Arabian Journal for Science and Engineering (2024) 49:12729–12742
https://doi.org/10.1007/s13369-024-08767-3

RESEARCH ARTICLE-COMPUTER ENGINEERING AND COMPUTER SCIENCE

Detecting Suicidality in Arabic Tweets Using Machine Learning and Deep Learning Techniques

Asma Abdulsalam[1] · Areej Alhothali[1] · Saleh Al-Ghamdi[2]

Received: 12 April 2023 / Accepted: 23 January 2024 / Published online: 5 March 2024

Abstract
Social media platforms have revolutionized traditional communication techniques by allowing people to connect instantaneously, openly, and frequently. As people use social media to share personal stories and express their opinions, negative emotions such as thoughts of death, self-harm, and hardship are commonly expressed, particularly among younger generations. Accordingly, the use of social media to detect suicidality may help provide proper intervention that will ultimately deter the spread of self-harm and suicidal ideation on social media. To investigate the automated detection of suicidal thoughts in Arabic tweets, we developed a novel Arabic suicidal tweet dataset, examined several machine learning models trained on word frequency and embedding features, and investigated the performance of pre-trained deep learning models in identifying suicidal sentiment. The results indicate that the support vector machine trained on character n-gram features yields the best performance among conventional machine learning models, with an accuracy of 86% and F1 score of 79%. In the subsequent deep learning experiment, AraBert outperformed all other machine and deep learning models with an accuracy of 91% and F1-score of 88%, significantly improving the detection of suicidal ideation in the dataset. To the best of our knowledge, this study represents the first attempt to compile an Arabic suicidality detection dataset from Twitter and to use deep learning to detect suicidal sentiment in Arabic posts.

Keywords Suicidality · Suicidal ideation · Suicidal thoughts · Natural language processing · Twitter · Social media · Machine learning · Deep learning · AraBERT · Arabic tweets · Arabic text classification

1 Introduction

Approximately 3.96 billion people actively access the internet worldwide [1], with millions using social media platforms such as chat rooms, social networking sites, and blogs. Social networks such as Facebook, Twitter, and Snapchat allow users to exchange information and interact with others. Among these, Twitter is a free broadcast site that allows registered users to communicate through 280-character messages called tweets, allowing users to express themselves freely. As many users utilize social media to convey their feelings, experiences, thoughts, difficulties, and concerns [2], thoughts pertaining to self-harm, death, and suicidal ideation have become among the most popular topics discussed on these networks.

As the intentional attempt by a person to end their own life, suicide is a phenomenon that arises from a complex interaction of social, biological, cultural, psychological, and spiritual variables [3]. Suicide is a manifestation of underlying suffering that is brought on by a variety of events, such as underlying mental illnesses that create psychological pain [4]. Suicidality encompasses three behavioral patterns: suicidal behavior, suicidal planning, and suicidal ideation [3–5]. Suicidal ideation refers to the thought or imagined intent of taking one's own life, whereas a suicide attempt is an act of self-harm with the intended purpose of dying. In contrast, a suicide plan, is a specific method one may adopt to terminate their life [3–5]. Suicide affects individuals, families, com-

✉ Asma Abdulsalam
aabdulsalam0012@stu.kau.edu.sa

Areej Alhothali
aalhothali@kau.edu.sa

Saleh Al-Ghamdi
syalghamdi@kau.edu.sa

[1] Department of Computer Science, Faculty of Computing and Information Technology, King AbdulAziz University, 21589 Jeddah, Saudi Arabia

[2] Department of Psychology, Faculty of Educational Graduate Studies, King AbdulAziz University, 21589 Jeddah, Saudi Arabia

munities, and even countries [4]. It has become the second leading cause of death among young people internationally, causing more fatalities than diabetes, liver disease, stroke, or infection [6]. Due to the stigma associated with mental disorders, more than 40% of people who seek primary care are unwilling or hesitant to discuss their depressive symptoms. Although suicidal thoughts and actions require immediate attention, there is no effective method for assessing, managing, and preventing suicide [6].

Traditional methods for suicidality risk assessment rely on professional psychological expertise and self-report questionnaires [4]. The Patient Health Questionaire-9 (PHQ-9) and Columbia Suicide Severity Rating Scale (C-SSRS) are two examples of tools used for screening suicide and identifying depressive symptoms [6]. Although these methods are quick and efficient, they are vulnerable to misleading results stemming from participant concealment and challenging to carry out over an extended period of time or on a massive scale [6]. Accordingly, the suicidality detection task has attracted researchers from a variety of fields to investigate linguistic and psychological indicators, as well as other factors that may help diagnose and identify individuals with suicidal thoughts [4].

Social media posts can offer valuable insights into an individual's emotional and psychological well-being of individuals [2]. Many individuals are unable to share their personal experiences or express their emotions in person, choosing instead to write about their feelings, including possible suicidal intentions, through blogs or social media posts. Unfortunately, these posts are often disregarded or overlooked. However, this information can be useful in conducting large-scale screenings for suicidal behaviors. Studies in the field investigated suicidality content written in various languages, including Chinese, Spanish [7], Russian [8], Japanese [9], and Tagalog [1]. English was the most frequently examined language [10–18], followed by Chinese [19, 20] and Spanish. However, the sole Arabic language study on this topic used a translated English dataset [21].

As millions of Twitter users are from Arab countries, with 15 million being from Arab countries. A total of 15 million Twitter users are from Saudi Arabia alone [22], many tweets are written in Arabic. The Arabic language is the predominant language of 422 million people in over 27 nations [23], having a large lexicon and different varieties—including Classical Arabic and Modern Standard Arabic (MSA)—as well as dialects and colloquialisms [23, 24]. MSA is primarily used in formal speech and writing, whereas dialects and colloquialisms often vary between nations [23]. Although all of these varieties have certain properties in common, each has its own lexicon, grammar, and morphology [24]. The morphology and orthography of Arabic language text are the primary challenges associated with the language in the context of natural language processing (NLP) [23, 25]. Most MSA texts lack orthographic representation for short vowel letters, as these letters are generally represented by diacritical marks not used in MSA. Consequently, the meanings of individual words may ambiguous [23, 24]. Morphology plays an important role in Arabic because it is a derivational and highly structured language [24], with many words differing according to morphological elements such as root, stem, part-of-speech (POS), and affix [23]. Even Arabic letters may have different shapes based on their positions in words [24].

A review of related works in the field revealed a lack of research pertaining to suicidality on Arabic social media, and especially the automatic detection and identification of Arabic suicidal ideation using machine learning. The development of machine learning techniques to extract Arabic linguistic indicators of suicidality is expected to help develop accurate and effective mechanisms to extensively screen social media platforms with the objective of detecting suicidal ideation and potentially preventing suicidal behaviors. To the best of our knowledge, this study represents the first attempt to construct a machine or deep learning model to detect suicidal content in Arabic.

We propose an Arabic suicidal ideation detection framework for the analysis tweets with various features to explore the potential of monitoring suicidal behaviors. The contributions of this study are threefold. First, we have compiled the first Arabic suicidality detection dataset, comprising 5,719 Arabic tweets collected between August 23, 2021 and April 21, 2022 from various countries using different Arabic dialects; 1,429 were labeled as Suicidal tweets, and 4,290 were labeled as Non-Suicidal. The labeling process was conducted by two annotators, one of whom holds a Ph.D. in psychology and has been working on several cyberpsychology projects. Second, several machine learning models were adopted to automatically detect suicidal sentiment with different types of feature representations including word frequency (BOW, unigram TF-IDF, n-gram TF-IDF, and character n-gram TF-IDF) and word-embedding (Arabic word2vec and FastText) features. Third, we utilized pre-trained deep learning models with transformer-based architectures and attention mechanisms (AraBert, AraELECTRA, and AraGPT2) to predict suicidal ideation from Arabic tweets.

The rest of this paper is organized as follows: Sect. 2 reviews prior studies pertaining to the detection of suicidality. The methodology used for dataset, as well as feature extraction and the algorithms used in the classification process, are discussed in detail in Sect. 3. The results of this study are examined in Sect. 4. Section 5 concludes the study and looks ahead toward future works.

