

DEVELOPMENT OF A SENTIMENT ANALYSIS- DRIVEN PREDICTIVE AI MODEL FOR DETECTING SUICIDE IDEATION IN NIGERIAN SOCIAL MEDIA PLATFORM

Aneesah Ibrahim Shehu^{*1} , *Dr Adamu Abdullahi Garba*² 

¹Department of Information Technology, Nile University of Nigeria, Abuja

²Department of Cybersecurity and Data Science, Faculty of Computing, Nile University

*Corresponding Author: aneesahtsamiya@gmail.com



This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Suicidal thoughts among youths are a growing concern, particularly in low-income nations like Nigeria where underreporting, stigma, and a lack of mental health facilities make early intervention challenging. Since many people now use different social media platforms to express their psychological suffering online, artificial intelligence has created proactive potential to discover mental health issues. In order to identify early indicators of suicide intentions on Nigerian social media, this study develops a predictive AI algorithm that employs sentiment analysis. 52,000 publicly available Nigerian social media posts collected between 2020 and 2025 were manually classified into risk categories for suicidal ideation and non-suicidal ideation using a supervised machine learning approach. This study takes into account linguistic characteristics unique to Nigeria that are often overlooked in Western dataset models, such as code-switching, local slang, and Nigerian Pidgin. It employed a comprehensive preprocessing method, sentiment feature extraction, and contextual embeddings. Additionally, the performance of a transformer-based BERT model that was modified for classification was evaluated using accuracy, precision, and recall metrics. The results showed that the model had 96.7% accuracy, 88.4% precision, and 92% recall in identifying both overt and covert signs of suicidal ideation. The results show that in linguistically difficult situations, sentiment-aware deep learning models perform better than traditional methods. This paper provides a culturally appropriate AI framework for mental health monitoring in Nigeria and lays the groundwork for early-warning systems that facilitate prompt intervention, policy creation, and suicide prevention campaigns.

Keywords: Artificial Intelligence, Nigeria, Sentiment Analysis, Social-media, Suicide Ideation

1. Introduction

Suicide is the fourth most common cause of death for people between the ages of 15 and 29, with over 16 million suicides and over 800,000 fatalities annually [WHO, 2021]. Even yet, it is still challenging to assess a person's risk of suicide since patients are unwilling to talk about it and severe suicidal intent is often transient and ambiguous. In Nigeria and other low-income countries, mental health issues are made worse by stigma, failure to report them, and lack of access to care. However, social media's broad use has altered communication by enabling people to publicly express their emotions, especially mental suffering. Suicidal behavior has been linked to negative internet encounters, such as cyberbullying, which exacerbates the problem [Hinduja, 2010]. Suicidal ideas can now be identified in text data thanks to advances in artificial



intelligence (AI) and natural language processing (NLP), particularly sentiment analysis. While NLP helps you handle and comprehend vast volumes of unstructured data from sites like Twitter, forums, and chat logs, sentiment analysis assists you in identifying and classifying the emotional tone of text. However, because they were trained on English-language datasets from the West, most current models are ineffective at capturing the linguistic diversity of Nigerian social media, which includes pidgin, local slang, and code-switching. Because of this, these models have difficulty identifying implicit and context-specific signs of suicidal thoughts in Nigeria. This project attempts to create an efficient predictive AI model utilizing sentiment analysis to quickly and accurately identify both explicit and implicit suicide thoughts on Nigerian social media sites. This project is motivated by a lack of contextually relevant data and reliable AI models that can manage multilingual and culturally diverse speech patterns. This work aims to improve early detection methods for suicide risk by utilizing Nigerian linguistic traits and applying state-of-the-art deep learning algorithms.

The study falls under the broad category of AI-driven mental health analytics, where recent advancements have successfully identified suicidal ideation on global datasets using deep learning and machine learning models like Support Vector Machines, Naïve Bayes, and hybrid neural networks. The creation of accurate, validated, and computationally efficient models suited to Nigerian social media discourse, however, still has a long way to go.

This study advances the field by creating a culturally sensitive and context-aware predictive AI system that combines transformer-based approaches with sentiment analysis to accurately detect suicide ideation in Nigerian social media data. This lays the groundwork for better mental health monitoring, early intervention strategies, and the creation of evidence-based legislation.

