



Thalassemia Screening by Sentiment Analysis on Social Media Platform Twitter

Wadhah Mohammed M. Aqlan¹, Ghassan Ahmed Ali^{2,*}, Khairan Rajab², Adel Rajab²,
Asadullah Shaikh², Fekry Olayah², Shehab Abdulhabib Saeed Alzaeemi^{3,*}, Kim Gaik Tay³,
Mohd Adib Omar¹ and Ernest Mangantig⁴

¹School of Computer Sciences, Universiti Sains Malaysia, USM, 11800, Penang, Malaysia

²College of Computer Science and Information Systems, Najran University, Najran, 61441, Saudi Arabia

³Faculty of Electrical and Electronic Engineering, Universiti Tun Hussein Onn Malaysia, 86400, Johor, Malaysia

⁴IPPT, Universiti Sains Malaysia, USM, 11800, Penang, Malaysia

*Corresponding Authors: Ghassan Ahmed Ali. Email: Alhabeab@gmail.com; Shehab Abdulhabib Saeed Alzaeemi.
Email: shehab@uthm.edu.my

Received: 16 January 2023; Accepted: 12 April 2023; Published: 09 June 2023

Abstract: Thalassemia syndrome is a genetic blood disorder induced by the reduction of normal hemoglobin production, resulting in a drop in the size of red blood cells. In severe forms, it can lead to death. This genetic disorder has posed a major burden on public health wherein patients with severe thalassemia need periodic therapy of iron chelation and blood transfusion for survival. Therefore, controlling thalassemia is extremely important and is made by promoting screening to the general population, particularly among thalassemia carriers. Today Twitter is one of the most influential social media platforms for sharing opinions and discussing different topics like people's health conditions and major public health affairs. Exploring individuals' sentiments in these tweets helps the research centers to formulate strategies to promote thalassemia screening to the public. An effective Lexicon-based approach has been introduced in this study by highlighting a classifier called valence aware dictionary for sentiment reasoning (VADER). In this study applied twitter intelligence tool (TWINT), Natural Language Toolkit (NLTK), and VADER constitute the three main tools. VADER represents a gold-standard sentiment lexicon, which is basically tailored to attitudes that are communicated by using social media. The contribution of this study is to introduce an effective Lexicon-based approach by highlighting a classifier called VADER to analyze the sentiment of the general population, particularly among thalassemia carriers on the social media platform Twitter. In this study, the results showed that the proposed approach achieved 0.829, 0.816, and 0.818 regarding precision, recall, together with F-score, respectively. The tweets were crawled using the search keywords, "thalassemia screening," thalassemia test, "and thalassemia diagnosis". Finally, results showed that India and Pakistan ranked the highest in mentions in tweets by the public's conversations on thalassemia screening with 181 and 164 tweets, respectively.



Keywords: Social media platform; Twitter; screening; thalassemia; lexicon-based; VADER

1 Introduction

Thalassemia is a significant global health problem that posed a heavy burden on the affected individual's quality of life and healthcare [1]. Patients with severe forms of thalassemia typically suffer from severe anemia, enlargement of the spleen, retardation of growth, and endocrine systems complications due to the body's inability to produce hemoglobin, a substance that transports oxygen to the whole-body tissue. Moreover, severe thalassemia patients need a long-lasting therapy of blood transfusion, in addition to costly iron chelation for survival [2]. Worldwide estimates revealed that every year more than 50,000 new affected people are inherently born with a severe thalassemia form (i.e., beta-thalassemia major, as well as HbE beta-thalassemia) and about 80 percent of the infected births take place in several developing countries [3].

For high-risk couples who are both carriers of thalassemia mutation, prenatal thalassemia screening is a must for the expecting mother. However, despite the health care effort for prenatal and premarital thalassemia screening, the incidence of new thalassemia births continues to rise [4]. Accordingly, to control this disease more effectively, screening should be promoted to the general population before marriage. This indicates that more effort is needed to improve thalassemia awareness in the general population. Therefore, understanding the public's opinion and sentiment toward thalassemia screening is important in formulating strategies to promote thalassemia screening to the public. Social media represents a crucial source, whereby the general public's opinion is grasped and, therefore, it has been progressively growing in recent years [5]. Around the globe, people share thoughts and communicate a wide range of topics like health conditions and issues related to public health on social media platforms. Sharing these pieces of information gives advantages to researchers over traditional data sources, such as real-time data availability, ease of access, and low cost. Sentiment analysis is often conducted on these thoughts to help the healthcare sector analyze people's interests and opinions, thereby helping them in making decisions and providing effective solutions [6].

Sentiment Analysis represents the most employed tool of text classification (also known as opinion mining), which analyses textual materials and classifies an underlying sentiment into positive or negative opinions, and neutral sentiments. The principal objective of sentiment analysis involves identifying the user's or audience's viewpoint on a target object by analyzing a vast amount of text from several sources. Sentiment analysis approaches are widely used in various fields, namely marketing, political, and sociological [7]. They are classified into two main methods, including the Lexicon-Based approach and the Machine Learning approach [8]. The following Table 1 to explains all abbreviations and symbols that used at this study.

Table 1: List of abbreviations and symbols

Abbreviations	Explanation	Abbreviations	Explanation
NLP	Natural Language Processing	HTML	Hyper Text Markup Language
NLTK	Natural Language Toolkit	VADER	Valence Aware Dictionary for Sentiment Reasoning

(Continued)

Table 1: Continued

Abbreviations	Explanation	Abbreviations	Explanation
API	Application Programming Interfaces	TWINT	Twitter Intelligence Tool
RegEx	Regular Expression	POST	Part of Speech Tagging

2 Lexicon-Based Approach

The Lexicon-Based approach is unsupervised as it aims to perform the analysis using the lexicon and scoring method for evaluating opinions. A sentiment lexicon is a lexical list of features, which is labeled based on the semantic orientation of such features [9,10], interpreted as positive sentiments, negative sentiments, or neutral sentiments.

This lexicon-based method has been introduced to compute polarity scores at two levels, involving the word level and the sentence level using a manually created corpus of patient-provided medical drug reviews [11]. The results showed that the proposed technique obtained 79 percent accuracy and achieved 81 percent accuracy on average at the word level and the sentence level, respectively.

Next, a sentiment lexicon is built to provide comprehensive coverage of health-related words through a hybrid approach that combines two strategies: bootstrapping and lexicon-based [12]. Furthermore, a proper polarity class can be defined for every word by suggesting a specific count-based probability measurement. The results obtained by this approach were 0.89, 0.79, and 0.83 scores in precision and recall, in addition to F1, respectively.

In [13], an approach for the analysis of the aspect-level opinion is introduced according to the SentiWordNet (SWN) lexicon and the ontologies within the domain of diabetes. The N-gram techniques were applied by the authors (i.e., N-gram-after and N-gram-before, as well as N-gram around) for calculating a given aspect-level sentiment by heeding the words around the aspect. A Twitter dataset corpus has been collected and manually labeled at the aspect level to ternary classes for evaluating the proposed approach. Furthermore, it emerges that the N-gram-around method achieves better results by 0.819, 0.811, and 0.812 in precision and recall, as well as *F*-measure, correspondingly.

Another lexicon-based method has been employed for analyzing people's thoughts and insights communicated via Twitter about Cardiovascular diseases [14]. About two million Tweets, including one of the terms cardiovascular disease or heart disease, were extracted using the Twitter application programming interfaces (API).

Lastly, Wong and his team have proposed a lexicon-based sentiment analysis model called VADER for classifying the public's attitudes on Twitter about breast cancer screening in the US. [15], linking the sentiments on Twitter to an actual screening of breast cancer patterns from the behavioral risk factor surveillance system (BRFSS) to investigate how sentiments can be possibly related to screening uptake behaviour. Based on the results, the projected method in this study achieved an accuracy of 77.2%.

3 Machine Learning Approach

Machine learning requires the application of feature extraction, as well as model training with the help of a feature set and some dataset. Indeed, this approach utilizes the classification technique for categorizing the text into classes.

A sentiment analysis approach is proposed on various medical forums devoted to Hearing Loss (HL) [16]. Evaluations were made using three different supervised learning-based classifiers like Naïve Bayes and support vector machine (SVM), as well as the algorithms of logistics regression. A feature extraction method was introduced by the authors depending on data about the speech parts. A set of experiences has been conducted using a manually labeled dataset as a positive or negative sentiment; otherwise, it is a neutral sentiment to assess the performance of this approach. The results showed that the logistics regression's performance is the best of the features selected, with an average of 0.685 on the F1 score.

