[1]B Bhaskar Rao

[2]Shiva Prasad M S

[3]Ganga Shirisha M S

[4]Sanjeevkumar Chetti

[5]Suresh L

[6]Chandrakant Naikodi

Deep Belief Networks and Hybrid Term-Frequency Inverse Document Frequency (TF-IDF) Based Suicide Identification in Online Social Networks

Journal of Electrical Systems

Abstract: - With the growing popularity of Online Social Networks (OSNs) like Facebook, Instagram, Twitter and YouTube, most individuals believe that writing about their emotions and sharing their preferences and views on social media is more convenient than doing it in person. These networks are improving associated to a dissimilar of societal events, such as illness, unhappiness, and even suicide. In this scenario, the cooperative of people who have suicidal thoughts (ST) and use OSNs rapid their intentions. Our method combines Deep Belief Networks (DBN), a type of deep learning model, with Hybrid Term-Frequency Inverse Document Frequency (HTF-IDF), a popular text mining method for collecting feature vectors from textual data. The DBNs are trained utilizing TF-IDF features extracted from user postings, permitting for the detection of subtle patterns and semantic subtleties that designate suicidal intent. The HTF-IDF component assistance in highlighting significant yet unusual terms across postings, which are important users' emotional and psychological states. The project began with the collection of a dataset containing dissimilar kinds of communication from numerous main online social networks, which was then preprocessed account for the linguistic and syntactic characteristics of online communication. The proposed DBN-TF-ID model was tested against many benchmarks and outperformed existing machine learning algorithms in terms of accuracy and sensitivity in detecting probable suicidal material. To conclude, the use of DBN-HTF-IDF to examine social media material provides a viable path for developing suicide prevention tools, potentially leading to more prompt and effective interventions in mental health emergencies.

Keywords: Online Social Networks, Social Media, Suicidal Thoughts, Suicide Prevention, Deep Belief Networks, Hybrid Term-Frequency Inverse Document Frequency.

1.Introduction

A situation that seems too problematic to handle or bear is one that individuals frequently try to escape due to stressful life conditions like death, illness, loneliness, and relationship problems. These individuals therefore attempt to find ways to communicate their despair. Suicidal ideation, an early warning sign of future suicide attempts, can regrettably result from an inability to manage depression during problematic periods. People are progressively discussing their thoughts of suicide in their public profiles on online social networks (OSNs), which are becoming more popular. This trend has been expanding concomitantly. It has even occurred that people who committed suicide wrote their last thoughts on social media pages. The World Health Organization (WHO) estimates that over 800,000 people attempt suicide annually. Moreover, research displays that suicide is a leading cause of death worldwide [1]. As a result, a variation of activities and research efforts are essential to decrease suicide risks and urge people who are considering suicide to seek help. Suicide prevention opportunities have improved due to the increasing popularity and accessibility of social networks. To that end, our study will concentrate on the subject of suicide as an extremely difficult problematic that affects millions of individuals worldwide annually. To uncover suicidal users and understand their thoughts, this study purposes to identify high-risk user profiles and analyze user behavior in great depth.

[1] Research Scholar, Cambridge Institute of Technology, Bengaluru-560036

[#2,3]Research Scholar, Department of Studies in Computer Science, Davangere University- 577007 India

[#4] Principal Director, Ministry of MSME, Government of India, Mumbai, Maharashtra, India

[#5] Professor, and Head, Department of Information Science and Engineering, RNS Institute of Technology, Bengaluru.

[#6]Professor and Chairman, Department of Studies in Computer Science, Davangere University, Davangere-577007.

With numerous profiles available in many OSNs, it appears that information gleaned from a single social network is not enough for a whole user identity [2]. To create it comprehensive, more user data from other OSNs must be included to create an online persona that creates personality predictions about the user possible. This process, known as matching profiles, entails gathering data on a single person from all of their registered accounts across several data sources. Matching profiles specifically target the creation of a comprehensive and detailed profile for suicidal users to better identify them and determine whether or not they are contemplating suicide [3].

To provide prompt interventions to stop suicidal behavior, there are still insufficiently effective ways to identify prospective sufferers of suicidal ideation as soon as feasible. Researchers have recently looked at mental health issues from two primary angles. One traditional perspective uses standard scales and questionnaires to measure suicidal thoughts based on clinical contacts between patients and medical staff [4]. This method's disadvantage is that individuals frequently engage in purposeful concealment and misreporting, are hesitant or bashful to speak with a psychologist or counselor, and do not reveal their plans to commit suicide before they actually do. On the other hand, the field of suicide screening techniques is growing and is based on the retrieval and analysis of social media data. Due to the extensive use of the Internet, prior research has demonstrated that young individuals with suicidal ideation may express such ideas or look for information or assistance on social media [5]. Young people with suicidal thoughts are increasingly leaving suicide notes on social media due to the rise in popularity of platforms like Facebook and Twitter. Multiple studies have shown a correlation between psychologically rated suicide risk and the online manifestation of suicidal ideation, although the degree of similarity to physician-derived suicide risk is still undetermined [6].

Social media ideation detection may enable psychologists or public health experts to spot suicidal ideas in users and take immediate action. Psychiatrists have previously evaluated this concept on an actual dataset and provided positive comments. Chiang et al. [7] generated a rapid-response technique to assistance psychologists identify suicide ideation on social networking sites such as Facebook. This method efficiently classifies those who display indicators of suicidal ideation. Additionally, a case study highlights the fact that psychologists and clinical medical professionals are worried about the effects of Facebook and similar social media platforms. They employ linguistic indicators from user posts to detect suicidal intent and respond rapidly.

This study introduces a new method to online social network suicide detection using a HTF-IDF model and DBNs. Due to its layered architecture and capacity to learn hierarchical representations, Deep Belief Networks, a subset of deep learning models, are perfect for classifying intricate patterns in large datasets. Collective with the HTF-IDF model, this method rises the recall and precision of classifying suicidal content by weighting terms in a document according to their relevance to a corpus. Our suggested method overcomes the gap between classic text analysis methods and state-of-the-art deep learning models to progress an efficient framework for early identification of suicide intents in social media posts. Through the use of these potent computational methods, we seek to support individuals who are at risk and assistance them promptly, eventually lowering the suicide rate through preventive measures.

