Computer Systems
Science and Engineering

Tech Science Press

DOI: 10.32604/csse.2023.026915
Article

Web Intelligence with Enhanced Sunflower Optimization Algorithm for Sentiment Analysis

Abeer D. Algarni*

Department of Information Technology, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, P.O.Box 84428, Riyadh, 11671, Saudi Arabia
*Corresponding Author: Abeer D. Algarni. Email: adalqarni@pnu.edu.sa
Received: 06 January 2022; Accepted: 11 February 2022; Published: 26 May 2023

Abstract: Exponential increase in the quantity of user generated content in websites and social networks have resulted in the emergence of web intelligence approaches. Several natural language processing (NLP) tools are commonly used to examine the large quantity of data generated online. Particularly, sentiment analysis (SA) is an effective way of classifying the data into different classes of user opinions or sentiments. The latest advances in machine learning (ML) and deep learning (DL) approaches offer an intelligent way of analyzing sentiments. In this view, this study introduces a web intelligence with enhanced sunflower optimization based deep learning model for sentiment analysis (WIESFO-DLSA) technique. The major intention of the WIESFO-DLSA technique is to identify the expressions or sentiments that exist in the social networking data. The WIESFO-DLSA technique initially performs pre-processing and word2vec feature extraction processes to generate a meaningful set of features. At the same time, bidirectional long short term memory (BiLSTM) model is applied for classification of sentiments into different class labels. Moreover, an enhanced sunflower optimization (ESFO) algorithm is exploited to optimally adjust the hyperparameters of the BiLSTM model. A wide range of simulation analyses is performed to report the better outcomes of the WISFO-DLSA technique and the experimental outcomes ensured its promising performance under several measures.

Keywords: Sentiment analysis; web intelligence; deep learning; social networking; natural language processing

1 Introduction

The integration of Web Intelligence with data mining, business analytics, infrastructure data visualization, best practices, and data tools, assist organizations to make further data-driven decisions [1]. Practically, current company intelligence has a common assumption of our organization's information and makes use of that information to eliminate inefficiencies, quickly adapt, and drive change to supply or market changes. Web intelligence constitutes the use of worldwide web (WWW) as a phenomenon of data retrieval from data storage in a smarter way. Generally, Sentiment Analysis (SA) is the method of categorizing and identifying the polarization of a text at phrase, document, and sentence levels [2]. This method has been utilized in several domains such as politics, e-commerce, healthcare, entertainment, and so on. For instance, SA is beneficial to businesses for monitoring user opinion about the product, and consumer select the good products according to public opinion [3]. The major process in Twitter SA is to describe the opinions of Twitter as negative or positive opinions. The major problems of twitter SA are: (a) abbreviations and acronyms are commonly utilized on twitter (b) tweet posts are usually written in informal language and (c) short message shows limited cues regarding their sentiments [4,5]. Fig. 1 showcases the process involved in SA.

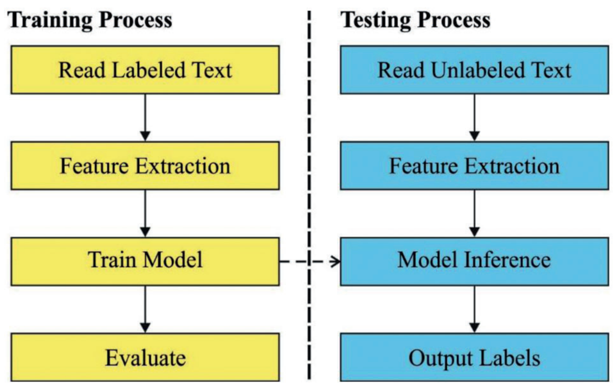

Figure 1: Process in sentiment analysis

Because of the increasing textual data, there has been a necessity for analyzing the concepts of expressing sentiment and calculating the insight to explore businesses [6]. Advertising companies and Business owners frequently employ SA to determine advertising campaigns and new business strategies. The machine learning (ML) algorithm is highly useful in classifying and predicting documents that represent positive or negative sentiments. The ML method is considered as supervised and unsupervised ML methods. Supervised approach employs labeled datasets in which all the documents of training set are labeled using relevant sentiments [7]. While unsupervised learning includes unlabelled datasets in which text isn't labeled using relevant sentiments. The research mostly considered supervised learning techniques on labeled datasets. Usually, SA can be performed on document, aspect, and sentence levels. Sentence level sentiment classification considers the polarity of single sentence of a document. Document Level sentiment classification focus on categorizing the whole topic or document as negative or positive. Aspect level sentiment classification recognizes the divergent aspect of corpus and for all the documents, the polarity is estimated by the attained aspect [8].

Conventional methods based on automatic feature engineering, i.e., time consuming. At the same time, deep learning (DL) method is a promising alternative to conventional models [9]. It proved outstanding efficiency in NLP tasks, including SA. The primary concept of DL methods is to learn complicated features extracted from information with minimal external contributions with deep neural network (DNN) model. This algorithm doesn't want to be passed automatically crafted feature: they manually learn new feature [10]. Nonetheless, a representative feature of DL method is that they require massive number of information to be well performed. The availability of resources and automated feature extraction is vital while comparing the conventional ML and DL approaches.

Yoo et al. [11] presented the Polaris, the system to analyze and predict users' sentimental trajectory to event analyzing from real time out of huge social media content, and demonstrate the outcomes of initial validation work is completed. Araque et al. [12] progressed a DL based sentiment classification utilizing a word embedded technique and linear ML technique. This classifier serves as baseline for comparing with following outcomes. Secondary can be presented 2 ensemble approaches that aggregate our baseline classifiers with other surface classifiers extremely utilized in SA.

Jawad et al. [13] related the efficiency of distinct ML and DL techniques, besides presenting a novel hybrid technique that utilizes text mining and NNs to sentimental classification. The data set utilized during this work has more than 1 million tweets gathered from 5 domains. Chen et al. [14] utilized the Military life PTT board of Taiwan's main online forum as source of their experimental data. The drive of this analysis is for constructing a SA structure and procedures to social media for proposing a self-developed military sentiment dictionary to increase sentimental classification and analyzing the efficacy of several DL techniques with different parameter calibration combinations. The experimental outcomes illustrate that the accuracy and F1-score of this technique which integrates present sentiment dictionary and the self-developed military sentimental dictionaries are superior to the outcomes in utilizing present sentimental dictionary only.