In medical forums, a new approach for analyzing sentiments has been proposed by Arbane et al., [17]. They created a new lexicon called HealthAffect and used two algorithms of classification: Naïve Bayes and k -Nearest Neighbors (KNN), for solving a multiclass sentiment classification problem. The results showed that the used Naïve Bayes algorithm provided better performance compared with the KNN algorithm by achieving a 0.518 F1 score.

Another sentiment analysis method is proposed by creating a vast corpus of manually labeled data [18]. Data is collected from medical forums related to depression, anxiety, asthma, and allergy. Additionally, they built a deep Convolutional Neural Network (CNN) model for evaluation based on the medical sentiment analysis system for predicting a potential medical sentiment classification for two categories, including 'medical condition' and 'medication' categorization schemes. They achieved a significant performance with a precision of 0.86 and 0.68, respectively.

Lastly, an enhanced machine learning method has been proposed for extracting sentiments from people's tweets related to the HPV vaccine [19]. The authors manually annotated these tweets, and they completed a hierarchical categorization utilizing the SVM standard. According to the results, better performance with a 0.744 F-score was achieved compared with different baseline models.

Although the machine learning approach was successfully applied in many domains, it still has some disadvantages compared to the lexicon-based, including (1) Depending on a labeled dataset. (2) Requires a massive amount of training set with many features that are tricky to obtain on social media with short, sparse text. (3) High computational operation (time, memory, and process) is needed for the training and testing process [20,21]. On the other side, despite the lexicon-based approach demands linguistic resources, which are limited for some languages, studies showed that the analysis of the lexicon-based method outperformed the supervised machine learning techniques not just in performance but also in the economy of time and effort [22].

Therefore, this work aims to use the lexicon-based sentiment analysis method by using the VADER lexicon, which is successfully applied in many works [23–26]. Moreover, according to the prior studies, all papers have focused on sentiment analysis in particular health domains such as diabetes [13], Cardiovascular Disease [14], breast cancer screening [15], and SARS-CoV-2 [27,28].

Besides, there is a lack of published studies focusing on the thalassemia screening domain. Consequently, this paper focuses on identifying the public's sentiment polarity on thalassemia screening on Twitter using a lexicon-based approach. The subsequent sections of this paper outline in detail the architecture and components of the introduced method.

4 Methodology

4.1 Proposed Method

This study falls into three main stages. These include (1) the first stage involves acquiring data on Twitter; (2) the second stage concentrates on the initial operation of preprocessing, which was

conducted for cleaning and filtering out irrelevant information (i.e., punctuation, stop words, and retweet symbols) from the tweets; (3) the third stage involves utilizing NLTK’s VADER classifier and identifying the most frequent terms used in public conversations on thalassemia screening. The application of the scoring method was carried out to the results of VADER to assess the method’s capability of classifying the collected tweets to a three-point measurement (either positive or negative, and neutral). The architecture of the introduced approach is shown in Fig. 1.

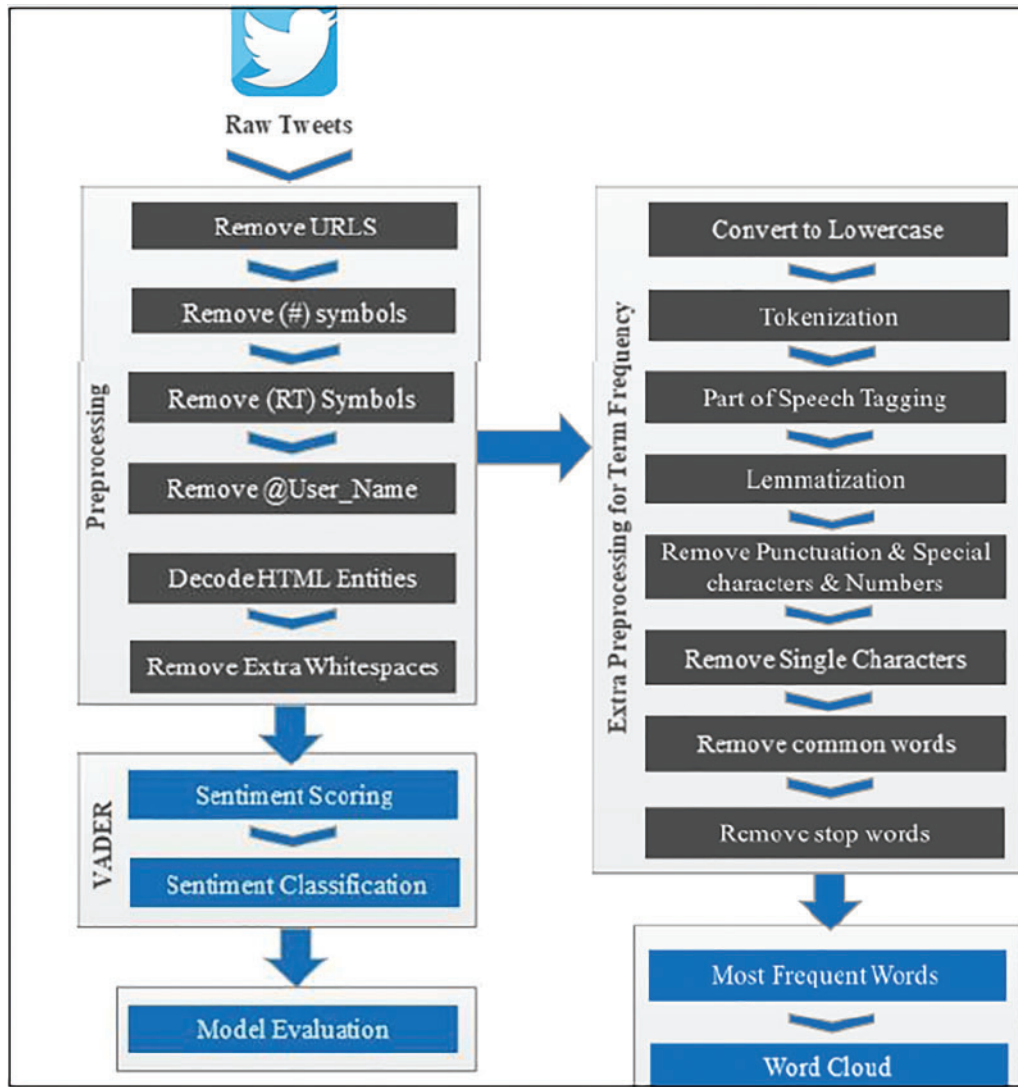


Figure 1: Proposed approach

4.2 Data Acquisition

The data in the context of this study comprise a total of 3,376 English tweets, which were gathered between February 2009 and September 2020, applying the TWINT application. The collected tweets were crawled utilizing several search keywords, e.g., “thalassemia screening.” Fig. 2 shows the number of tweets posted each year, starting from 2009 until 2020. It is observed from the Figure that the number

of tweets published related to thalassemia increases over time. This confirms how people are aware of thalassemia screening and its impact on society.

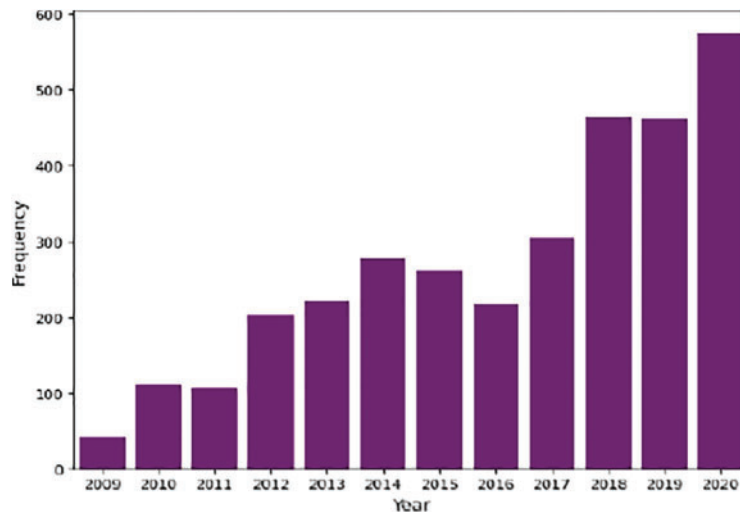


Figure 2: Number of posted tweets every year

4.3 Data Preprocessing

A tweet can be defined as a specific microblog message, which is posted on the Twitter platform and is confined to a total of 280 characters. Often, users do not apply the proper language structure when they post opinions or sentiments about a specific topic. Instead, slang, misspelling, different emoticons, and abbreviations, sometimes puns, complicates the analysis of these structures. Complex textual data significantly influence the performance of analyzing sentiments because the quality of the output depends on the input [29]. Therefore, a preprocessing steps series have been performed for removing irrelevant data from the collected tweets because the cleaner the information is, the more appropriate this information will be for mining, as well as feature extraction, thereby improving the findings' accuracy and precision [30,31]. For the preprocessing, we have utilized Python's natural language processing (NLP) toolkit (NLTK). Initially, a given regular expression, i.e., (Regex) in Python is applied for detecting and eliminating uniform resource locator (URL) links, retweet symbols (RT), user mentions (@), and decoding hypertext markup language (HTML) Entities with equivalent characters. Hashtags (#) describe the subject of the tweet. They carry useful information about the topic of the tweet, included as a particular part of a given tweet, except for the symbol "#", which is deleted [31]. Once the preprocessing steps are complete, a dataset is now ready for the classification of sentiments by using the algorithm of VADER. Various NLTK functions are also used as additional preprocessing steps and are required to calculate terms' frequency and thus to visualize in Word Cloud. These steps include:

1. Convert the tweets to lowercase to reduce redundancy.
2. Tokenize these tweets into separate tokens or words.
3. Apply part of speech (POS) tagging (i.e., adjectives, adverbs, verbs, and nouns). The reason behind the application of POS tagging prior to the lemmatization and elimination operation involves retaining the structure of linguistic categories and maintaining the quality of the analysis.

4. Lemmatize the tweets by eliminating inflectional endings; thus, obtaining the original dictionary or base form of this word utilizing the NLTK wordnet lemmatizer.
5. Remove the special characters and numbers.
6. Remove single characters.
7. Remove the common query words that emerge in almost all tweets (i.e., thalassemia, screening, test, and diagnosis).
8. Remove stop words that do not express any meaning in tweets (i.e., a, the, he, them, etc.).

4.4 Sentiment Classification

The analysis of sentiment can be generally used for examining the context polarity, whether a positive or negative opinion and a neutral sentiment. During this phase, a largely approved human-validated sentiment analysis instrument was used, i.e., Valence Aware Dictionary for Sentiment Reasoning (VADER). This analysis tool was used because it represents a lexicon-based sentiment analysis engine, which is specially tailored for expressions on social media sentiments [22]. It is a wholly open-sourced lexicon under the MIT License. We use the VADER analyzer to examine both the sentiment's polarity and intensity in each tweet. This operation results in four sentiment scores, including positive, negative, neutral, and compound. The compound score is a unidimensional metric with a specific value between -1 , referring to (extremely negative), in addition to 1 , referring to (extremely positive), which is a very helpful metric to measure the overall sentiments in each tweet. [Table 2](#) demonstrates the compound threshold value utilized for classifying tweets as positive or negative tweets and neutral tweets.

Table 2: Sentiment type according to the compound score threshold

Type of sentiment	Compound score
Positive	≥ 0.05
Negative	≤ -0.05
Neutral	$-0.05 < \text{score} < 0.05$

Based on [Table 2](#), 0.05 and -0.05 were used as threshold values to explain compound values, i.e., positive values, negative values, as well as neutral values. If the compound value is above or equal to 0.05 , it signifies a positive value. If it is less than or equivalent to -0.05 , it signifies a negative value. Otherwise, it signifies a neutral value.

4.5 Modification of VADER

Although VADER has been substantiated by researchers on typical tweets according to Hutto et al. [22], VADER's performance in categorizing tweets linked to public health issues, especially thalassemia screening, requires additional validation. Accordingly, validation has been performed by selecting a randomly selected subset of a total of 320 tweets obtained from an original set of thalassemia-screening tweets. The sentiment polarity of each of the 320 tweets is interpreted by domain experts as actual results. However, the poor performance of the F1 score (<0.7) has been observed based on the classification of VADER, whereby the primary reason is detected.

In this original lexical dictionary of VADER, some words appeared in tweets that can reverse the polarity score from negative to positive and vice versa, relying on their score values in VADER. For example, In the tweet "I did thalassemia test; unfortunately, it was positive," this tweet shows a

positive compound score of 0.296. In fact, it is originally a negative sentiment. Its polarity score has been changed to positive due to the word ‘positive’ appearing in the tweet with a high sentiment value (i.e., 2.6) which has affected the final compound score to be positive at the end. Similarly, the word ‘negative’ refers to thalassemia screening results containing a high negative sentiment value (i.e., -2.7).

This can affect the sentiment polarity towards the negative side once it is a part of thalassemia screening conversations, affecting sentiment classification accuracy. In addition, we have also removed the word ‘cancer’ from the original lexical dictionary since these words do not contain any positive or negative connotations as they are a part of the health domain. Consequently, that has resulted in a more favorable F1 score of 0.818. Thus, other classification variations between using VADER and utilizing a human rater refer to challenging obstacles faced in classifying sentiments, including ambiguity, sarcasm, as well as mixed sentiments, which seem quite difficult for the human rater to identify. Eventually, this study presented the VADER analyzer’s adjusted version to measure thalassemia screening tweets’ sentiment scores.

4.6 Tools

TWINT, NLTK (Natural Language Toolkit), and VADER constitute the three main tools, which were applied in this study. TWINT has been applied for collecting the raw data from Twitter. NLTK, a text analysis Python library, has been utilized for data pre-processing activities to prepare it for sentiment classification. VADER, a sentiment lexicon, has been used to categorize the relevant tweets’ sentiments into ternary classes.

4.7 TWINT

TWINT represents a sophisticated Twitter scraping instrument, which is written in Python and aims at enabling scholars to scrape stored tweets collected from profiles on Twitter without using API authentication. TWINT utilizes the search operators of Twitter for scraping the tweets. Additionally, it aims to scrape tweets, which are linked to hashtags, topics, and trends. It also aims to sort out sensitive data from the Tweets’ emails and users’ phone numbers in a helpful way. Moreover, TWINT provides queries to Twitter, and enables scraping the users’ followers on Twitter, special tweets, which are loved by users, and their likes and follows without using an API authentication or even browser emulation [32]. TWINT is a standard Python library used in various text analysis research as a primary tool for data acquisition [26,33].

4.8 NLTK

Natural Language Toolkit, i.e., NLTK represents a Python library. It provides a specific base for building Python programs, as well as data classification. The toolbox also plays a key role in converting textual data to a certain format, where sentiments are extracted. NLTK mainly aims to perform natural language processing via performing analysis of human language data. Also, NLTK provides different functions for pre-processing data to perform all NLP methodologies, such as part-of-speech tagging, tokenizing, lemmatizing, stemming, parsing performing sentiment analysis for specified datasets. By doing so, the available data can be fit for feature extraction and mining [34].

4.9 VADER

Vader represents a gold-standard sentiment lexicon; it is mainly attuned to contexts on social media platforms. It aims to combine several lexical features considering 5 generalizable rules: (1) Punctuation, (2) Capitalization, (3) Degree modifiers, (4) Polarity shift due to Conjunctions, and (5)

Catching Polarity Negation. The rules are syntactic and grammatical conventions used by people when they highlight or express the intensity of sentiment. Hutto et al. [22] compared VADER efficiency to eleven benchmark models, involving linguistic inquiry and word count (LIWC), affective norms for english words (ANEW), the General Inquirer, SentiWordNet, and other techniques of machine learning, including Naive Bayes, Maximum Entropy, as well as SVM algorithms. They concluded that the VADER sentiment lexicon generalized more favorably over contexts than other models. Botchway et al. [26] and Kumaresh et al. [25] examined various lexicons in text classification and concluded that VADER had beaten other lexicons in terms of accuracy. Indeed, the VADER sentiment lexicon discriminates itself from other models as being more sensitive to some sentiment phrases in certain social media texts. VADER enjoys the capability of generalizing more efficiently in other domains. Besides common dictionary words, it also gives sentiment scores on emoticons, slang (nah, meh, giggly, etc.), and acronyms (LOL, OMG, etc.) [22].

5 Results

This section provides the results of the sentiment analysis on Twitter with the application of the following tools: VADER and NLTK.

5.1 Descriptive Analysis

Primarily, there were 3,375 tweets were found during the data collection period associated with the thalassemia screening topic. The baseline monthly thalassemia screening tweets volume has fluctuated between (20) and (50) tweets, whereby an explosive number of tweets were observed in May due to world thalassemia day, celebrated every year on May 8th. The tweets memorialized thalassemia victims and encouraged those who are struggling to live with the disease, raising awareness among vulnerable people against thalassemia risks and encouraging them to do screening early. The volume of tweets gradually dropped down and returned to the baseline, as illustrated in Fig. 3.

