A NOVEL SECURITY FRAMEWORK FOR PREVENTING ATTACKS IN SENTIMENT ANALYSIS

V. Laxmi Narasamma
PhD Scholar, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India-522502
laxmi8866@gmail.com

Dr. M. Sreedevi
Professor , Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation,Vaddeswaram, Guntur, Andhra Pradesh, India-522502
msreedevi_27@kluniversity.in

Abstract

Nowadays, Twitter data-based sentiment analysis is the mainly common topic in Natural Language Processing (NLP). Nevertheless, security attacks on Twitter data are increased day by day because hackers and the attacks will reduce the performance of sentiment analysis. Many kinds of research are developed to overcome this problem, but there are no accurate results found. So this current research proposed a novel Ant Lion honeypot with Regression (ALHR) for detecting the attacks and continuous monitoring of data. Moreover, the fitness function of the introduced replica is used for preventing attacks and continuous monitoring. Also, this model utilizes Twitter-based data about the corona disease 2019 (COVID-19) for detecting attacks and enhances the classification of sentiments using continuous monitoring. For verifying the effectiveness of ALHR technique, launch attacks in classification layer. The developed technique is executed in Python, and the achieved performance metrics are compared with another existing replica regarding the accuracy, recall, precision, F-measure, and error rate. Finally, the ALHR technique enhances the sentiment analysis and provides continuous monitoring.

***Keywords*: Tweets, sentiment analysis, continuous monitoring, DoS attacks, preprocessing, features extraction, and detect attacks**

1. Introduction

Sentiment Analysis (SA) uses text analysis, language processing, and statics for analyzing customer sentiments or customer reviews [1]. Moreover, good businesses should understand customer reviews. The business organization understands people's opinion, what they are saying to our products etc. [2]. Thus, the customer sentiments are gathered in reviews, tweets, and comments [3]. Furthermore, sentiment analysis is the understanding of emotions of people with the help of software, and it will be helpful to the business leader for the growth of the workplace; it is the idea to express feeling in words [4]. Additionally, sentiment analysis extracts feelings to people in positive, negative, and neutral [5]. Thus, the positive, neutral, and negative statements depend on the user's issues or comments [6]. Using sentiment analysis in business areas, they do not need nonstop hours to classify customer data like reviews, social media comments, survey responses, tweets, and support tickets [7]. It helps the organization monitor their brand reputation on social media, customer needs understanding and earn imminent in customer feedback, etc. [8]. In addition, sentiment analysis uses Natural language processing (NLP) replica and algorithms that contain hybrid systems, automatic systems, and rules-based systems [9].
Generally, the sentiment analysis technique focuses on polarization such as positive, neutral, and negative emotions and feelings like anger, sadness, happiness, etc. [10]. This kind of sentiment analysis is used to detect emotions. Thus, a lot of emotion detection is identified using lexicons [11]. Consequently, sentiment analysis generally tackles the task of classification. Thus, the user-decided classification algorithm is handled with the help of Support Vector Machine (SVM) [12]. It is one of the most famous classifiers. The fast-developing online platform is Twitter; they provide posts, create, read short messages and update [13]. They are also called tweets. Thus, the platform of Twitter may ultimately influence by traditional media agenda [14]. They are surroundings mainly in dangerous events and collect data from rewet valuable messages and tweets shared through users [15]. Moreover, systematic learning of semantic contented tweets is known as SA, and this is the technique of categorizing and identifies the polarity of text. The qualitative dimension of tweets is fundamentally used for the critical situation [16]. The most critical event is natural disasters such as hurricanes, earthquakes, wildfires, floods, or political transitions like coups, movements, terrorist attacks, and revolution [17]. Consequently, Twitter is the

most powerful source for gathering information, and they contribute to allocating organization, human, resource materials and preventing accessing other people affected zones [18].
The most critical task in sentiment analysis is analyzing emotion during critical situations, and it is more complicated [19]. Thus the critical situation is personality through the experience of threshold trespassing and socially unknown state [20]. The main issue of the sentiment analysis is malicious activity and wrong classification [21]. Generally, several techniques are introduced to classify the sentiment analysis in the correct way and detect malicious attacks during sentiment analysis. The existing techniques are frequency technique [23], machine learning [24], and multiple neutrosophic sets [26] but still having the issues of malicious attacks, high false rate, and improper classification [22]. Consequently, in this paper a novel ALHR technique is proposed to prevent the malicious activity and continuous monitoring. Thus, the developed framework achieves high privacy also continuous monitoring of attacks. So, it will enhance the sentiment analysis and achieved better performance in emotion identification.
This paper summarizes the structure of this paper: The related work of this research is summarized in section 1, and the problem definition and system model are explained in section 2. Moreover, in section 3, the proposed techniques have been highly structured. Consequently, the result and discussion, along with the comparative analysis, are described in section 4. Finally, the paper has been concluded in section 5.

2. Related Works

Samah Mansour [23] has proposed a Twitter APT and frequency method to search and collect tweets and conduct text sentiment analysis. Moreover, obtained results are the same words, transferring the same negative and positive words from people's opinions. Thus, the developed technique views a lot of users ISIS threats and queries regarding text mining. But the classification of sentiment analysis did not process correctly because of large data.
Nowadays, sentiment analysis by machine learning with the use of Twitter data is the most popular topic. Gonzalo et al. [24] have addressed sentiment analysis issues in critical events like social movements and natural disasters. Additionally, introduced Bayesian network classification is developed to better performance in sentiment analysis. Two datasets are used for the experimental purpose that is Chilean earthquake and Catalan independence. Finally, it will identify the word relation and quantities information. But error occurrence rate is high in the classification layer.
Muhammad Asif et al. [25] have proposed sentiment analysis of multilingual textual data of social media for discovering the intensity of sentiments. This paper studied incorporated textual views of four types: moderate, high extreme, neutral, low extreme, and so on. Initially developed multilingual lexicon intensity weight and achieved results of replica attain 88% inaccuracy. Hence the developed model attains a long time to classify the emotion, so computation time is large.
Generally, sentiment analysis of Twitter data gauge's public opinion, police, social movements, and legislation. Vasantha et al. [26] have developed multiple neutrosophic sets with two positives, two negatives, and three neutral. Based on the sentiment analysis easily validated the results using single, triple, and multiple neutrosophic sets. The proposed replica gains better performance in the refinement of presented indeterminacy data.
Zidong Jiang et al. [27] have proposed the Sina Weibo platform to troll recognition through SA and other user movement data. Additionally, it introduced a new method in terms of word embedding, Chinese sentence segmentation, and sentence score measurement. The main aim of the developed technique is to enhance the test sentiment analysis to detect trolls using machine learning. The gained outcomes of troll detection id high, but malicious attacks are gathered because of large data. The overall summary of the paper of the related work has a detailed Table 1.

Table.1 Summary of related works

Writer	Technique	Advantage	Disadvantage
Samah Mansour [23]	Twitter APT and frequency technique	It will identify many threats and	During classification occur error due to presented attack
Gonzalo et al. [24]	Bayesian network classification	Better performance in sentiment analysis, identify the word relation and quantities information	The error rate is high
Muhammad Asif et al. [25]	SA of multilingual textual data in social media	Discovering the intensity of sentiments and accuracy rate is high	Large computation time because of many data.
Vasantha et al. [26]	MRNS	The correct identification rate is high.	Less F-Measure
Zidong Jiang et al. [27]	Troll detection through sentiment analysis	Improve test sentiment analysis and troll detection rate high	The performance of sentiment analysis is slow because of detection trolls.

The key metrics of the developed replica is declared as follows:

- At first, download Twitter data established from user comments regarding COVID-19.
- Moreover, collected data such as tweets are trained to the system

- Furthermore, introduced a new Ant Lion Honeypot with regression (ALHR) framework for continuous monitoring of attacks and high privacy
- Thus, the proposed technique detects and neglects the attacks also identifies current sentiment analysis of collected twitter data.
- At next, launch malicious attacks in the sentiment analysis for checking the strength of the ALHR technique.
- Hereafter, the measured fitness function of the ant-lion and honeypot is updated to the sentiment analysis for continuous monitoring and prevent attacks.
- Subsequently, the key metrics of the proposed technique are evaluated with other existing replicas in terms of accuracy, F-measure, recall, error rate, and precision.

