Journal of Theoretical and Applied Information Technology
31st January 2022. Vol.100. No 2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

A NOVEL SENTIMENT CLASSIFICATION MODEL USING GRASSHOPPER OPTIMIZATION ALGORITHM WITH BIDIRECTIONAL LONG SHORT TERM MEMORY

[1]D. ELANGOVAN, [2]V. SUBEDHA

[1]Research Scholar, Department of Computer Science and Engineering, Sathyabama Institute of Science and Technology.

[2]Professor & Head, Department of Computer Science and Engineering, Panimalar Institute of Technology, Chennai-600123.

Email: [1]elangovan.durai@yahoo.com, [2]subedha@gmail.com

ABSTRACT

In recent times, social media has received great attention among the research communities towards the domain of sentiment analysis (SA). The proficient design of SA is needed to improve the service and product qualities for the marketing and financial schemes for increasing the company's profit and user satisfaction. Although several SA techniques are available in the literature, it is needed to further enhance the classification results of the user review which helps to comprehend the user reviews, thereby quality of the products can be improved. This study devises an effective SA and classification technique using grasshopper optimization algorithm (GOA) with bidirectional long short term memory (Bi-LSTM), named GOA-BiLSTM. The GOA-BiLSTM model involves word2vec based feature extraction process to derive a useful set of features. In addition, Bi-LSTM based classifier is applied to determine the optimal class label of the extracted features. Moreover, GOA is utilized for the hyperparameter optimization of the Bi-LSTM model. To ensure the better outcome of the GOA-BiLSTM model, an extensive set of simulations were carried out on four datasets. The simulation outcome verified the superiority of the GOA-BiLSTM model by accomplishing a higher accuracy of 99.57%, 99. 71%, 99. 06%, and 98.98% on the applied Canon, Nokia DVD, and iPod dataset respectively.

Keywords: *Sentiment analysis, Classification, Deep learning, Parameter optimization, Bi-LSTM*

1. INTRODUCTION

Presently, the exponential development of large number of distinct kinds of data such as image, video, audio and documents have resulted in the generation of Big Data. When it is compared with other data resources, textual data act as a vital role to perform analysis. Particularly, text classification process is found to be interesting where the text is allocated to the categorical labels such as sentiments and language classes. For example, the user in the social networking sites expresses the emotion and recommendation based on regular news updates with their friends and public. The identification of emotions like joy, sad, anger and surprise takes place by using the sentiment analysis (SA) model. For instance, product feedbacks are available on e-commerce websites regarding the quality of the purchased products. If this information is identified accurately, it is applicable for developing distinct industrial application areas such as movie recommendations and customized news feed. In addition, worldwide movie advertisements are also developed continuously through online media like Netflix and Hotstar.

Several studies have been presented based on machine learning (ML) models to classify the textual data. Although the ML models are applied obviously with better competence, they are mainly based on handcrafted features where more efforts are needed by the feature definition requirements. On the other hand, Deep Learning (DL) models are also found to be popular owing to the simple nature and low complexity in achieving optimal outcomes. At this point, the sentiments are considered as emotions that can be expressed in several scenarios. The class labels of the sentiments can be defined by the polarities (positive, neutral, and negative) or numerous class labels of emotions (angry, happy, sad, and pleased). The various works have defined a greater number of sentiment labels such as opinion rating value and emotional feeling [1], and limited models make use of 2-dimension classes. Although the outcome can be improved in SA, the binary classification of sentiments is considered to be the crucial problem.

In [2], a new SA is developed by the use of unigram features in addition to support vector machine (SVM). In [3], a novel set of unigram and bigram features are employed for the classification of sentiments related to movie reviews. In [4], binary classification model is developed by the use of unigram and bigram features and is ensured with respect to accuracy on Amazon product reviews. In [5], an effective SA model is developed by the use of SVM with bigram features for the classification of five types of sentiments. In [6], a new SVM with semantic analytics model is developed to classify the sentiments from Twitter data. Besides, in [7], an SVM with n-gram features model is presented to accomplish reasonable performance in categorizing the sentiments. In [8], a new method for naïve Bayes (NB) with unigram features are developed to compute sentiments for Urdu tweets with effective outcome. In addition, the manual set of features are used to classify the sentiments. In [9], an affective lexicon, mis-spelling and emoticons are utilized as the main features to classify the data using SVM. In addition, [10] has explained a set of 3 values namely positivity, negativity and objectivity as characteristics which carried out the binary classification process using Logistic Regression (LR). [11] handled the classification process using SVM for Twitter data in which a set of two types of target-independent features namely twitter content and sentiment lexicon features. [12] has applied a set of positive and negative words as features and accomplish binary classification on Twitter data. In [13], a generalized sequence of words is considered as features for the classification process using SVM model.

