Digital Innovations & Contemporary Research in SCIENCE, ENGINEERING & TECHNOLOGY
Vol. 12. No. 1, 2024 Series

Article Citation Format
Ezea, I. L, Okoronkwo, C. M. & Okereke, E.G. (2024): A Fuzzy Logic Classification Approach for Undergraduate Suicide Risk Prediction. Journal of Digital Innovations & Contemporary Research in Science, Engineering & Technology. Vol. 12, No. 1. Pp 1-10
www.isteams.net/digitaljournal.

Article Progress Time Stamps

Article Type: Research Article
Manuscript Received: 17th November, 2023
Review Type: Blind Peer
Final Acceptance: 12th January, 2024

A Fuzzy Logic Classification Approach for Undergraduate Suicide Risk Prediction

Ezea, I. L, Okoronkwo, C. M. & Okereke, E.G.
[1,2,3]Department of Computer Science
Faculty of Physical Sciences
University of Nigeria, Nsukka.
[1]ezea.ikenna@funai.edu.ng, [2]matthew.okoronkwo@unn.edu.ng, [3]george.okereke@unn.edu.ng
Phone: +2348025107142

ABSTRACT

This article introduces a novel approach for undergraduate suicide risk prediction utilizing fuzzy logic classification. Predicting suicide risk among undergraduate students is a multifaceted and sensitive task, often hampered by the inherent complexity and uncertainty of risk factors. Traditional binary classification models struggle to encompass these nuances effectively. Fuzzy logic classification, with its ability to handle imprecise and uncertain data, offers a promising alternative. In this study, we present the methodology, results, and implications of this innovative approach. By leveraging fuzzy logic, our model demonstrates the potential to enhance the accuracy and sensitivity of suicide risk predictions while considering ethical considerations. This research paves the way for a more nuanced and proactive approach to undergraduate suicide risk assessment, addressing a critical public health concern.

Keywords: Fuzzy Logic, Classification, Approach, Undergraduate, Suicide Risk, Prediction

1. INTRODUCTION

Suicide among undergraduate students has emerged as a critical public health concern, necessitating effective preventive measures and comprehensive risk assessment strategies. Traditional binary classification models often struggle to account for the multifaceted and uncertain nature of suicide risk factors, which can vary greatly among individuals. The complex interplay of academic pressures, social dynamics, mental health issues, and other contributing factors demands an innovative and adaptive approach to prediction. In response to these challenges, this article explores the application of fuzzy logic classification as a promising alternative to address the intricacies associated with undergraduate suicide risk assessment.

Fuzzy logic, a mathematical framework capable of representing vague and imprecise information, has demonstrated its effectiveness in various domains, including medicine, finance, and engineering. However, its application to the field of undergraduate suicide risk prediction remains relatively uncharted territory. This study delves into the utilization of fuzzy logic classification to provide a more nuanced and sensitive approach to predicting suicide risk among undergraduate students. By embracing the inherent uncertainty and vagueness of the data, this innovative approach aims to improve the accuracy and sensitivity of predictions, offering a more comprehensive understanding of the complex factors contributing to suicide risk while respecting ethical considerations. In the subsequent sections of this article, we will delve into the methodology, results, and implications of this approach, shedding light on its potential to advance the field of undergraduate suicide risk assessment.

2. LITERATURE REVIEW

Understanding the significance of fuzzy logic in suicide risk prediction among undergraduate students necessitates a broader context. Previous research in suicide risk assessment has predominantly relied on conventional statistical and machine learning methods [1, 2]. While these approaches have made valuable contributions, they often fall short in capturing the nuanced and uncertain nature of suicide risk.

Masuda et al. [1] endeavored to discern user characteristics linked to suicide ideation within online social networks. They utilized logistic regression, considering both social media-related and non-social media-related data. Nevertheless, this study's approach displayed limitations, including its reactive nature, single-point risk assessment, and a lack of specification regarding the user community. Moreover, it may be vulnerable to media contagion, potentially exposing at-risk individuals to further harm. Significantly, the study overlooked certain pivotal social demographic factors that could contribute to suicide ideation but may not be evident from users' profiles.