3. Problem definition and system model

Sentiment analysis is more popular because of the increasing use of social media, and they gain the popularity of many people through various motivations and interests. In the world, all people can express their feelings or emotions in various concepts related to education, politics, commercial products, culture, and so on. The most common thing is to classify the polarity based on user satisfaction, neutrality, and dissatisfaction. But the issues of the sentiment analysis are a sentence started with happy mode but ending with sad mode. In these types, they are very critical to classify the emotion. Also, a large amount of data generates malicious activities in the classification layer. The process of sentiment analysis with the problem has illustrated in Fig.1.

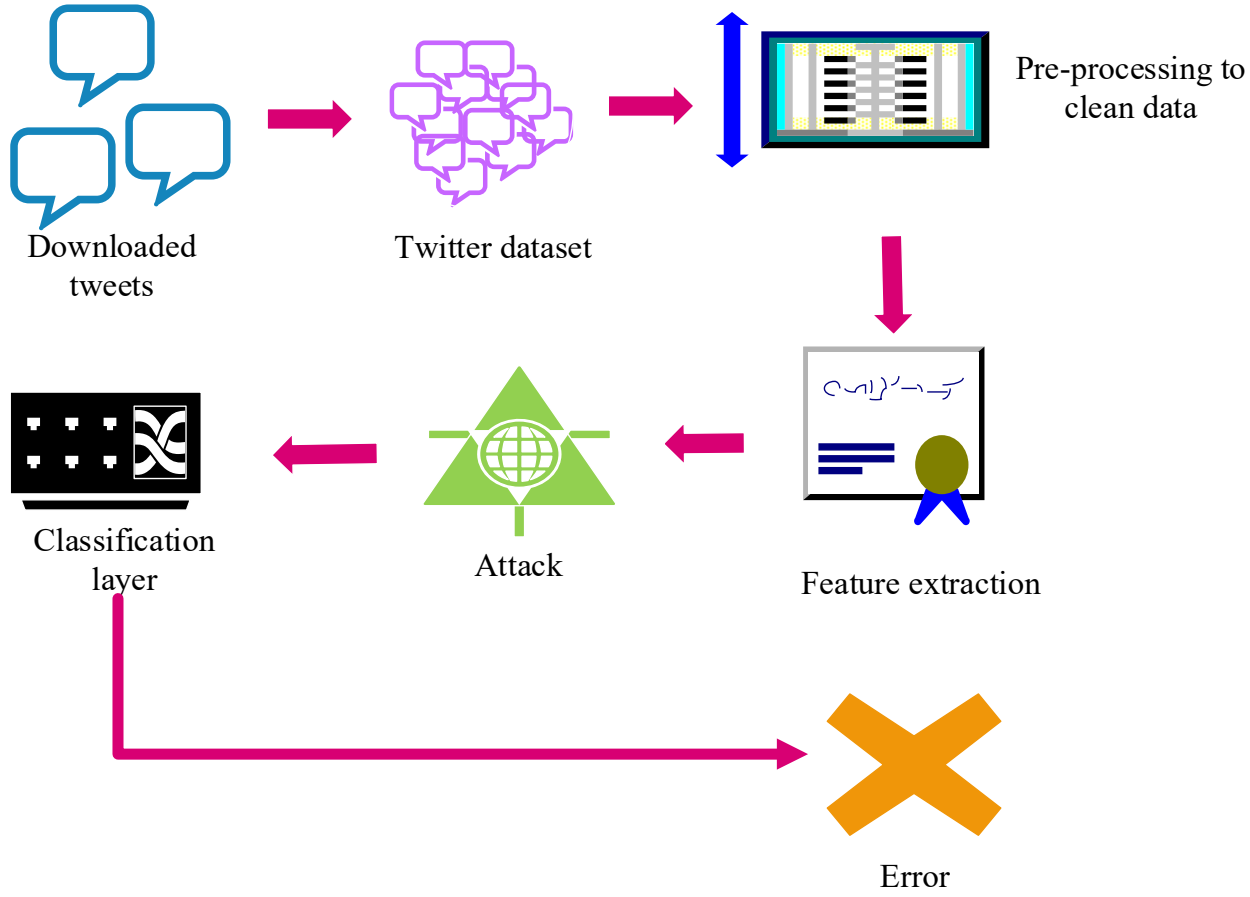

Fig. 1. System model and problem statement

Because of the attacks, the classification layer becomes weak and does not execute the correct job, and the identification of user emotion is incorrect. So, the new ALHR technique is proposed to preserve high privacy of sentiment analysis and prevent attacks, also continuously monitoring malicious activities. It will secure the sentiment analysis from attacks and detect and neglect attacks presented in the classification layer

4. Proposed Methodology

The detection of malicious activities presented in the sentiment analysis and classification of user emotions is the most critical task. Moreover, a novel Ant Lion Honeypot with Regression (ALHR) framework is proposed to monitor attacks and high privacy rates. This paper measured the fitness function of ant lion and honey pot values

to protect the sentiment analysis and continues monitoring. Thus, the developed replica detects and neglects the attacks using the ALHR framework. Additionally, developed replica gains better performance in classify twitter data into positive, neutral, and negative. Consequently, launch attacks in sentiment analysis for validating the strength of the developed technique.