2 Related Works

The number active users on social media is continuously growing, with many using these services to express their sentiments on a daily basis. This trend of using social media as a modern-day diary can help reveal and analyze the personalities and mental states of users. Researchers have investigated various machine learning and deep learning techniques along with various types of features to identify suicidal content and determine suicidal risk severity based on content generated by social media users. Classification approaches used in the field include traditional machine learning models, such as the support vector machine (SVM), naive Bayes (NB) classifier, random forest (RF), logistic regression (LR), decision tree (DT), and artificial neural network (ANN); ensembling techniques such as voting, stacking, boosting, bagging XGBoost, and AdaBoost; and deep learning models such as long short-term memory (LSTM), convolutional neural networks (CNNs), recurrent neural networks (RNNs), gated recurrent units (GRUs), and transformer networks including Bidirectional Encoder Representations from Transformers (BERT). The features used in the field can be categorized as statistical (e.g., length of post, number of words and characters), temporal, linguistic (e.g., emotion and sentiment using the Linguistic Inquiry and Word Count (LIWC) dictionary [26]), syntactic (e.g., parts of speech tagging information), word frequency (e.g., bag-of-words (BOW) models, Term Frequency-Inverse Document Frequency (TF-IDF)), word embedding (e.g., word2vec [27], Glove [28]), and topical (e.g., Latent Dirichlet allocation (LDA) [29]). We examined the latest research on suicidal detection with a primary focus on text classification.

Several studies have employed supervised learning to detect suicidal ideation O'Dea et al. [10] examined the possibility of determining levels of concern from Twitter post content, developing two text classifiers based upon the LR and SVM algorithms. The word frequency and weighting word frequency (i.e., TF-IDF) with and without filtration of the feature space were used to test these techniques. TF-IDF may use a filter to eliminate terms that appear more than a specified number of times in the document, whereas conventional word frequency does not involve weighting. The results indicate that the SVM with unfiltered TF-IDF achieved the highest accuracy. Another study [30] used four machine classifiers—namely DT, NB, RF, and SVM—on the same dataset [11]. Data were divided into two subsets-one for binary classification (suicide or flippant references to suicide), and the other for multi-class classification (suicide, flippant, and non-suicide)—and the part-of-speech (POS), BOW, and IDF features were adopted for classification. The results show that DT achieves the highest accuracy for multi-class classification. A study by Chiroma et al. [17] on the same dataset [11] using the same preprocessing techniques was conducted to evaluate the performance of the Prism algorithm compared to standard machine learning algorithms (SVM, DT, NB, and RF) with BOW features. The results demonstrate that the Prism algorithm outperformed the other classifiers in all performance measures.

Huang et al. [31] developed a method to detect suicidal content on the Sina Weibo microblogging platform with three sets of linguistic features: an automated machine learning dictionary, a Chinese suicide dictionary, and the Simplified Chinese Micro-Blog Word Count (SCMBWC). These feature sets were separately employed with the SVM, DT, and LR algorithms to classify content between six classes. The results demonstrate that the SVM algorithm with the feature set extracted using an automated machine learning dictionary from real blog data with N-grams achieved the highest accuracy. Moulahi et al. [13] proposed a probabilistic framework based on conditional random fields (CRFs) to track suicidal ideation. They used three sets of features: syntactic features (i.e., POS), linguistic features (i.e., psychological and emotional lexicon features), and contextual features (i.e., posts observed at previous and upcoming sessions). The CRF model was observed to outperform the other machine learning methods, namely the SVM, NB, J48, RF, and ANN. Rajesh Kumar et al. [14] utilized Vader sentiment analysis to assign a sentiment score to each word in their study, employing various classifiers—including NB, RF, XGBoost, and LR—to assign sentences into positive or neutral categories. Several preprocessing techniques were employed, including statistical features, tokenization stemming, BOW, and word frequency, after cleaning the data. The highest accuracy was achieved using the RF method when considering all sets of features.

Ensemble machine learning models have also been employed for suicidal ideation detection. Burnap et al. [11] incorporated SVM and NB in an ensemble approach known as rotation forest, trained on lexical, structural, emotive, and psychological features extracted from Twitter posts to identify concerning content pertaining to suicide including ideation, suicide reporting, memorials, campaigning, and support. The rotation forest approach was compared to three classifiers—namely SVM, NB, and J48—and achieved superior performance.

Sakib et al. [32] developed a machine learning model to predict suicidal ideation within tweets. The dataset compassed 9119 tweets, with 5121 non-suicidal tweets and 3998 suicidal tweets. The algorithms used in this study included machine learning models, including SVM, DT, LR, NB, KNN, and ensemble models, namely AdaBoost, Gradient Boost, Bagging, CatBoost, XGBoost, and the voting classifier (VC). The results indicate that VC achieves the best accuracy. Accordingly, VC was used as an estimator in conjunction with the three best-performing classifiers, namely LR, SVM, and DT. A dataset of individuals who expressed

suicidal thoughts on Twitter, compiled by [33], consists of 1897 tweets gathered using keywords extracted from a previous study [34] and various web forums. Labeling was performed by a human annotator and a psychiatric expert. Several machine learning techniques were used, including SVM, LR, MNB, BNB, RF, and DT. Three ensemble learning techniques, including Voting Ensemble and AdaBoost, were also used. The results indicate that LR outperforms all other models.

Huang et al. [19] utilized a real-time system that employed machine learning and a psychological lexicon dictionary to detect suicidal ideation. They analyzed the social media platform Weibo and identified 53 users who had posted suicidal content before their deaths. They employed various classifiers including SVM, NB, LR, J48, rotation forest, and sequential minimal optimization (SMO) with three N-gram features and a psychological lexicon dictionary. Of these, the SVM classifier achieved the best performance. Rezig [35] aimed to improve the machine learning model that uses the grey wolf optimizer (GWO) to identify users contemplating suicide. Specifically, GWO was combined with a machine learning algorithm to predict suicidality in Twitter users. Tweets mentioning depression, self-harm, and anxiety were gathered, and only those with indicators of suicidality were subsequently selected. A total of 193,720 tweets were collected and annotated. The machine learning models used in the study included NB, LR, SVM, and DT. The results show that LR-GWO trained on unigrams achieved the best accuracy among these models.

Supervised deep learning has also been previously used to identify suicidal content. Ji et al. [16] used different classifiers from both traditional machine learning models and deep learning models to identify suicidality through the online content of users. The specific algorithms used included SVM, RF, gradient boost classification tree (GBDT), XGBoost, multilayer feed-forward neural network (MLFFNN), and LSTM. Several sets of features were used, including statistical, linguistic, syntactic, topical, and word-embedding features. Furthermore, some of these features were combined to increase model accuracy. Tadesse et al. [15] proposed a model to detect suicidality on the Reddit platform, combining LSTM and CNN models to classify Reddit content into multiple classes. TF-IDF, BOW, and statistical features were used with four machine learning models (SVM, NB, RF, and GXBoosting), whereas word2vec was used with the two deep learning models. The proposed LSTM-CNN hybrid model outperformed all other algorithms used in this study.

Metzler et al. [36] examined the relationship between social media content consumption and suicidal behavior. They created a dataset of 3202 English tweets and manually classified them into 12 categories, such as personal stories of suicidal ideation, coping and recovery, spreading awareness, prevention information, and irrelevant tweets. They developed several machine learning models to perform multi-class and binary classification tasks. The majority of classifiers used TF-IDF with a linear SVM, as well as two deep learning models, namely BERT and XLNet. The first task involved categorizing posts into six key content categories related to suicide prevention, whereas the second task used binary classification to distinguish between posts in the 12 categories that referred to actual suicide and those in off-topic categories that used terms related to suicide in a different context. Both deep learning models performed similarly in both tasks, outperforming the SVM with TF-IDF except for the suicidal ideation and attempts category. The BERT model achieved the highest overall scores in binary classification.