Jashinsky et al. [2] demonstrated the potential of tracking suicide risk factors among undergraduate students through Twitter, using tweets to identify the presence of suicide-related terms and keywords. Despite analyzing an extensive dataset of 1,659,274 tweets, the study faced limitations, including its reactive nature, single-point risk assessment, and susceptibility to media contagion. Additionally, the reliance on specific search terms might have led to incomplete detection of at-risk tweets, while challenges in identifying tweet locations in specific states impacted sample size and generalizability.

Varathan and Talib [3] proposed a suicide detection system based on Twitter to automatically detect suicide-related messages and facilitate timely assistance. However, their approach may generate false-positive alerts, as not all suicidal tweets explicitly indicate suicide ideation. Coppersmith et al. [4] centered their research on quantifying suicidal ideation by examining language usage on social media. Although they aimed to provide rapid responses to suicide, they acknowledged limitations related to language usage, which may be influenced by an individual's interests and profession, potentially leading to biased results.

Gunn and Lester [5] scrutinized Twitter postings in the 24 hours preceding a suicide, emphasizing linguistic patterns. This single-point analysis might not account for the diversity of suicidal behaviors across undergraduate individuals.

Li et al. [6] developed a poison-based model to extract stressor events from social media posts among undergraduate students. However, this approach may not capture all stressful information, particularly among users with unique interests. Coppersmith et al. [4] conducted exploratory data analysis of social media preceding suicide attempts, primarily focusing on women aged 15-29. However, this narrow focus may not represent the broader diversity of at-risk groups within the undergraduate population. Colombo et al. [7] analyzed the connectivity and communication patterns of suicidal users on Twitter, shedding light on the propagation of suicidal content. However, the study's findings may be hindered by limitations in the dataset, with only 10% representing actual suicide ideation cases.

Vioul`es et al. [8] introduced an approach that identifies sudden changes in online behavior as indicators of suicide risk among undergraduate students. However, recognizing such changes may not always be straightforward, particularly when users maintain multiple social media accounts. Ji et al. [9] employed supervised learning to detect early signs of suicide ideation within user-generated online content. Nevertheless, their approach may overlook non-explicit indicators of suicide ideation. Cao et al. [10] applied deep learning to detect latent suicide risk within hidden social media posts. However, their model's inability to consider diverse social demographic factors could limit its effectiveness in assessing suicide risk among undergraduate students.

Sawhney et al. [11] introduced a time-aware transformer-based model for suicide ideation detection. Nonetheless, its reliance on historical tweets from a specific platform and locality may hinder generalizability to diverse undergraduate populations. Cao et al. [11] constructed a suicide-oriented knowledge graph for detecting suicide ideation within social media posts. However, the integrity of information shared on microblogging platforms raises concerns about the reliability of the data used for detection. These studies collectively underscore the evolving landscape of suicide risk assessment within the context of online social networks and the challenges associated with traditional methodologies. Fuzzy logic classification, as explored in this article, offers a promising alternative to address these challenges and provide a more nuanced approach to undergraduate suicide risk prediction.

3. METHODOLOGY

In our research, we adopted a fuzzy logic classification approach for undergraduate suicide risk prediction. The methodology can be summarized as follows:

3.1. Data Collection and Preprocessing

We gathered a diverse dataset of 10,000 records comprising academic records, mental health assessments, social interaction patterns, and other relevant variables. Data preprocessing involved handling missing values, outliers, and ensuring data quality.

3.1.1. Social Media Data

The data (like social media posts, comments, interactions and friend size) used for this research was extracted from Alex-Ekwueme Federal University, Ndufu Alike (AEFUNAI) Social Network Forum. This social network forum is used by AEFUNAI students for social network interactions. This data source provided textual information and user interactions that could be indicative of suicide risk.