Valencia et al. [15] presented the procedure of general ML tools and accessible social media information to forecast the price effort of Bitcoin, Ethereum, Ripple, and Litecoin cryptocurrency market movement. It can be related to the consumption of NN, SVM, and RF but utilizing components in Twitter and market data as input features. Vashishtha et al. [16] compute the sentimental of social media posts utilizing a novel group of fuzzy rules containing several lexicons and data sets. The presented fuzzy system combines NLP approaches and Word Sense Disambiguation utilizing a novel unsupervised 9 fuzzy rule based systems for classifying the posts as positive, negative, or neutral sentimental classes.

This article develops a novel web intelligence with enhanced sunflower optimization based deep learning model for sentiment analysis (WIESFO-DLSA) technique. The WIESFO-DLSA technique initially performs pre-processing and word2vec feature extraction processes to generate a meaningful set of features. Besides, bidirectional long short term memory (BiLSTM) model is applied for classification of sentiments into different class labels. In addition, an enhanced sunflower optimization (ESFO) algorithm is exploited to optimally adjust the hyperparameters of the BiLSTM model and the ESFO algorithm is devised by integrating the concepts of SFO algorithm with Hill climbing (HC) concept. An extensive set of experiments were carried out to highlight the betterment of the WISFO-DLSA technique.

The rest of the paper is arranged as follows. Section 2 offers the proposed model, Section 3 provides the experimental analysis, and Section 4 draws the conclusion.

2 The Proposed Model

In this study, a new WIESFO-DLSA technique has been developed to identify the expressions or sentiments that exist in the social networking data. The WIESFO-DLSA technique initially performs pre-processing and word2vec feature extraction processes to generate a meaningful set of features. Followed by, the ESFO with BiLSTM model is applied for classification of sentiments into different class labels where ESFO algorithm is exploited to optimally adjust the hyperparameters of the BiLSTM model. Fig. 2 demonstrates the block diagram of WIESFO-DLSA technique.

2.1 Data Pre-processing

Frequently, the dataset has noise from the procedure of unwanted information which doesn't give near the classifier and requires that cleaned. The data pre-processed is the technique utilized for removing noisy and insufficient data. The pre-processed roles an important play in improving the accuracy [17]. The dataset utilized under this analysis involves a huge count of unwanted data which is not playing some role in the forecast. As training as well as testing time improves once the dataset is superior, thus, eliminating unwanted data is speeds up the trained model as well. The pre-process contains the phases implemented for cleaning the data thus the learning efficacy of this technique is improved. Therefore, the natural language tool kit (NLTK) of Python was employed. It can be set of text processing libraries that are utilized for a variety of process tasks and it can be utilized NLTK 3.5b1 with Python 3. In the primary phase, every review with missing value is recognized and distant as missing data is reduce the efficiency of the classifier. Afterward, a numerical value is eliminated in the text as it does not give to learned classifiers. It decreases the effort of trained classifiers. Sometimes, analyses have special symbols as hear sign, thumb sign, and so on, which require that eliminate for decreasing feature dimensional and higher efficiency. Then, the subsequent punctuation []() /|, ; . ' has been eliminated in the review considering the detail which it doesn't give to text analysis. In the next stage, word is changed to lowercase as the text study has case sensitive. When this stage is not applied, the ML techniques are counted for instance 'Excellent' and 'excellent' as 2 distinct words that is eventually move the classifier efficacy. Finally, stemming was implemented. It can be essential pre-processed stage which eliminates the affixes in the words. It alters the extended word as to their base procedures. i.e., 'loves', 'loved', and 'loving' is the altered procedures of 'love'. The stemming alters these words as to original or root method and uses for increasing the efficiency of classifier.