The major contribution of this research

- To improve a combined DBN, a kind of deep learning model, with TF-IDF, a popular text mining method for collecting feature vectors from textual data.
- The DBNs are trained using TF-IDF features extracted from user postings, permitting for the detection of subtle patterns and semantic subtleties that designate suicidal intent.
- The development began with the collection of a dataset containing dissimilar kinds of communication from numerous main online social networks, which was then preprocessed to account for the linguistic and syntactic characteristics of online communication.
- Lastly, the recommended DBN TF-IDF model was tested against many benchmarks, and it outperformed existing machine learning algorithms in terms of accuracy and sensitivity in detecting probable suicidal material.

1.1 Motivation of this research

- Using deep learning to detect suicide risk on social media can save lives by enabling timely interventions.
- Using social networks as early warning systems for mental health emergencies is the aim of our research.
- Integrating text analysis and deep learning provides a very precise and sensitive technique for suicide risk identification.

The rest of our research follows this structure: Section 2 discusses the literature review, whereas Section 3 summarizes the proposed research. Section 4 provides the datasets, evaluation criteria, and experimental results. Finally, Section 5 offers a summary of this work's conclusions.

2. Literature survey

This review of the literature purposes to provide a thorough summary of the research conducted on suicide detection in online social networks, with key conclusions, methods, and theoretical frameworks used by researchers in this emerging field. Through an important analysis of the literature, identification of information gaps, and a description of future research determinations, this survey appearance for to advance the present discussion about the role of digital platforms in suicide prevention and intervention programs.

Using machine learning and deep learning methods, Nikhileswar et al. [8] appearance for to distinguish texts showing suicidal ideation, helping in the early diagnosis of suicide. The 'Push shift' API was used to collect content from Reddit's 'teenagers' and 'Suicide Watch' subreddits. Their suggested model outperformed previous methods in accuracy by using a Fully Connected Neural Network (FCNN) for text classification into categories of suicide and non-suicide and a Universal Sentence Encoder for text encoding.

An ongoing study on the automatic identification of suicide posts is described by Tadesse et al. [9]. They inspect the application of machine learning and deep learning-based classification approaches for the initial identification of suicidal ideation on Reddit. Exactly, they use an LSTM-CNN hybrid model to assess and associate its effectiveness with other classification models. Our experiment establishes that optimal relevance classification consequences can be found by combining neural network architecture with word embedding methods.

The objective of this research, according to Chatterjee et al. [10], is to evaluate online social media in order to provide early detection of suicidal thinking. A study identified six unique feature types using a fully labeled dataset of suicide-related Reddit and Twitter posts. These included both online social media behavior and clinical signs of suicidal inclinations. Researchers established a multimodal method for categorizing suicidal thoughts reported on social media by using these feature clusters. With an accuracy rate of 87%, the Logistic Regression classifier recognized impressive performance connected to other approaches.

Using publicly accessible Reddit datasets, Aldhyani et al. [11] offer an experimental technique to increasing a system to categorize suicidal thoughts. For text demonstration, TF-IDF and Word2Vec methods are used, and for classification tasks, deep learning and machine learning algorithms are used. The researchers classified social media posts as either not suggestive or suggestive of suicidal inclinations in two dissimilar trials. These concerns combine complex machine learning models such as Convolutional Neural Networks (CNN), Bidirectional Long Short-Term Memory (BiLSTM) models, and the XGBoost algorithm with textual data and features extracted from LIWC-22 analysis. These methods performance is measured with the use of metrics including F1-scores, recall, accuracy, and precision. The CNN–BiLSTM model recognized suicidal thoughts using textual features with 95% accuracy, associated to the XGBoost model's 91.5% accuracy. XGBoost overtook CNN-BiLSTM while using LIWC features.

A novel technique that automatically categorizes suicidal users by examining their OSN profile establishments was introduced by Mbarek et al. [12]. Our contribution focuses on categorizing suicidal individuals by investigative their shared content across OSNs and considering profiles from frequent data sources. They develop a comprehensive profile that aids in predicting high-risk suicidal behavior by utilizing several feature types extracted from user posting data. To differentiate among suicidal and non-suicidal features, they use supervised machine learning methods. This study, based on a dataset of suicide deaths, establishes that patterns in user profiles from multiple data sources can reveal suicidal tendencies.

In order to measure social media data for the precise and efficient detection of individuals experiencing suicidal thoughts, Chadha et al. [13] developed an efficient learning model. They extensively optimized the hyper parameters of our proposed ACL model through grid search optimization, including long short-term memory, convolution, and attention components. Experiments showed that the finely tuned ACL model, enhanced by Glove embeddings, outperformed the state-of-the-art benchmarks with impressive metrics: 88.48% accuracy, 87.36% precision, 90.82% F1 score, 79.23% specificity, and 94.94% recall.

A novel ensemble learning method called LSTM-Attention-BiTCN was presented by Choi et al. [14]. It syndicates a self-attention mechanism with LSTM and BiTCN models to detect suicidal inclinations in social media posts. With impressive metrics, the LSTM-Attention-BiTCN model established its advantage over baseline methods in categorizing and recognizing suicidal ideation. It originate an F1-score of 0.9405, recall of 0.9424, precision of 0.9385, and accuracy of 0.9405.As part of an effort to lower the suicide rate, our recommended method can assistance medical professionals precisely classify suicidal inclinations between social media users.

To track and evaluate the occurrence of suicidal expressions on digital platforms, Cao et al. [15] advanced a sophisticated knowledge graph in addition to deep learning approaches. This technique uses a dual-layer attention approach to highlight crucial indications and the explicit mental procedures that drive suicide ideation in individuals. By applying a content analysis technique to Reddit and other micro-blogging platforms, they discovered that personality traits and user-generated content were the most influential factors among six domains that could designate suicidal ideation: text analysis, imagery analysis, emotional distress, depressive experiences, personality traits, and personal experiences.

In order to detect indications of suicidal thoughts in unstructured Electronic Health Records (EHRs), Cusick et al. [16] regarded at the application of a crudely trained algorithm. They annotated training and validation notes for natural language processing using a rule-based framework. The CNN, SVM, and LR models were trained using these annotations as the basis. This method should be additional examined in clinical records as it has great potential to progress suicide prevention efforts in clinical information systems.

Li et al. [17], for instance, developed a Deep Hierarchical Ensemble model called DHE-SD exactly for suicide detection. This model was trained on a SinaWeibo dataset with about 550 thousand posts composed by 4,521 users using a hierarchical ensemble technique. They tested its effectiveness using 7,329 public Weibo postings. On both the generated and public datasets, the recommended model performs better. To additional progress the model's generalizability, they additionally employ the sentence-level mask method to remove user postings that exhibit strong suicidal thoughts. Experiments show that the suggested technique may detect suicidal ideation in social media users even when baseline models perform inadequately.