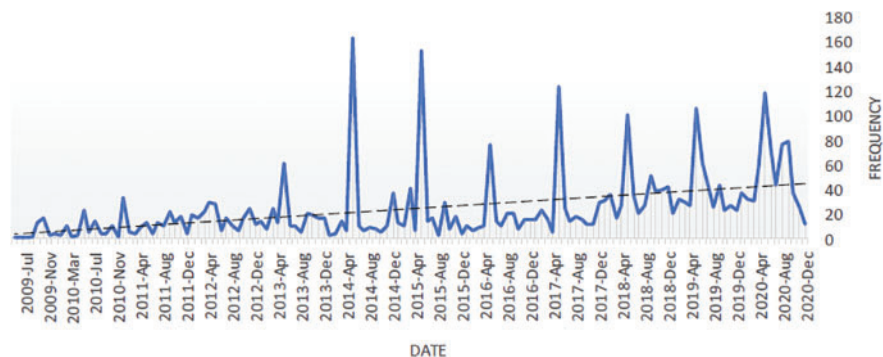


Figure 3: Temporal trends of thalassemia screening tweet volume

Fig. 4 displays the most 25 unigram words that are frequently used while people were having conversations about thalassemia screening. As observed in Fig. 4 above, many words are relevant to blood, screen, get, marriage, and patient. Similarly, Fig. 5 displays the most 25 frequent bigrams and common words, such as (make, mandatory), (blood, marriage), (pregnant, women), (get, marry), and (blood, transfusion). For example, the terms “Getting Married”, “Pregnant Woman”, “Genetic”, “Carrier” and “Prevent” were observed in both bigram and unigram graphs, emphasizing the dire need for thalassemia screening for people who will get married and for pregnant women, which allowed

them to access their status of thalassemia at an early stage. Therefore, they have adequate time to identify the risks and prevent the disease from being passed from parents to children through genes. Moreover, the terms, “Blood Transfusion”, “Blood Donation”, “Blood Group” and “Iron Deficiency” emphasized an urgent need for blood donation because of the lack of iron and the periodic need for thalassemia patients for blood transfusions.

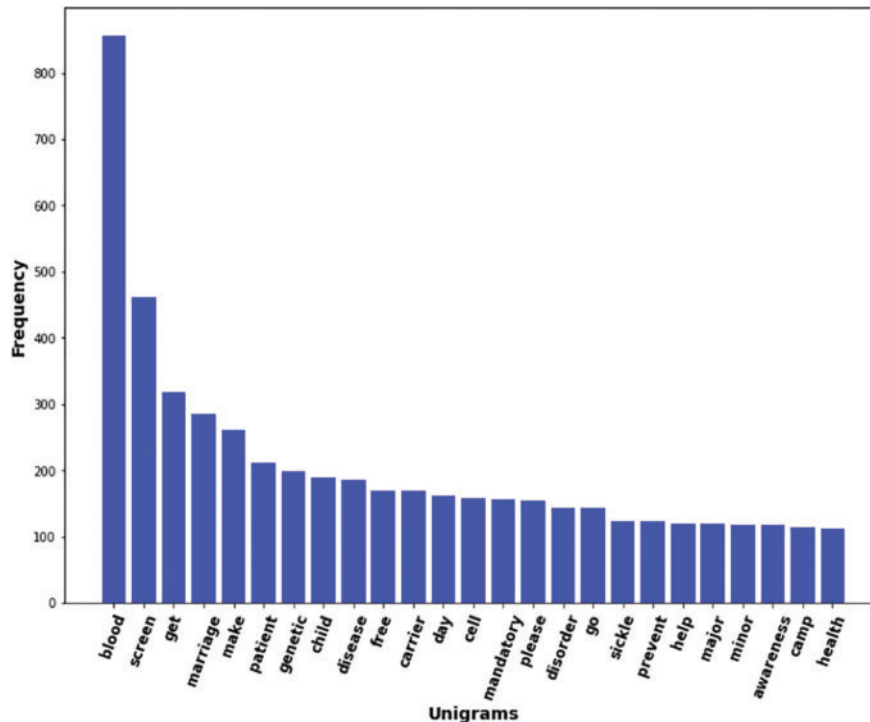


Figure 4: The most 25 unigrams frequent words

Fig. 6 displays the top co-occurring bigrams visualization in the dataset as a network diagram with the help of a Python package named NetworkX. Based on the graph in Fig. 6, it is easier to understand the relationship between the words (nodes), which frequently appeared together by drawing the edge that determines the words' connectedness to each other. For instance, combinations of connected terms in sickle-cell-anemia disease were observed, indicating that most patients of thalassemia have severe forms of the disease and suffer from chronic anemia (sickle cell disease). Additionally, a combination of “hiv-hepatitis” terms was observed with the co-occurrence of words for three main reasons:

- Some tweets complained that most people do not pay attention to thalassemia screening as much as they are concerned about HIV and hepatitis tests.
- Another group of users demanded a mandatory thalassemia test before marriage besides HIV and Hepatitis tests.
- Some tweets mentioned that some patients with thalassemia suffered from HIV and Hepatitis positive results due to allegedly receiving infected blood during the transfusion stage. Therefore, examining the donated blood is a priority to prevent thalassemia people from having infectious diseases, thus alleviating their pain.

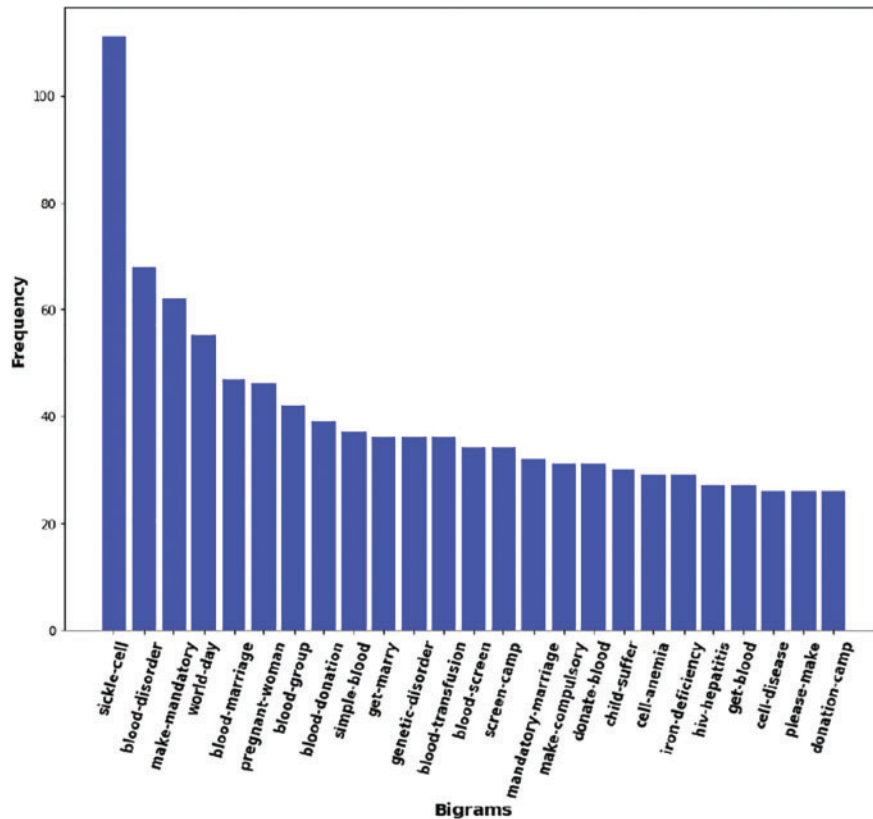


Figure 5: The most 25 bigrams frequent words

Fig. 7 maps the countries mentioned in public conversations regarding thalassemia screening. This map helped determine the correlation between countries with a high rate of thalassemia prevalence and active conversations over thalassemia screening on Twitter. Table 3 demonstrates the 10 most countries mentioned during people’s conversations about thalassemia screening.

As remarked in [35], more than 70,000 babies are born with thalassemia worldwide each year. This defect is observed more often in the Indian subcontinent, the Mediterranean, Southeast Asia, and West Africa. Also, most children with thalassemia are born to women in countries, where people receive a low income. Thalassemia was first confined to the tropics and subtropics. Currently, it is commonly found worldwide, and that is because of the substantial population migrations. Moreover, Beta-thalassemia is most prevalent in Mediterranean, African, and Southeast Asian descent, whereas Alpha thalassemia is much more widespread among African and Southeast Asian descent.