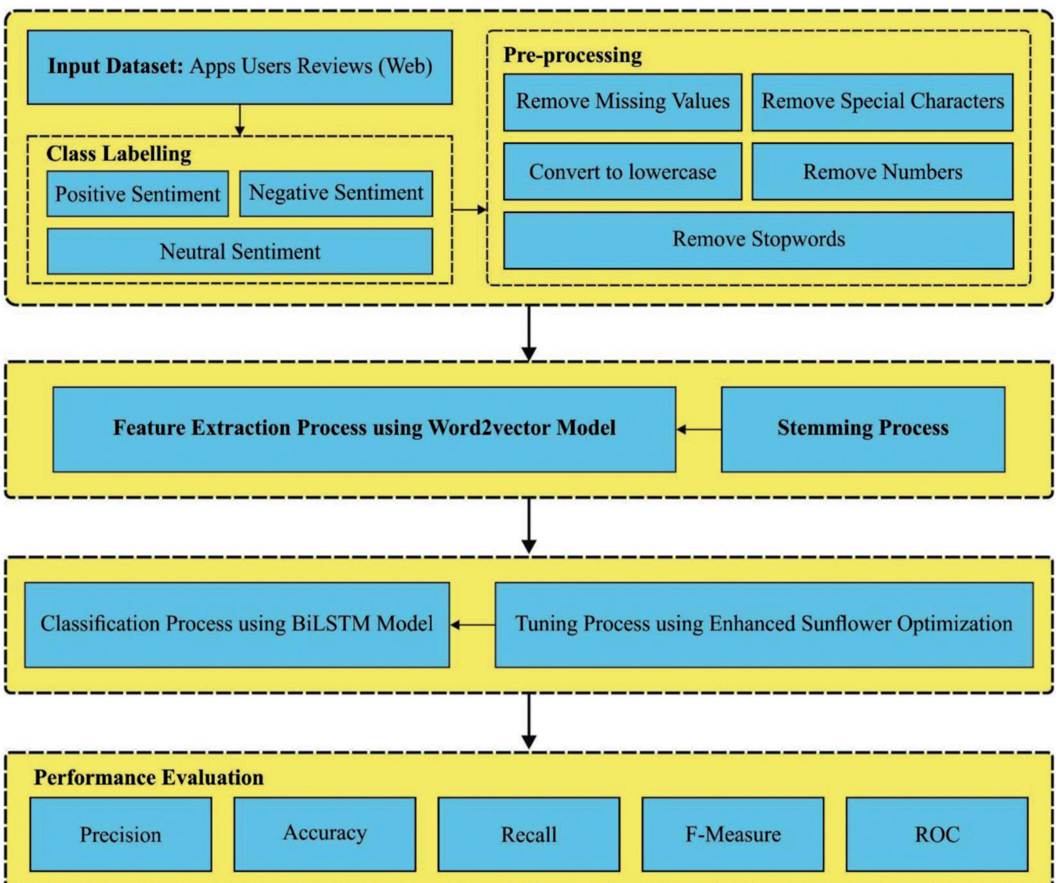

Figure 2: Overall process of WIESFO-DLSA technique

2.2 Word2Vec Model for Feature Extraction

The similarity measure (SM) is a function in which the data of 2 ontology entities are utilized as input and real values among [0, 1] is outputted to characterize the similarity [18]. SM is a significant portion of the ontology matching procedure. Exploiting distinct SMs affect the outcomes of ontology alignment. In the study, we used two classes of SMs to compute the similarity value of two entities, that is., cosine similarity and linguistic-based measures utilizing the Word2Vec models. Word2Vec is a language model of Natural Language Processing (NLP) in which phrases or words are characterized as real value vectors. Usually, similar words are mapped to the same region and have the proximity of vectors [18]. Regarding this, the ontology depiction in vector space, it implies that property or class of ontologies is characterized in dimension of the vector space. Especially, distinct properties or classes are characterized uniquely in the vector space. The vector space covers each class and property in ontology. In the study, the dimension of the vector space is defined by each class and property in the two ontologies. It can be determined by the following equation:

$$(V_{w_1},\ \ V_{w_2}) = \frac{V_{w_1} \cdot V_{w_2}}{V_{w_1} \cdot V_{w_2}} \tag{1}$$

Whereas V_{w_1} and V_{w_2} correspondingly denotes the vector of w_1 and w_2 words as well as V_{w_1} and V_{w_2}, correspondingly, denotes the norms. The linguistic similarity among two words is estimated using semantic relationships (synonyms and antonyms), i.e., usually performed by lists and dictionaries of synonym. WordNet, vocabulary databases build semantic network based semantic data of word, which is utilized for calculating similarity. The linguistic similarity of w_1 and w_2 words are 1 while w_1 and w_2 represents synonym in WordNet; the similarity is 0.5 while w_1 and w_2 are hypernym in WordNet; in another case, the similarity is 0. In the study, it utilizes the maximal approach to incorporate the SM, that is, the large one of two similarity values are chosen as the last similarity values that assist in ensuring the comprehensiveness of the alignments.

2.3 Sentiment Classification Using BiLSTM Model

Once the features are derived, they are passed into the BiLSTM model to classify the sentiments. The BiLSTM method [19] is a discrete version of RNN. LSTM is particularly proposed to avoid long-term dependency problems. It comprises of constant memory cell, input, forget, and output gates. The LSTM based sequential method follows this equation in which i_t, f_t and 0_t represents the input, forget, and output gates, correspondingly:

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \tag{2}$$

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \tag{3}$$

$$0_t = \sigma(W_0[h_{t-1}, xt] + b_0) \tag{4}$$

$$g_t = tanh(W_g[h_{t-1},\ \ x_t] + b_g) \tag{5}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \mathrm{g_t} \tag{6}$$

$$h_t = 0_t \odot \tanh\ (c_t) \tag{7}$$

Now tanh denotes the nonlinear hyperbolic function and $\odot$ characterizes the element-wise multiplication. When the input sequence $(x_1, \ldots, x_t)$, is fed into the LSTM, it calculates the hidden control unit $(h_1, \ldots, h_t)$, the cell memory sequence $(c_1, \ldots, c_t)$, b signifies the bias vector. But, the sequential model of LSTM focuses on existing sequences in the temporal order. In this presented method, we employed *bi*-LSTM as an encoder that considers the existing and future states. It enhanced the typical LSTM network by adding other layers. The two layers are in opposite directions and forecast data from the preceding and upcoming through forward and backward pass.

$$h_i = [\overrightarrow{h_i} \oplus \overleftarrow{h_i}] \tag{8}$$

The decoder comprises the sequential LSTM that takes input from the encoder. Now, $\oplus$ characterizes the element-wise summation to chain the backward and onward pass. The output y_t and hidden layer h_t update based early location y_{t-1} and h_{t-1} based time step t.

$$\begin{bmatrix} y_t \\ h_t \end{bmatrix} = Dec(h_{t-1}, y_{t-1}, \quad V) \tag{9}$$

Then, an attention-based *bi*-LSTM has been proposed as an encoder.

2.4 Design of ESFO Algorithm for Hyperparameter Tuning

For boosting the classification efficiency of the BiLSTM technique, the ESFO algorithm is derived. The sunflower life process is consistent: as the needles of clock, it rises and follows the sun daily. It can be moved alternate way at night for awaiting its disappearance the subsequent morning [20]. The inverse square law radiation is other key nature based optimized. The amount of heat Q obtained by all plants are provided as:

$$Q_i = \frac{P}{4\pi r_i^2} \tag{10}$$

where P implies the source powers and r_i stands for the distance between the current paramount and the plant i. The direction of sunflowers near the sun is as:

$$\vec{S_i} = \frac{X^* - X_i}{\|X^* - X_i\|},\ i = 1,\ 2,\ \cdots,\ n_p \tag{11}$$

The sunflowers stride from the way s has computed as:

$$d_i = \lambda \times P_i(\|(X_i - X_{i-1})\|) \times \|X_i - X_{i-1}\| \tag{12}$$

where λ indicates the perpetual value that determines a. "inertial" dislocation of plants, $P_i(\|X_i - X_{i-1}\|)$ has the probability of pollinations. The maximal stage is determined as:

$$d_{\max} = \frac{\|X_{\max} - X_{\min}\|}{2 \times N_{pop}} \tag{13}$$

where $X_{\max}$ and $X_{\min}$ represents the upper and lower boundary, and N_{pop} stands for the entire amount of plants. The latest plantation as:

$$\vec{X}_{i+1} = \vec{X}_i + d_i \times \vec{s}_i \tag{14}$$

The technique begins with generating the population of people that is even or arbitrary. The respective individual rating assists us in choosing that one is altered as to sun. But, it can be supposed to comprise the capability to function with many suns from a future edition, it can be now restricted to separate under this study. Afterward, similar to the sunflowers, every further entity places itself near the sun and move from an arbitrary controlled method. The paramount plant is pollinated nearby the sun.