The influence of stigma and support for suicide in the progression of college students' suicidal ideation to actual suicide attempts was examined by Lyu et al. [18]. The average age of the 944 Chinese college students in the study was 20.97 years. The researchers assessed suicidal thoughts, attempts, attitudes toward suicide, and the stigma associated with suicide. The Suicidal Behaviors Questionnaire-Revised (SBQ-R), the Suicidal Ideation Attributes Scale (SIDAS), and the Stigma of Suicide Scale (SOSS) were all used in the study.An investigation of the links between suicide attempts, stigma, and acceptability of suicide revealed that suicidal ideation mediated some of those associations.

2.1 Limitations of existing systems

- Accurately recognizing actual suicide risks is a challenge for current methods. Because human emotions are complicated and online language can be subtle, false positives and negatives are prevalent.
- Significant privacy concerns can arise while monitoring suicidal behaviours in internet activity. Ethical and legal issues could arise if users feel that their personal space is being violated.
- Social media content is very dynamic and context-dependent. Algorithms may struggle to keep up with shifting trends, terminology, and communication styles.

3. Proposed System

In this section, we describe a DBN, a type of deep learning model, with TF-IDF, a popular text mining technique for collecting feature vectors from textual data. The DBNs are trained using TF-IDF features extracted from user postings, allowing for the detection of subtle patterns and semantic subtleties that indicate suicidal intent. The TF-IDF component aids in highlighting important yet uncommon terms across postings, which are critical in determining users' emotional and psychological states. Fig.1 shows the block diagram of DBN-HTF-IDF method.

3.1 Dataset

Since data is necessary for training our classifiers, the analytical procedure starts with data collection. However, identifying suicidal ideation remains an enormous obstacle in the absence of a publicly available dataset. Because mental diseases and suicidal thoughts are stigmatized in society, it has historically been difficult to obtain statistics on them. On the other hand, an increasing number of people are using the Internet to talk about mental health difficulties, vent their frustrations, and look for assistance. Because Twitter has been established to be useful in evaluating mental health issues, including suicidal ideation, we used it as our main source of data.

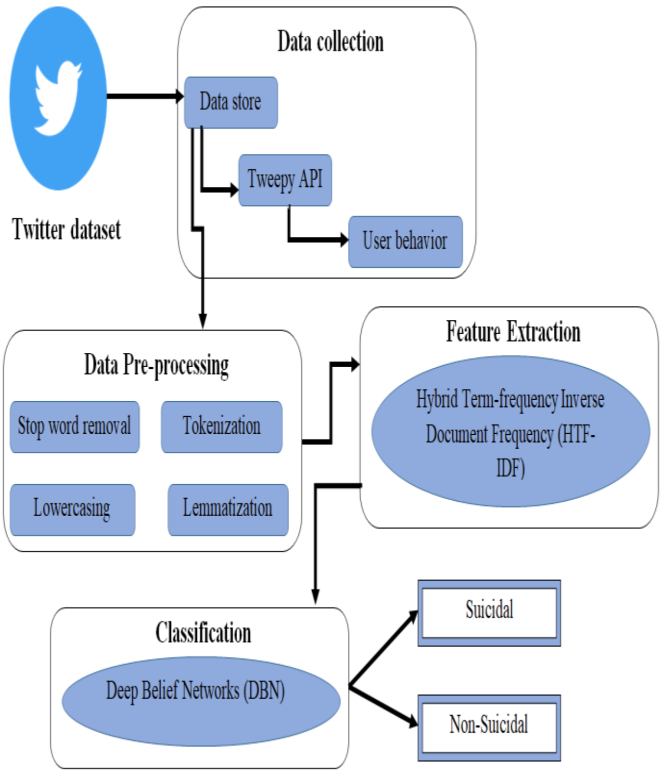


Figure 1: Block diagram of DBN-HTF-IDF method

Gathering various kinds of messages related to suicide, aside from those that explicitly convey suicidal thoughts, is the major goal here. No direct quotes from any data or identifiable information are presented to protect the privacy of the persons in the dataset. A unique ID is generated in place of the user's personal data to protect their privacy.

We may regulate historical tweets with tokenized terms due to Tweepy, an API for extracting tweets. Generating a list of suicide-related terms to query Twitter data was our objective. We used a two-stage process to accomplish this. Primary, we inspected a large number of tweets about suicide to become acquainted with phrases expressive of suicidal ideation, including "want to die" and "kill myself." Our objective was to compile regularly used expressions that designate suicide ideation or intent. We then reviewed pertinent research papers on suicide, which provided us with significant information about suicidal behavior. Integrating knowledge from these two approaches, we generated a comprehensive list of Twitter terms connected to self-harm or suicide. Utilizing a curated keyword list, we collected tweets in real-time via the Tweepy API [19]. We then meticulously reviewed these tweets, adding, deleting, and modifying terms from our list as needed. Concluding that the majority of the tweets we gathered were related to suicide, we deemed the approach a success. After roughly two to three weeks, the result was the compilation of an all-inclusive list of suicide-related key phrases. These phrases were found in 65,516 tweets retrieved between February 20, 2021, and May 13, 2021. The following are comprehensive lexicons pertaining to suicide:

Emotional upheaval, hopelessness, desire for eternal rest, thoughts of ending it all, dealing with depression, feeling overwhelmed, crying uncontrollably, desiring release, inner conflict, self-loathing, and the weight on my shoulders are all part of my unhappy existence.

3.2 Data Annotation

The tweets were collected without any indication of their sentiment. While gathering tweets using a phrase, for instance, it was unclear if the tweet was meant to raise awareness or prevent suicide, whether the user was discussing suicide-related topics like ways to end one's own life, whether the tweet mirrored the suicide of a third party, or whether the tweet used suicide as a plot point. Despite the fact that a number of the collected tweets contained suicidal language, it's possible that they were talking about a suicide movie or advertisement that avoided suggesting suicide. Annotating the gathered tweets was done in two phases. To measure sentiment polarity (positive, neutral, or negative), we first applied TextBlob, a Python tool. Then, we applied VADER for additional annotation. We next manually examined and edited the labeled tweets using Table 1's annotation rules. Out of a total of 65,516 tweets, 24,458 of them, or around 38%, contained suicide sentiments in our Twitter dataset. Predictive performance was poor due to the imbalanced classes in the training sample, especially the minority group (those indicating suicidal thoughts). The dataset's unequal class distribution was the main cause of this. In order to solve this problem, we eliminated 16,338 tweets that were not suicidal, which significantly reduced the imbalance and increased prediction accuracy. Ultimately, out of 49,178 tweets in our dataset, 50.3% and 49.7% of the tweets were classified as suicidal or non-suicidal.