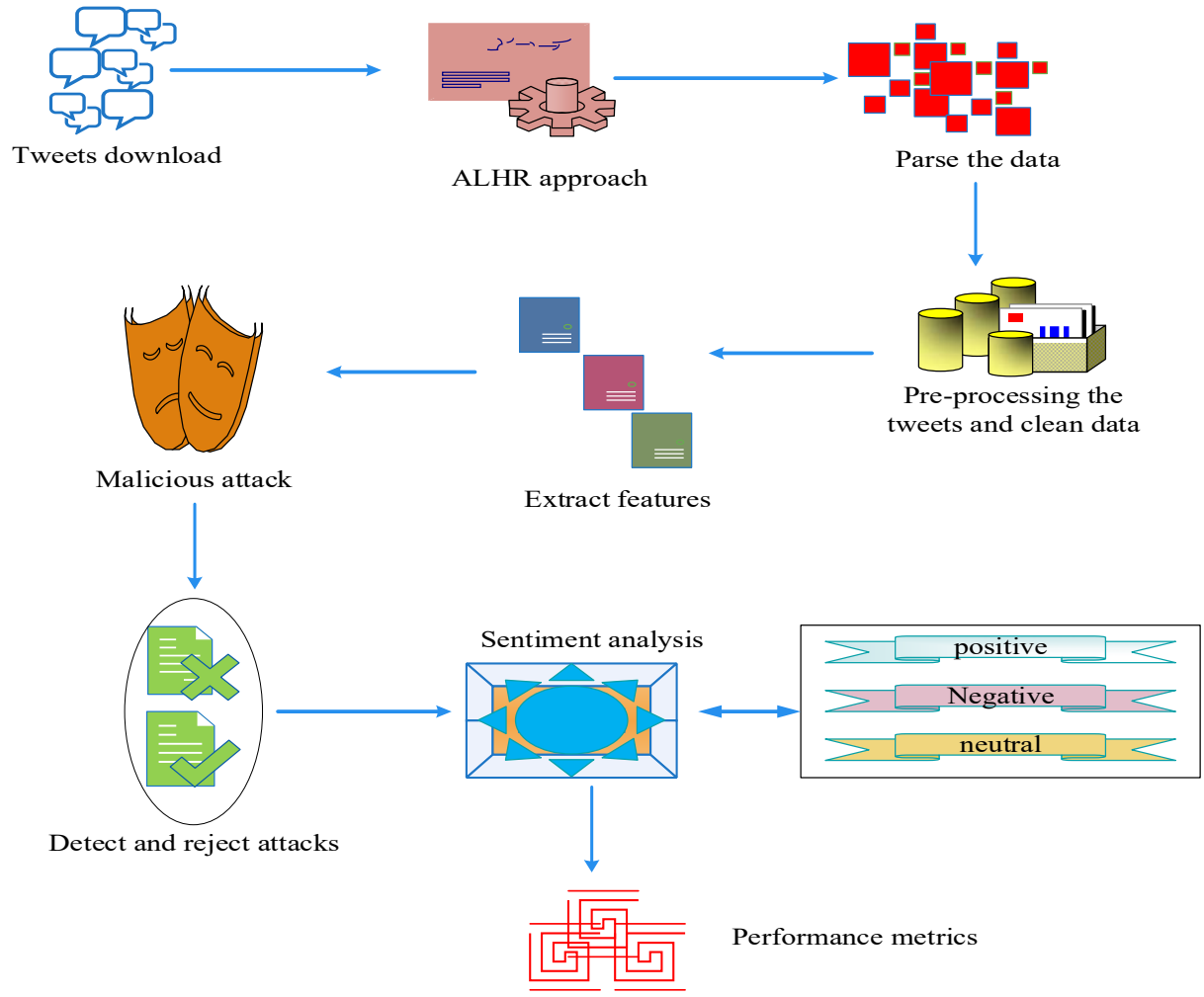

Fig. 2. Proposed ALHR methodology

Thus, the overall process of the ALHR framework is illustrated in Fig.2. The proposed ALHR approach parses and preprocesses the trained data. Next, the attacks are detected and neglected with the updated fitness function of Ant Lion Honeypot (ALH). At last, the sentiment analysis classifies as positive, neutral, and negative.

4.1. *Description of dataset*

The performance of sentiment analysis of the developed approach is executed using Twitter data or tweets about COVID-19 disease, and the tweets are collected from Kaggle.com. Moreover, the employed dataset contains 38460 quantities of users, including user name, ID, location, followers, friends, likes, and date. The data are used to classify the sentiment analysis such as positive review, negative review, and neutral review. Furthermore, collected datasets are transferred to the proposed ALHR technique for performing sentiment analysis.

4.2. *Ant Lion Honeypot with Regression (ALHR)*

The developed ALHR framed work is introduced to enhance privacy at a high rate and continuous monitoring of attacks. Thus, the proposed ALHR framework performs several processes such as parse the data, preprocessing, features extraction, detect malicious attacks, and sentiment analysis classification. In this technique used data set is COVID-19 disease. Moreover, the introduced technique enhances the performance of sentiment analysis and detects malicious activity that occurs during classification. Also, the fitness function of ant-lion and honey pot prevents malicious activity and continuous monitoring of sentiment analysis. Initially, the collected, trained data set are updated to the input layer of the proposed ALHR technique that is given as eqn. (1).

$$y(t) = [d(1), d(2), d(3), \ldots\ldots, d(n)] \quad (1)$$

Let $y(t)$ is denoted as the total quantity of updated dataset, and d (1), d (2) ……d(n) is represented as the dataset of tweets downloaded from Twitter.

4.2.1. Parse and preprocessing

After finishing the training process, the next process is started that is parsing and preprocessing of data which means cleaning the data because tweets contain many opinions that are expressed in various ways or by various users. In the cleaning process, the proposed ALHR technique removes special characters, URLs, hashtags, all punctuations, numbers, non-Spanish tweets, emoji ideograms, repeated words, and symbols presented in the tweets. Also, it will convert the tweets into the lower case for making the dataset uniform. Thus, the preprocessing and parsing of the data are done by eqn. (2).

$$T_d = \frac{d_i - min(d)}{max(d) - min(d)} \tag{2}$$

Where T_d is expressed as the performance of parse and preprocessing data and d_i is represented as an aspect of noise present in the dataset of tweets such as hashtags, punctuations, repeated words, special characters, and so on are removed. Moreover d is denoted as the lower case of the dataset.

4.2.2. Extract features

Consequently, preprocessed datasets are sent to the features extraction. Thus, the trained preprocessed tweets are converted into a numerical symbol. Based on the numerical value, easily identify the positive, negative, and neutral comments. Furthermore, it will extract the features in the aspects of the dataset and operate to identify sentence polarity. Thus, the feature extraction is mentioned in eqn. (3).

$$Y_p^t = \frac{(Y_p^t - T_d) \times (d_p - p_a)}{d_p - T_d} \tag{3}$$

Where, Y_p^t is denoted as denoted as feature extraction of tweets data and p_a is considered as positive features of tweets also$d_p$is denoted as negative features of tweets extracted from sentiment analysis. At next, the extracted features are sent to the sentiment analysis to classify the tweets based on positive, neutral, and negative reviews.

4.2.3. Detection of malware activity

Afterward, attacks are generated during the classification of sentiment analysis. So measured fitness function of the ALH framework is updated to the regression for preventing the attacks and provide continuous monitoring. Thus, the attacks are identified using eqns. (4) and (5)

$$GA_j^t = GP^t \tag{4}$$

$$\text{If} f\left(GA_j^t\right) > f(GP^t) \tag{5}$$

Let, GA_j^t is denoted as the continuous monitoring of data and GP^t is expressed as sentiment analysis dataset value. Whether, the obtained value of continuous monitoring is greater than the dataset means there will be a presented attack. The detected attacks are neglected with the help of the ALH fitness function. But the value of continuous monitoring is equal to the dataset means there are no presented attacks.

4.2.4. Classification of sentiment analysis

After neglecting the attacks, they start the classification process to classify the positive, neutral, and negative sentiment analysis. Thus, the fitness function of ALHR is given in eqn. (6).

$$f(H) = \frac{HA^t + Y_p^t + T_d}{2} \tag{6}$$

Where, $f(H)$is represented as the fitness function of ALH and HA^t is denoted as the polarity of sentiment analysis and. However, the proposed ALHR technique increases the privacy rate of sentiment analysis, detects attacks, and provides continuous monitoring.

Algorithm:1 ALHR for sentiment analysis and continuous monitoring
Start
{
Initialization
Update the datasets

//COVID-19 disease tweets from Twitter data
Trained the dataset in ALH

```
Parse and preprocessing
        Clean the data and remove unwanted data
        // remove special characters, URLs, hashtags, punctuations, numbers, symbols, stop words, emojis, and
repeated words
Extract features
        Extract the features of the words in the dataset
        // features are extracted based on the numerical value
Attack detection
        Update the fitness function of ALH
        Launch attack
        Continuous monitoring of data while its present attack
        If( )

        {
        No attack
        }
        then
        If( )

        {
        Presented attack
        }
        End if
        Neglect the attacks with proposed ALH
Sentiment analysis classification
        Analyze the tweets using eqn.6
        If

        {
        Sentiment is positive (+1)
        // tweets is positive comments
        }
        Else if

        {
        Sentiment is negative (-1)
        //tweets are negative comments
        }
        otherwise
        {
        Sentiment is neutral (0)
        //tweets are common, either positive or negative
        }
        End if
        High privacy rate in classification
        Measured performance metrics
}
End
```

Generally, the classification of sentiment analysis is identified in terms of aspects. While the tweets contain a more positive aspect, they are considered positive tweets, and if the tweets contain a more negative aspect, they are denoted as negative tweets. Moreover, the remaining tweets are represented as neutral tweets. Based on the positive and negative quantity, tweets are classified. They launched attacks such as Daniel of service (DoS) attacks and neglected them by developed the ALHR framework. Thus, the fitness function of ALR is used to prevent attacks and continuous monitoring. Additionally, the proposed ALHR technique attains a high rate of insecurity, continuous monitoring, and detect malicious activities. The overall process of the developed ALHR method has illustrated in Fig.3 and algorithm.1.