Approximately 50,000 tweets were collected by Haque et al. [37] to detect suicidal ideation using machine learning and deep learning classifiers. Their results demonstrate that although the RF model achieved the highest classification score among conventional machine learning methods, the deep learning models exhibited enhanced performance owing to word-embedding training, with the BiLSTM model achieving higher precision and F1-score. Chatterjee et al. [38] developed a deep learning algorithm to analyze tweets for characteristics of suicidal tendencies. To develop the model, 188,704 tweets were extracted from 1169 users and manually annotated into two classes: suicide and non-suicidal posts. The extracted features included sentiment analysis, emoticons, TF-IDF, statistics, topic-based features, and temporal features. The results indicate that LDA with trigrams, TF-IDF, statistics, temporal features, emoticons, and sentiment analysis features achieves the best performance in detecting suicidal ideation.

As mentioned earlier, most studies pertaining to this task were conducted on English datasets. Only one study, conducted by Benlaaraj et al. [21], compiled an Arabic suicide dataset translated from an English dataset called ASuiglish. ASuiglish was constructed by combining several English data sources, namely Twitter data provided by [39], Suicide Notes from the Kaggle dataset [40], Victoria Suicide Data [40], and Suicidal Phrases from the Kaggle dataset [41]. The final dataset consists of 1960 posts with 980 passages per class. The dataset was then pre-processed and cleaned, and abbreviations from the dataset were deleted. Subsequently, the data were first translated into Arabic using the Google and Microsoft Translate APIs. Then, each entry was examined manually to ensure accuracy and correctness. The dataset was then vectorized using TF-IDF and subsequently fed into the following algorithms: RF, MNB, SVM, and LR. Of these algorithms, SVM achieved the highest accuracy.

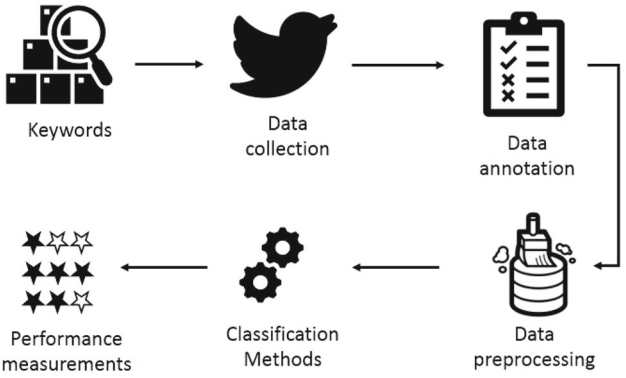


Fig. 1 Architecture of suicide detection methodology

3 Methodology

The task of classifying suicide-related posts aims to identify individuals with suicidal tendencies. Accordingly, the detection of suicidal ideation in social media content is often formulated as a supervised learning classification problem. Different machine learning and deep learning techniques have been previously employed for this task. To identify Arabic suicidal sentiment on social media, we initially collected Twitter data using suicidality keywords extracted in previous studies. After removing duplicate tweets, we annotated and labeled the data, applied feature extraction and preprocessing, and deployed machine learning and deep learning models trained on different feature representation sets to identify suicidal content. Finally, we analyzed the performance of the utilized approaches. Figure 1 illustrates the procedure used in this study, which follows that of most studies discussed in Sect. 2, and in particular the studies [1, 10, 16].

3.1 Data Collection

The data collection process was performed using Tweepy, an open-source Python library used to access the Twitter API provided by the developers of Twitter. An Arabic tweet corpus was developed from Arabic tweets written in different dialects (e.g., عايز اموت، ابى اموت، بدي موت). Between August 23, 2021 and April 21, 2022, 47,292 tweets were gathered using Arabic suicidal keywords translated from previous English-language studies, as shown in Table 1. Following the removal of duplicated tweets, our dataset encompassed 5,719 tweets.

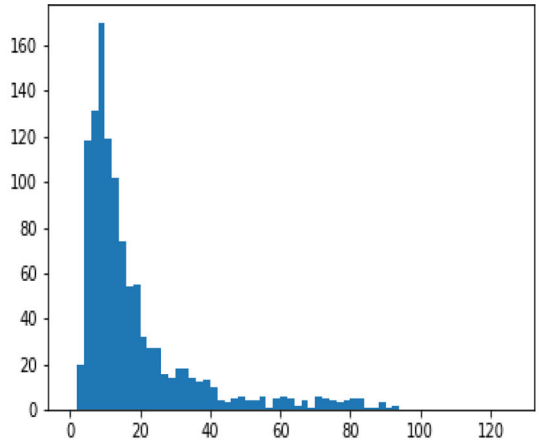
Table 1 Suicidal keywords

Reference	Keywords
[9]	I want to kill myself
	I want to die
	I want to disappear
[10]	Suicidal; suicide
	Kill myself;end my life
	Never wake up; sleep forever
	Want to die; be dead
	Better off dead
	Tired of living
	Don't want to be here
	Die alone
[42]	Just want to sleep forever
	Kill myself
	Life is so meaningless
	Tired of being alone
	Don't want to exist
	Life is worthless
	Don't want to live
	My life is pointless
	My life is this miserable
	My life isn't worth
	Want to be dead
	Hate my life
	Want to disappear
	Hate myself
	Suicidal/suicide suicida
	Isn't worth living
[16]	I want to end my life
	I'm feeling so bad
	I'm going to kill myself

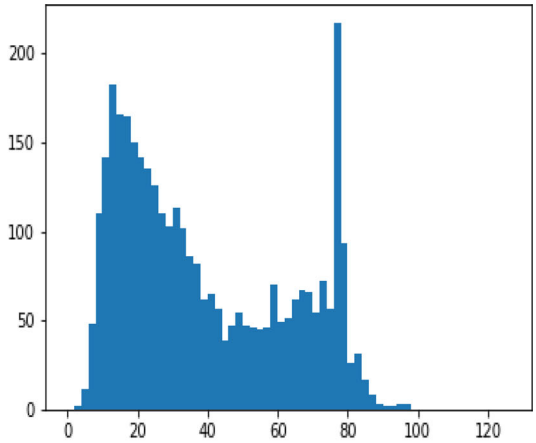
3.2 Data Annotation

All collected tweets were labeled manually by two annotators, one of whom is an expert in cyberpsychology. The annotators were asked to read only the textual content of each tweet and rate the corresponding level of concern for suicidality. Two levels of concern were defined: suicidal tweets were labeled with 1, and non-suicidal tweets were labeled with 0 [10].

- Suicidal: tweet shows serious suicidal ideation; the person expresses a deep and personal intention to commit suicide, e.g., "I will kill myself," "I want to die," "I think of commit suicide," أقتل نفسي، نفسي اموت، افكر انتحر. In contrast, a suicide risk will not be considered if the content is relevant to a conditional event; e.g., "I will kill myself if my team doesn't win today," "I wish to die once the doctor says quiz," "I'm thinking of committing suicide if the third part doesn't come out," حقتل نفسي لو ما فاز الاتحاد اليوم ، افكر انتحر



(a) Average Sentence Lengths for Suicidal Tweets



(b) Average Sentence Lengths for Non-Suicidal Tweets

Fig. 2 Average sentence lengths for suicidal and non-suicidal tweets

Table 2 Labeled tweets and their weights

Class	%of Datasets
Suicidal	25%
Non-suicidal	75%

Table 3 Quantified agreement in terms of Cohen's kappa

	A	B	Total
A	4258	8	4293
B	39	1387	1426
Total	4324	1395	5719

لو مانزلوا جزء ثالث , unless this event is a serious suicidal risk factor such as abuse, drug use, or bullying, e.g., "Her words are beyond my power, I want to die." كلامها تجاوز قوتي ابى اموت.