3.3 Preprocessing

This step is used to reduce noise in text messages before using feature extraction and embedding methods to create word vectors for classification. It includes tokenization, lemmatization, converting text to lowercase, stop word removal, and punctuation removal.

Table 1: Rules for Data Annotation.

Label	**Examples**	**Rule**
Suicidal	‘I have nothing more to live for’ "I sometimes hate my life" "I want to kill myself. I despise both humanity and the earth. ‘I want to shoot myself’	Suicidal thoughts expressed Possibility of suicide thoughts
Non- Suicidal	‘"He wanted to die, but his partner wanted to live. Please keep an eye on him. His life is worthless." 'Please end my back ache, I'm tired.' 'My boyfriend's phone is slowly dying; this would be ideal for him.	Discussing suicide Irrelevant to suicide

The Natural Language Toolkit (NLTK) [20] was utilized to carry out fundamental preprocessing tasks on the dataset.

- Eliminating punctuation, emoticons, and numeric digits: This method creates the text easier to interpret by removing characters like "?", "!", ":", ";", "'", and emoticons.
- Stop word removal: Words like "the," "a," "an," and "in" that are unconnected to the functioning of the model are removed using this method.
- Lowercasing: This is the procedure of creating all words lowercase.
- Tokenization: This method entails breaking down a phrase or sentence into its individual words, phrases, and other parts.
- Lemmatization: With this method, inflected word forms are combined with their root form.
- Post-padding sequence method: A deep learning neural network must allocate an equal real-value vector to each text sequence in the dataset with the objective to differentiate among suicidal and non-suicidal statements. This is achieved by applying the post-padding sequence method.

3.4 Hybrid Term-frequency Inverse Document Frequency

The approach known as HTF-IDF, or Hybrid -Term Frequency-Inverse Document Frequency, is frequently used in text classification models to help them retrieve information and understand the subtleties of real language. [21].This statistical technique is especially useful for assessing a text's pattern significance. The initial constituent, HTF, ascertains the frequency of particular terms to determine their resemblance, as demonstrated below:

$$HTF(wo)_d = \frac{n_{wo}(d)}{|d|} \qquad (1)$$

One document, d ∈ D, is indicated by the set D, which also corresponds to a collection of documents. Words and phrases are the fundamental units of any document. The frequency of a word in a document is indicated by the variable $n_{wo}(d)$. Therefore, we calculate the size of document D using the formula below:

$$|d| = \sum_{wo \in d} n_{wo}(d) \qquad (2)$$

The frequency of a term's occurrence in the text is indicated by equation (2). Using TF-IDF's IDF (Inverse Document Frequency) component, one can find the number of documents in a text corpus that include a given phrase. This is the way it operates:

$$IDF(wo)_d = 1 + \log\left(\frac{|D|}{\{d : D \mid wo \in d\}|}\right) \qquad (3)$$

The following formula can be used to calculate the TF-IDF for the word "w" in document "d" within corpus "D":

$$HTF - IDF = TF(wo)_d \times IDF(wo)_D \qquad (4)$$

Using a document-term matrix, TF-IDF generally supports a variety of text classification methods.

3.5 Deep Belief Networks

The Deep Belief Network is an advanced generative model that uses several levels of processing to reveal complex patterns and abstractions in data. It is made up of multiple stacked Restricted Boltzmann Machines (RBMs) that have all been extensively trained on different datasets. An unsupervised phase lays the basis for comprehensive learning and precedes the supervised training stage for the RBMs in a DBN.

An embedded Restricted Boltzmann Machine (RBM) in a Deep Belief Network (DBN) usually makes use of the 'visible' and 'hidden' processing layers. The data's observable features are represented by the visible layer, and latent or hidden representations are captured and interpreted by the hidden layer. An RBM's units that are in the similar layer are not directly associated to one another. Moderately, the procedures of making and reconstruction are made possible by the interconnectivity among layers [22].

Each RBM is repeatedly trained layer-by-layer during the training of a DBN. The visible layer of the subsequent RBM is created by activations from the hidden layer of the previous RBM. Iteratively, this process is carried out until all RBMs have received training. This strategy, known as layer-wise pretraining, overcomes the issues that arise when the network is initialized with random, untrained link weights. Using unsupervised learning approaches, such as probability models, makes it simpler to train generative stochastic neural networks.

Fig.2 depicts an RBM network with two separate processing layers, "visible" and "hidden." While there are no direct connections between units within a single layer, the interconnections between levels help with both production and restoration operations. The network's visible layer (v) is made up of multiple hidden entities $(h_1, h_2, ..., h_j)$ and observable entities $(v_1, v_2, ..., v_i)$ which are trained on the provided unlabeled pattern structures. Nodes in the network that are not visible have binary values and help to rebuild the hidden (h) patterns by receiving data from visible nodes.

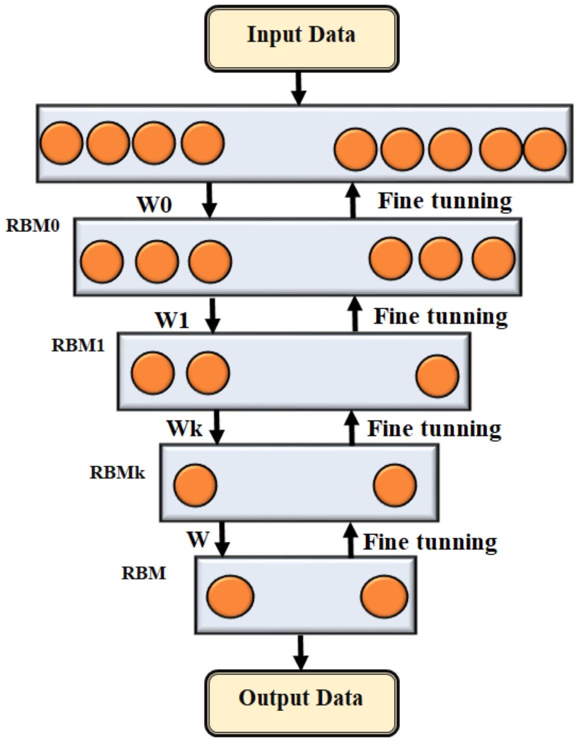


Figure 2: The architecture of RBM

A symmetrical two-way matrix of weight (S_{ij}) controls the interaction between all evident nodes and the prior biases (q_i) and (p_j).

$$K(v,h) = \sum_{i \in vis} \frac{(v_i - q_i)^2}{2\sigma_i^2} - \sum_{j \in hid} p_j h_j - \sum_{ij} \frac{v_i}{\sigma_i} h_j . S_{ij} \tag{5}$$

Here, the symbol σ signifies the distribution of Gaussian noise in the i-th visible dimension.