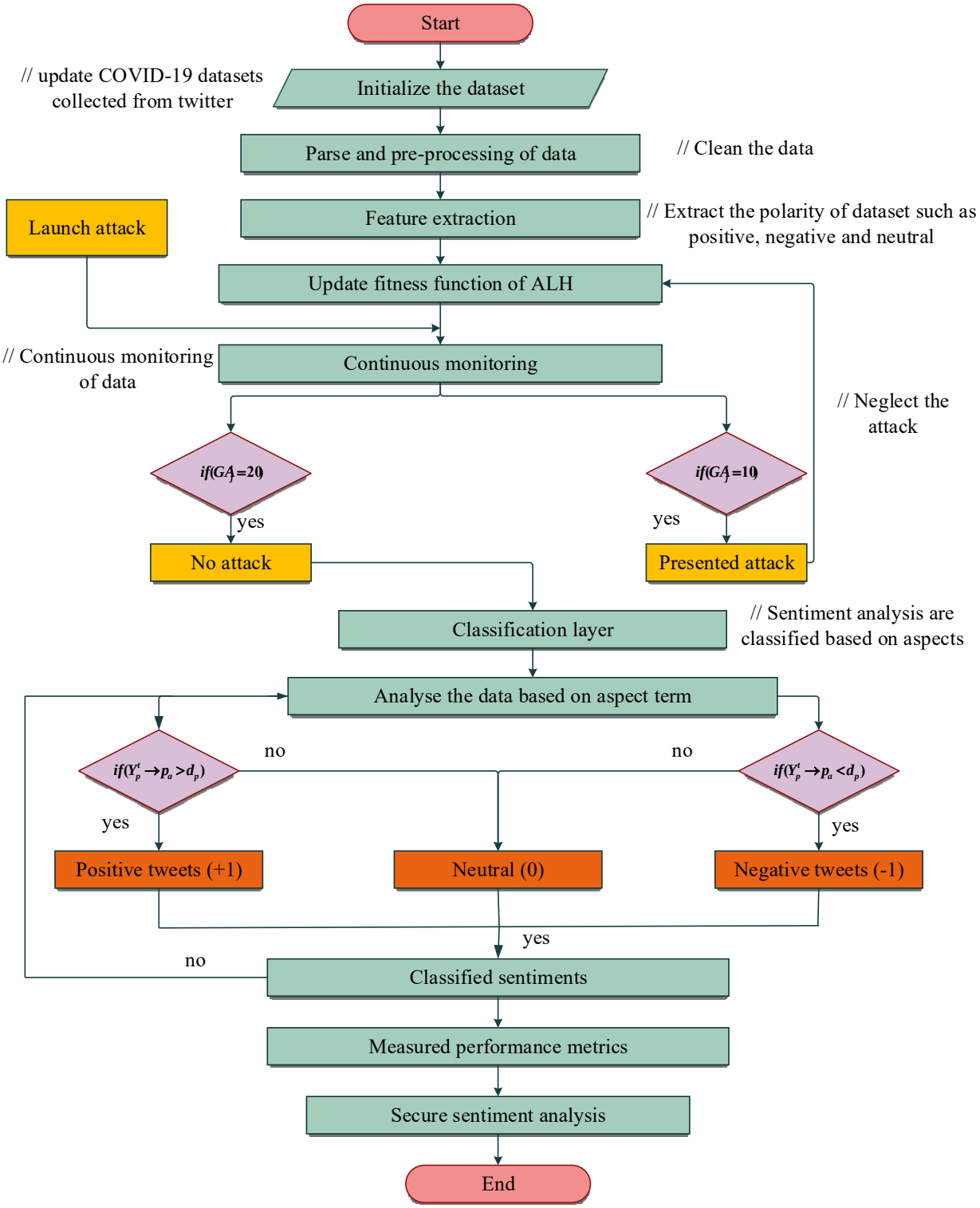

Fig. 3. Flowchart of ALHR technique

Moreover, the proposed ALHR technique successfully classified the sentiment analysis and ALR fitness function to prevent attacks and continuous monitoring. Thus, the developed technique enhances the security of sentiment analysis and reduces the error rate.

5. Results and Discussion

Initially, the proposed ALHR framework is implemented in the python tool, and the efficiency of the developed replica is measured with other existing techniques regarding the accuracy, recall, F-measures, precision, and error rate. Furthermore, the introduced method detects malicious activity with continuous monitoring and provides a high privacy rate in sentiment analysis.

5.1. *Case Study*

Generally, the classification of sentiment analysis and the detection of attacks are identified using downloaded tweets through the developed ALHR framework. Thus, the tweets are downloaded from twitter based upon user opinion, user reviews, and user comments about COVID-19. Let us more than 1500 tweets are collected and trained to the system. Thus, the trained tweets are initialized by the proposed ALHR replica using eqn.1. Hereafter, updated datasets enter into the parsing and preprocessing process. In this process, they remove all unwanted punctuation, errors, full stops, special character, stop words, hashtags, and repeated words present in the tweets. Subsequently, cleaned datasets enter into the next level that is the feature extraction process. The function of feature extraction is to extract the features based upon the polarity of the sentence. Thus, the aspects terms of the sentence identify the positive, negative and neutral. Moreover, positive aspects are considered, and negative aspects are denoted, and common aspects are considered neutral. The positive aspects are classified in terms of vaccine, immunity, mask, healthy, sanitizer, etc. Negative aspects are characterized in terms of death, symptoms, side effects, lockdown, spread, etc. Thus, the workflow of the developed ALHR approach has been illustrated in Fig. 4.