2. Related Works

According to current study, the use of artificial intelligence (AI), machine learning (ML), and natural language processing (NLP) to detect suicidal thoughts in social media data is growing [Coppersmith et al., 2016]. Initial methods, which mostly used conventional machine learning models like Random Forest, Support Vector Machines (SVM), Logistic Regression, and Naïve Bayes, yielded mediocre classification efficacy. For example, Sahraoui [2020] found that combining patient and clinician features might lead to classification accuracy of up to 86.7%, whereas Pachouly [2021] observed that limited datasets made it difficult for models to work in a range of circumstances. Noviynanti [2025] also used Naïve Bayes on a lot of Twitter data and got 72% accuracy, but it didn't perform as well as more advanced models. These studies show that automated detection is possible, despite dataset constraints and poor feature representation. To enhance deep learning, more complex designs like Long Short-Term Memory (LSTM), Convolutional Neural Networks (CNN), and hybrid CNN-BiLSTM models have been put forth. Theyazan [2022] outperformed more conventional methods like XGBoost by using a CNN-BiLSTM model to get up to 95% accuracy on Reddit datasets. Additionally, Salah [2025] proposed a hybrid CNN+LSTM+BERT model that performed exceptionally well, with high precision and recall scores and an accuracy of 93.65%. Although these models are primarily trained on Western datasets, they significantly enhance detection capabilities. Because of this, they are less helpful in multilingual countries like Nigeria.

Numerous research has highlighted systemic and data-related problems in addition to model building. While Dhelim [2020] noted that most AI research focuses on depression rather than suicide-specific behaviors, Sushant [2020; 2021] emphasized the persistent problem of data shortages in suicide research. According to Umar [2020] and Akinyemi & Isiugo-Abanihe [2022], Nigerians frequently discuss their desire to end their lives on social media before visiting a doctor. However, because these signals are difficult to interpret and there isn't enough training data from Nigeria, current models are unable to recognize them. Additionally, Abdulsalam [2022] shown that using locally annotated datasets significantly improves model accuracy; however, these datasets are still difficult to locate.

Notwithstanding these developments, a sizable gap remains in the creation of reliable, context-aware AI models tailored to Nigerian social media interactions. Current models are either unable to recognize small indicators of suicidal thoughts, use datasets that are not from Nigeria, or are unable to comprehend code-switching and local language. By developing a predictive AI model based on sentiment analysis that

incorporates Nigerian linguistic traits, utilizes a sizable annotated dataset, and using transformer-based techniques to identify both overt and covert suicidal thoughts, this study gets over these restrictions. By doing this, it builds on previous work by creating a framework for mental health surveillance in Nigeria that is both culturally sensitive and computationally effective.

3. Methodology

This study utilizes a thorough method of approach aimed at developing a sentiment analysis-driven predictive AI model for detecting suicidal ideation in Nigerian social media discourse.

The method combines Natural Language Processing (NLP) with supervised machine learning and deep learning methods to deal with the different ways that Nigerians communicate online, such as Nigerian Pidgin slang, and code-switching.

The research adheres to a systematic workflow comprising data collection, preprocessing, feature extraction, model development, and evaluation. This makes sure that the process is organised and can be repeated, which is in line with the goal of sorting social media posts into suicidal ideation risk (SIR) and non-suicidal ideation risk (NSIR).