Having both visible and hidden units with Gaussian distributions can make learning journey management more challenging. To ensure stability, the coefficients supervising the quadratic 'containment' elements, which limit activities within permissible ranges, are computed using the standard deviations of expected noise levels. Equation (6) illustrates the particular form that the energy function takes as a result.

$$K(v,h)=\sum_{i\in vis}\frac{\left(v_i-q_i\right)^2}{2\sigma_i^2}+\sum_{j\in hid}\frac{\left(h_j-p_j\right)^2}{2\sigma_i^2}-\sum_{ij}\frac{v_ih_j}{\sigma_i\sigma_j}.S_{ij} \tag{6}$$

These predictions were visually represented, and the probabilities for the hidden units were estimated using the training set data (7).

$$C\left(h_j=1\right)=l\left(p_j+\sum v_iw_{ij}\right) \tag{7}$$

At the observable level, we can infer the latent variable v using only an h sample. Step 8 of the Gibbs algorithm requires acquiring a new set of hidden activations, indicated as h'.

$$C\left(v_i=1\right)=l\left(q_i+\sum h_jS_{ij}\right) \tag{8}$$

The application of h to v from outside the system was critical in resolving this (negative phase). Suggested Adjustments to the Weight Matrix Law (9).

$$\Delta Sij=\gamma\left(\left(v_i\bullet h_j\right)_{data}-\left(v_i^{'}\bullet h_j^{'}\right)_{model}\right) \tag{9}$$

Where the assumed learning speed is γ. Equations (10) and (11), where $(\bullet)$ is a logistic activation function, should have q_i and hi changed, respectively.

$$q=q+l\left(v-v^{'}\right) \tag{10}$$

$$p=p+l\left(h-h^{'}\right) \tag{11}$$

$$\varnothing\left(x\right)=\frac{1}{1+e^{-x}} \tag{12}$$

Finally, the logistic activation function that has been applied to each processing node is explained and shown (12). This function applies the logistic transformation to an input value (x) in order to squash the output between 0 and 1.

4. Result and Discussion

Using the Tesla A100 GPU and 84 GB RAM provided by the Google Colab Notebook service, we established and assessed all baseline approaches for our experimental configuration. We used a variety of techniques, such as state-of-the-art deep learning techniques, ensemble models with attention mechanisms included, and traditional machine learning algorithms. Python 3.10 was used for our development and testing. We used Keras, an open-source toolkit with a lot of strength, to design and train our deep learning models. We used the Natural Language library (NLTK), a powerful Python library designed for natural language processing (NLP) activities.

4.1 Data Analysis Results

We evaluated language variations by determining the frequency of suicidal thoughts through analyzing the complete pre-processed text sample. We looked at the quantity of tweets with and without suicidal themes appeared. We looked into the relationships between suicidal thoughts and the top 200 phrases in each category using Python's WordCloud visualization tool. The top 200 single terms from both non-suicidal and suicidal tweets are highlighted in Figure 3, which is a WordCloud representation of the dataset.

Twitter users identified for expressing suicidal intent frequently use terms like "overwhelmed," "exhausted," "despair," "alone," "can't go on," and "worthless," combined with negative statements such as "don't care," "never again," and "nothing matters." Additionally, terms directly associated with death, such as "death," "want to die," and "kill," also reveal the user's suicidal thoughts. Conversely, unigrams found in posts that are not suicidal primarily convey good experiences, attitudes, and feelings. They include the following: "I'm happy," "I crave enjoyment," "laugh out loud," and "experience beauty."Furthermore, phrases like "get better" and "job" imply a tendency to interact with people or have a positive outlook.

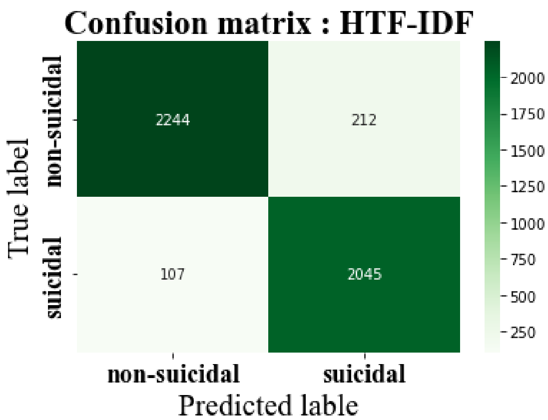
a).

b)

Figure 3: Suicidal vs non-suicidal tweets in a WordCloud

The confusion matrix is among the most illuminating and simple techniques for assessing the precision and comprehensiveness of a deep learning system. Its main application is in classification tasks where multiple class forms may be included in the result. The confusion matrix of the DBN model, which was tested with 9,302 tweets, shows (see Figure 4a) that it is more accurate at predicting non-suicidal tweets than suicidal ones. On the other together, looking at the HTF-IDF confusion matrix (Figure 4b) from a different angle, this model performs better than

the non-suicidal ones in identifying suicidal tweets in a test dataset with 4,608 tweets.



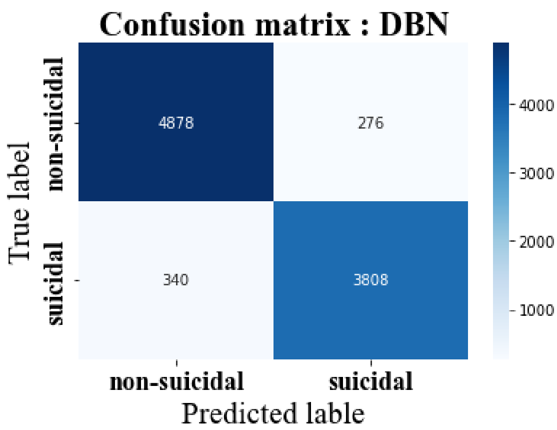


Figure 4: Confusion Matrix of the best performing model of (a) DBN and (b) HTF-IDF.