- Non-Suicidal: no evidence to indicate a possibility of suicide, e.g., "I lost my mind when I signed myself up as a leader," "I thought of committing suicide." وين كان عقلي يوم سجلت ليدر؟ افكر انتحر

After annotating each tweet, 1,426 tweets were labeled as Suicidal, with the remaining 4,293 labeled Non-Suicidal. Table 2 lists the dataset classes and their weights. We observed most suicidal tweets to be short, ranging from 2 to 20 words, as shown in Fig. 2a, while non-suicidal tweets varied from 2 to 80 words, as shown in Fig. 2b. The most frequent words used in each class are depicted as word cloud representations in Fig. 3a and b. As shown in Fig. 4a and b, the most frequent word used in suicidal tweets was أموتْ, with more than 800 occurrences. As shown in Fig. 5a and b, suicidal tweets peak after ten o'clock at night and continue to increase, with a clear decrease at five o'clock in the afternoon.

3.3 Annotators Agreements

After labeling the dataset, the inter-rater agreement was assessed using Cohen's kappa to measure the pairing agreement between annotators after accounting for the likelihood of chance agreement. Cohen's kappa is calculated using the following formula:

$$K = \frac{P(A) - P(E)}{1 - P(E)} \tag{1}$$

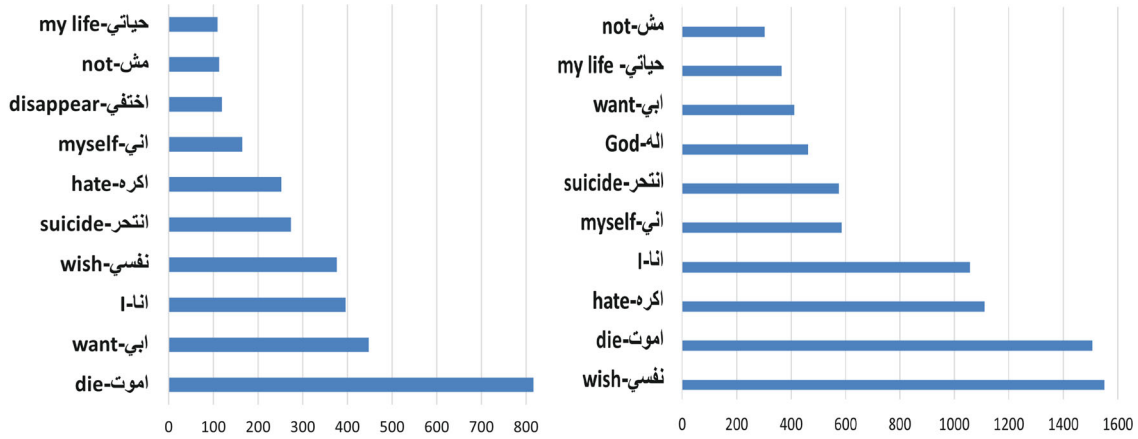
where $P(A)$ is the number of times the annotators agreed, and $P(E)$ is the number of times the annotators were expected to agree, calculated in accordance with the aforementioned intuitive argument [43]. Thus, K equals zero if there is no agreement beyond the level expected by chance. Conversely, K is one in cases of complete agreement. In this study, the level of agreement between the annotators was 0.978, indicating almost perfect reliability. Table 3 shows the quantified agreement between the two annotators.

3.4 Feature Extraction

We elected not to use any data preparation methods on our dataset, as our objective was to examine the unfiltered emotions and subtleties conveyed by tweets, highlighting the authenticity of data. Social media chats pertaining to suicide frequently encompass a wide range of emotions, slang terms, and unusual phrase structures that are essential for interpreting the mental health of participants. Accordingly, we avoided data preparation to retain the text's validity and

(a) Word cloud for Suicidal tweets (b) Word cloud for Non-Suicidal tweets

Fig. 3 Word clouds for suicidal and non-suicidal tweets



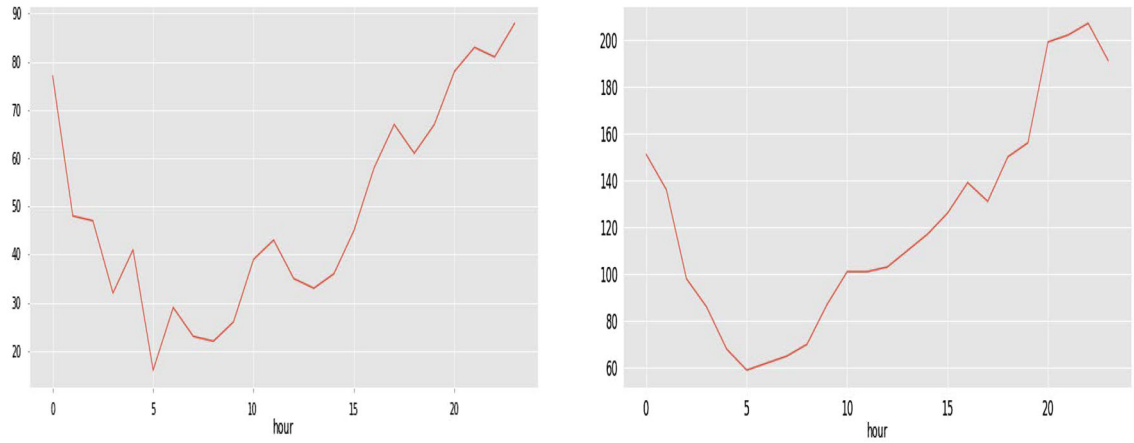
(a) Common Words Found in Suicidal Tweets (Without Stop Words) (b) Common Words Found in non-suicidal Tweets (Without Stop Words)

Fig. 4 Common words found in suicidal and non-suicidal tweets (without stop words)

all crucial information. Although preprocessing is frequently applied to enhance model performance and reduce noise, we considered the preservation of unfiltered expressions to be crucial in obtaining a profound understanding of emotions in the context of mental health. Because we recognize that preprocessing may provide advantages such as standardization and noise reduction, we intend to investigate the impact of preprocessing on classification performance in subsequent studies. However, we omitted the preprocessing stage in the present study to faithfully reflect the full spectrum of attitudes conveyed in suicidal tweets.

Feature extraction is a critical task in machine learning and NLP, transforming textual data to numerical representations. The features used in this study include word frequency features (i.e., BOW and TF-IDF), (Generic), word-embedding features (i.e., word2vec, FastText), and contextualized word embeddings for deep learning models (i.e., AraBert, Ara-ELECTRA, and AraGPT2).

- Bag-of-words: BOW is a standard word frequency model that converts textual data into numerical representations based on their occurrence in the document. The BOW model represents each document in a dataset as a feature vector of |V|, V is the unique vocabulary of the corpus and each unit in the vector represents the occurrence of known words. The BOW model is one of the most commonly used text-modeling approaches in NLP and information retrieval. However, this model has several limitations, as it ignores grammatical structure and word order, and represents textual data very sparsely [44].



(a) Trends of active hours for Suicidal tweets. (b) Word cloud for Non-Suicidal tweets

Fig. 5 Trends of active hours for suicidal and non-suicidal

- TF-IDF: TF-IDF is a word frequency model that weights word counts by the number of times each word appears in the corpus. TF-IDF is used to enhance the performance of the standard word frequency approach by assigning lower weights to frequently occurring words using the following formula:

$$\mathrm{TF} - \mathrm{IDF} = \mathrm{TF}_{i,j} * \mathrm{IDF}_i \tag{2}$$

where $\mathrm{TF}_{i,j}$ is the number of times word i appears in document j divided by the total number of words in j and IDF_i is the logarithm of the total number of documents in the corpus divided by the number of documents containing i.
All n-gram features are normalized by TF-IDF value to determine their relative importance in the corpus.

  - Unigram: The basic type of textual feature, consisting of an individual word TF-IDF value.
  - N-gram: a combination of bigrams and trigrams.
  - Character n-gram: based on single characters, or sequences of characters, rather than whole words or sentences. To calculate the frequency of each character in each document, each character is treated as a "word." Consequently, a matrix is created where each row corresponds to a document and each column corresponds to a character. The standard TF-IDF formula is then applied to this matrix to obtain a weighted representation of the character-level features. This technique is particularly important for morphologically rich languages such as Arabic, as it can identify the morphological components of words. It is also useful for identifying misspellings and alternative spellings that are common in online communications [45]. The character n-gram has previously achieved state-of-the-art performance on several text classification tasks [46]

- Word-embedding features: Word embeddings have achieved a significant advancement in numerous NLP applications, including sentiment analysis, topic segmentation, text mining and, recommendation [47–49]. Most studies in this field used one of two common word-embedding models: word2vec and FastText. A study by Kaibi et al. [50] demonstrated that Fasttext, followed by word2vec, exhibits the best performance in Twitter sentiment analysis [50].