[14] developed a 5-class SA and classification model based on the domain-free feature set for Twitter data. [15] presented an experimental result of Named Entity (NE) based feature extractor. It is reported that the integration of handcrafted features with n-gram features. Although ML models are found to be proficient in handcrafted and n-gram features, the works are restricted with respect to the definition of features which needs expert knowledge to gain effective outcome. In addition, these limitations are resolved in the data fusion techniques of SA integrating distinct sources namely ontology and lexicons owing to the fact of high cost and time. To resolve these issues, DL models find useful to offer effective outcome owing to the standing of taking arbitrary patterns regularly. Similarly, in [16], a DL model based SA tool is developed and meta-level feature representation is presented for application generalization.

This study develops an effective SA and classification technique using grasshopper optimization algorithm (GOA) with bidirectional long short term memory (Bi-LSTM), named GOA-BiLSTM. The GOA-BiLSTM model involves word2vec based feature extraction process to derive a useful set of features. Furthermore, Bi-LSTM based classifier is utilized for determining the optimum class labels of the extracted features. Besides, the GOA is exploited for the hyperparameter optimization of the Bi-LSTM model. To guarantee the improved outcome of the GOA-BiLSTM model, a widespread set of simulations were carried out on four datasets.

2. THE PROPOSED GOA-BILSTM MODEL

Fig. 1 illustrates the working principle of the GOA-BiLSTM model. The figure demonstrates that the input reviews are initially preprocessed to remove the unwanted details. This is followed by word2vec based feature extraction using GOA-BiLSTM based classification processes.

2.1. Data Pre-Processing

The preprocessing is employed to discard the noise that exists in the input data which is applied for enhancing the classifier performance. The data involved is comprised of a large amount of irregular data that restricts the overall performance. The mathematical values are discarded in the first stage and then punctuation marks are deleted from the input data. At last, the stemming process is determined by eliminating the affixes that exist in the words.

2.2. Feature Extraction

Word2vec is mainly employed for feature extraction process where it receives the input as corpus and outcome as a set of vectors. The Word2vec predicts the words depending upon the context with two distinct neural techniques namely Continuous Bag of Words (CBOW) and Skip-Gram. Since the CBOW technique determines the present work using the context, the skip-gram model predicts the rest of the words using the present word. The CBOW technique will compare the word with output for correcting the word representation with respect to the backpropagation (BP) of the error gradient. Actually, the CBOW has tried for the maximization of Eq. (1):

$$\frac{1}{V}\sum_{t=1}^{v}\log p\left(m_t \middle| m_{t-\frac{c}{2}}\dots m_{t+\frac{c}{2}}\right) \quad (1)$$

At the same time, Skip-Gram model searches the predictability of the context provided a word and tried maximization of Eq. (2):

$$\frac{1}{V}\sum_{t=1}^{V}\sum_{j=t-c,j\neq t}^{t+c} logp(m_j|m_t) \quad (2)$$

When the feature vectors for all the words are generated, the resemblance among the words is determined by the use of cosine similarity. Consider $a(x_1, y_1)$ and $b(x_2, y_2)$ which are the two points provided in 2-dimensional space, the cosine similarity among the two points are defined below (3):

$$\cos\theta = \cos(a,b) = \frac{a \cdot b}{\|a\|\|b\|} = \frac{x_1x_2 + y_1y_2}{\sqrt{x_1^2 + x_2^2} \times \sqrt{y_1^2 + y_2^2}} \quad (3)$$

Simultaneously, once the size gets increased, the vector a and b are represented as $a(a_1, a_2, a_3, .. a_n)$ and $b(b_1, b_2, b_3, \dots b_n)$. The above equation can be rewritten as follows:

$$\cos\theta = \cos(a,b) = \frac{\sum_1^n (a_i \times b_i)}{\sqrt{\sum_1^n a_i^2} \times \sqrt{\sum_1^n b_i^2}} \quad (4)$$

From(4), cosθ is in the range [0,1] and 0 implies no semantic relation among 2 words, 1 refers that the word has a similar meaning [17].