4.2 Evaluation Metrics

The objective is to use the DBN-HTF-IDF model in conjunction with basic methods to classify social media posts for indications of thoughts of suicide. Equations (13)–(16) provide the specifics of how we assess the performance of these models using common classification measures, such as accuracy, precision, recall, and the F1-score.

$$\mathrm{Pr}\,ecision = \frac{Truel_{positive}}{Ture_{positive} + False_{positive}} \quad (13)$$

$$\mathrm{Re}\,call = \frac{Truel_{positive}}{Ture_{positive} + False_{negative}} \quad (14)$$

$$F1 - score = 2 \times \frac{\mathrm{Re}\,call \times \mathrm{Pr}\,ecision}{\mathrm{Re}\,call + \mathrm{Pr}\,ecision} \quad (15)$$

$$Accuracy = \frac{Ture_{positive} + True_{negative}}{Total\ \ predictions} \quad (16)$$

4.2.1 The proposed model's performance in comparison to existing systems [23]

Table 2: Comparative Analysis of Proposed Model with Existing Systems

Methods	**Precision**	**Recall**	**F1-Score**	**Accuracy**
LSTM	76.12	77.52	75.73	75.94
LSTM-CNN	75.46	75.34	76.47	76.23
BiLSTM	76.23	75.61	75.92	76.10
CNN	63.44	86.49	73.19	68.43
LSTM-Attention-BiTCN	93.85	94.24	94.05	94.05
Our Model	94.23	95.16	95.65	96.38

In Fig. 5 and Table 2, the precision of the DBN-HTF-IDF technique is compared to other frequently used approaches. The graph illustrates the precise performance improvement that results from the deep learning approach. In contrast, the precision values of the LSTM, LSTM-CNN, BiLSTM, CNN, and LSTM-Attention-BiTCN models are 76.12%, 75.46%, 76.23%, 63.44%, and 93.85%, respectively, while the DBN-HTF-IDF model achieves 94.23%.

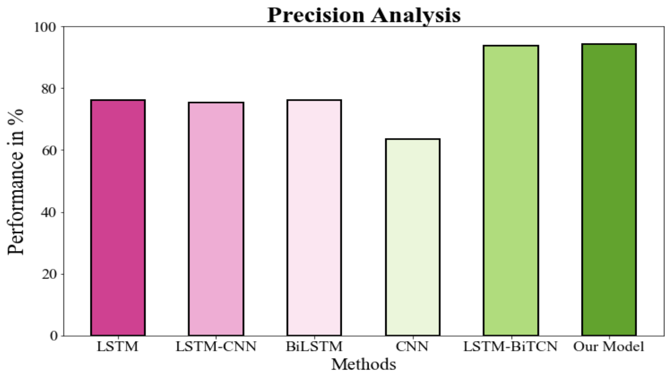


Figure 5: Precision Analysis for DBN-HTF-IDF method

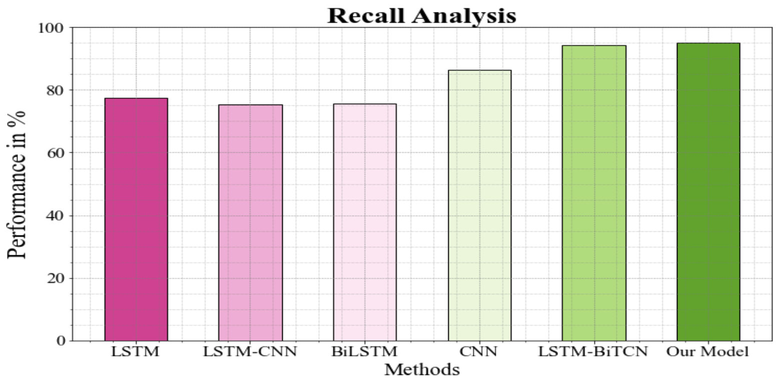


Figure 6: Recall Analysis for DBN-HTF-IDF method

In Fig. 6 and Table 2, the recall of the DBN-HTF-IDF method is compared to other frequently used approaches. The graph displays how the deep learning strategy outperforms in recall. For example, the DBN-HTF-IDF model has a recall value of 95.16%, whereas the LSTM, LSTM-CNN, BiLSTM, CNN, and LSTM-Attention-BiTCN models have recall values of 77.52%, 75.34%, 75.61%, 86.49%, and 94.24%, respectively.

The F1-scores of the DBN-HTF-IDF approach are compared to those of other commonly used methods in Fig. 7 and Table 2. The graph demonstrates the way F1-scores are used to demonstrate that the deep learning method enhances performance. For instance, the DBN-HTF-IDF model has an F1-score of 95.65%, whereas the LSTM, LSTM-CNN, BiLSTM, CNN, and LSTM-Attention-BiTCN models have F1-scores of 75.73%, 76.47%, 75.92%, 73.19%, and 94.05%, respectively.