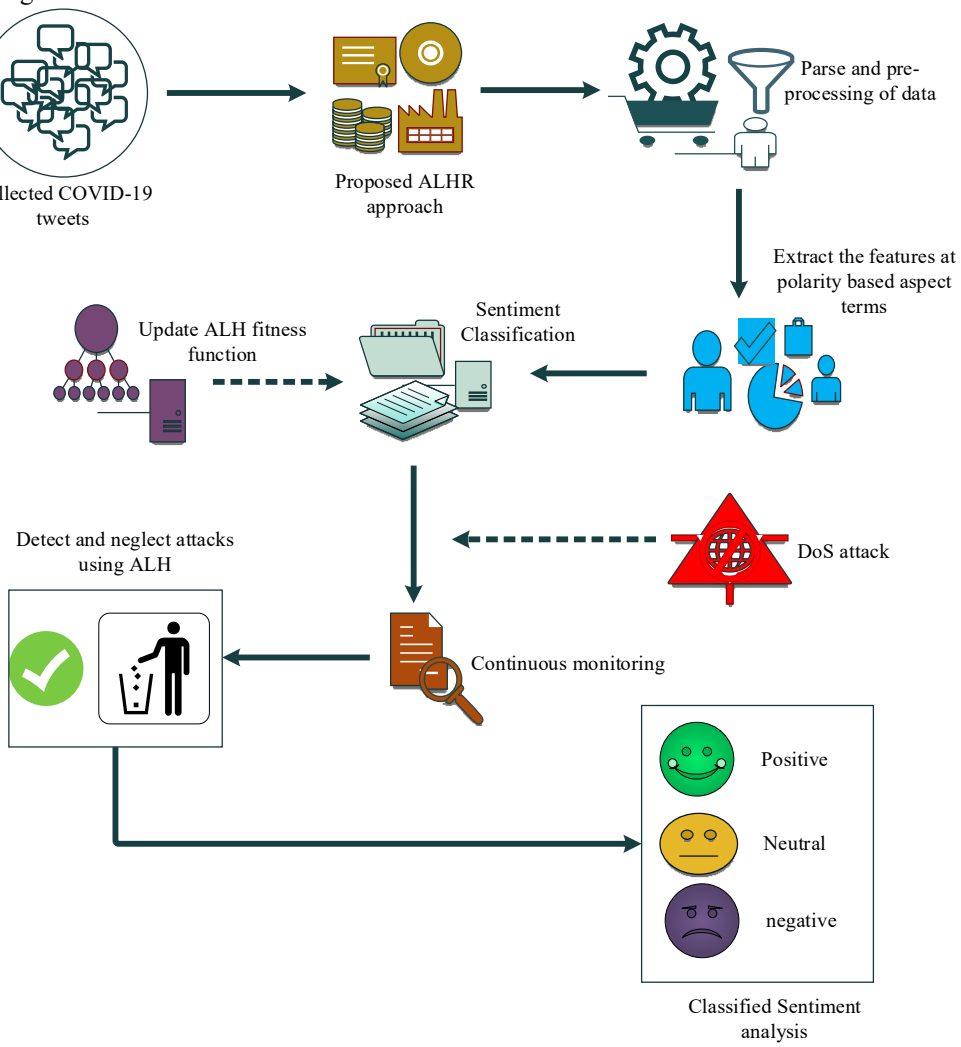

Fig.4. Workflow of Proposed ALHR technique

At next, extracted features are sent to the sentiment classification. In this section, they classify the data based upon positive, negative, and neutral. The measured fitness function of ALH is updated to the sentiment layer for the purpose of continuous monitoring and detecting attacks. During sentiment analysis, DoS attacks are launched for testing the efficiency of the developed replica. The generated attacks are detected and neglected with updated ALH. Finally, classified sentiment analyses are gained in an accurate manner, and the proposed ALHR model

continuously monitors the data; if the malware is entered, then it is automatically neglected. Thus, the proposed approach has accomplished better results for performing detecting malware and continuous monitoring of data.

5.2. *Performance Metrics*

The sentiment analysis of the proposed ALHR technique is implemented in Python. Furthermore, gained performance metrics are validated with respect to the accuracy, recall, error rate, etc. Thus, the achieved performance is compared with other existing techniques such as Multi-Class Sentiment Analysis (MCSA) [28], Troll Detection (TD) [27], Bayesian Network Classifier (BNC) [24], and Sentiment Analysis of Extremism (SAE) [25].

5.2.1 Accuracy

The measurement of accuracy (A) is developed to establish the effectiveness of the developed technique through detecting attacks and continuous monitoring. Also, it detects the success of the ALHR technique to categorize sentiments. Moreover, accuracy measurements are calculated depending on classified tweets using eqn. (7),

$$6 \quad A = \left(\frac{A_c + A_r}{A_c + Q_c + Q_r + A_r}\right) \tag{7}$$

Where, $A_c$is denoted as true positive, which is the measurement of the total amount of appropriately categorizing positive tweets, $A_r$is represented as true negative, which is the entire amount of correctly categorizing negative tweets. Moreover,$Q_c$is expressed as false positive, which signifies the total amount of unacceptably categorizing positive tweets, and $Q_r$is called a false negative that denotes the entire quantity of unacceptably categorizing negative tweets.

Table.2 Accuracy comparison

Number of tweets	Accuracy (%)				
	MCSA	TD	BNC	SAE	ALHR (proposed)
1000	91.2	82	78	75	98
2000	88	79	74.2	73.2	96.4
3000	84	75.3	70	70	93
4000	82.5	72	64.3	66	90
5000	70.1	68	60	63.2	88.5

The achieved accuracy rate of the proposed ALHR technique is compared with other existing replicas such as MCSA, SAE, TD, and BNC. Thus the BNC replica attained 78% accuracy with 1000 tweets, and the TD method gained 82% accuracy. Moreover, the MCSA method attained 91.2 % accuracy, and the SAE technique achieved 75% with 1000 tweets. The accuracy comparison of the exciting technique is detailed in table 2 and Fig.5.

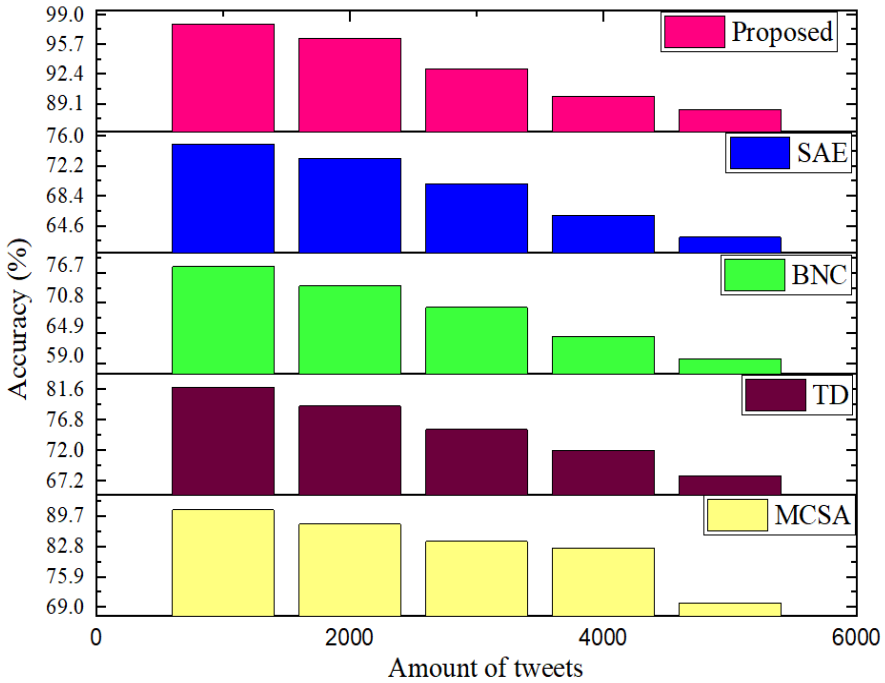

Fig.5. Comparison of accuracy with the existing method

Furthermore proposed replica achieved 98% accuracy for using 1000 tweets, and it has high accuracy rate than other existing replicas. Also, it will establish the effectiveness of the ALHR technique.

5.2.2 Precision

The computation of precision (P) is operated for recognizing the success of the proposed ALHR technique while detecting attacks and continuous monitoring. In addition, the measurement of precision rate is obtained using eqn. (8) and comparison of precision has shown in Fig. 6.