Additionally, cosine similarity is used. The Euclidean, Manhattan, Minkowski and Chebyshev vector distance dimensions are used for measuring results. In addition to a and b multi-dimensional vectors, d is the distance among 2 vectors. Euclidean distance is expressed as (5):

$$d(a,b) = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \cdots + (a_n - b_n)^2} \quad (5)$$

In this case, the formula is expressed as (6):

$$d(a,b) = \sqrt{\sum_{i=1}^{n}(a_i - b_i)^2} \quad (6$$

The Manhattan distance follows a grid-like path among 2 points and is written as (7):

$$d(a,b) = \sum_{i=1}^{n}|a_i - b_i| \quad (7)$$

Minkowski distance is assumed as to generalized version of Manhattan and Euclidean distances and p is order of among 2 points, thus equation is written as (8):

$$d(a,b) = \left(\sum_{i=1}^{n}|a_i - b_i|^p\right)^{1/p} \quad (8)$$

At last, the Chebyshev distance called chessboard distance is written as follows:

$$d(a,b) = \max_i(|a_i - b_i|) \quad (9)$$

2.3. Bi-LSTM based Classification

The LSTM is reliable on the classical Recurrent Neural Network (RNN) model. However, it exploits distinct models for calculating the hidden states to resolve the issue of Recurrent Artificial Neural Network (ANN) model which could not handle the long distance dependencies. But the good abilities of LSTM are not learnt by the use of techniques. However, the intrinsic benefits are offered to the architecture of the model. In addition, the LSTM holds a sequence of iterative memory units, where each unit is comprised of three gates with distinct functions. By using text feature vector S as input, and the t^{th} word as sample, the corresponding stage value of the LSTM of the t^{th} word can be defined below. Fig. 2 illustrates the structure of LSTM and BiLSTM methods. The particular computation follows is defined in Eq. (10) where the σ and ⊙ indicates the sigmoid function and dot multiplication respectively.

The forget gate f_t can be defined by:

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

$$f_t = \sigma\left(W_f w_t + U_f h_{t-1} + b_f\right) \quad (10)$$

Then, the input gate i_t can be equated as:

$$i_t = \sigma(W_i w_t + U_i h_{t-1} + b_i) \quad (11)$$

The $\widetilde{c_t}$ denotes the candidate memory cell position at the present timestep, where tanh is the tangent hyperbolic function;

$$\widetilde{c_t} = tanh\,(W_c w_t + U_i h_{t-1} + b_c) \quad (12)$$

The c_t defines the state values of the present time in the memory cell, the values of f_t and i_t ranges between [0, 1]. The computation of $i_t \odot \hat{C}_t$ representing that the new data is saved in c_t from the candidate unit $\widetilde{c_t}$. The design of $f_t \odot c_{t-1}$ demonstrating that the information is engaged and eliminated in the earlier memory cell c_{t-1}.

$$c_t = i_t \odot \hat{c}_t + f_t \odot c_{t-1} \quad (13)$$

Then, the output gate o_t can be represented as follows:

$$o_t = \sigma(W_o w_t + U_o h_{t-1} + b_o) \quad (14)$$

h_t is the hidden layer state at time t:

$$h_t = o_t \odot tanh(c_t) \quad (15)$$

The LSTM assumes the previous data of the sequence which it is inadequate. When the upcoming data is accessible, then it is advantageous for the series of processes. The Bi-LSTM model is comprised of forward and backward LSTMs whose basic principle is given. The forward layer will capture the past data of the sequence and the backward layer holds the upcoming details of the sequence [18]. Each and every layer is linked to the identical output layer. A major benefit of this model is that the sequence context information is completely taken. Assume the input of time t is the word embedding w_t, at time $t-1$, the outcome of the forward and backward hidden units are $\vec{h}_{t-1}$ and $\overleftarrow{h}_{t+1}$. Afterward, the outcome of the backward and hidden units at time t can be defined by

$$\vec{h}_t = L\left(w_t, \vec{h}_{t-1}, c_{t-1}\right) \quad (16)$$

$$\overleftarrow{h}_t = L\left(w_t, \overleftarrow{h}_{t+1}, c_{t+1}\right) \quad (17)$$

where $L(\cdot)$ signifies the hidden layer process of the LSTM hidden layer. The forward and backward outcome vectors are $\vec{h}_t \in R^{1\times H}$ and $\overleftarrow{h}_t \in R^{1\times H}$ respectively, which needs to be integrated for obtaining the text features. It needs to be defined that the H is the hidden layer cell count:

$$H_t = \vec{h}_t || \overleftarrow{h}_t \quad (18)$$

GOA is a novel metaheuristic algorithm that is inspired by the large swarm of every creature. The grasshoppers are herbivores which affect the crop productivity. The swarming nature of the grasshopper is based on the nymph and adult. The nymph moves by rolling on the ground and feeds on succulent and soft plant. An adult grasshopper jumps higher in the food searching process and has large exploration region. Consequently, slow and fast movements are noticed representing exploration and exploitation. The swarming nature of the grasshopper is defined by

$$X_i = S_i + G_i + A_i, \quad (19)$$

Journal of Theoretical and Applied Information Technology
31st January 2022. Vol.100. No 2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