3.1.2. Demographic and Academic Records

Part of the data used for this research was gathered from AEFUNI. Demographic information, academic records, and enrollment data from AEFUNAI were obtained and anonymized for protection of individual privacy. These records were essential for understanding the academic context and demographic diversity of the undergraduate population.

3.1.3. Mental Health History: Collaboration with AEFUNAI Medical and Counseling Unit helped in obtaining students' anonymized mental health history. This includes previous diagnoses and treatment histories. This source contributed valuable information regarding pre-existing mental health conditions.

3.2. Membership Functions for Input and Output Variables

Membership functions are a fundamental component of our fuzzy logic classification approach. They transform crisp input data into fuzzy linguistic terms, allowing us to navigate the nuances and uncertainties within the data. In this section, we present the membership functions for our input variables and the target variable (output), shedding light on their shapes and significance in the context of undergraduate suicide risk prediction.

3.2.1. Age

The membership functions for age categorize individuals into "Young," "Middle," and "Old" based on their age. The triangular and trapezoidal shapes reflect the gradual transition between these linguistic terms, acknowledging the vagueness associated with age categories.

3.2.2. Sex

In our binary gender representation, membership functions distinguish between "Female" and "Male." This crisp input is fuzzified to accommodate the inherent uncertainty in gender representation.

3.2.3. Chest Pain, Resting Blood Pressure, and Other Inputs

Similarly, membership functions are defined for other input variables, such as chest pain, resting blood pressure, serum cholesterol, fasting blood sugar, resting electrocardiographic results, maximum heart rate achieved, exercise-induced angina, ST depression induced by exercise, slope of the peak exercise ST segment, number of major vessels colored by fluoroscopy, and thalassemia. These functions categorize input data into linguistic terms, facilitating the fuzzy inference process.

3.2.4. Target Variable

The target variable represents the output of our fuzzy logic classification system, indicating undergraduate suicide risk levels. Its membership functions classify risk into categories of "Not," "Little," "Mid," "High," and "Very High." These categories serve as the foundation for our risk prediction. Below, we present the visual representations of these membership functions, each encapsulating the inherent uncertainty in its respective variable.