$$P = \left(\frac{A_c}{A_c+Q_c}\right) \tag{8}$$

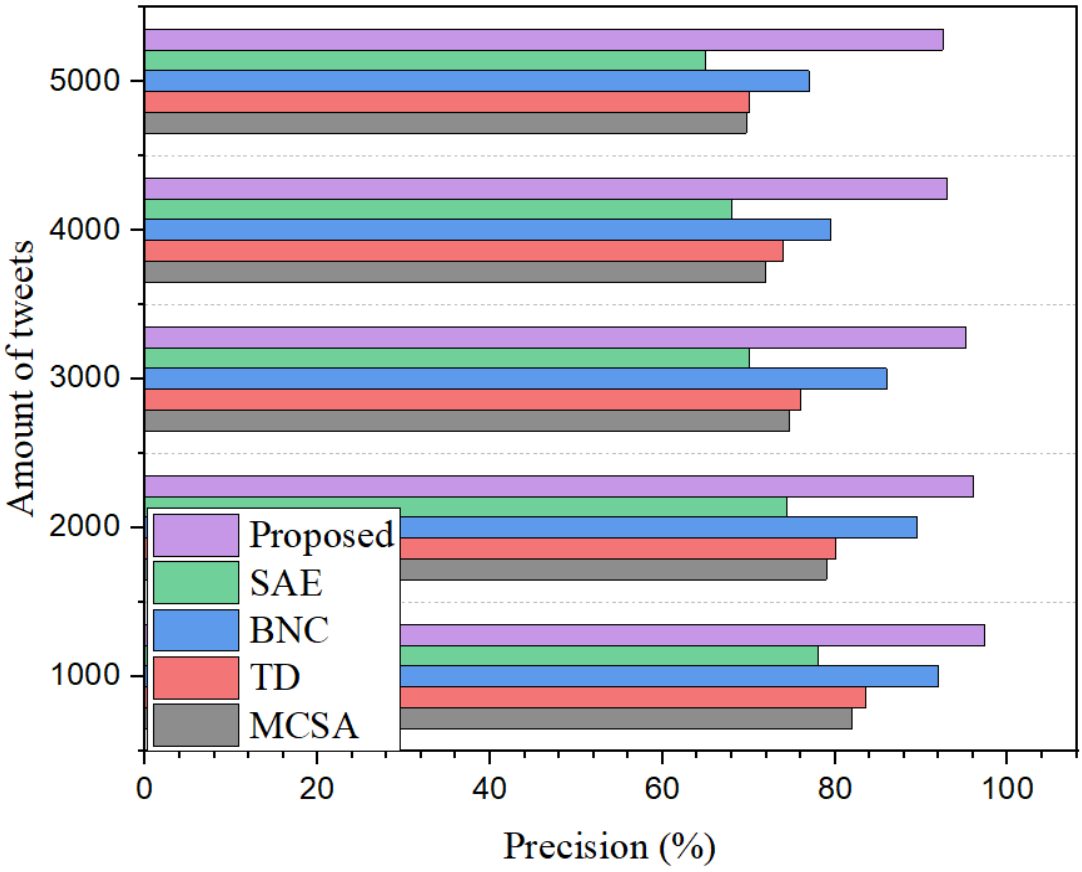

Fig.6. Precision comparison

The achieved performance of the ALHR technique with 1000 tweets precision rate is 97.4%, and the gained precision value is compared with other existing replicas. Moreover, MCSA and TD techniques attained 82% and 83.5% precision rates. Also, the SAE replica gained a precision rate of 1000 tweets is 78%, and the BNC method achieved 92% precision.

Table.3 Comparison of precision

Number of tweets	Precision (%)				
	MCSA	TD	BNC	SAE	ALHR (proposed)
1000	82	83.5	92	78	97.4
2000	79	80	89.5	74.4	96
3000	74.7	76	86	70	95.2
4000	72	74	79.5	68	93
5000	69.7	70	77	65	92.5

The overall comparison of other existing techniques, proposed ALHR method attain a high rate of precision, and the comparison of precision values with other existing methods are explained in table.3.

5.2.3 Recall

Measurement of recall (R) is developed to categorize the sensitivity of the developed ALHR technique. Additionally, recall is the term of true positive value to the addition of false-negative and true positive value. Moreover, the recall calculation of the ALHR method was obtained using eqn. (9),

$$R = \left(\frac{A_c}{A_c + Q_r} \right) \tag{9}$$

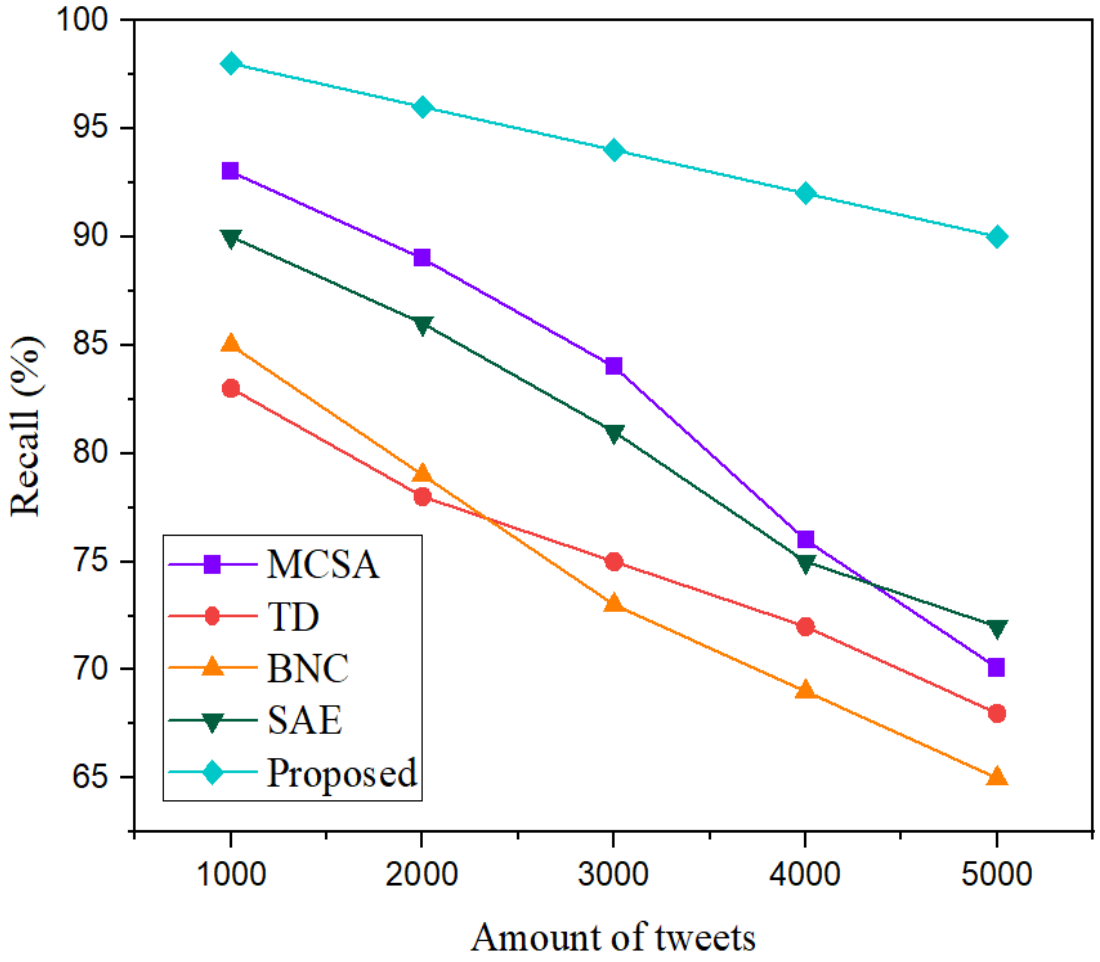

Fig.7. Comparison of recall

The achieved recall rate of the proposed ALHR technique is compared with other existing replicas such as MCSA, SAE, TD, and BNC. Thus the BNC replica attained 85% in recall with 1000 tweets, and the TD method gained 83% recall. Moreover, the MCSA method attained 93 % recall, and the SAE technique achieved 90% with 1000 tweets. The recall comparison of the exciting techniques is detailed in table 4 and fig. 7.

Table.4 Recall Comparison

Number of tweets	Recall (%)				
	MCSA	TD	BNC	SAE	ALHR (proposed)
1000	93	83	85	90	98
2000	89	78	79	86	96
3000	84	75	73	81	94
4000	76	72	69	75	92
5000	70.1	68	65	72	90

Furthermore, proposed replica achieved 98% recall for using 1000 tweets, and it has high recall rate than other existing replicas. Also, it will establish the effectiveness of the ALHR technique.

5.2.4 F-Measure

The F-measure calculation is called the arrangement of computed precision value and recall value, which is calculated using eqn. (10),

$$F\_measure = \left(2 \frac{P*R}{P+R}\right) \tag{10}$$

Table.5 Comparison of F-measure with existing technique

Number of tweets	F-measure (%)				
	MCSA	TD	BNC	SAE	ALHR (proposed)
1000	86	84.3	89.9	85	97
2000	80	81	84	80.9	96.3
3000	77.7	78.5	79.5	76	96
4000	73.4	73	75	71.1	94.3
5000	69.9	70	72	67	93

The achieved performance of the ALHR technique with 1000 tweets F-measure is 97%, and the gained F-measure value is compared with other existing replicas. Moreover, MCSA and TD techniques attained 86% and 84.3% in F-measure. Also, the SAE replica gained an 85% F-measure for 1000 tweets, and the BNC method achieved 89.9% in F-measure.