The presented ESFO approach is named local search, which is a basic version of local improvement method. It starts with one arbitrary solution (x), proceed repeatedly by shifting in the existing solution to optimal neighboring solution until it reaches a local optimal (that is., the local optimum solution doesn't have optimal neighboring solution, no development in FF). It takes downhill development in which the FF of neighboring solution must be superior to the existing solution [21]. Subsequently, it converges to the local optimal suddenly and fasts. But it rapidly gets stuck in local optimal that in most situations isn't acceptable. Afterward producing the early solution x and the iterative development method, a collection of neighboring solutions is produced by employing the process *Improve* ($N(x)$). Algorithm 1 presents the pseudocode of the HC approach. This process search for discovering the better neighboring solution from the collection of neighbors using satisfactory rules like random walk, first improvement, best improvement, and sidewalk. However, all the above rules are terminated in local optimal.

Algorithm 1: Hill climbing approach

The first solution x

$x_i = LB_i + (UB_i - LB) * \mathrm{U(O,I)}, \forall_i = (1,\ 2,\ \ N)$

Evaluate fitness function $F(x)$

while (termination criteria isn't fulfilled) do

$$x' = Improve((N(x))$$

if $F(x') \leq F(x)$ then

$x = x'$

end if

end while

return x

3 Results and Discussion

In this section, the performance validation of the WIESFO-DLSA technique takes place using the dataset, comprising user reviews from the Google apps in the English language. It is available at https://www.kaggle.com/lava18/google-playstore-apps. It holds 64,295 instances with 23,998 positive reviews, 8271 negative and 5158 neutral reviews in the dataset.

Fig. 3 demonstrates three confusion matrices produced by the WIESFO-DLSA technique under distinct epochs. With 500 epochs, the WIESFO-DLSA technique has identified 23877 instances into positive, 8072 instances into negative, and 4883 instances into neutral class. Similarly, with 1500 epochs, the WIESFO-DLSA technique has identified 23875 instances into positive, 8111 instances into negative, and 4947 instances into neutral class.

Table 1 provides a comprehensive classification result analysis of the WIESFO-DLSA technique under distinct class labels and epochs. The results notified that the WIESFO-DLSA technique has the ability to classify all the instances effectively.

Overall classifier results of the WIESFO-DLSA technique under distinct epochs are offered in Table 2 and Fig. 4. The experimental values denoted the betterment of the WIESFO-DLSA technique under all epochs. For instance, with 500 epochs, the WIESFO-DLSA technique has resulted to $prec_n$, $reca_l$, $accu_y$, F_{score}, and MCC of 98.30%, 97.25%, 98.94%, 97.76%, and 96.80% respectively. At the same time, with 1000 epochs, the WIESFO-DLSA technique has accomplished $prec_n$, $reca_l$, $accu_y$, F_{score}, and MCC of 98.30%, 97.50%, 98.97%, 97.89%, and 96.95% respectively. Furthermore, with 1500 epochs, the WIESFO-DLSA technique has led to $prec_n$, $reca_l$, $accu_y$, F_{score}, and MCC of 98.51%, 97.82%, 99.12%, 98.16%, and 97.36% respectively.

Fig. 5 demonstrates the ROC analysis of the WIESFO-DLSA technique under epoch of 500. The figure exposed that the WIESFO-DLSA technique has reached enhanced outcome with the minimum ROC of 99.7821.