  - Word2vec: One of the most popular word-embedding models, created by Mikolov et al. [51], word2vec is based on two neural network architectures: skip-gram and continuous bag of words (CBOW). Upon accepting the word as input, skip-gram predicts the surrounding words as output, whereas CBOW uses the context words to predict the word. The Arabic word2vec (AraVec) model provides powerful and free-to-use word embeddings for Arabic NLP research. It is built on three different Arabic content domains including Tweets, World Wide Web pages, and Arabic Wikipedia articles, encompassing more than 3.3 billion tokens [52]. AraVec has been proven to be effective in capturing semantic and syntactic relationships between Arabic words, achieving state-of-the-art results in various language processing tasks [53, 54].
  - FastText: An extension of the Skip-gram model, wherein each word is represented by a sum of n-gram vectors [55]. FastText was adapted for the Arabic language generate word embeddings for sentiment analysis tasks. The model outperformed other Arabic word embedding models on various NLP tasks, such

as sentiment analysis and named entity recognition [56].

3.5 Models and Hyperparameters

In this experiment, the performance of three deep learning models—AraBert, AraELECTRA, and AraGPT2—was compared with that of five popular machine learning models—Gaussian NB, SVM, KNN, RF, and XGBoost—trained on different sets of textual feature representations to classify. We applied hyperparameter smoothing to the Gaussian NB model to address the zero probability issue. The smoothing parameter ranged between 1E–11 and 1E–7 according to [57, 58].

The SVM model was implemented with an RBF kernel, with C values ranging from 1 to 10 and gamma values ranging from 0.1 to 1 to ensure optimal performance. These ranges were selected based on previous Arabic sentiment analysis studies, which included comprehensive comparative analyses of hyperparameter tuning techniques [58]. For the KNN algorithm, the Euclidean distance was used as a similarity measure, with k-nearest values selected in the range between 1 and 31. The best k values were 2 in the case of BOW, 3 in the case of unigrams, and 30 in the cases of n-grams and character n-grams. For the RF model, the parameters "max_features" and "n_estimators" were set to log2 and 1000, respectively, as they improve accuracy compared to the default values, as also reported in [58]. We set "max_features" based on the highest performance from ['auto,' 'sqrt,' 'log2'], whereas "n_estimators" was chosen from [100,200,300,1000]. In the case of XGBoost, "n_estimators" was set to 200 and 300, as in [59]

AraBERT is a pre-trained Arabic language model trained utilizing a masked language modeling objective on a large corpus of Arabic textual data. AraBERT has previously delivered state-of-the-art results on a variety of Arabic language processing tasks [60]. Other recent pre-trained Arabic language models include AraELECTRA and ARAGPT2. AraELECTRA is a variant of the ELECTRA model, which pre-trains a language model which is pre-trained using a generator-discriminator architecture. It has been demonstrated to achieve competitive performance on a number of Arabic natural language processing tasks [61]. ARAGPT2 is a variant of the transformer-based GPT-2 model pre-trained on a large corpus of Arabic text data. Previously, ARAGPT2 has demonstrated state-of-the-art performance on several Arabic language processing tasks, including text generation and language modeling [62]. A batch size of 16, epoch size of 5, adam_epsilon value of 1e–8, and learning rate of 2e–5 were set to fine-tune the models.

3.6 Performance Evaluation

To evaluate system performance and evaluate model capacity, we considered the measurements of precision, recall, F1 score, and accuracy defined as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{3}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{4}$$

$$F1 = \frac{2 * (\text{Precision} * \text{Recall})}{\text{Precision} + \text{Recall}} \tag{5}$$

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

where TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively. Precision is the result of dividing the number of true positives by the total number of positive instances, reflecting the model's efficacy in identifying positive instances. Recall is the result of dividing the number of true positive outcomes by the total number of instances, representing the capture of positive instances. The F1 score is the harmonic mean of recall and precision. Accuracy, represented by the percentage of correctly classified results among all data instances, reflects the overall correctness of model prediction.

In addition to using the aforementioned conventional evaluation metrics, our study represents the first attempt to utilize Arabic textual data to classify suicidality sentiment using machine learning. Our dataset represents a new and unexplored resource for the research community, having been specially curated for the detection of suicidal tendencies in Arabic social media content. We achieved satisfactory performance by utilizing this dataset and putting forth a comprehensive methodology that integrates deep learning. In addition to providing a baseline for Arabic tweet categorization, our study sheds light on the recognition and mitigation of risk factors associated with self-harm among Arabic-speaking individuals.

4 Result Analysis

We adopted the BOW, TF-IDF with different levels, and word-embedding features for representation. Five machine learning models were deployed and compared against deep learning models. We measured accuracy, precision, recall, and F1 score as evaluation metrics. Table 4 presents a comparison of the classifiers. The results show that character n-gram TF-IDF features were associated with the best performance for all machine learning models. Specifically, the SVM with character n-gram features achieved the highest accuracy among the conventional machine learning models.

Table 4 Performance analysis of different algorithms in classifying suicidal tweets

Classifier	Feature	Precision	Recall	F1-score	Accuracy
NB	Bag of word	73%	73%	73%	81%
	Unigram	75%	72%	74%	82%
	N-Gram	73%	**75%**	74%	80%
	CharLevel*	**81%**	72%	**75%**	**84%**
	Word2Vec	77%	54%	51%	76%
	Fasttext	76%	62%	63%	78%
SVM	Bag of word	82%	**78%**	**79%**	**86%**
	Unigram	84%	76%	**79%**	86%
	N-Gram	**86%**	72%	76%	85%
	CharLevel*	85%	76%	**79%**	**86%**
	word2vec	78%	66%	68%	80%
	Fasttext	77%	72%	74%	81%
KNN	Bag of word	63%	68%	60%	63%
	Unigram	75%	76%	75%	81%
	N-Gram	75%	69%	71%	81%
	CharLevel*	**78%**	75%	**76%**	**84%**
	Word2Vec	64%	71%	65%	77%
	Fasttext	71%	**77%**	73%	81%
RF	Bag of word	77%	76%	**77%**	83%
	Unigram	83%	72%	75%	85%
	N-Gram	77%	73%	75%	85%
	CharLevel*	**85%**	74%	**77%**	**86%**
	Word2Vec	81%	65%	67%	81%
	Fasttext	69%	**81%**	72%	82%
XGBoost	Bag of word	79%	67%	69%	82%
	Unigram	**83%**	69%	72%	84%
	N-Gram	79%	60%	62%	80%
	CharLevel*	82%	**75%**	78%	**85%**
	Word2Vec	64%	78%	66%	79%
	Fasttext	81%	68%	**80%**	70 %
AraBert	–	**88%**	**89%**	**88%**	**91%**
AraELECTRA	–	85%	84%	85%	88%
AraGPT2	–	87%	85%	86%	89%

*Indicates the best-performing feature extraction technique
Bold Indicates highest score for each algorithm in each metric

Figure 6 presents the receiver operating characteristic (ROC) curve obtained for the five machine learning algorithms.

Furthermore, all deep learning models outperformed the machine learning models in identifying suicidal ideation from Arabic tweets. The AraBert model achieved the best overall performance with 88% precision, 89% recall, 88% F1-score, and 91% accuracy. The confusion matrix of AraBert is shown in Fig. 7. Based on these results, it is important to use advanced deep learning models to accurately identify suicidal sentiment in Arabic tweets. With regard to identifying and categorizing content associated with mental health, the superior performance exhibited by AraBert emphasizes the importance of contextual data obtained by pre-trained transformers.