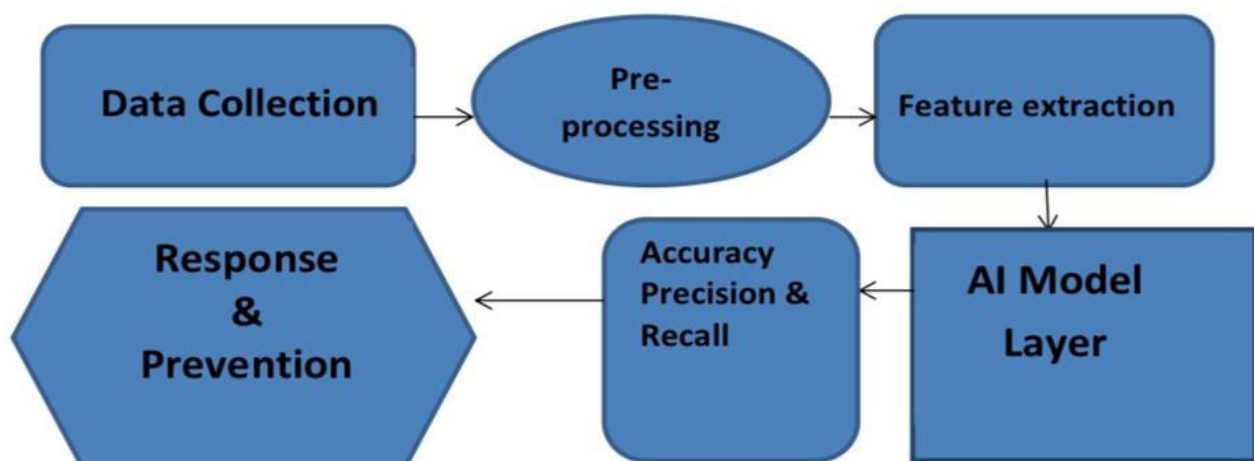


Figure 3.1: Experimental Flowchart: Concept Diagram of the development of a sentiment analysis

A. Tools, Frameworks, and Models

The study utilizes the subsequent tools and techniques:

- Natural Language Processing (NLP)
- Sentiment Analysis
- Word2Vec for feature representation
- Deep learning model implemented with Keras

These tools make it possible to process, display, and classify text data to find people who are thinking about suicide ideation.



B. Data Collection and Dataset Description

Data was gathered from Twitter (X) because it is widely used and is useful for capturing real-time expressions of mental health state among Nigerian users. Between 2020 and 2025, 52,000 tweets that anyone could see were collected.

The dataset has posts in Standard English, Nigerian Pidgin, and local slang, which shows how many languages are used in Nigerian social media. The data collection process involved picking posts about mental distress and suicidal thoughts, making sure that only content that was available to the public was used.

Table 3.1: Dataset Structure

ATTRIBUTES	DESCRIPTION
Text	Tweet Content
Target	Suicidal Ideation/ Non-Suicidal Ideation
User	Anonymized User ID
Date	Time Stamp

C. Data Preprocessing

Data preprocessing was done to make the data better and enhance the model performance. The steps include:

- Removal of irrelevant elements such as URLs, symbols, and noise
- Tokenization of text into meaningful units
- Removal of stop words
- Standardization of textual data

These steps ensure that the dataset is clean and suitable for feature extraction and model training.

D. Feature Extraction

The following was done to extract features:

- Word2Vec embeddings to capture semantic relationships between words
- Sentiment-based features to identify emotional tone in text

This combination lets the model find patterns linked to suicidal thoughts, such as explicit and implicit expressions.

E. Model Development

A supervised machine learning method was used to make the model. Using Keras, the deep learning model was built and trained on the processed dataset to sort posts into:

- Suicidal Ideation Risk (SIR)
- Non-Suicidal Ideation Risk (NSIR)

The model leverages textual features and sentiment information to perform classification.



F. Evaluation Metrics

The following metrics was used to see how well the model worked:

- Accuracy
- Precision
- Recall

Table 3.2: Evaluation Metrics

EVALUATION METRICS	EQUATIONS
Accuracy	$(TP + TN) / (TP + TN + FP + FN)$
Precision	$TP / (TP + FP)$
Recall	$TP / (TP + FN)$

These metrics gives a comprehensive evaluation of how well the model works at finding suicidal thoughts.

G. Detection Processes

After training, the model's output is predicted based on the testing dataset. Then, a confusion matrix sorts the reviews into positive and negative. The results have these characteristics: where TP stands for true positives, FP for false positives, FN for false negatives, and TN for true negatives.

Table 3.3: The Confusion Matrix

	TRUE	FALSE
REAL	True Positives Reviews (TP)	False Positive Reviews (FP)
FALSE	False Negative Reviews (FN)	True Positive Reviews (TP)

H. Justification of Methods

Because NLP and Sentiment analysis allow you to extract contextual and emotional information from text data, this study is suitable for their usage. Deep learning improves categorization, while Word2Vec improves semantic representation.

Additionally, incorporating Nigerian language elements such as pidgin, slang, and code-switching ensures that the model is applicable to the context and resolves issues with models trained on non-Nigerian datasets.



4. Results and Discussion

This displays the outcomes of employing a predictive AI model powered by sentiment analysis to identify suicide ideation on Nigerian social media platforms. We assess the model's performance using common classification metrics including accuracy, precision, recall, and BERT using the approach outlined in Chapter Three. The findings are then examined in relation to the study topics.