From the figure above, we can observe the highest countries mentioned in tweets, including India and Pakistan. They were expressed in the extracted data with 181 and 164 tweets, respectively. According to the study conducted by [36], most of the available info about thalassemia in the South Asian region originated from studies in India. Because of severe heterogeneity, the variable frequency of the beta-thalassemia heterozygote or carrier ranging between 1%–10% was registered in different areas of India. Nevertheless, the general prevalence of beta-thalassemia carriers was in India between 2.78% and 4%. This represents nearly 30–48 million of India’s population, in other words, approximately 5–12 million carriers of thalassemia with a rate, ranging between 5% and 7% in Pakistan [37–39].

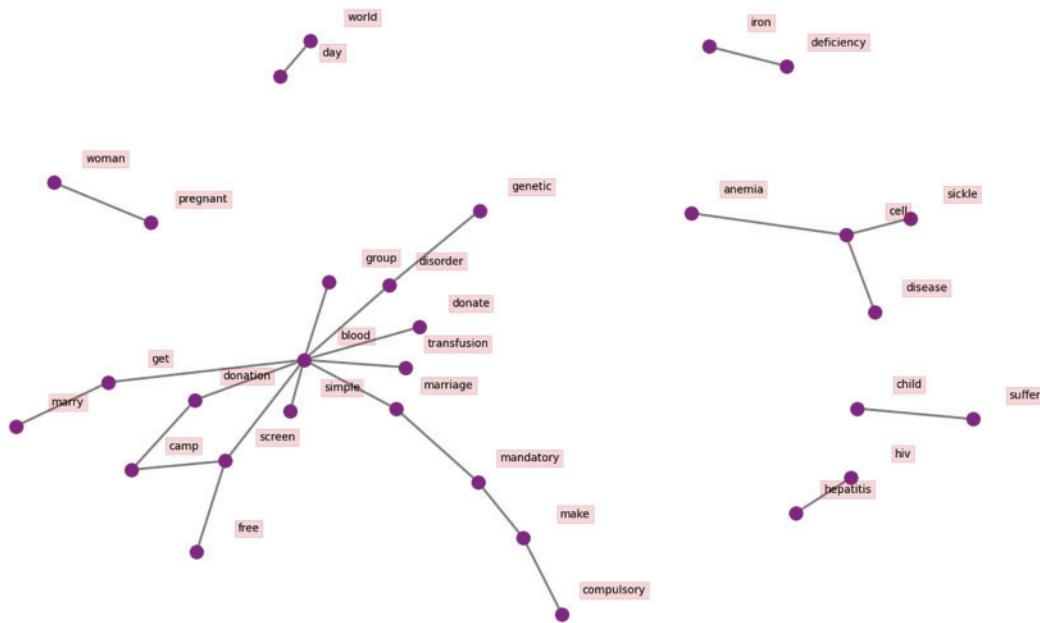


Figure 6: Networks of 25 co-occurring bigrams words

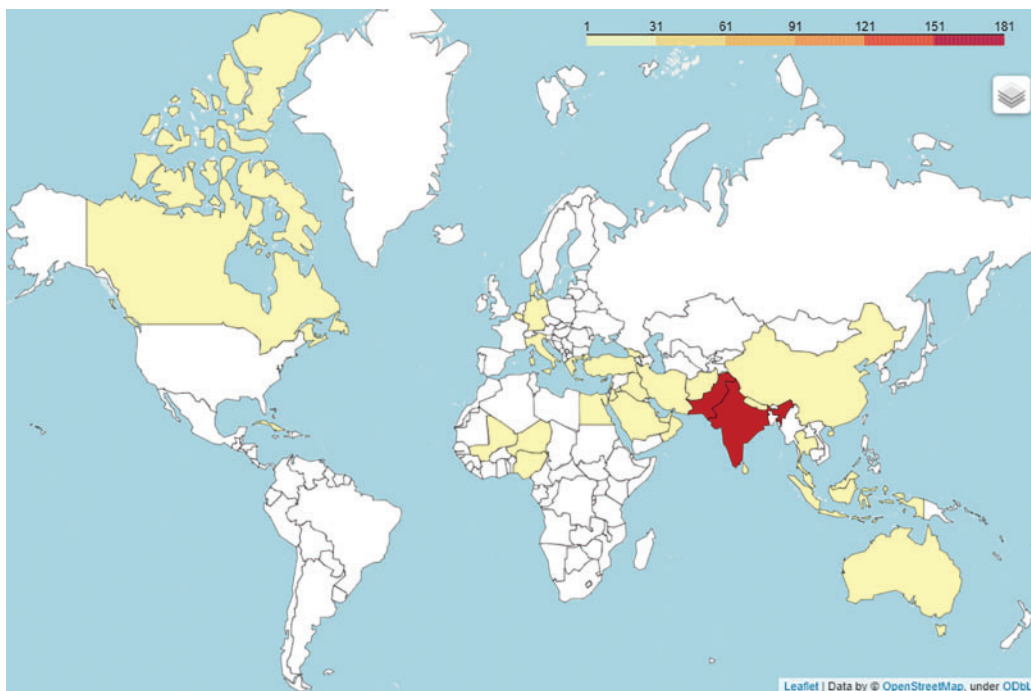


Figure 7: Geospatial distribution of mentioned countries

Table 3: Top 10 countries mentioned by people with their frequencies

Country	Frequency mentioned
India	181
Pakistan	164
UAE	22
Cyprus	17
Oman	13
Thailand	12
Nepal	12
Maldives	11
Egypt	11
Mali	10

5.2 VADER Sentiment Classification

The algorithm of VADER lexicon-based was utilized for extracting features via importing SentimentIntensityAnalyzer from vaderSentiment in the NLTK library. The utilized method “polarity_scores ()” from SentimentIntensityAnalyzer () module has provided four sentiment scores as the output scores, comprising positive, negative, as well as neutral scores, with the compound scores. Regarding the compound scores, every sentiment orientation of the tweet has been identified in accordance with its values. These scores are in a dictionary in Python, whereas the compound scores are extracted for measuring whether each tweet is a positive tweet, a negative tweet, or a neutral tweet with a value ranging from $[-1$ to $+1]$. Table 4 shows the tweets’ sentiments with corresponding values after using the threshold (Table x) to classify tweets into three predefined categories. The values of $[1, 0, -1]$ were set to refer to positive public sentiments, neutral, as well as negative public sentiments [40].

Table 4: Classification of tweets by sentiment

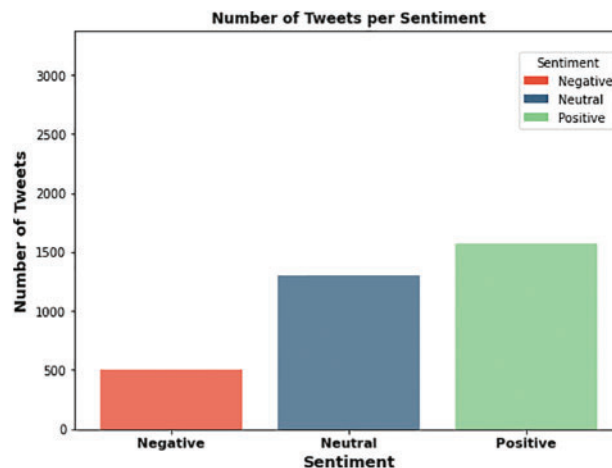
Orginal_Tweet	Tweet_SA	Neg	Neu	Pos	Compound	Sentiment
@BorealPenguin Yeah, that helps! The thing the doc doesn't know where the blood came from, so a diet test (since I have thalassemia) and retina scans are needed	Yeah, that helps! The thing is the doc doesn't know where the blood came from, so a diet test (since I have thalassemia) and retina scans are needed	0.00	0.836	0.164	0.6239	1

(Continued)

Table 4: Continued

Original_Tweet	Tweet_SA	Neg	Neu	Pos	Compound	Sentiment
@narcissists_the yeh, but I think they do test for thalassemia, and they should—it's a legit concern bc some of my red blood cells aren't shaped normally. But def I would lie about the gay thing . . .	yeh, but I think they do test for thalassemia, and they should—it's a legit concern bc some of my red blood cells aren't shaped normally. But def I would lie about the gay thing or living in Ger . . .	0.00	1.00	0.00	0.000	0
@ftrhshrf Hi Dr I want to ask why malaysia does not require thalassemia test before marriage like HIV, hepatitis B n blood group. I understand that is a choice to every couple to do the test or . . .	Hi Dr I want to ask why malaysia does not require thalassemia test before marriage like HIV, hepatitis B n blood group. I understand that is a choice to every couple to do the test or not but wh . . .	0.07	0.877	0.053	-0.4497	-1

Fig. 8 displays the sentiment results by the number of tweets of each class in a bar chart using the matplotlib.pyplot Python library.