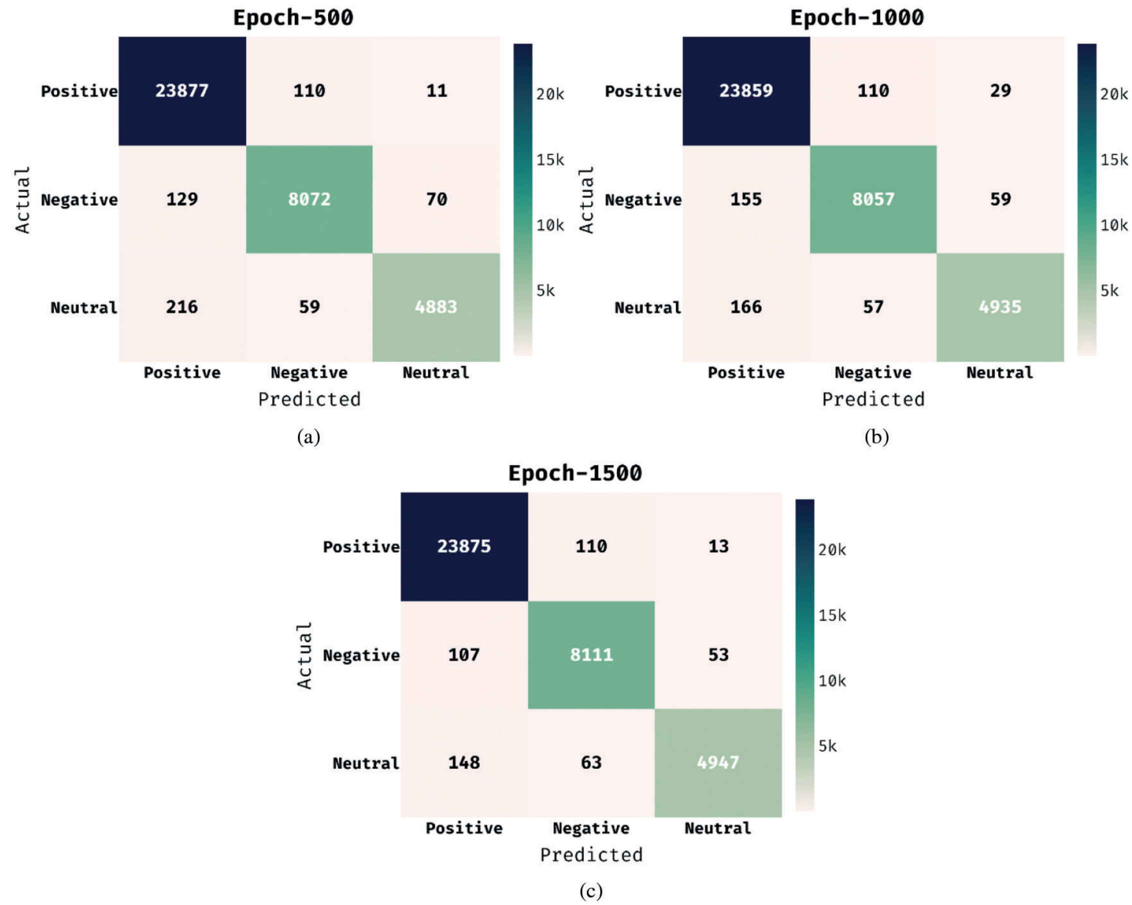

Figure 3: Confusion matrix analysis of WIESFO-DLSA technique

Table 1: Result analysis of WIESFO-DLSA technique with respect to different measures

Methods	Precision	Recall	Accuracy	F-Score	MCC
Epoch-500					
Positive	98.58	99.50	98.75	99.03	97.29
Negative	97.95	97.59	99.02	97.77	97.14
Neutral	98.37	94.67	99.05	96.48	95.96
Average	98.30	97.25	98.94	97.76	96.80
Epoch-1000					
Positive	98.67	99.42	98.77	99.05	97.33
Negative	97.97	97.41	98.98	97.69	97.04
Neutral	98.25	95.68	99.17	96.95	96.48
Average	98.30	97.50	98.97	97.90	96.95
Epoch-1500					
Positive	98.94	99.49	98.99	99.21	97.80
Negative	97.91	98.07	99.11	97.99	97.42
Neutral	98.68	95.91	99.26	97.28	96.86
Average	98.51	97.82	99.12	98.16	97.36

Table 2: Overall classifier analysis of WIESFO-DLSA technique with varying measures

No. of epochs	Precision	Recall	Accuracy	F-score	MCC
Epoch-500	98.30	97.25	98.94	97.76	96.80
Epoch-1000	98.30	97.50	98.97	97.89	96.95
Epoch-1500	98.51	97.82	99.12	98.16	97.36

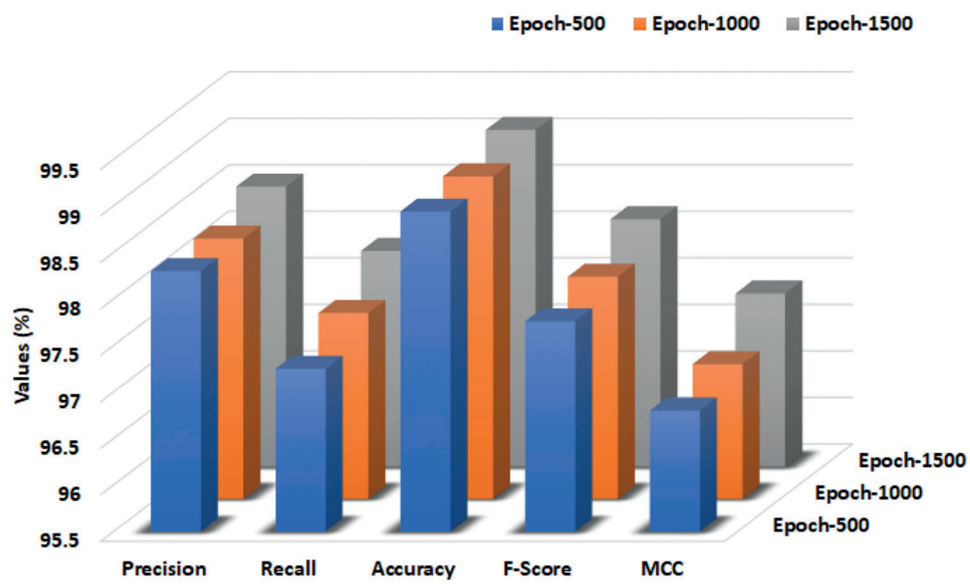

Figure 4: Overall analysis of WIESFO-DLSA technique

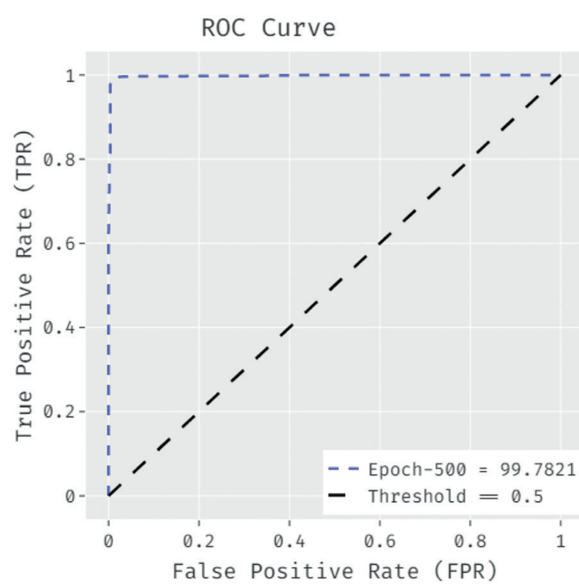

Figure 5: ROC analysis of WIESFO-DLSA technique under epoch of 500

Fig. 6 depicts the ROC analysis of the WIESFO-DLSA approach under epoch of 1000. The figure demonstrated that the WIESFO-DLSA algorithm has attained increased results with the minimum ROC of 99.6295.