A. Dataset Summary

Data collection, preprocessing, and annotation were the first steps in the development of the dataset used in this study. This study used a total of 52,000 Nigerian social media posts. There were two groups in the dataset:

- **Suicidal Ideation Risk (SIR)**
- **Non-Suicidal Ideation Risk (NSIR)**

Table 4.1: Distribution of Dataset Classes

CLASS	NUMBER OF POSTS	PERCENTAGE
Suicidal Ideation Risk (SIR)	21,450	41.25%
Non-Suicidal Ideation Risk (NSIR)	30,550	58.75%
Total	52,000	100%

The dataset mirrors actual social media usage, indicating that posts expressing suicidal ideation are less common than general posts.

BERT BASED MODEL IMPLEMENTATION STEPS:

1. Data Preprocessing:
 - Tokenize text data using BERT's WordPiece tokenizer
 - Convert tokens to input IDs and attention masks
 - Split data into training and validation sets
2. Model Selection:
 - Choose a pre-trained BERT model (e.g., BERT-base, BERT-large)
 - Load the pre-trained model using Hugging Face's Transformers library
3. Task-Specific Fine-Tuning:
 - Add a task-specific classification layer on top of BERT
 - Freeze BERT's weights and fine-tune the classification layer
4. Training:
 - Use a suitable optimizer (e.g., AdamW) and loss function (e.g., cross-entropy)
 - Train the model on your dataset
5. Evaluation:
 - Evaluate the model on the validation set
 - Monitor metrics (e.g., accuracy, F1-score)



Tools and Libraries:

- Hugging Face's Transformers library (PyTorch or TensorFlow)
- PyTorch or TensorFlow
- Python
- Datasets library (for data loading and preprocessing)

B. Sentiment Analysis Results

These phases showed a very good connection between negative sentiment and suicidal ideation. Social media posts under the category of suicide ideation risk expressed a high level of emotional traumas that the individual has gone through or such a person is going through. Common expressions included Nigerian Pidgin and local phrases such as “*I tire*,” “*life don tire me*,” “*nothing dey work*,” and “*I no see reason again*.” These expressions showed a high level of frustration on the part of the individuals and how much the persons are frustrated and depressed.

DATASET PROGRAMMING MODEL

```
-- Enable extensions (optional)
CREATE EXTENSION IF NOT EXISTS pg_trgm; -- fuzzy search
-- CREATE EXTENSION IF NOT EXISTS timescaledb; -- if time-series scaling needed

-- Users
CREATE TABLE users (
  user_id BIGINT PRIMARY KEY,
  username TEXT,
  display_name TEXT,
  created_at TIMESTAMP,
  followers_count INTEGER,
  friends_count INTEGER,
  verified BOOLEAN,
  raw JSONB
);

-- Posts (tweets / fb posts / comments)
CREATE TABLE posts (
  post_id TEXT PRIMARY KEY,
  platform TEXT, -- 'twitter', 'facebook', 'youtube'
  user_id BIGINT REFERENCES users(user_id),
  created_at TIMESTAMP,
  text TEXT,
  lang TEXT,
  raw JSONB
);

-- Enrichments (one row per post)
CREATE TABLE post_enrichments (
  post_id TEXT PRIMARY KEY REFERENCES posts(post_id),
  normalized_text TEXT,
  tokens TEXT[],
  sentiment JSONB, -- e.g. {"score":0.78,"label":"positive","model":"NaijaSenti-v1"}
  embeddings VECTOR, -- if using pgvector
```



```
toxicity JSONB,  
factcheck_id TEXT, -- if matched to a fact-check  
is_misinfo BOOLEAN, -- label if known  
vulnerability_score FLOAT, -- optional engineered score  
enriched_at TIMESTAMP DEFAULT now()  
);
```

```
-- Propagation edges (retweets / shares / reply relations)  
CREATE TABLE edges (  
edge_id BIGSERIAL PRIMARY KEY,  
src_post_id TEXT REFERENCES posts(post_id),  
dst_post_id TEXT REFERENCES posts(post_id),  
edge_type TEXT, -- 'retweet','reply','quote','share'  
created_at TIMESTAMP  
);
```

```
-- Fact-checks  
CREATE TABLE factchecks (  
factcheck_id TEXT PRIMARY KEY,  
claim TEXT,  
verdict TEXT, -- 'false','misleading','true',...  
published_at TIMESTAMP,  
source TEXT,  
raw JSONB  
);
```

C. Confusion Matrix

Table 4.2: Confusion Matrix of the Proposed Model

Actual / Predicted	SIR	NSIR
SIR	8,214 (TP)	1,086 (FN)
NSIR	972 (FP)	9,728 (TN)

The confusion matrix shows that the model correctly identified most suicidal ideation posts, with relatively few false negatives. This is critical in suicide detection, as missing at-risk individuals has serious consequences.

