International journal of environmental research and public health*, *17*(16), 5929.https://doi.org/10.3390/ijerph17165929

[15] Chadha, A., & Kaushik, B. (2022). Performance evaluation of learning models for identification of suicidal thoughts. *The Computer Journal*, *65*(1), 139-154.https://doi.org/10.1093/comjnl/bxab060

[16] Ji, S., Yu, C. P., Fung, S. F., Pan, S., & Long, G. (2018). Supervised learning for suicidal ideation detection in online user content. *Complexity*, *2018*.https://doi.org/10.1155/2018/6157249

[17] Cox, C. R., Moscardini, E. H., Cohen, A. S., & Tucker, R. P. (2020). Machine learning for suicidology: A practical review of exploratory and hypothesis-driven approaches. *Clinical psychology review*, *82*, 101940.https://doi.org/10.1016/j.cpr.2020.101940

[18] Vioules, M. J., Moulahi, B., Azé, J., &Bringay, S. (2018). Detection of suicide-related posts in Twitter data streams. *IBM Journal of Research and Development*, *62*(1), 7-1.https://doi.org/10.1147/JRD.2017.2768678

[19] Sharma, A., Sanghvi, K., &Churi, P. (2022). The impact of Instagram on young Adult's social comparison, colourism and mental health: Indian perspective. *International Journal of Information Management Data Insights*, *2*(1), 100057.https://doi.org/10.1016/j.jjimei.2022.100057

[20] Mbarek, A., Jamoussi, S., &Hamadou, A. B. (2022). An across online social networks profile building approach: Application to suicidal ideation detection. *Future Generation Computer Systems*, *133*, 171-183.https://doi.org/10.1016/j.future.2022.03.017

[21] 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.https://doi.org/10.1016/j.jjimei.2022.100103

[22] Aldhyani, T. H., Alsubari, S. N., Alshebami, A. S., Alkahtani, H., & Ahmed, Z. A. (2022). Detecting and analyzing suicidal ideation on social media using deep learning and machine learning models. *International journal of environmental research and public health*, *19*(19), 12635.https://doi.org/10.3390/ijerph191912635

[23] Rabani, S. T., Khanday, A. M. U. D., Khan, Q. R., Hajam, U. A., Imran, A. S., &Kastrati, Z. (2023). Detecting suicidality on social media: Machine learning at rescue. *Egyptian Informatics Journal*, *24*(2), 291-302.https://doi.org/10.1016/j.eij.2023.04.003

[24] Cheng, Q., Li, T. M., Kwok, C. L., Zhu, T., & Yip, P. S. (2017). Assessing suicide risk and emotional distress in Chinese social media: a text mining and machine learning study. *Journal of medical internet research*, *19*(7), e243.https://doi.org/10.2196/jmir.7276

[25] Reimers, N., &Gurevych, I. (2019). Sentence-bert: Sentence embeddings using siamesebert-networks. *arXiv preprint arXiv:1908.10084*.https://doi.org/10.18653/v1/D19-1410

[26] Wu, B., Liu, L., Yang, Y., Zheng, K., & Wang, X. (2020). Using improved conditional generative adversarial networks to detect social bots on Twitter. *IEEE Access*, *8*, 36664-36680.https://doi.org/10.1109/ACCESS.2020.2975630