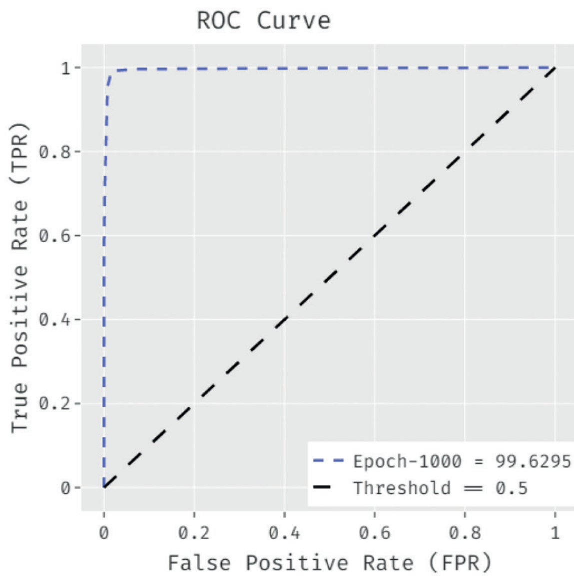

Figure 6: ROC analysis of WIESFO-DLSA technique under epoch of 1000

Fig. 7 illustrates the ROC analysis of the WIESFO-DLSA approach under epoch of 1500. The figure revealed that the WIESFO-DLSA method has gained improved outcomes with the minimal ROC of 99.8517.

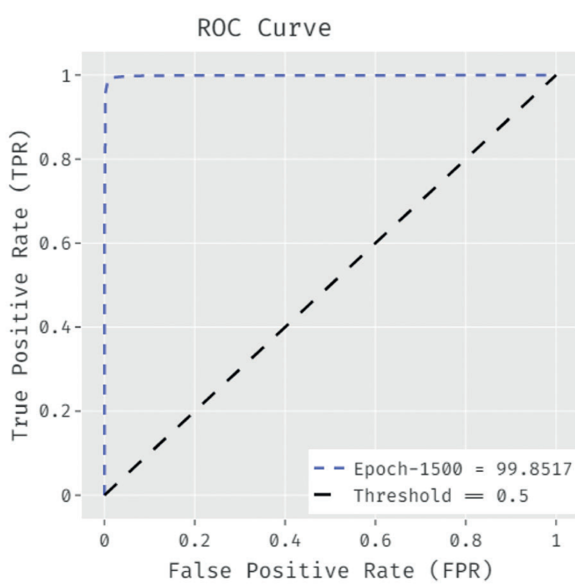

Figure 7: ROC analysis of WIESFO-DLSA technique under epoch of 1500

Table 3 and Fig. 8 illustrate the comparative result analysis of the WIESFO-DLSA technique with recent methods on the classification of positive instances. The experimental results stated that the LOGR-CM technique has exhibited least outcome with the minimal values of $prec_n$, $reca_l$, and F_{score}. In addition, the GBM-CM and SVM-CM techniques have resulted in slightly enhanced values of $prec_n$, $reca_l$, and F_{score}. In line with, the GBSVM-CM and RANDFO-CM techniques have accomplished reasonable values of $prec_n$, $reca_l$, and F_{score}. However, the WIESFO-DLSA technique has obtained effective results with the $prec_n$, $reca_l$, and F_{score} of 98.94%, 99.49%, and 99.21% respectively.

Table 3: Comparison study of WIESFO-DLSA technique under positive class instances

Positive class			
Methods	Precision	Recall	F-score
GBSVM-CM	76.00	92.00	83.00
GBM-CM	71.00	94.00	81.00
SVM-CM	74.00	97.00	84.00
LOGR-CM	67.00	99.00	80.00
RANDFO-CM	75.00	93.00	83.00
WIESFO-DLSA	98.94	99.49	99.21

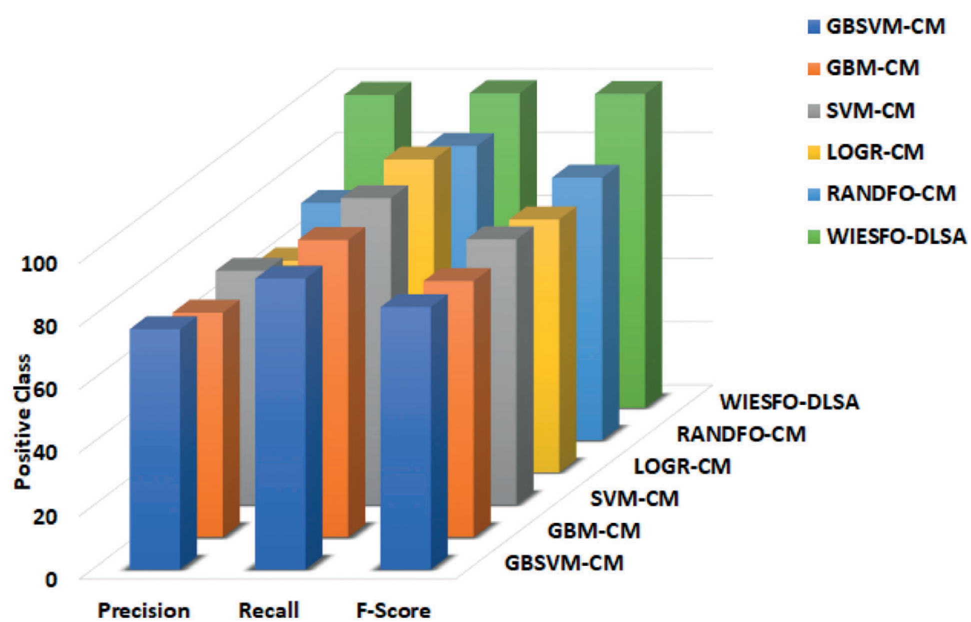

Figure 8: Comparative analysis of WIESFO-DLSA technique under positive class

Table 4 and Fig. 9 depict the comparative outcome analysis of the WIESFO-DLSA approach with existing approaches on the classification of negative instances. The experimental outcomes revealed that the LOGR-CM technique has exhibited least outcome with the minimal values of $prec_n$, $reca_l$, and F_{score}. Also, the GBM-CM and SVM-CM methods have resulted in somewhat increased values of $prec_n$, $reca_l$, and F_{score}. Followed by, the GBSVM-CM and RANDFO-CM techniques have accomplished reasonable values of $prec_n$, $reca_l$, and F_{score}. At last, the WIESFO-DLSA approach has obtained effective outcomes with the $prec_n$, $reca_l$, and F_{score} of 97.91%, 98.07%, and 97.99% correspondingly.