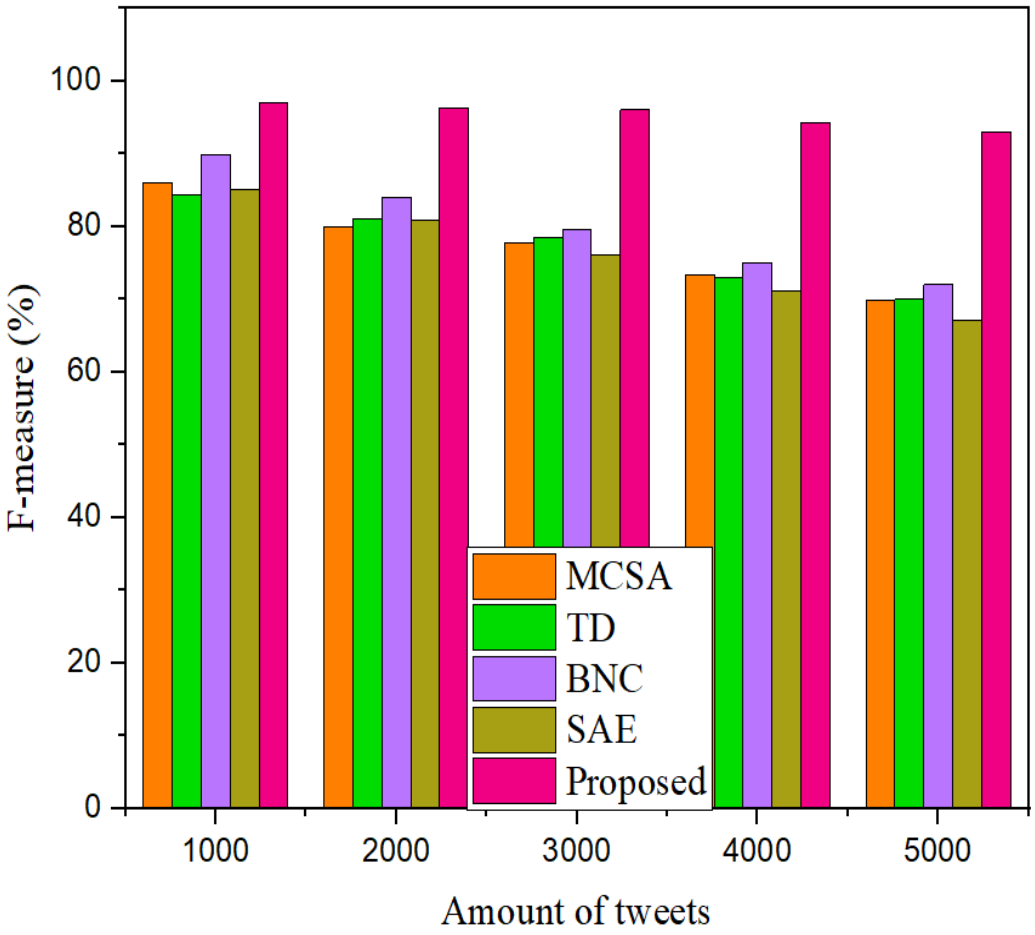

Fig.8 F-Measure comparison

From the overall comparison of other existing techniques, the proposed ALHR method attained a high rate of F-measure, and the comparison of F-measure with other existing is explained in table 3 and Fig.8.

5.2.5 Error Rate

Error Rate (ER) is the ratio of the number of errors in the tweets to the quantity of all classified tweets that are obtained using eqn. (11). The measurement of error rate calculation is operating for identifying sentiment errors in the developed ALHR technique, and the evaluation of error rate is described in Fig. 9.

$$ER = \left(\frac{F_P + F_N}{T_P + T_N + F_P + F_N}\right) \tag{11}$$

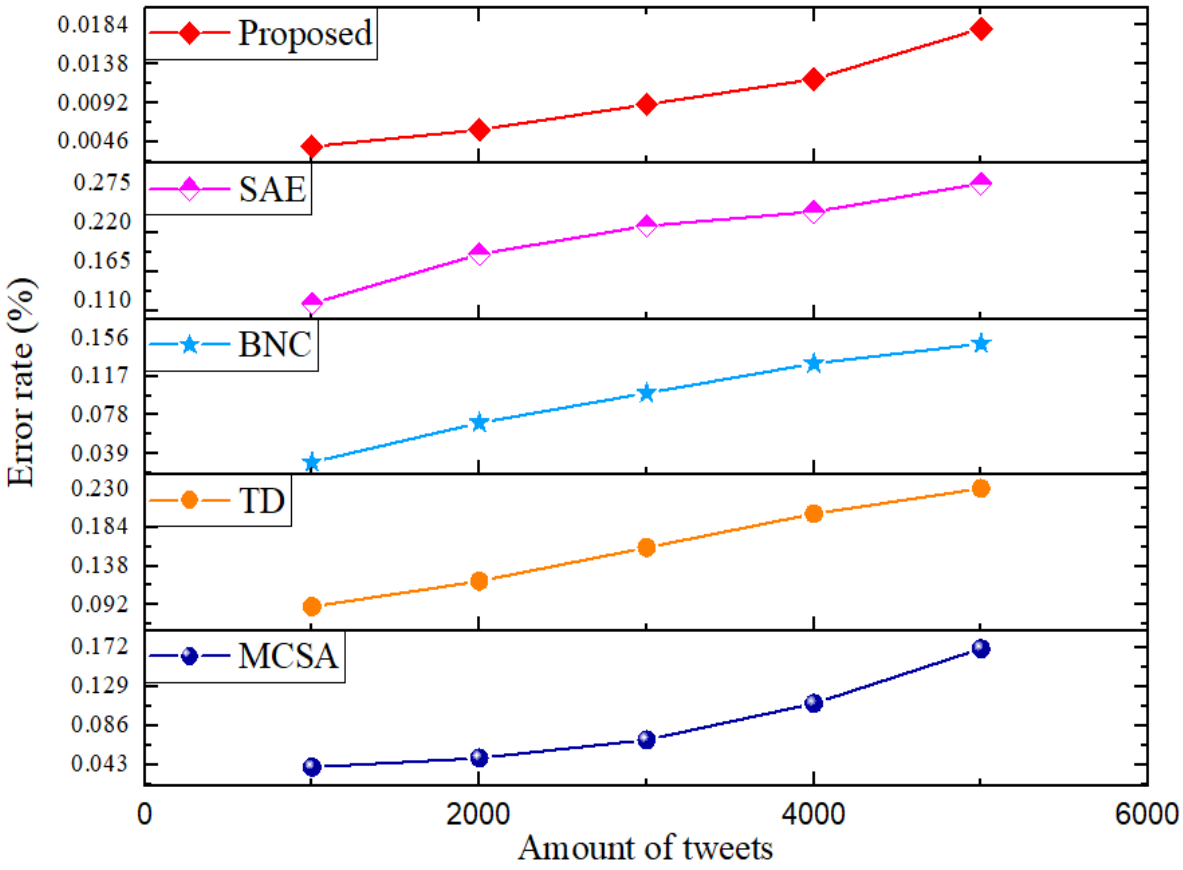

Fig. 9 Comparison of error rate

The gained error rate value is compared to other existing techniques such as MCSA, SAE, BNC, and TD. Initially, the error rate of 1000 tweets of SAE obtain 0.12%, BNC replica achieved 0.03%, and MCSA model attained 0.04%, similarly proposed ALHR approach attained 0.004%.

Table.6 error rate comparison

Number of tweets	Error rate (%)				
	MCSA	TD	BNC	SAE	ALHR (proposed)
1000	0.04	0.09	0.03	0.12	0.004
2000	0.05	0.12	0.07	0.19	0.006
3000	0.07	0.16	0.10	0.23	0.009
4000	0.11	0.20	0.13	0.25	0.012
5000	0.17	0.23	0.15	0.29	0.018

While comparing to other existing techniques proposed ALHR framework achieved a low error rate, and the overall comparison is shown in table.6. Thus, the developed framework minimizes the error rate presented in the dataset, and it will indicate the efficiency of the ALHR technique.