**Figure 8:** Number of the tweets by sentiment

Based on the number of tweets by sentiment in Fig. 8 and Table 5, there are about 3376 tweets, incorporating all sentiment categories. Most tweets referred to positive and neutral sentiments regarding thalassemia screening. Remarkably, (1569) 46.5% of the obtained tweets indicated positive sentiments, while (1302) 38.6% of these tweets indicated neutral opinions, and (503) 14.9% referred to negative sentiments. The positive tweets were the highest in number compared with other classifications as the value of the given compound threshold expressed positive sentiments regarding thalassemia screening.

Table 5: Results of the vader algorithm

Total tweets	3376
Positive tweets	1569
Neutral tweets	1302
Negative tweets	503

Fig. 8 and Table 5 illustrates the word cloud of the negative and positive tweets labeled through VADER to manifest the dominant words for each class. The search keywords (i.e., thalassemia and screening), and stop words (i.e., they, I, do, them) were removed because they appeared in almost all tweets. This helped Wordcloud focus on terms only and be more productive in gaining insights precisely. It was observed that the most frequent words appearing in both Figures are domain-related, such as blood, disease, disorder, alpha, beta, genetic, etc. as we expected. By looking at the Wordcloud of positive tweets, some words were observed like prevent, awareness, free, support, mandatory, marriage, camp, and pregnancy. The results showed considerable solidarity and support among the thalassemia community on Twitter translated by organizing free test camps, spreading awareness among the public, and making mandatory tests before marriage and during pregnancy. The Wordcloud of negative tweets showed the following words: child, suffer, HIV, and negligence, which highlighted the suffering of children from thalassemia during blood transfusion who were infected with HIV due to negligence in blood screening.

A Word Cloud also called a tag cloud, or a text cloud is a tool for visually summarizing a vast amount of text [25]. In this study, Fig. 9 showed the word cloud to visualize the tweets for every sentiment classification based on the VADER scores after removing the meaningless information.



Figure 9: Wordcloud of positive (right) and negative (left) tweets

Fig. 10 demonstrates the distribution of the negative and positive compound scores as a histogram plot after neglecting the neutral ones.

Based on Fig. 11, the negative opinions do not correspond to the positive and negative sentiments. The positive tweets have a significant numerical advantage over the negative ones in terms of the compound score. The mean of the positive tweets is 0.4758, as opposed to negative tweets, which have an average of about -0.4234 .

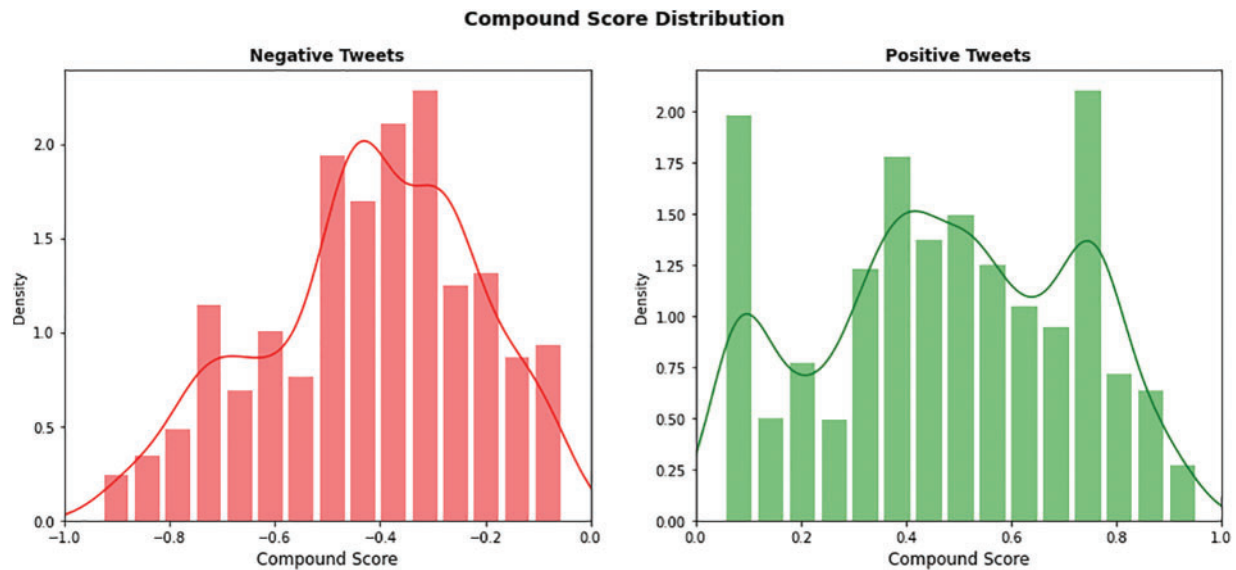


Figure 10: Compound score distribution: negative tweets (left) and Positive Tweets (right)

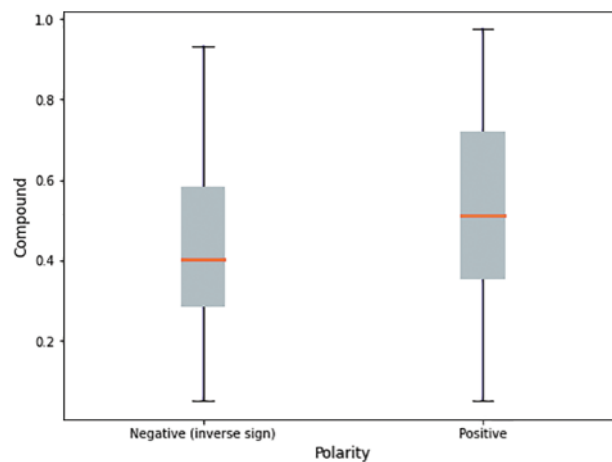


Figure 11: Compound score distribution in boxplot

As shown in Fig. 11, a box plot gives graphical information concerning the positive and negative tweets such as batch location, dispersion, and the data set's skewness. Also, a boxplot drew attention to specific possible outliers. Therefore, it is easier to compare and reflect on these 4 features of our data sets.

- Assessment of batch location: The figure above indicates that the median compound score of the obtained positive tweets was larger compared with the negative tweets.
- Assessment of data dispersion: These interquartile ranges were not parallel, as indicated by the length of boxes for positive tweets and negative tweets and the data set's total range for positive tweets is greater. This is illustrated through the distance between the two whiskers' ends for every boxplot.

- The difference in skewness: Even though both data batches appeared to be left skewness and the negative tweet batch is a little more skewed compared with positive tweets. The skewness of the sample of a compound score for positive tweets recorded -0.184 , while negative tweets recorded a compound score of -0.287 . Thus, both skewnesses were negative, and the negative tweets' value was slightly larger, which corresponds to a further apparent absence of symmetry; however, neither of the skewnesses was especially large.
- Assessment of potential outliers: Neither of the data sets showed any questionable far-out rates.

5.3 Model Evaluation

Regarding the model evaluation of the proposed approach, we used four performance measures, such as True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). These four performance measures represent the confusion matrix and are shown in Fig. 12. The confusion matrix represents a mixture of predicted and actual observations to measure the current algorithm's effectiveness by determining precision and recall, in addition to the F1 Score.

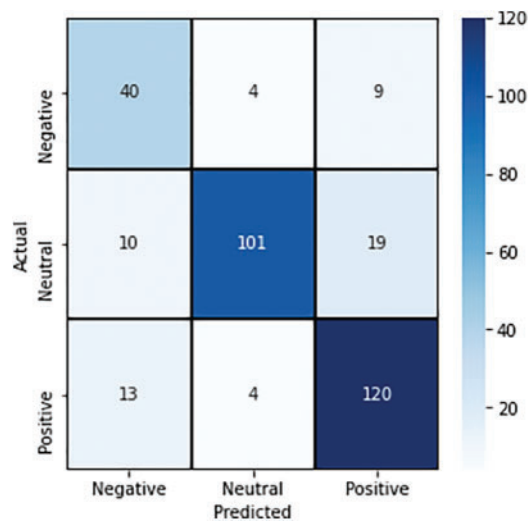


Figure 12: VADER confusion matrix

■ Precision

Precision represents the observations' rate, which correctly forecasted positive to total forecasted positive ones.

$$\text{Precision} = TP / (TP + FP) \quad (1)$$

■ Recall

Recall, which is referred to as sensitivity, signifies the observations' rate, which is correctly forecasted positive to the entire real class observations.

$$\text{Recall} = TP / (TP + FN) \quad (2)$$

■ F-measure

F-measure or F-score refers to a performance measure, which combines precision with recall by determining the weighted harmonic mean covering accuracy flaws with the skewed data.

$$F - \text{measure} = 2 * (\text{Recall} * \text{Precision}) / (\text{Recall} + \text{Precision}) \quad (3)$$

Fig. 12 shows the confusion matrix of 320 randomly selected sentiments. They were manually labeled by three domain expert annotators at the tweet level for identifying ternary classes (actual labels) to be compared to the results of the VADER classifier (predicted labels). According to the obtained results, the total number of the actual positive sentiments, the neutral sentiments, as well as the negative sentiments is 137, 130, 53, correspondingly. When the VADER lexicon was used, the model anticipated that the positive attitudes number will be 148, 109 for neutral sentiments, and 63 for negative opinions.