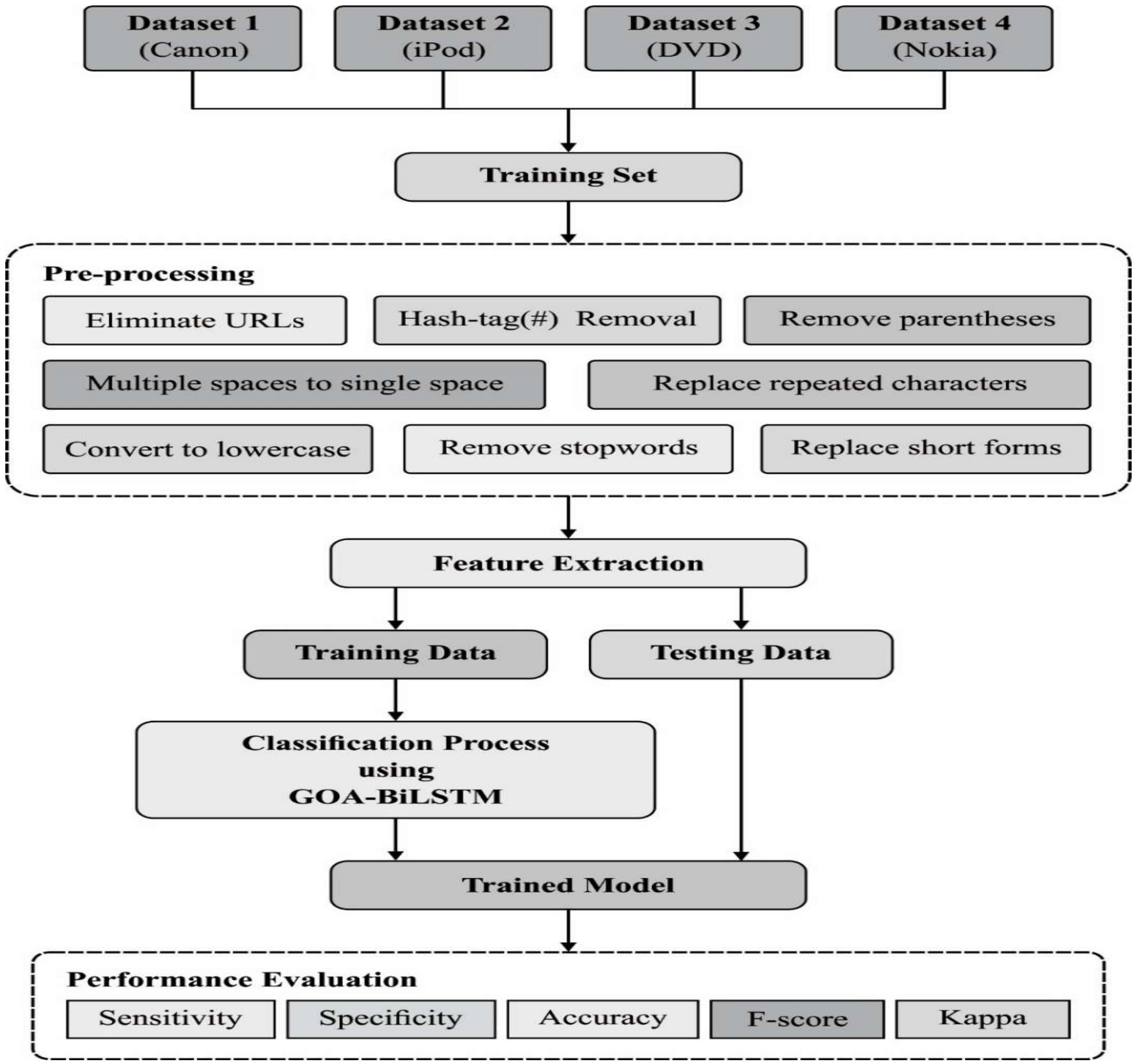

Figure 1: Overall Working Process of GOA-BiLSTM Model

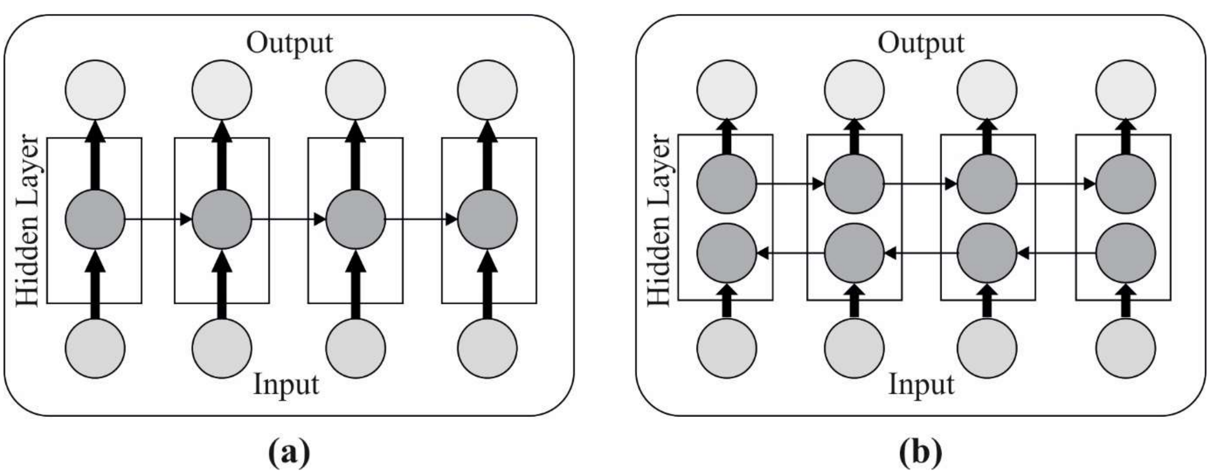

Figure 2: Structure of a) LSTM b) BiLSTM

2.4. Parameter Optimization

In order to effectually elect the hyperparameters of the Bi-LSTM, GOA is applied to improve the overall outcome.