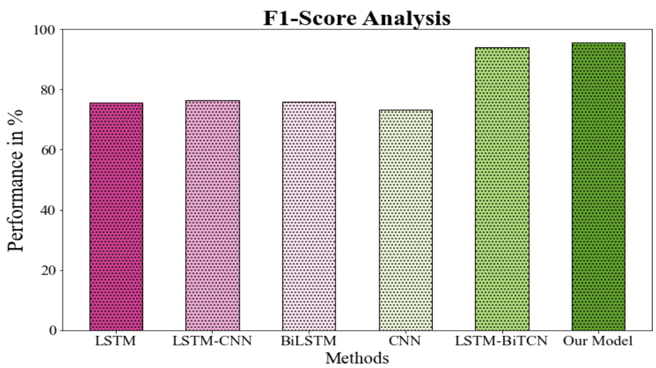


Figure 7: F1-Score Analysis for DBN-HTF-IDF method

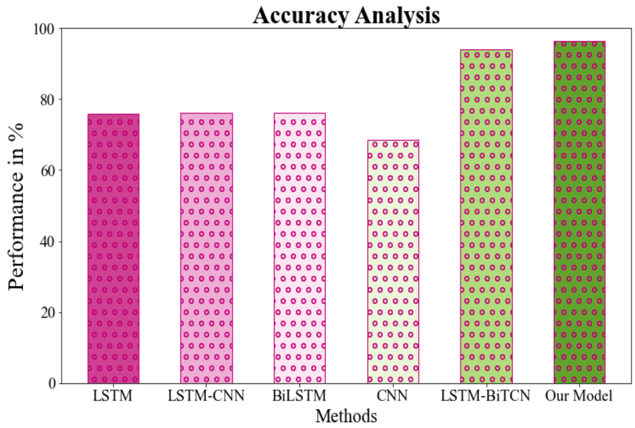
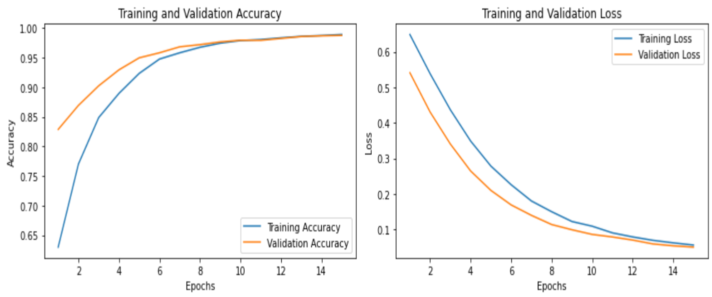


Figure 8: Accuracy Analysis for DBN-HTF-IDF method

In Fig. 8 and Table 2, the accuracy of the DBN-HTF-IDF technique is compared to other frequently used methods. The graph displays how the deep learning technique progresses in performance with accuracy. For example, the DBN-HTF-IDF model has an F1-score of 96.38%, while the LSTM, LSTM-CNN, BiLSTM, CNN, and LSTM-Attention-BiTCN models have accuracy values of 75.94%, 76.23%, 76.10%, 68.43%, and 94.05%, respectively.

Figures 9a and 9b display the validation accuracy and loss for each epoch of our DBN-HTF-IDF model. Interestingly, we see an increasing trend in validation accuracy and a decreasing trend in loss as the number of epoch's increases. Because we used a model checkpoint to keep the best-performing model, our selection method allowed us to use the model from the 11th epoch for additional testing on the test dataset. Furthermore, we applied the ReduceLROn Plateau method to modify the learning rate to decrease over fitting. Notably, validation loss and accuracy show stability after the twelfth epoch.



(a) (b)

Figure 9: Performance comparisons of DBN-HTF-IDF models during training and validation. (a) Training and validation accuracy; (b) Training and validation loss per iteration.

4.3 Discussion

An effective method for identifying suicide in online social networks is to use HTF-IDF in combination with DBNs. Our research shows how well this hybrid methodology works to identify suicidal language patterns within the large volume of social media data. Through the use of HTF-IDF's nuanced text representation and DBNs' hierarchical feature learning capabilities, our model is able to detect at-risk individuals with improved sensitivity and specificity. The findings provide important new information for future study and intervention tactics by demonstrating the potential of cutting-edge machine learning tools to support attempts to prevent suicide in the context of online social networks.

4.4 Ablation study

In the suggested model, every module is essential. In this section, we will appearance at the logic behind the proposed DBN-HTF-IDF method and other methods such as LSTM, LSTM-CNN, BiLSTM, CNN, and LSTM-Attention-BiTCN through a series of ablation tests displayed on the Twitter dataset. To inspect the performance development and found the motivations of our proposed DBN-HTF-IDF, dissimilar features extracted by the proposed technique are additional step by step.

4.4.1 Influence of DBN

In the field of suicide identification inside online social networks, DBNs have become a powerful tool. Using deep learning algorithms, DBNs are able to sort through huge volumes of social media data and classify subtleties and patterns that point to suicidal conduct. By detecting alarming messages or signals that could then go ignored, this technology permits for major action and support, which could potentially save lives. Additionally, DBNs are always learning and increasing, augmenting their accuracy over time and keeping up with the changing face of online communication. Therefore, the incorporation of DBNs into initiatives intended at preventing suicide marks a notable development in the use of technology to address mental health problems in the digital age. Lastly, the DBN-HTF-IDF model accomplished a superior performance of 96.38% for our input data. In assessment, the existing LSTM, LSTM-CNN, BiLSTM, CNN, and LSTM-Attention-BiTCN found accuracy performances of 75.94%, 76.23%, 76.10%, 68.43%, and 94.05% correspondingly.

Table 3: 10-fold cross validation of LSTM-AOA analysis.

K-folds	DBN-HTF-IDF Accuracy
1-Fold	0.98
2-Fold	0.95
3-Fold	0.95
4-Fold	0.98
Fold-5	0.97
Fold-6	0.97
Fold-7	0.93
Fold-8	0.98
Fold-9	0.95
Fold-10	0.97
10-Fold Mean	**0.96**

4.4.2 Influence of the K-fold cross validation:

The resilience and reliability of models in Deep Belief Networks and Hybrid TF-IDF based suicide diagnosis inside online social networks are importantly enhanced by the use of 10-fold cross-validation. Cross-validation is a statistical method wherein the dataset is separated into ten segments, or subsets, of which one is selected for validation and nine are used for training determinations.

Table 4: A contrast of the proposed model to new validation analysis strategies.

Models	**Evaluation methods**	**Accuracy (%)**
MNB	10-fold cross validation	83.15
LR	10-fold cross validation	91.76
SGD	10-fold cross validation	91.49
RF	10-fold cross validation	93.38
SVC	10-fold cross validation	91.67
Our proposed model	10-fold cross validation	96.38

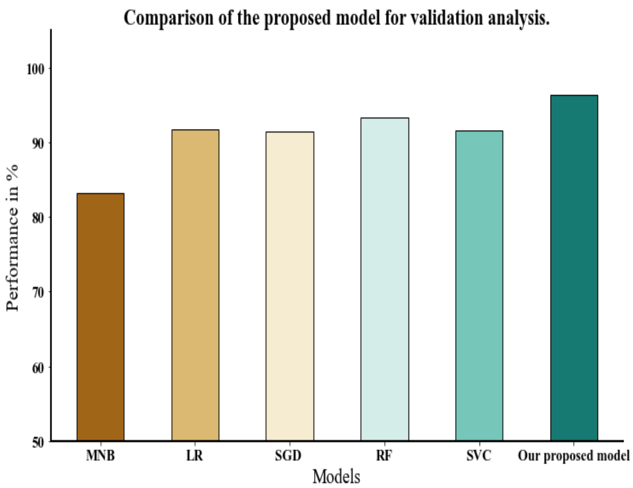


Figure 10: A comparison of the proposed model to innovative validation analysis techniques.