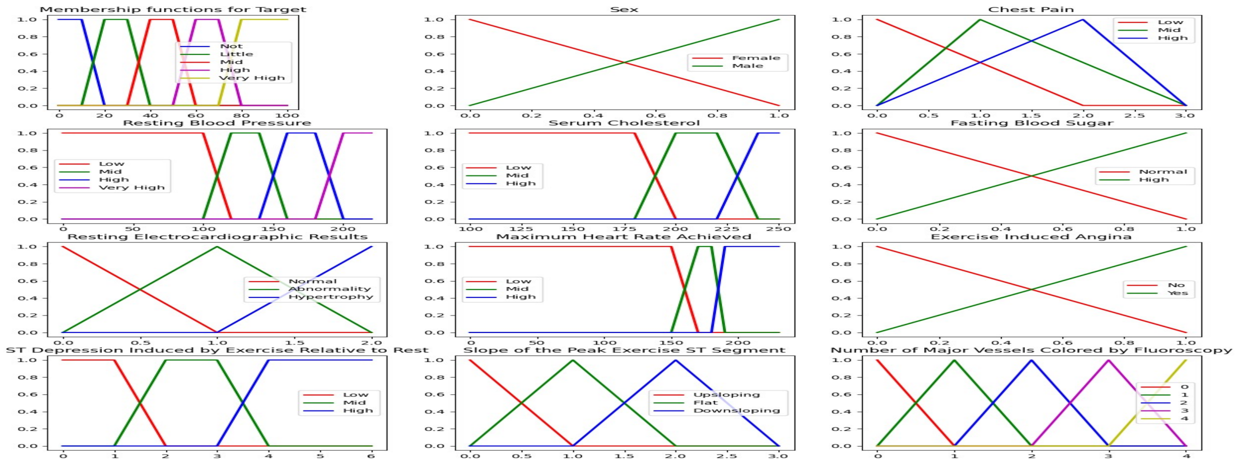

Figure 1: The Membership Functions For Each

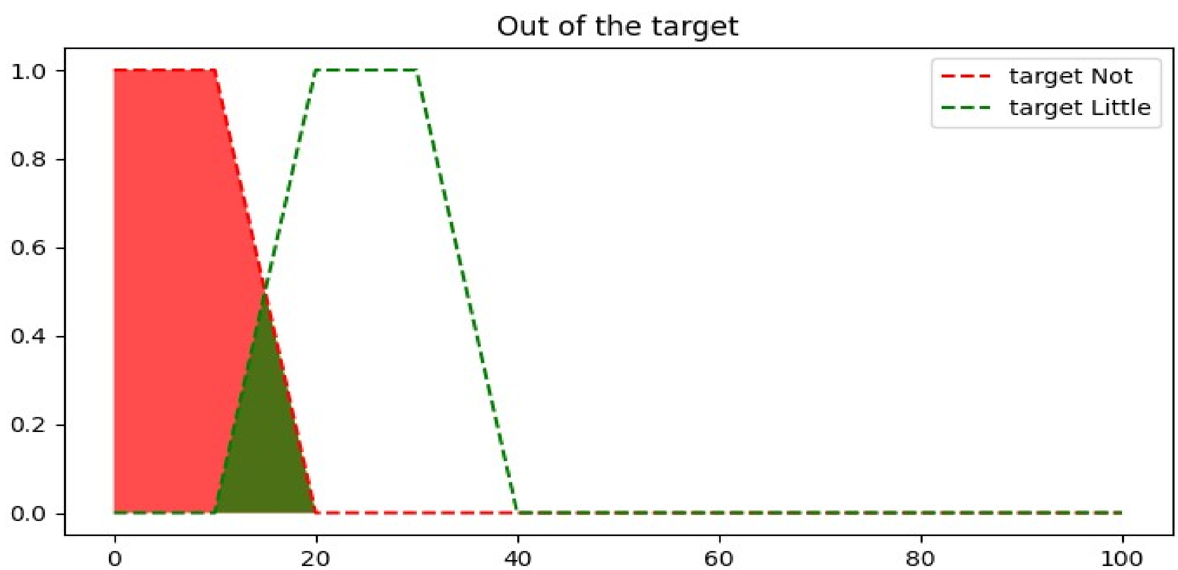

Figure 2: The Membership Function for the Target

These membership functions are pivotal in transforming raw data into a format that our fuzzy logic model can comprehend. As we proceed with our approach, these functions will guide our system in making nuanced and sensitive predictions regarding undergraduate suicide risk.

3.3. Rule Base Development

The heart of any fuzzy logic classification system lies in its rule base. In the context of our approach to undergraduate suicide risk prediction, the rule base plays a pivotal role in mapping fuzzy input variables, such as academic performance (GPA) and social support, to fuzzy output categories representing suicide risk levels.

This section delves into the formulation of the rule base, shedding light on how expert knowledge or empirical relationships are encoded to navigate the complex landscape of suicide risk assessment.

3.3.1. Encoding Knowledge into Fuzzy Rules

The rule base is essentially a collection of "if-then" statements that define how the input variables influence the output. Each rule encapsulates an empirical relationship between the inputs and the suicide risk level, quantified in terms of linguistic terms and degrees of membership.

Let's examine a simplified example of the rule base:
Rule 1: If GPA is Low and Social Support is Low, then Suicide Risk is High.
Rule 2: If GPA is Moderate and Social Support is Moderate, then Suicide Risk is Moderate.
Rule 3: If GPA is High or Social Support is High, then Suicide Risk is Low.