This study reveals important information for the development of suicidality detection systems in Arabic social media, demonstrating the superior performance of deep learning models and SVM with character n-gram TF-IDF features. It is important to note that by expanding the dataset and investigating different machine learning and deep learning strategies, the performance of the aforementioned models can be further improved. Subsequent studies may be conducted to enhance the resilience and generalizability of detection systems. In conclusion, our findings highlight the accuracy achieved by deep learning and character n-gram features in

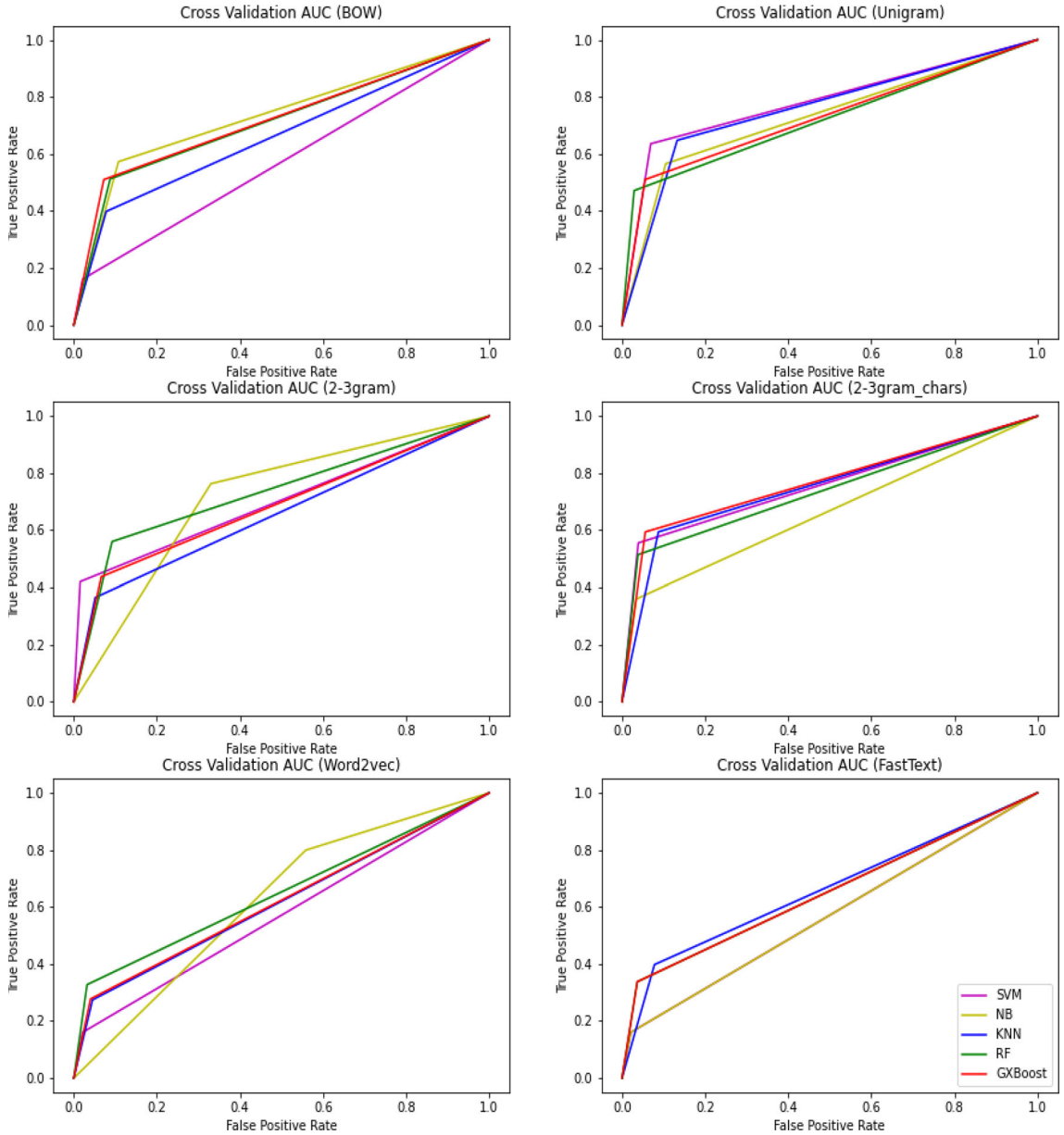


Fig. 6 ROC curves depicting performance of machine learning algorithms

the detection of suicidal sentiment within Arabic textual data. The results pave the way for proactive interventions within the Arabic-speaking community and beyond by furthering our understanding of machine learning techniques for the early diagnosis and prevention of self-harm and suicide.

5 Conclusion

Our study demonstrates that deep learning models and machine learning algorithms—especially pre-trained transformer models including AraBert, AraELECTRA, and AraGPT2—can be used to identify suicidal ideation in Arabic textual data. We evaluated the performance of our proposed strategy through comprehensive experiments, including the creation and annotation of a fresh Arabic suicidality dataset. Pre-trained transformer models were observed to outperform conventional machine learning models trained on various feature sets. In particular, AraBert detected suicidal intent within textual data with exceptional accuracy, recall, precision, and F1 score.

These findings demonstrate the effectiveness of deep learning methods, as well as the significance of using contextual data incorporated in pre-trained transformer models. Owing to its effectiveness, our strategy may offer opportunities for the early identification and treatment of self-harm

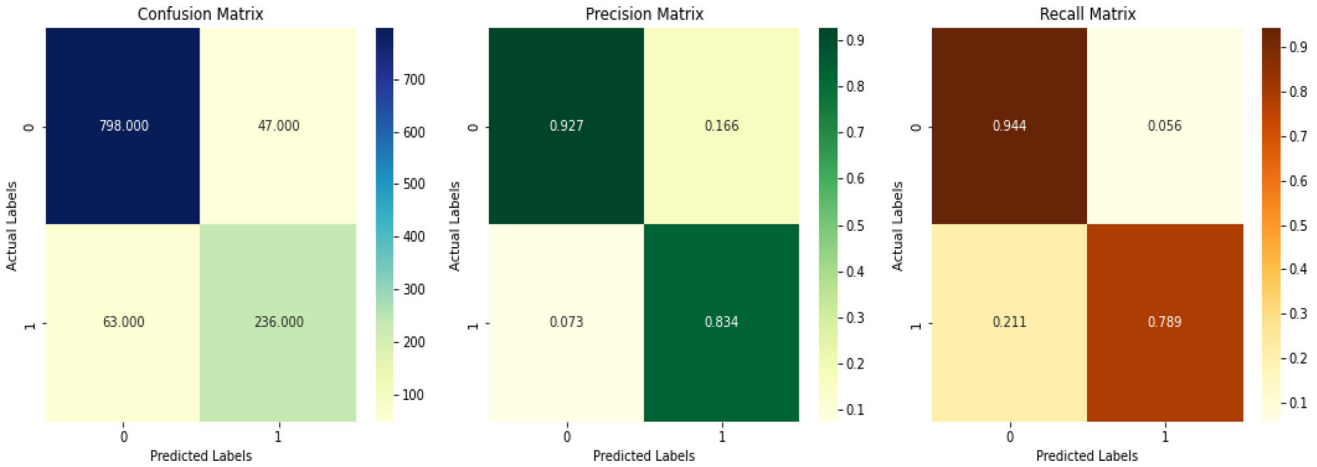


Fig. 7 Confusion matrix for best-performing model (AraBert)

concerns among Arabic-speaking individuals. Additionally, our work adds to the expanding corpus of information on the use of machine learning in social media to identify and treat mental health issues. The development of the Arabic suicide dataset and the use of machine learning offer important insights into the challenges and linguistic subtleties involved in detecting suicidal tendencies within Arabic textual data. Subsequent research endeavors may concentrate on augmenting the dataset dimensions and investigating alternate methodologies to account for disparate feature selection strategies, further improving the robustness and generalizability of our detection system. Ultimately, our research advances the study of online mental health expression and uses machine learning to achieve the greater goal of minimizing suicide and self-harm. We aim to make a positive impact on individuals' lives and advance mental health within the Arabic-speaking community and beyond by providing proactive treatments and support.