The proposed model DBN-HTF-IDF achieved superior performance of 96.38% for our input data by applying 10-fold cross-validation as shown in Table 5 and Figure 10. In assessment, the existing MNB, LR, SGD, RF, and SVC found accuracy performance of 83.15%, 91.76%, 91.49%, 93.38%, and 91.67%, correspondingly, as displayed in table 3 and table 4.

Table 5: A comparison of the suggested model's performance with various existing techniques.

Author	**Dataset description**	**Model**	**Results**
Tadesse et al.[24]	There were 3549 suicide posts and 3652 nonsuicidal ones.	LSTM	92%
Singh et al.[25]	3549 suicidal postings and 3652 non-suicidal	LSTM-CNN	93%
Alada ˘g et al.[26]	785 suicidal posts and 785 nonsuicidal	SVM	92%
Our propose model	There are 24,458 suicide posts and 24720 non-suicidal	DBN-HTF-IDF	96.38%

5. Conclusion

In conclusion, a useful framework for classifying suicidal tendencies in online social networks is the hybrid TF-IDF and DBNs combination. The hybrid TF-IDF progresses the procedure of feature extraction by capturing the meaning of terms in the context of the full corpus. The combined method creates use of the deep learning capabilities of DBNs to model difficult patterns in textual data. Since of this synergy, a more sophisticated and exact detection system is produced, one that may classify minor signs of suicidal ideation that conventional methods could miss. Experiments show that this hybrid model is more effective than standalone approaches in terms of precision, recall, and total accuracy. According to this technology, researchers and mental health professionals dedicated to suicide prevention can quickly intervene and gain valuable insights from massive volumes of social media data. We aim to develop in two directions in the future. To develop accuracy and F1-score, we wish to train our framework/model on balanced ternary labelled data. Since multi-classification enables us to grasp the true context and subject of the tweet, it is important. Furthermore, research can employ depression severity ratings to evaluate the disorder's severity.

References

[1] 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

[2] Mbarek, A., Jamoussi, S., &BenHamadou, A. (2020). Tuser 3: A profile matching based algorithm across three heterogeneous social networks. In *Advanced Data Mining and Applications: 16th International Conference, ADMA 2020, Foshan, China, November 12–14, 2020, Proceedings 16* (pp. 191-206). Springer International Publishing.https://doi.org/10.1007/978-3-030-65390-3_16

[3] Rabani, S.T., Khan, Q.R. and Khanday, A.M.U.D., 2021. Quantifying suicidal ideation on social media using machine learning: A critical review. *Iraqi Journal of Science*, pp.4092-4100.https://doi.org/10.24996/ijs.2021.62.11.29

[4] Nesi, J., Burke, T. A., Bettis, A. H., Kudinova, A. Y., Thompson, E. C., MacPherson, H. A., ... & Liu, R. T. (2021). Social media use and self-injurious thoughts and behaviors: A systematic review and meta-analysis. *Clinical psychology review*, *87*, 102038.https://doi.org/10.1016/j.cpr.2021.102038

[5] Fahey, R. A., Boo, J., & Ueda, M. (2020). Covariance in diurnal patterns of suicide-related expressions on Twitter and recorded suicide deaths. *Social Science & Medicine*, *253*, 112960.https://doi.org/10.1016/j.socscimed.2020.112960

[6] 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.https://doi.org/10.1109/TCSS.2020.3021467

[7] Chiang, W. C., Cheng, P. H., Su, M. J., Chen, H. S., Wu, S. W., & Lin, J. K. (2011, June). Socio-health with personal mental health records: suicidal-tendency observation system on Facebook for Taiwanese adolescents and young adults. In *2011 IEEE 13th International Conference on e-Health Networking, Applications and Services* (pp. 46-51). IEEE.

[8] Nikhileswar, K., Vishal, D., Sphoorthi, L., &Fathimabi, S. (2021, October). Suicide ideation detection in social media forums. In *2021 2nd International Conference on Smart Electronics and Communication (ICOSEC)* (pp. 1741-1747). IEEE.https://doi.org/10.1109/ICOSEC51865.2021.9591887

[9] 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

[10] 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

[11] 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

[12] 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

[13] Chadha, A., &Kaushik, B. (2022). A hybrid deep learning model using grid search and cross-validation for effective classification and prediction of suicidal ideation from social network data. *New Generation Computing*, *40*(4), 889-914.https://doi.org/10.1007/s00354-022-00191-1

[14] Choi, H. S., & Yang, J. (2024). Innovative Use of Self-Attention-Based Ensemble Deep Learning for Suicide Risk Detection in Social Media Posts. *Applied Sciences*, *14*(2), 893.https://doi.org/10.3390/app14020893

[15] 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

[16] 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.https://doi.org/10.1016/j.jpsychires.2021.01.052

[17] 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

[18] Lyu, S., & Li, Y. (2023). The roles of endorsement and stigma in suicidal ideation and behavior among Chinese college students. *International journal of environmental research and public health*, *20*(1), 877.https://doi.org/10.3390/ijerph20010877

[19] Hswen, Y., Naslund, J. A., Brownstein, J. S., & Hawkins, J. B. (2018). Monitoring online discussions about suicide among Twitter users with schizophrenia: exploratory study. *JMIR mental health*, *5*(4), e11483.https://doi.org/10.2196/11483

[20] 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

[21] 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

[22] Alqahtani, N., Alam, S., Aqeel, I., Shuaib, M., Mohsen Khormi, I., Khan, S. B., &Malibari, A. A. (2023). Deep belief networks (DBN) with IoT-based alzheimer's disease detection and classification. *Applied Sciences*, *13*(13), 7833.https://doi.org/10.3390/app13137833

[23] Choi, H. S., & Yang, J. (2024). Innovative Use of Self-Attention-Based Ensemble Deep Learning for Suicide Risk Detection in Social Media Posts. *Applied Sciences*, *14*(2), 893.https://doi.org/10.3390/app14020893

[24] 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

[25] Singh, C., Imam, T., Wibowo, S., &Grandhi, S. (2022). A deep learning approach for sentiment analysis of COVID-19 reviews. *Applied Sciences*, *12*(8), 3709.https://doi.org/10.3390/app12083709

[26] Aladağ, A. E., Muderrisoglu, S., Akbas, N. B., Zahmacioglu, O., &Bingol, H. O. (2018). Detecting suicidal ideation on forums: proof-of-concept study. *Journal of medical Internet research*, *20*(6), e9840.https://doi.org/10.2196/jmir.9840