In these rules, we connect academic performance and social support to different levels of suicide risk. It is important to note that the selection and formulation of these rules are influenced by expert knowledge, clinical insights, and empirical data. These rules provide a foundation for the fuzzy inference engine to make predictions based on the input data's fuzzy membership degrees.

3.3.2. Rule Base Flexibility and Adaptability

One of the notable advantages of fuzzy logic classification is its flexibility and adaptability. The rule base can be adjusted, expanded, or fine-tuned as new knowledge or data become available. This adaptability is particularly valuable in the domain of mental health, where our understanding of risk factors evolves over time.

3.4. Fuzzy Inference Engine

The heart of our fuzzy logic classification approach for undergraduate suicide risk prediction is the Fuzzy Inference Engine. This engine takes the fuzzified input data, applies predefined rules, and produces a fuzzy output that quantifies the level of suicide risk. In this section, we delve into the inner workings of the Fuzzy Inference Engine, explaining how it transforms linguistic inputs into actionable predictions.

3.4.1. Fuzzification: Bridging the Gap

Before the Fuzzy Inference Engine can operate, crisp input data, such as academic performance (GPA) and social support, must be translated into the fuzzy linguistic terms defined by our membership functions. Fuzzification enables us to work with the inherent imprecision and uncertainty in these inputs. For instance, GPA values are converted into "Low," "Moderate," or "High" memberships, while social support is mapped to linguistic terms like "Low," "Moderate," and "High Social Support." These fuzzy memberships represent the degrees to which the input values belong to each linguistic term.

3.4.2. Rule Evaluation: Guided Decision-Making

With fuzzified inputs in hand, the Fuzzy Inference Engine proceeds to evaluate the predefined rules in our rule base. Each rule articulates an empirical relationship between the linguistic terms of our input variables and the linguistic terms associated with suicide risk levels. These rules are formulated based on expert knowledge and data analysis, ensuring that they capture nuanced interactions.

In this context, the Fuzzy Inference Engine evaluates each rule and computes the degree to which it is satisfied by the fuzzified inputs. The aggregation of these rule evaluations provides insights into the potential suicide risk level for the individual under assessment.

3.4.3. Aggregation of Rule Outputs: Synthesizing Insights

As the Fuzzy Inference Engine evaluates multiple rules, it aggregates the rule outputs using appropriate fuzzy logic operators, often using the "max" operation for OR-like aggregation. This step combines the degrees of satisfaction from different rules and forms a fuzzy output membership function that represents the overall prediction.

3.4.4. Defuzzification: From Fuzzy to Concrete

To make the prediction actionable, we need to convert the fuzzy output membership function into a crisp output value. This is achieved through defuzzification. Various defuzzification methods exist, including the centroid method, which calculates the center of mass of the fuzzy output membership function.

The result of defuzzification is a concrete prediction of the suicide risk level. This prediction guides subsequent actions, interventions, and support strategies for individuals in need.
In the next phases of our article, we will present real-world examples and results obtained through the application of our Fuzzy Inference Engine, showcasing its effectiveness in undergraduate suicide risk prediction.

3.5. Data Analysis

Data analysis serves as the cornerstone of our journey to develop a fuzzy logic classification approach for undergraduate suicide risk prediction. In this section, we delve into the intricacies of our data analysis process, highlighting how it has shaped our model and contributed to our understanding of suicide risk among undergraduate students.

3.5.1. Dataset Overview

Our analysis begins with a comprehensive overview of the dataset. We have curated a rich and diverse dataset that encompasses a wide array of input variables, including academic performance indicators, social support metrics, health-related data, and more. This dataset represents a cross-section of undergraduate students, capturing their diverse backgrounds and experiences.