Table 6 compares several related studies about public health regarding precision (P), recall (R), and F -score ($F1$). It is, therefore, concluded that the proposed approach achieved promising results, which outperformed many proposals with 0.829 precision, 0.816 recall, and 0.818 $F1$ score. However, holding a comparison between approaches is a difficult task for a couple of reasons. First, these approaches are oriented toward various sentiment analysis stages [41]. Second, the proposed methods are based on various approaches (i.e., machine learning or lexicon-based), topics, as well as domains, whereby the linguistic resources utilized vary in size and context, making the comparison between the proposals so tricky. Therefore, to perform an appropriate comparison between the proposals, the same corpus in all evaluations is necessitated.

Table 6: Comparison of results of previous related works

Work	Dataset	P	R	F1	Approach
[12]	Health reviews	0.89	0.79	0.83	Hybrid (UMLS + SWN)
[13]	Diabetes	0.819	0.811	0.812	SWN
[16]	Hearing loss	0.688	0.686	0.685	Linear Reg.
[17]	Health reviews	0.527	0.541	0.518	Naïve Bayes
[18]	Medical Condition	0.68	0.60	0.63	CNN
	Medication	0.86	0.77	0.82	CNN
[19]	HPV vaccines	–	–	0.744	SVM
Our approach	Thalassemia screening	0.829	0.816	0.818	VADER

Vader is a gold-standard sentiment lexicon that is particularly developed for microblogs such as Twitter. It combines the lexical features with consideration for five generalizable rules (punctuation, capitalization, degree modifiers, constructive conjunction, and Tri-gram examination to identify negation), which are grammatical and syntactic conventions that humans use to express and emphasize sentiment intensity. Indeed, VADER has outperformed individual human raters at correctly classifying the sentiment of tweets into positive, neutral, or negative classes by 0.96 to 0.84 of the $F1$ measure (Hutto & Gilbert, 2014). The VADER sentiment lexicon discriminates itself from others in that it is more sensitive to sentiment expressions in social media contexts and generalizes better to other domains. Besides common dictionary words, it also gives information on emoticons, slang (nah, meh, giggly, etc.), and acronyms (LOL, OMG, etc.) (Hutto & Gilbert, 2014). I have chosen VADER in the

present project because it is becoming broadly adopted as it was even implemented as a component of the NLTK python library. However, several limitations can be outlined in this study. Firstly, the introduced approach handles English tweets only although a wealth of information about thalassemia is adequately available in different languages. It is, therefore, highly recommended that this method is applied to other languages, e.g., Arabic. Secondly, the general opinion lexicon is inadequate to capture the meanings of health-related texts.

6 Conclusion

Due to the lack of relevant studies about the utilization of sentiment analysis in the public health domain, particularly thalassemia screening, this study utilized one of the social media platforms, i.e., Twitter platform. The study explored people's sentiments regarding thalassemia screening with the application of the VADER lexicon-based algorithm. The results revealed that the suggested approach had achieved results of 0.829, 0.816, and 0.818 for the corresponding precision and recall, as well as the F1 measure. It was found that the results of the VADER analysis are remarkably promising. This study showed by digging into tweets and based on the most frequent words obtained, there is a significant amount of support and solidarity among the thalassemia Twitter community, with numerous popular terms related to positive emotions and sentiments observed such as 'make,' 'mandatory,' 'marriage,' 'free,' 'child,' and 'prevent'. Also, the final results showed the correlation between countries with a high prevalence of thalassemia and the active conversations about thalassemia screening on Twitter where India and Pakistan ranked the highest mention in tweets by the public's conversations on thalassemia screening with 181 and 164 tweets, respectively. In the future, the system can be developed by establishing a domain-related lexicon along with the current one to achieve better results.

Acknowledgement: The authors are thankful to the Deanship of Scientific Research at Najran University and registrar of Universiti Tun Hussein Onn Malaysia.

Funding Statement: The authors are thankful to the Deanship of Scientific Research at Najran University for funding this work under the Research Collaboration Funding program grant coder NU/RC/SERC/11/5.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] M. Angastiniotis and S. Lobitz, "Thalassemias: An overview," *International Journal of Neonatal Screen*, vol. 5, no. 1, pp. 1–11, 2019. <https://doi.org/10.3390/ijns5010016>
- [2] M. T. Riquier, "The ethics of genetic screening for beta thalassemia in Vietnam," *Developing World Bioethics*, vol. 22, no. 1, pp. 44–52, 2022.
- [3] A. Kantharaj and S. Chandrashekar, "Coping with the burden of thalassemia: Aiming for a thalassemia free world," *Global Journal of Transfusion Medicine*, vol. 3, no. 1, pp. 1, 2018. https://doi.org/10.4103/gjtm.gjtm_19_18
- [4] H. Al Sabbah, S. Khan, A. Hamadna, L. A. Ghazaleh, A. Dudin *et al.*, "Factors associated with continuing emergence of β -thalassemia major despite prenatal testing: A cross-sectional survey," *International Journal of Women's Health*, vol. 9, no. 9, pp. 673–679, 2017. <https://doi.org/10.2147/IJWH.S141936>
- [5] N. Yadav, O. Kudale, A. Rao, S. Gupta and A. Shitole, "Twitter sentiment analysis using supervised machine learning," in *Intelligent Data Communication Technologies and Internet of Things*, vol. 57. Singapore: Springer, pp. 631–642, 2021.