Table 4: Comparison study of WIESFO-DLSA technique under negative class instances

Negative class			
Methods	Precision	Recall	F-Score
GBSVM-CM	75.00	56.00	64.00
GBM-CM	76.00	36.00	49.00
SVM-CM	83.00	49.00	61.00
LOGR-CM	88.00	21.00	34.00
RANDFO-CM	78.00	49.00	60.00
WIESFO-DLSA	97.91	98.07	97.99

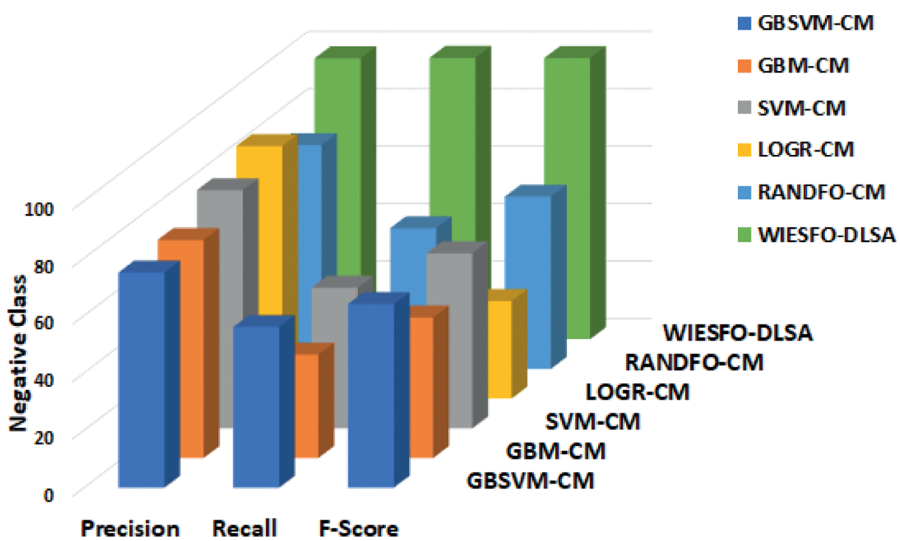

Figure 9: Comparative analysis of WIESFO-DLSA technique under negative class

Table 5 and Fig. 10 demonstrates the comparative result analysis of the WIESFO-DLSA technique with recent methods on the classification of neutral instances. The experimental outcomes revealed that the LOGR-CM technique has exhibited worse outcome with the minimal values of $prec_n$, $reca_l$, and F_{score}. Furthermore, the GBM-CM and SVM-CM approaches have resulted in slightly enhanced values of $prec_n$, $reca_l$, and F_{score}. Similarly, the GBSVM-CM and RANDFO-CM algorithms have accomplished reasonable values of $prec_n$, $reca_l$, and F_{score}. Finally, the WIESFO-DLSA methodology has achieved effectual outcomes with the $prec_n$, $reca_l$, and F_{score} of 98.68%, 95.91%, and 97.28% correspondingly.

Finally, a comparative $accu_y$ analysis of the WIESFO-DLSA technique with recent approaches [17] are offered in Table 6 and Fig. 11. The results notified that the LOGR-CM and GBM-CM techniques have obtained lower $accu_y$ values of 68% and 70% respectively. Also, the RANDFO-CM, GBSVM-CM, and SVM-CM techniques have resulted in moderately closer $accu_y$ values of 73%, 75%, and 75% respectively. However, the WIESFO-DLSA technique has accomplished improved outcomes with the maximum $accu_y$ of 99.12%. By looking into the abovementioned tables and figures, it can be ensured that the WIESFO-DLSA technique has appeared as an effective tool for SA.

Table 5: Comparison study of WIESFO-DLSA technique under neutral class instances

Neutral class			
Methods	Precision	Recall	F-Score
GBSVM-CM	60.00	26.00	36.00
GBM-CM	46.00	18.00	26.00
SVM-CM	90.00	20.00	32.00
LOGR-CM	92.00	30.00	60.00
RANDFO-CM	55.00	26.00	35.00
WIESFO-DLSA	98.68	95.91	97.28

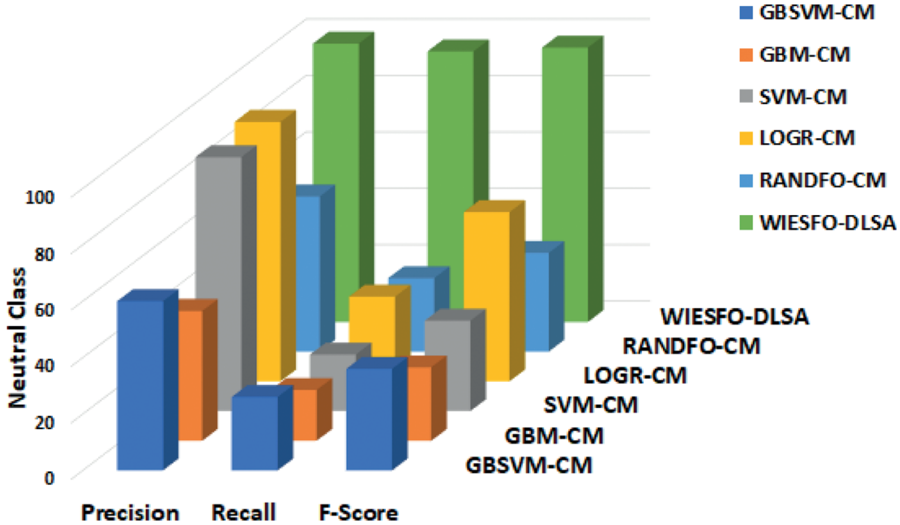

Figure 10: Comparative analysis of WIESFO-DLSA technique under neutral class

Table 6: Comparative $\boldsymbol{accu_y}$ analysis of WIESFO-DLSA technique

Methods	Accuracy (%)
GBSVM-CM	75.00
GBM-CM	70.00
SVM-CM	75.00
LOGR-CM	68.00
RANDFO-CM	73.00
WIESFO-DLSA	99.12