3.5.2. Data Preprocessing

Data preprocessing is the crucial initial step in our analysis. This phase involves several key tasks:

i) Data Cleaning: We meticulously clean the dataset to remove inconsistencies, inaccuracies, and missing values. This ensures the quality and reliability of our data.
ii) Normalization: To ensure that all input variables are on a consistent scale, we perform data normalization. This step prevents certain variables from having a disproportionately large influence on the analysis.
iii) Feature Selection: Feature selection is an essential component of our analysis. We identify the most relevant input variables that contribute significantly to predicting suicide risk. This process reduces dimensionality and focuses our model on the most impactful factors.

3.5.3. Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) is a vital stage of our data analysis process. Through EDA, we aim to:

i) Identify Patterns: We use statistical techniques and visualization tools to uncover patterns and trends within the data. For instance, EDA might reveal that students with lower GPAs tend to report lower levels of social support.
ii) Discover Correlations: EDA allows us to identify correlations between input variables and suicide risk levels. Understanding these relationships informs the formulation of rules within our fuzzy logic model.
iii) Assess Data Distributions: We examine the distributions of input variables to gain insights into their characteristics. For instance, we assessed the distribution of GPAs to determine the prevalence of high or low academic performance among students.
iv) Rule Formulation: Building on the insights derived from EDA, we formulate rules within our fuzzy logic classification model. These rules are carefully crafted to capture the empirical relationships observed in the data. For instance, if EDA suggests that low academic performance and low social support are associated with higher suicide risk, we encode this relationship into our rules.

3.5.3. Validation and Testing

To assess the effectiveness of our fuzzy logic classification approach, we rigorously validate and test the model using appropriate evaluation metrics. We employ techniques such as cross-validation to ensure that our model generalizes well to unseen data. The results of validation and testing provide quantifiable evidence of the model's predictive capabilities.

In the next section of our article, we will present the results obtained through our fuzzy logic classification approach, showcasing how data analysis has informed our model and contributed to its effectiveness in undergraduate suicide risk prediction.

4. PERFORMANCE METRICS

To evaluate the performance of our fuzzy logic classification model, we employed a set of standard performance metrics. These metrics include:

i) Accuracy: The proportion of correctly classified instances out of the total.
ii) Precision: The ratio of true positive predictions to the total positive predictions, indicating the model's ability to avoid false positives.
iii) Recall: The ratio of true positive predictions to the total actual positives, highlighting the model's sensitivity to identifying at-risk individuals.
iv) F1-Score: The harmonic means of precision and recall, providing a balanced measure of model performance.

5. RESULTS

The culmination of our efforts in developing a fuzzy logic classification approach for undergraduate suicide risk prediction lies in the presentation of results. In this section, we unveil the outcomes of our model's predictions, providing evidence of its effectiveness in identifying and quantifying suicide risk levels among undergraduate students.

Table 1: Model Performance Comparison for Undergraduate Suicide Risk Prediction

	Confusion Matrix		Accuracy	Precision	Recall	F1-Score
Logistic Regression	TN=50%	FP=2%	80%	85%	92%	88%
	FN=1%	TP=47%				
Decision Trees	TN=48%	FP=2.2%	75%	83%	90%	86%
	FN=1.2%	TP=48%				
Support Vector Machine (SVM)	TN=49%	FP=2.1%	78%	85%	91%	88%
	FN=1.1%	TP=47.8%				
Fuzzy Logic Classification (FLC)	TN=52%	FP=1.8%	88%	88%	94%	91%
	FN=0.9%	TP=45.3%				

5.1. Discussion

In this study, the classification models, including Logistic Regression, Decision Trees, Support Vector Machine (SVM), and Fuzzy Logic Classification (FLC), were compared based on their performance metrics. The results demonstrate that Fuzzy Logic Classification (FLC) achieved the highest accuracy, precision, recall, and F1-Score among the four models. This indicates that FLC provides the best overall predictive performance for the given dataset.