5.3. *Discussion*

The proposed model ALHR has pertained to a good performance from the overall result assessment by attaining high accuracy, precision, recall, F-measure, etc. Thus, the developed scheme detects attacks present in sentiment analysis and provides continuous monitoring. The collected dataset is downloaded from Twitter, and the dataset is trained to the developed ALHR framework. The updated dataset are parsed, preprocessing, feature extracted, and classified by the developed framework. Finally, classify the sentiment analysis and provide security also continuous monitoring of data.

6. Conclusion

Monitoring malicious attacks during sentiment analysis and providing continuous monitoring is the most critical task in sentiment analysis. So in this paper, a novel Ant Lion Honeypot with regression (ALHR) framework is proposed. This technique used dataset is COVID-19 disease that is collected from Twitter, and the collected

datasets are trained and updated to the developed technique. Moreover, the developed ALHR approach performs parse and preprocessing processes. It will clean the data for removing unwanted errors, hashtags, URLs, special characters, and repeated words. At next, extraction of features is a function based on aspects terms. Thus the fitness function of ALH is updated to the classification layer for detecting attacks and provides continuous monitoring. Launch attacks in the classification layer for identifying the efficiency of the developed technique. The generated attacks are detected and neglected with the proposed ALHR technique. Finally, classified sentiments are categorized based upon positive, negative, and neutral. As a result, the developed ALHR technique attained a high accuracy rate of 98% and a low error rate compared to other existing techniques, which is 0.004%. Furthermore, it will enhance the security of sentiment analysis and provide continuous monitoring.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] Riaz, Sumbal, et al. "Opinion mining on large scale data using sentiment analysis and k-means clustering." Cluster Computing 22.3 (2019): 7149-7164.
[2] Toussaint, Mariana, Pablo Cabanelas, and Tania E. González-Alvarado. "What about the consumer choice? The influence of social sustainability on consumer's purchasing behavior in the Food Value Chain." European Research on Management and Business Economics 27.1 (2021): 100134.
[3] Alsaeedi, Abdullah, and Mohammad Zubair Khan. "A study on sentiment analysis techniques of Twitter data." International Journal of Advanced Computer Science and Applications 10.2 (2019): 361-374.
[4] Shayaa, Shahid, et al. "Sentiment analysis of big data: Methods, applications, and open challenges." IEEE Access 6 (2018): 37807-37827.
[5] Öztürk, Nazan, and Serkan Ayvaz. "Sentiment analysis on Twitter: A text mining approach to the Syrian refugee crisis." Telematics and Informatics 35.1 (2018): 136-147.
[6] Calefato, Fabio, et al. "Sentiment polarity detection for software development." Empirical Software Engineering 23.3 (2018): 1352-1382.
[7] Hamad, Mhd Mousa, Marcin Skowron, and Markus Schedl. "Regressing Controversy of Music Artists from Microblogs." 2018 IEEE 30th International Conference on Tools with Artificial Intelligence (ICTAI). IEEE, 2018.
[8] Appel, Gil, et al. "The future of social media in marketing." Journal of the Academy of Marketing Science 48.1 (2020): 79-95.
[9] Zeng, Zexian, et al. "Natural language processing for EHR-based computational phenotyping." IEEE/ACM transactions on computational biology and bioinformatics 16.1 (2018): 139-153.
[10] Poria, Soujanya, Amir Hussain, and Erik Cambria. Multimodal sentiment analysis. Cham: Springer International Publishing, 2018.
[11] Chaturvedi, Iti, et al. "Distinguishing between facts and opinions for sentiment analysis: Survey and challenges." Information Fusion 44 (2018): 65-77.
[12] Ghosh, Monalisa, and Goutam Sanyal. "An ensemble approach to stabilize the features for multi-domain sentiment analysis using supervised machine learning." Journal of Big Data 5.1 (2018): 1-25.
[13] Acker, Amelia, and Dhiraj Murthy. "What is Venmo? A descriptive analysis of social features in the mobile payment platform." Telematics and Informatics 52 (2020): 101429.
[14] Linvill, Darren L., et al. "“THE RUSSIANS ARE HACKING MY BRAIN!” investigating Russia's internet research agency twitter tactics during the 2016 United States presidential campaign." Computers in Human Behavior 99 (2019): 292-300.
[15] Donratanapat, N., et al. "A national scale big data analytics pipeline to assess the potential impacts of flooding on critical infrastructures and communities." Environmental Modelling & Software 133 (2020): 104828.
[16] Shaukat, Zeeshan, et al. "Sentiment analysis on IMDB using lexicon and neural networks." SN Applied Sciences 2.2 (2020): 1-10.
[17] Scheffran, Jürgen. "2 Climate change and weather extremes as risk multipliers." Climate Change, Security Risks, and Violent Conflicts (2020): 19.
[18] Rouhanizadeh, Behzad, and Sharareh Kermanshachi. "Post-disaster reconstruction of transportation infrastructures: Lessons learned." Sustainable Cities and Society 63 (2020): 102505..
[19] Chakraborty, Koyel, et al. "Sentiment Analysis of COVID-19 tweets by Deep Learning Classifiers—A study to show how popularity is affecting accuracy in social media." Applied Soft Computing 97 (2020): 106754.
[20] Fombelle, Paul W., et al. "Customer deviance: A framework, prevention strategies, and opportunities for future research." Journal of Business Research 116 (2020): 387-400.
[21] Thelwall, Mike. "Sentiment analysis for tourism." Big Data and Innovation in Tourism, Travel, and Hospitality (2019): 87-104.
[22] Usman, Nighat, et al. "Intelligent dynamic malware detection using machine learning in IP reputation for forensics data analytics." Future Generation Computer Systems 118 (2021): 124-141).
[23] Mansour, Samah. "Social media analysis of user’s responses to terrorism using sentiment analysis and text mining." Procedia Computer Science 140 (2018): 95-103.
[24] Ruz, Gonzalo A., Pablo A. Henríquez, and Aldo Mascareño. "Sentiment analysis of Twitter data during critical events through Bayesian networks classifiers." Future Generation Computer Systems 106 (2020): 92-104.
[25] Asif, Muhammad, et al. "Sentiment analysis of extremism in social media from textual information." Telematics and Informatics 48 (2020): 101345.
[26] Kandasamy, Ilanthenral, et al. "Sentiment analysis of tweets using refined neutrosophic sets." Computers in Industry 115 (2020): 103180.
[27] Jiang, Zidong, Fabio Di Troia, and Mark Stamp. "Sentiment Analysis for Troll Detection on Weibo." Malware Analysis Using Artificial Intelligence and Deep Learning. Springer, Cham, 2021. 555-579.
[28] Bouazizi, Mondher, and Tomoaki Ohtsuki. "A pattern-based approach for multi-class sentiment analysis in Twitter." IEEE Access 5 (2017): 20617-20639.
[29] V. Laxmi Narasamma and M. Sreedevi, “Twitter based Data Analysis in Natural Language Processing using a Novel Catboost Recurrent Neural Framework” International Journal of Advanced Computer Science and Applications(IJACSA), 12(5), 2021. http://dx.doi.org/10.14569/IJACSA.2021.0120555

[30] V. Laxmi Narasamma and M. Sreedevi, "DETECTING MALICIOUS ACTIVITIES ON TWITTER DATA FOR SENTIMENT ANALYSIS USING A NOVEL OPTIMIZED MACHINE LEARNING APPROACH" Journal of Theoretical and Applied Information Technology 15th December 2021. Vol.99. No 23 © 2021 Little Lion Scientific

Authors Profile

V. Laxmi Narasamma, Research Scholar, Department of CSE, Koneru Lakshmaiah Education Foundation, Green Fields, Vaddeswaram, Guntur, Andhra Pradesh 522502, India, Email: laxmi8866@gmail.com

Dr. M. Sreedevi, Professor, Department of CSE, Koneru Lakshmaiah Education Foundation, Green Fields, Vaddeswaram, Guntur, Andhra Pradesh 522502, India, Email: msreedevi_27@kluniversity.in