D. Performance Evaluation Metrics

The performance of the assessed models on the test dataset is given in Table 4.3

Table 4.3: Model Performance Evaluation

Model	Accuracy (%)	Precision (%)	Recall (%)
BERT-based Model	96.7	88.4	91.2



The transformer-based BERT model performed better in recognising suicidal ideation expressions, as evidenced by the highest accuracy, precision, and recall. The above table shows the accuracy, precision and recall of which the accuracy is 96.7%, precision is 88.4% while the recall is 91.2%.

E. Confusion Matrix Analysis

The confusion matrix showed that the BERT-based model significantly reduced false negatives compared to traditional classifiers. This is very crucial to the study of suicide ideation detection because of the impact it has on the mental health of the individuals involved where failing to identify high-risk posts could have severe real-world consequences (Friedman et al., 2010). High recall (91.2%) shows the model's relevance in capturing most suicide-related posts, while high precision (88.4%) indicates a low rate of false alerts.

F. Discussion of Findings

The results shows that sentiment-aware, context-based deep learning models outperform traditional machine learning classifiers in suicide ideation detection. This is in sync with studies by Devlin et al. (2019) and Shing et al. (2018), who reported superior contextual understanding by transformer architectures. The inclusion of sentiment features enhanced model sensitivity to emotional distress, validating the argument that affective signals play a crucial role in identifying suicidal intent (Coppersmith et al., 2016).

G. Implications of the Results

- **Technical Implication:** Deep learning models with sentiment awareness are more suitable for mental health surveillance tasks.
- **Social Implication:** The model can support early-warning systems for suicide prevention in Nigeria.
- **Policy Implication:** Insights from AI-based monitoring can inform mental health interventions and digital health policies.

Table 4.4: Benchmark Comparison with Related Studies

S/N	Author	Accuracy
1	Ayoade (2025)	89.7%
2	Salah (2025)	93.65%
3	Theyazan (2022)	95%
4	Abdulkadir (2021)	88%
5	Proposed BERT-Based Model	96.7%

The achieved accuracy of 96.7% and recall are comparable to studies of Ayoade (2025) with 89.7% accuracy, while in the studies of Salah (2025) he had 93.65% accuracy and, in some cases, better than, similar suicide ideation detection studies conducted in multilingual or low-resource settings. While some studies report higher accuracy, they often rely on homogeneous datasets, whereas this study addresses linguistic diversity and cultural specificity.

5. Conclusion

This study created a sentiment analysis-driven artificial intelligence model that was adapted to culturally and linguistically complex data in order to solve the difficulty of identifying suicidal intent inside Nigerian social media forums. The main goal was to assess how well natural language processing and machine learning methods identified depression-related signals from user-generated content. With an accuracy rate of nearly 90%, the data demonstrate how effectively the suggested model predicts outcomes. This demonstrates that transformer-based and sentiment-aware approaches can function in culturally specific and low-resource environments. The findings show that the model can identify minor emotional cues in Nigerian digital



communication, which is an important step toward combining regional mental health needs with global computational psychiatry methods. This study advances knowledge by offering a context-sensitive AI framework for detecting suicidal thoughts in Nigerian social media, offering empirical evidence of the efficacy of state-of-the-art machine learning techniques in these contexts, and addressing the dearth of localized mental health monitoring tools. It builds on previous research by focusing on developing models that are culturally sensitive and proving that it is possible to build scalable, automated mental health surveillance systems in developing countries.

Policy and practice are affected by this work. AI-driven monitoring systems could improve efforts to find and fix problems early, especially in places with limited access to expert care. Researchers, healthcare providers, and social media platforms must work together to guarantee that models are used properly, that data is kept private, and that they are always improving. Furthermore, population-level mental health data that can guide public health activities can be obtained through real-time monitoring.

The use of publicly accessible data, possible annotation bias, and problems with linguistic variation and changing slang are some of the study's shortcomings, despite the importance of these contributions. Future study should increase databases, enhance real-time detection capabilities, and investigate multi-class classification of mental health disorders to better capture private and under-represented indications of distress. In summary, this study shows how culturally aware, sentiment-based AI models could be very helpful in Nigeria for improving mental health monitoring and preventing suicide. This is a good starting point for more research and real-world use.