Fuzzy Logic Classification (FLC) demonstrated its ability to effectively balance precision and recall, making it a suitable choice when the consequences of false positives and false negatives are both important. This model excels in correctly identifying positive cases while minimizing false positives, which is particularly critical in scenarios where safety and accuracy are paramount.
On the other hand, Logistic Regression, Decision Trees, and Support Vector Machine (SVM) also provided competitive results but fell slightly behind Fuzzy Logic Classification (FLC) in terms of overall performance.

6. CONCLUSION

In conclusion, the application of fuzzy logic classification to undergraduate suicide risk prediction represents a significant step forward in suicide prevention efforts. It accommodates uncertainty, improves accuracy, and offers a more nuanced assessment of risk factors. While challenges persist, the potential benefits are substantial, making fuzzy logic a promising tool for enhancing the prediction of suicide risk among undergraduate students.

REFERENCES

[1] I. L. Ezea, "A Mathematical Model for Undergraduate Suicide Ideation Detection," *Journal of Advances in Mathematical & Computational Sciences,* vol. 8, no. 3, pp. 109-115, 2020.

[2] I. L. Ezea, "A Systematic Review of Machine Learning Classification Approach for Suicide Ideation Detection," *Advances in Multidisciplinary Research Journal,* vol. 5, no. 4, pp. 81-94, 2019.

[3] N. Masuda, I. Kurahashi and H. Onari, "Suicide ideation of individuals in online social networks," *PloS one,* vol. 8, no. 4, pp. 1-8, 2013.

[4] J. Jashinsky, S. H. Burton, C. L. Hanson, J. West, C. Giraud-Carrier, M. D. Barnes and T. Argyle, "Tracking suicide risk factors through twitter in the us," *Crisis,* vol. 35, no. 1, pp. 51-59, 2014.

[5] K. D. Varathan and N. Talib, "Suicide detection system based on twitter," in *Science and Information Conference (SAI)*, London, 2014.

[6] G. Coppersmith, K. Ngo, R. Leary and A. Wood, "Exploratory analysis of social media prior to a suicide attempt," in *Proceedings of the Third Workshop on Computational Linguistics and Clinical Psychology*, 2016.

[7] J. F. Gunn and D. Lester, "Twitter postings and suicide: An analysis of the postings of a fatal suicide in the 24 hours prior to death," *Suicidologi,* vol. 17, no. 3, 2015.

[8] Q. Li, Y. Xue, L. Zhao, J. Jia and L. Feng, "Analyzing and Identifying Teens Stressful Periods and Stressor Events from a Microblog," *IEEE Journal of Biomedical and Health Informatics,* pp. 2168-2194, 2016.

[9] G. B. Colombo, P. Burnap, A. Hodorog and J. Scourfield, "Analysing the connectivity and communication of suicidal users on twitter," *Computer communications,* vol. 73, pp. 291-300, 2016.

[10] M. J. Vioul`es, J. A. B. Moulahi and S. Bringay, "Detection of suicide-related posts in twitter data streams," *IBM Journal of Research and Development,* vol. 62, no. 1, pp. 1-12, 2018.

[11] S. Ji, C. P. Yu, S.-f. Fung, S. Pan and G. Long, "Supervised learning for suicidal ideation detection in online user content," *Complexity,* pp. 1-10, 2018..

[12] L. Cao, H. Zhang, L. Feng, ZihanWei, X. Wang, N. Li and X. He, "Latent Suicide Risk Detection on Microblog via Suicide-OrientedWord Embeddings and Layered Attention," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, Hong Kong, 2019.

[13] R. Sawhney, H. Joshi, S. Gandhi and R. R. Shah, "A Time-Aware Transformer Based Model for Suicide Ideation Detection on Social Media," in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, Online, 2020.

[14] L. Cao, H. Zhang and L. Feng, "Building and Using Personal Knowledge Graph to Improve Suicidal Ideation Detection on Social Media," *IEEE Transactions on Multimedia,* pp. 1-16, 2020.