- [6] E. Smailhodzic, W. Hooijsma, A. Boonstra and D. J. Langley, "Social media use in healthcare: A systematic review of effects on patients and on their relationship with healthcare professionals," *BMC Health Services Research*, vol. 16, no. 1, pp. 1–15, 2016. <https://doi.org/10.1186/s12913-016-1691-0>
- [7] A. D'Andrea, F. Ferri, P. Grifoni and T. Guzzo, "Approaches, tools and applications for sentiment analysis implementation," *International Journal of Computer Applications*, vol. 125, no. 3, pp. 26–33, 2015. <https://doi.org/10.5120/ijca2015905866>
- [8] A. Mittal and S. Patidar, "Sentiment analysis on twitter data: A survey," in *Proc. of the 7th Int. Conf. on Computer and Communications Management*, New York, United States, pp. 91–95, 2019. <https://doi.org/10.1145/3348445.3348466>
- [9] R. K. Botchway, A. B. Jibril, Z. K. Oplatková and M. Chovancová, "Deductions from a sub-saharan African bank's tweets: A sentiment analysis approach," *Cogent Economics & Finance*, vol. 8, no. 1, pp. 1–19, 2020. <https://doi.org/10.1145/3348445.3348466>
- [10] Y. C. Phang, A. M. Kassim and E. Mangantig, "Concerns of thalassemia patients, carriers, and their caregivers in Malaysia: Text mining information shared on social media," *Healthcare Informatics Research*, vol. 27, no. 3, pp. 200–213, 2021.
- [11] M. Zubair, M. Qasim, B. Ahmad, S. Ahmad, A. Khan *et al.*, "Health miner: Opinion extraction from user generated health reviews," *International Journal of Academic Research*, vol. 5, no. 6, pp. 279–284, 2013. <https://doi.org/10.7813/2075-4124.2013/5-6/A.35>
- [12] M. Z. Asghar, S. Ahmad, M. Qasim, S. R. Zahra and F. M. Kundi, "SentiHealth: Creating health-related sentiment lexicon using hybrid approach," *Springerplus*, vol. 5, no. 1, pp. 1–23, 2016. <https://doi.org/10.1186/s40064-016-2809-x>
- [13] M. D. P. Salas-Zárate, J. Medina-Moreira, K. Lagos-Ortiz, H. Luna-Aveiga, M.Á. Rodríguez-García *et al.*, "Sentiment analysis on tweets about diabetes: An aspect-level approach," *Computational and Mathematical Methods in Medicine*, vol. 2017, no. 5, pp. 1–9, 2017. <https://doi.org/10.1155/2017/5140631>
- [14] L. Verma and V. Sapra, "Semantic analysis of cardiovascular disease sentiment in online social media," in *Proc. of Int. Conf. on Advancements in Computing & Management (ICACM)*, Dhaka, Bangladesh, pp. 1078–1082, 2019. <https://doi.org/10.2139/ssrn.3462426>
- [15] K. O. Wong, F. G. Davis, O. R. Zaïane and Y. Yasui, "Sentiment analysis of breast cancer screening in the United States using twitter," in *IC3K 2016—Proc. 8th Int. Jt. Conf. Knowl. Discov. Knowl. Eng. Knowl. Manag.*, UAlberta, Canadian, vol. 1, pp. 265–274, 2016.
- [16] T. Ali, D. Schramm, M. Sokolova and D. Inkpen, "Can i hear you? sentiment analysis on medical forums," in *Proc. of the Sixth Int. Joint Conf. on Natural Language Processing*, Nagoya, Japan, vol. 14–18, pp. 667–673, 2013.
- [17] M. Arbane, R. Benlamri, Y. Brik and A. D. Alahmar, "Social media-based COVID-19 sentiment classification model using Bi-LSTM," *Expert Systems with Applications*, vol. 212, no. 6, pp. 118710, 2023. <https://doi.org/10.1016/j.eswa.2022.118710>
- [18] S. Yadav, A. Ekbal, S. Saha and P. Bhattacharyya, "Medical sentiment analysis using social media: Towards building a patient assisted system," in *Proc. of the Eleventh Int. Conf. on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan, pp. 2790–2797, 2019.
- [19] J. Du, J. Xu, H. Song, X. Liu and C. Tao, "Optimization on machine learning based approaches for sentiment analysis on HPV vaccines related tweets," *Journal of Biomedical Semantics*, vol. 8, no. 1, pp. 1–8, 2017. <https://doi.org/10.1186/s13326-017-0120-6>
- [20] N. Mukhtar, M. A. Khan and N. Chiragh, "Lexicon-based approach outperforms supervised machine learning approach for Urdu sentiment analysis in multiple domains," *Telematics and Informatics*, vol. 35, no. 8, pp. 2173–2183, 2018. <https://doi.org/10.1016/j.tele.2018.08.003>
- [21] A. Jurek, M. D. Mulvenna and Y. Bi, "Improved lexicon-based sentiment analysis for social media analytics," *Security Informatics*, vol. 4, no. 1, pp. 9166, 2015. <https://doi.org/10.1186/s13388-015-0024-x>
- [22] C. J. Hutto and E. Gilbert, "VADER: A parsimonious rule-based model for sentiment analysis of social media text," in *Proc. of the Int. AAAI Conf. on Web and Social Media*, Atlanta, Georgia Institute of Technology, pp. 216–225, 2014.

- [23] V. L. Narasamma, M. Sreedevi, G. V. Kumar and A. Pradesh, "TweetShort text data analysis on COVID-19 outbreak," in *Smart Technologies in Data Science and Communication*, vol. 29. Singapore: Springer, pp. 183–193, 2021.
- [24] L. He and K. Zheng, "How do general-purpose sentiment analyzers perform when applied to health-related online social media data?," *Studies in Health Technology and Informatics*, vol. 264, pp. 1208–1212, 2019.
- [25] N. Kumares, V. Bonta and N. Janardhan, "A comprehensive study on lexicon based approaches for sentiment analysis," *Asian Journal of Computer Science and Technology*, vol. 8, no. S2, pp. 1–6, 2019. <https://doi.org/10.51983/ajcst-2019.8.S2.2037>
- [26] R. K. Botchway, A. B. Jibril, M. A. Kwarteng, M. Chovancova and Z. K. Oplatková, "A review of social media posts from Unicredit bank in Europe: A sentiment analysis approach," in *Proc. of the 3rd Int. Conf. on Business and Information Management*, New York, United States, pp. 74–79, 2020.
- [27] J. Al-Garaady and M. Mahyoob, "Public sentiment analysis in social media on the SARS-CoV-2 vaccination using VADER lexicon polarity," *Humanities and Educational Sciences Journal*, vol. 22, no. 1, pp. 591–609, 2022.
- [28] T. Mustaqim, K. Umam and M. A. Muslim, "Twitter text mining for sentiment analysis on government's response to forest fires with vader lexicon polarity detection and k-nearest neighbor algorithm," *Journal of Physics: Conference Series*, vol. 1567, no. 3, pp. 8–15, 2020. <https://doi.org/10.1088/1742-6596/1567/3/032024>
- [29] E. Apostolova and R. Andrew Kreek, "Training and prediction data discrepancies: Challenges of text classification with noisy, historical data," *W-NUT*, 2018. <https://doi.org/10.18653/v1/w18-6114>
- [30] A. Krouska, C. Troussas and M. Virvou, "The effect of preprocessing techniques on Twitter sentiment analysis," in *2016 7th Int. Conf. on Information, Intelligence, Systems & Applications (IISA)*, Chalkidiki, Greece, IEEE, pp. 1–5, 2016.
- [31] S. Elbagir and J. Yang, "Twitter sentiment analysis using natural language toolkit and Vader sentiment," in *Proc. of the Int. Multiconference of Engineers and Computer Scientists*, Hong Kong, vol. 2239, pp. 12–16, 2019.
- [32] C. Zacharias and F. Poldi, "GitHubtwintproject/twint: An advanced Twitter scraping & OSINT tool written in Python that doesn't use Twitter's API, allowing you to scrape a user's followers, following, tweets and more while evading most API limitations," February 2020. Available: <https://github.com/twintproject/twint>
- [33] T. Mehta, G. Kolase, V. Tekade, R. Sathe and A. Dhawale, "Price prediction and analysis of financial markets based on news, social feed, and sentiment index using machine learning and market data," *International Research Journal of Engineering and Technology*, vol. 7, no. 6, pp. 483–489, 2020.
- [34] E. Loper and S. Bird, "nltk: The natural language toolkit," in *Proc. of the ACL-02 Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics*, Philadelphia, Pennsylvania, vol. 1, pp. 63–70, 2002. <https://doi.org/10.3115/1118108.1118117>
- [35] T. Banerjee and R. B. Aniyery, "Thalassemia and its management during pregnancy," *World Journal of Anemia*, vol. 1, no. 1, pp. 5–17, 2017. <https://doi.org/10.5005/jp-journals-10065>
- [36] M. S. Hossain, E. Raheem, T. A. Sultana, S. Ferdous, N. Nahar *et al.*, "Thalassemias in South Asia: Clinical lessons learnt from Bangladesh," *Orphanet Journal of Rare Diseases*, vol. 12, no. 1, pp. 1–9, 2017. <https://doi.org/10.1186/s13023-017-0643-z>
- [37] B. Cartwright, R. Frank, G. Weir and K. Padda, "Detecting and responding to hostile disinformation activities on social media using machine learning and deep neural networks," *Neural Computing and Applications*, vol. 34, no. 18, pp. 15141–15163, 2022. <https://doi.org/10.1007/s00521-022-07296-0>
- [38] S. He, Q. Qin, S. Yi, Y. Wei, L. Lin *et al.*, "Prevalence and genetic analysis of α - and β -thalassemia in Baige region, a multi-ethnic region in southern China," *Gene*, vol. 619, no. 1, pp. 71–75, 2017. <https://doi.org/10.1016/j.gene.2016.02.014>
- [39] A. Kathuria, A. Gupta and R. K. Singla, "AOH-senti: Aspect-oriented hybrid approach to sentiment analysis of students' feedback," *SN Computer Science*, vol. 4, no. 2, pp. 1–23, 2023. <https://doi.org/10.1007/s42979-022-01611-1>

- [40] H. Li, X. B. Bruce, G. Li and H. Gao, "Restaurant survival prediction using customer-generated content: An aspect-based sentiment analysis of online reviews," *Tourism Management*, vol. 96, no. 7, pp. 104707, 2023. <https://doi.org/10.1016/j.tourman.2022.104707>
- [41] R. Jain, A. Kumar, A. Nayyar, K. Dewan, R. Garg *et al.*, "Explaining sentiment analysis results on social media texts through visualization," *Multimedia Tools and Applications*, vol. 6, no. 6, pp. 1–17, 2023. <https://doi.org/10.1007/s11042-023-14432-y>