Future research can concentrate on a number of topics to enhance the model's identification capabilities. Expanding the dataset to capture a wider range of linguistic patterns is one such option, as is addressing class imbalance, looking into additional features, adopting different model architectures, using transfer learning, accounting for multimodal approaches, looking into domain adaptation techniques, using thorough evaluation metrics, and creating real-time monitoring systems. Further improvements may be achieved by addressing these elements to develop stronger and more accurate models for early intervention and prevention of suicide and self-harm in the Arabic-speaking community.

References

1. Astoveza, G.; Obias, R.J.P.; Palcon, R.J.L.; Rodriguez, R.L.; Fabito, B.S.; Octaviano, M.V.: Suicidal behavior detection on twitter using neural network. In: TENCON 2018–2018 IEEE Region 10 Conference, pp. 0657–0662 (2018). https://doi.org/10.1109/TENCON.2018.8650162
2. De Choudhury, M.; Kiciman, E.; Dredze, M.; Coppersmith, G.; Kumar, M.: Discovering shifts to suicidal ideation from mental health content in social media. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. CHI '16, pp. 2098–2110. Association for Computing Machinery, New York (2016). https://doi.org/10.1145/2858036.2858207
3. Beck, A.T.; Kovacs, M.; Weissman, A.: Assessment of suicidal intention: the scale for suicide ideation. J. Consult. Clin. Psychol. **47**(2), 343 (1979)
4. Liu, D.; Fu, Q.; Wan, C.; Liu, X.; Jiang, T.; Liao, G.; Qiu, X.; Liu, R.: Suicidal ideation cause extraction from social texts. IEEE Access **8**, 169333–169351 (2020)
5. Nock, M.K.; Borges, G.; Bromet, E.J.; Cha, C.B.; Kessler, R.C.; Lee, S.: Suicide and suicidal behavior. Epidemiol. Rev. **30**(1), 133–154 (2008)
6. Weber, A.N.; Michail, M.; Thompson, A.; Fiedorowicz, J.G.: Psychiatric emergencies: assessing and managing suicidal ideation. Med. Clin. **101**(3), 553–571 (2017)
7. Ramírez-Cifuentes, D.; Freire, A.; Baeza-Yates, R.; Puntí, J.; Medina-Bravo, P.; Velazquez, D.A.; Gonfaus, J.M.; Gonzàlez, J.: Detection of suicidal ideation on social media: multimodal, relational, and behavioral analysis. J. Med. Internet Res. **22**(7), 17758 (2020)
8. Narynov, S.; Mukhtarkhanuly, D.; Kerimov, I.; Omarov, B.: Comparative analysis of supervised and unsupervised learning algorithms for online user content suicidal ideation detection. J. Theor. Appl. Inf. Technol. **97**(22), 3304–3317 (2019)
9. Fahey, R.A.; Boo, J.; Ueda, M.: Covariance in diurnal patterns of suicide-related expressions on twitter and recorded suicide deaths. Soc. Sci. Med. **253**, 112960 (2020). https://doi.org/10.1016/j.socscimed.2020.112960
10. O'Dea, B.; Wan, S.; Batterham, P.J.; Calear, A.L.; Paris, C.; Christensen, H.: Detecting suicidality on twitter. Internet Interv. **2**(2), 183–188 (2015). https://doi.org/10.1016/j.invent.2015.03.005
11. Burnap, P.; Colombo, W.; Scourfield, J.: Machine classification and analysis of suicide-related communication on twitter. In: Proceedings of the 26th ACM Conference on Hypertext and Social Media. HT '15, pp. 75–84. Association for Computing Machinery, New York (2015). https://doi.org/10.1145/2700171.2791023

12. Vioules, M.J.; Moulahi, B.; Azé, J.; Bringay, S.: Detection of suicide-related posts in twitter data streams. IBM J. Res. Dev. **62**(1), 7–11 (2018)
13. Moulahi, B.; Azé, J.; Bringay, S.: Dare to care: a context-aware framework to track suicidal ideation on social media. In: International Conference on Web Information Systems Engineering, pp. 346–353. Springer (2017)
14. Rajesh Kumar, E.; Rama Rao, K.; Nayak, S.R.; Chandra, R.: Suicidal ideation prediction in twitter data using machine learning techniques. J. Interdiscip. Math. **23**(1), 117–125 (2020)
15. Tadesse, M.M.; Lin, H.; Xu, B.; Yang, L.: Detection of suicide ideation in social media forums using deep learning. Algorithms **13**, 1 (2020). https://doi.org/10.3390/a13010007
16. Ji, S.; Yu, C.P.; Fung, S.-f.; Pan, S.; Long, G.: Supervised learning for suicidal ideation detection in online user content. Complexity **2018** (2018)
17. Chiroma, F.; Liu, H.; Cocea, M.: Suiciderelated text classification with prism algorithm. In: 2018 International Conference on Machine Learning and Cybernetics (ICMLC), vol. 2, pp. 575–580. IEEE (2018)
18. Du, J.; Zhang, Y.; Luo, J.; Jia, Y.; Wei, Q.; Tao, C.; Xu, H.: Extracting psychiatric stressors for suicide from social media using deep learning. BMC Med. Inform. Decis. Mak. **18**(2), 77–87 (2018)
19. Huang, X.; Zhang, L.; Chiu, D.; Liu, T.; Li, X.; Zhu, T.: Detecting suicidal ideation in chinese microblogs with psychological lexicons. In: 2014 IEEE 11th International Conference on Ubiquitous Intelligence and Computing and 2014 IEEE 11th International Conference on Autonomic and Trusted Computing and 2014 IEEE 14th International Conference on Scalable Computing and Communications and Its Associated Workshops, pp. 844–849. IEEE (2014)
20. Huang, X.; Li, X.; Liu, T.; Chiu, D.; Zhu, T.; Zhang, L.: Topic model for identifying suicidal ideation in chinese microblog. In: Proceedings of the 29th Pacific Asia Conference on Language, Information and Computation, pp. 553–562 (2015)
21. Benlaaraj, O.; El Jaafari, I.; Ellahyani, A.; Boutaayamou, I.: Prediction of suicidal ideation in a new arabic annotated dataset. In: 2022 9th International Conference on Wireless Networks and Mobile Communications (WINCOM), pp. 1–5 (2022). https://doi.org/10.1109/WINCOM55661.2022.9966481
22. Alqurashi, S.; Alhindi, A.; Alanazi, E.: Large Arabic Twitter Dataset on COVID-19. arXiv (2020). https://arxiv.org/abs/2004.04315v1
23. Boudad, N.; Faizi, R.; Oulad Haj Thami, R.; Chiheb, R.: Sentiment analysis in Arabic: A review of the literature. Ain Shams Eng. J. **9**(4), 2479–2490 (2018). https://doi.org/10.1016/j.asej.2017.04.007
24. Farghaly, A.; Shaalan, K.: Arabic natural language processing: challenges and solutions. ACM Trans. Asian Lang. Inf. Process. **8**(4), 21 (2009). https://doi.org/10.1145/1644879.1644881
25. Elnagar, A.: Investigation on sentiment analysis for arabic reviews. In: 2016 IEEE/ACS 13th International Conference of Computer Systems and Applications (AICCSA), pp. 1–7 (2016). https://doi.org/10.1109/AICCSA.2016.7945623
26. Tausczik, Y.R.; Pennebaker, J.W.: The psychological meaning of words: Liwc and computerized text analysis methods. J. Lang. Soc. Psychol. **29**(1), 24–54 (2010). https://doi.org/10.1177/0261927X09351676
27. Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G.S.; Dean, J.: Distributed representations of words and phrases and their compositionality. In: Advances in Neural Information Processing Systems, vol. 26 (2013)
28. Pennington, J.; Socher, R.; Manning, C.D.: Glove: Global vectors for word representation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1532–1543 (2014)
29. Blei, D.M.; Ng, A.Y.; Jordan, M.I.: Latent dirichlet allocation. J. Mach. Learn. Res. **3**(Jan), 993–1022 (2003)
30. Chiroma, F.; Liu, H.; Cocea, M.: Text classification for suicide related tweets. In: 2018 International Conference on Machine Learning and Cybernetics (ICMLC), vol. 2, pp. 587–592 (2018). https://doi.org/10.1109/ICMLC.2018.8527039
31. Huang, Y.; Liu, X.; Zhu, T.: Suicidal ideation detection via social media analytics. In: Milošević, D., Tang, Y., Zu, Q. (eds.) Human Centered Computing, pp. 166–174. Springer, Cham (2019)
32. Sakib, T.H.; Ishak, M.; Jhumu, F.F.; Ali, M.A.: Analysis of suicidal tweets from twitter using ensemble machine learning methods. In: 2021 International Conference on Automation, Control and Mechatronics for Industry 4.0 (ACMI), pp. 1–7 (2021). https://doi.org/10.1109/ACMI53878.2021.9528252
33. Chadha, A.; Kaushik, B.: Machine learning based dataset for finding suicidal ideation on twitter. In: 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), pp. 823–828 (2021). https://doi.org/10.1109/ICICV50876.2021.9388638
34. Colombo, G.B.; Burnap, P.; Hodorog, A.; Scourfield, J.: Analysing the connectivity and communication of suicidal users on twitter. Comput.Commun. **73**, 291–300 (2016). https://doi.org/10.1016/j.comcom.2015.07.018
35. Rezig, A.A.: A novel optimizer technique for suicide prediction in twitter environment. In: 2021 International Conference on Information Systems and Advanced Technologies (ICISAT), pp. 1–5 (2021). https://doi.org/10.1109/ICISAT54145.2021.9678419
36. Metzler, H.; Baginski, H.; Niederkrotenthaler, T.; Garcia, D.: Detecting potentially harmful and protective suicide-related content on twitter: machine learning approach. J. Med. Internet Res. **24**(8), 34705 (2022). https://doi.org/10.2196/34705
37. Haque, R.; Islam, N.; Islam, M.; Ahsan, M.M.: A comparative analysis on suicidal ideation detection using nlp, machine, and deep learning. Technologies (2022). https://doi.org/10.3390/technologies10030057
38. Chatterjee, M.; Samanta, P.; Kumar, P.; Sarkar, D.: Suicide ideation detection using multiple feature analysis from twitter data. In: 2022 IEEE Delhi Section Conference (DELCON), pp. 1–6 (2022). https://doi.org/10.1109/DELCON54057.2022.9753295
39. Chadha, A.; Kaushik, B.: Performance evaluation of learning models for identification of suicidal thoughts. Comput. J. **65**(1), 139–154 (2021). https://doi.org/10.1093/comjnl/bxab060
40. Mashaly, M.: Suicide notes (2020). https://www.kaggle.com/mohanedmashaly/suicide-notes
41. Sonu, I.: Suicidal phrases (2020). https://www.kaggle.com/imeshsonu/suicideal-phrases
42. Valeriano, K.; Condori-Larico, A.; Sulla-Torres, J.: Detection of suicidal intent in Spanish language social networks using machine learning. Int. J. Adv. Comput. Sci. Appl. **11**(4), (2020)
43. Sidney, S.: Nonparametric statistics for the behavioral sciences. J. Nerv. Ment. Dis. **125**(3), 497 (1957)
44. Deepa, D.; Tamilarasi, A.; et al.: Sentiment analysis using feature extraction and dictionary-based approaches. In: 2019 Third International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC), pp. 786–790. IEEE (2019)
45. Alsafari, S.; Sadaoui, S.; Mouhoub, M.: Hate and offensive speech detection on arabic social media. Online Soc. Netw. Media **19**, 100096 (2020)
46. Zhang, X.; Zhao, J.; LeCun, Y.: Character-level convolutional networks for text classification. In: Advances in Neural Information Processing Systems, vol. 28 (2015)
47. Xu, Y.; Liu, J.; Yang, W.; Huang, L.: Incorporating latent meanings of morphological compositions to enhance word embeddings. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 1232–1242 (2018)