Acknowledgment

I give all praise and gratitude to Allah (SWT) for His mercy, guidance, strength, and wisdom throughout this work. I sincerely appreciate my supervisor, Dr. Adamu Abdullahi Garba, for his guidance, patience, constructive feedback, and invaluable support throughout the research. My heartfelt gratitude goes to my parents, Late Alhaji Sheikh Shehu Ibrahim and Hajiya Hadiza Mohammed, and my siblings, Abdul-Azeez, Qesma, Aisha, and Hajara Shehu Ibrahim, for their unwavering encouragement and support.

Ethical Consideration

Ethical approval was not required for this study as it utilized publicly available, anonymized archival data that did not involve direct interaction with human subjects or living organisms.

Authors' Declaration

- Conflicts of Interest: None.
- I/We hereby confirm that all the Figures and Tables in the manuscript are mine/ours. Furthermore, any Figures and images, that are not mine/ours, have been included with the necessary permission for re-publication, which is attached to the manuscript.
- No animal studies are present in the manuscript
- No human studies are present in the manuscript
- Ethical Clearance: The project was approved by the local ethical committee at Nile University of Nigeria, Abuja.

Authors' Contribution Statement

This Paper was drafted and written by Aneesah Ibrahim Shehu while it was Proof-read, validated by Dr Adamu Abdullahi Garba.



References

1. Al Maeni, N., Suarna, N., & Prihartono, W. (2024). Shopee User Sentiment Analysis Based on Tweet Data from Twitter Using the Naive Bayes Classifier Method. *JATI (Journal of Informatics Engineering Students)* ,8 (2), 1834–1840.
2. Coppersmith, G.; Ngo, K.; Leary, R.; Wood, A. Exploratory analysis of social media prior to a suicide attempt. In Proceedings of the Third Workshop on Computational Linguistics and Clinical Psychology, San Diego, CA, USA, 16 June 2016; pp. 106–117.
3. Devlin, J., Chang, M-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *NAACL-HLT*.
4. Fadjeri, A., Hidayat, K., & Handayani, DR (2021). Emotion Detection in Text Using the Naïve Bayes Algorithm. *Journal of Information and Computer Technology Research (JURISTIK)* ,1 (2), 1–4. <https://doi.org/10.53863/juristik.v1i02.365>
5. Hinduja and J.W.Patchin, "Bullying, cyberbullying, and suicide," *Archives of suicide research*, vol. 14, no. 3, pp. 206–221, 2010.
6. Hinduja, S., Afrin, M., Mistry, S., & Krishna, A. (2022). Machine learning-based proactive social-sensor service for mental health monitoring using twitter data. *International Journal of Information Management Data Insights*, 2(2), 100113
7. Liu, D., Feng, X. L., Ahmed, F., Shahid, M., & Guo, J. (2022). Detecting and measuring depression on social media using a machine learning approach: Systematic review. *JMIR Mental Health*, 9(3),
8. Pachouly, S.J.; Raut, G.; Bute, K.; Tambe, R.; Bhavsar, S. Depression Detection on Social Media Network (Twitter) using Sentiment Analysis. *Int. Res. J. Eng. Technol.* **2021**, 8, 1834–1839.
9. Stankevich, M.; Latyshev, A.; Kuminskaya, E.; Smirnov, I.; Grigoriev, O. Depression detection from social media texts. *CEUR Workshop Proc.* **2019**, 6, 2523.
10. Syed, T.R.; Qamar, R.K.; Akib, M.U. Khanday. Machine Classification for Suicide Ideation Detection on Twitter. *Int. J. Innov. Technol. Explor. Eng.* **2019**, 8, 4154–4160.
11. Tadesse, M.M.; Lin, H.; Xu, B.; Yang, L. Detection of Depression-Related Posts in Reddit Social-Media Forum. *IEEE Access* **2019**, 7, 44883–44893
12. Tadesse, M.M.; Lin, H.; Xu, B.; Yang, L. Detection of suicide ideation in social media forums using deep learning. *Algorithms* **2020**, 13, 7
13. Vioule, B. Moulahi, J. Aze', and S. Bringay, "Detection of suicide-related posts in twitter data streams," *IBM Journal of Research and Development*, vol. 62, no. 1, pp. 7:1–7:12, 2018
14. World Health Organization, *Preventing suicide: A global imperative*. World Health Organization, 2014