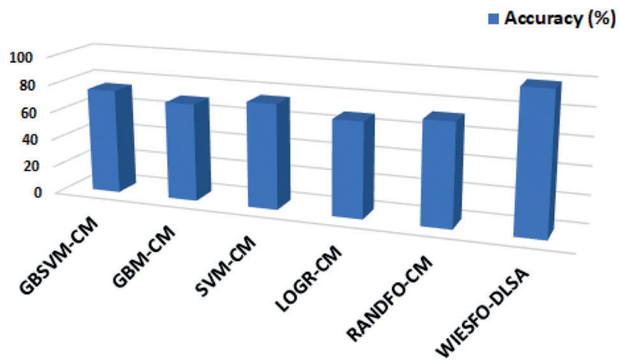

Figure 11: Accuracy analysis of WIESFO-DLSA technique with recent algorithms

4 Conclusion

In this study, a new WIESFO-DLSA technique has been developed for identifying the expressions or sentiments that exist in the social networking data. The WIESFO-DLSA technique initially performs pre-processing and word2vec feature extraction processes to generate a meaningful set of features. Followed by, the ESFO with BiLSTM model is applied for classification of sentiments into different class labels where ESFO technique was exploited to optimally adjust the hyperparameters of the BiLSTM model. Finally, the ESFO algorithm is devised by integrating the concepts of SFO algorithm with HC. An extensive set of experiments were carried out to highlight the betterment of the WISFO-DLSA technique. The experimental outcomes ensured its promising performance over the recent state of art approaches under several measures. Therefore, the WISFO-DLSA technique can be utilized as an effective tool to analyze sentiments. In the future, it can be deployed in smartphone applications to enable real time data processing.

Funding Statement: Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2023R51), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

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

References

[1] D. M. E. D. M. Hussein, "A survey on sentiment analysis challenges," *Journal of King Saud University-Engineering Sciences*, vol. 30, no. 4, pp. 330–338, 2018.

[2] M. Luca and G. Zervas, "Fake it till you make it: Reputation, competition, and yelp review fraud," *Management Science*, vol. 62, no. 12, pp. 3412–3427, 2016.

[3] F. Rustam, I. Ashraf, A. Mehmood, S. Ullah and G. Choi, "Tweets classification on the base of sentiments for us airline companies," *Entropy*, vol. 21, no. 11, pp. 1078, 2019.

[4] A. Ortigosa, J. M. Martín and R. M. Carro, "Sentiment analysis in Facebook and its application to e-learning," *Computers in Human Behavior*, vol. 31, pp. 527–541, 2014.

[5] R. Duwairi and M. El-Orfali, "A study of the effects of preprocessing strategies on sentiment analysis for Arabic text," *Journal of Information Science*, vol. 40, no. 4, pp. 501–513, 2014.

[6] V. Kalra and R. Aggarwal, "Importance of text data preprocessing & implementation in rapidminer," in *the Proc. of the First Int. Conf. on Information Technology and Knowledge Management*, pp. 71–75, 2018.

[7] D. Agnihotri, K. Verma, P. Tripathi and B. K. Singh, "Soft voting technique to improve the performance of global filter based feature selection in text corpus," *Applied Intelligence*, vol. 49, no. 4, pp. 1597–1619, 2019.

[8] S. Yang and H. Zhang, "Text mining of twitter data using a latent dirichlet allocation topic model and sentiment analysis," *World Academy of Science, Engineering and Technology International Journal of Computer and Information Engineering*, vol. 12, no. 7, pp. 525–529, 2018.
[9] R. Zhao and K. Mao, "Fuzzy bag-of-words model for document representation," *IEEE Transactions on Fuzzy Systems*, vol. 26, no. 2, pp. 794–804, 2018.
[10] Ankit and N. Saleena, "An ensemble classification system for twitter sentiment analysis," *Procedia Computer Science*, vol. 132, pp. 937–946, 2018.
[11] S. Yoo, J. Song and O. Jeong, "Social media contents based sentiment analysis and prediction system," *Expert Systems with Applications*, vol. 105, pp. 102–111, 2018.
[12] O. Araque, I. C. Platas, J. F. S. Rada and C. A. Iglesias, "Enhancing deep learning sentiment analysis with ensemble techniques in social applications," *Expert Systems with Applications*, vol. 77, pp. 236–246, 2017.
[13] M. H. A. E. Jawad, R. Hodhod and Y. M. K. Omar, "Sentiment analysis of social media networks using machine learning," in *2018 14th Int. Computer Engineering Conf. (ICENCO)*, Cairo, Egypt, pp. 174–176, 2018.
[14] L. C. Chen, C. M. Lee and M. Y. Chen, "Exploration of social media for sentiment analysis using deep learning," *Soft Computing*, vol. 24, no. 11, pp. 8187–8197, 2020.
[15] F. Valencia, A. G. Espinosa and B. V. Aguirre, "Price movement prediction of cryptocurrencies using sentiment analysis and machine learning," *Entropy*, vol. 21, no. 6, pp. 589, 2019.
[16] S. Vashishtha and S. Susan, "Fuzzy rule based unsupervised sentiment analysis from social media posts," *Expert Systems with Applications*, vol. 138, pp. 112834, 2019.
[17] M. Khalid, I. Ashraf, A. Mehmood, S. Ullah, M. Ahmad *et al.,* "GBSVM: Sentiment classification from unstructured reviews using ensemble classifier," *Applied Sciences*, vol. 10, no. 8, pp. 2788, 2020.
[18] X. Xue, H. Wang, J. Zhang, Y. Huang, M. Li *et al.,* "Matching transportation ontologies with word2vec and alignment extraction algorithm," *Journal of Advanced Transportation*, vol. 2021, pp. 1–9, 2021.
[19] T. Chen, R. Xu, Y. He and X. Wang, "Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN," *Expert Systems with Applications*, vol. 72, pp. 221–230, 2017.
[20] G. F. Gomes, S. S. da Cunha and A. C. Ancelotti, "A sunflower optimization (SFO) algorithm applied to damage identification on laminated composite plates," *Engineering with Computers*, vol. 35, no. 2, pp. 619–626, 2019.
[21] M. Shehab, H. Alshawabkah, L. Abualigah and N. AL-Madi, "Enhanced a hybrid moth-flame optimization algorithm using new selection schemes," *Engineering with Computers*, vol. 37, no. 4, pp. 2931–2956, 2021.