48. Li, Q.; Shah, S.; Liu, X.; Nourbakhsh, A.: Data sets: word embeddings learned from tweets and general data. Proc. Int. AAAI Conf. Web Social Media **11**(1), 428–436 (2017). https://doi.org/10.1609/icwsm.v11i1.14859
49. Naili, M.; Chaibi, A.H.; Ghezala, H.H.B.: Comparative study of word embedding methods in topic segmentation. Procedia Comput. Sci. **112**, 340–349 (2017)
50. Kaibi, I.; Satori, H.; et al.: A comparative evaluation of word embeddings techniques for twitter sentiment analysis. In: 2019 International Conference on Wireless Technologies, Embedded and Intelligent Systems (WITS), pp. 1–4. IEEE (2019)
51. Mikolov, T.; Chen, K.; Corrado, G.; Dean, J.: Efficient estimation of word representations in vector space. arXiv:1301.3781 (2013)
52. Soliman, A.B.; Eissa, K.; El-Beltagy, S.R.: Aravec: a set of arabic word embedding models for use in arabic nlp. Procedia Comput. Sci. **117**, 256–265 (2017). https://doi.org/10.1016/j.procs.2017.10.117
53. Al-Rfou, R.; Kulkarni, V.; Perozzi, B.; Skiena, S.: POLYGLOT-NER: massive multilingual named entity recognition, pp. 586–594. https://doi.org/10.1137/1.9781611974010.66. https://epubs.siam.org/doi/abs/10.1137/1.9781611974010.66
54. Heikal, M.; Torki, M.; El-Makky, N.: Sentiment analysis of arabic tweets using deep learning. Procedia Comput. Sci. **142**, 114–122 (2018). https://doi.org/10.1016/j.procs.2018.10.466
55. Athiwaratkun, B.; Wilson, A.; Anandkumar, A.: Probabilistic FastText for multi-sense word embeddings. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 1–11. Association for Computational Linguistics, Melbourne, Australia (2018). https://doi.org/10.18653/v1/P18-1001. https://aclanthology.org/P18-1001
56. Elhassan, N.; Varone, G.; Ahmed, R.; Gogate, M.; Dashtipour, K.; Almoamari, H.; El-Affendi, M.A.; Al-Tamimi, B.N.; Albalwy, F.; Hussain, A.: Arabic sentiment analysis based on word embeddings and deep learning. Computers (2023). https://doi.org/10.3390/computers12060126
57. John, G.H.; Langley, P.: Estimating continuous distributions in Bayesian classifiers. In: Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence. UAI'95, pp. 338–345. Morgan Kaufmann Publishers Inc., San Francisco (1995)
58. Elgeldawi, E.; Sayed, A.; Galal, A.R.; Zaki, A.M.: Hyperparameter tuning for machine learning algorithms used for arabic sentiment analysis. Informatics (2021). https://doi.org/10.3390/informatics8040079
59. Amoudi, G.; Albalawi, R.; Baothman, F.; Jamal, A.; Alghamdi, H.; Alhothali, A.: Arabic rumor detection: a comparative study. Alex. Eng. J. **61**(12), 12511–12523 (2022). https://doi.org/10.1016/j.aej.2022.05.029
60. Antoun, W.; Baly, F.; Hajj, H.: Arabert: Transformer-based model for arabic language understanding. In: LREC 2020 Workshop Language Resources and Evaluation Conference 11–16 (May 2020), p. 9
61. Antoun, W.; Baly, F.; Hajj, H.: AraELECTRA: Pre-training text discriminators for Arabic language understanding. In: Proceedings of the Sixth Arabic Natural Language Processing Workshop, pp. 191–195. Association for Computational Linguistics, Kyiv, Ukraine (Virtual) (2021). https://www.aclweb.org/anthology/2021.wanlp-1.20
62. Antoun, W.; Baly, F.; Hajj, H.: AraGPT2: Pre-trained transformer for Arabic language generation. In: Proceedings of the Sixth Arabic Natural Language Processing Workshop, pp. 196–207. Association for Computational Linguistics, Kyiv, Ukraine (Virtual) (2021). https://www.aclweb.org/anthology/2021.wanlp-1.21