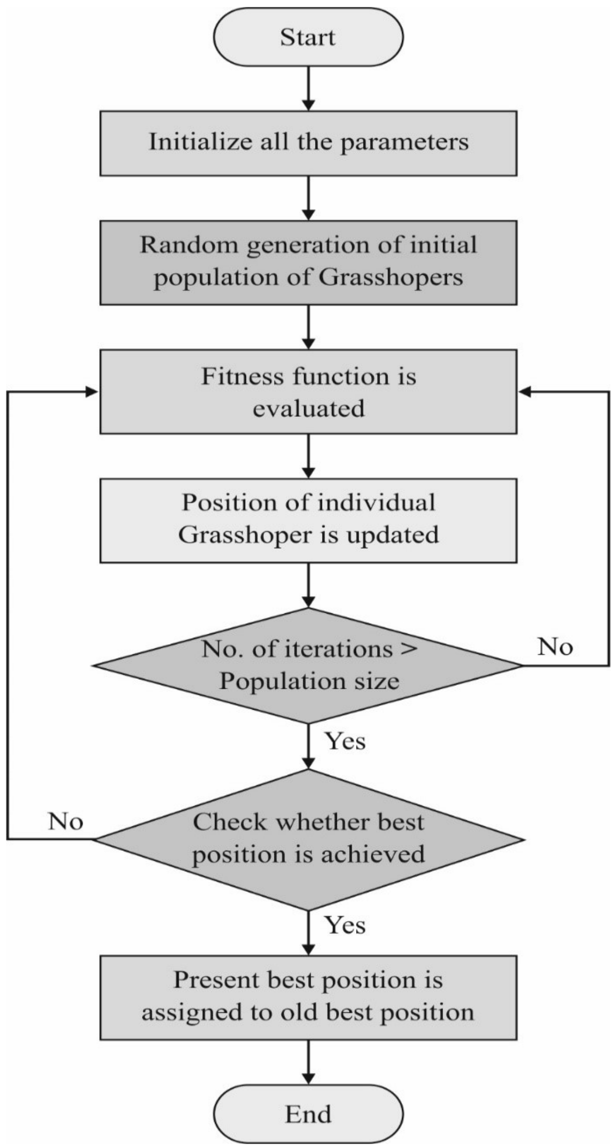

Figure 3: Flowchart of GOA Algorithm

where X_i indicates the location of the ith grasshopper, S_i is the social interaction, G_i is the gravity force in the ith grasshopper, and A_i is the wind advection. Then, the social interaction S_i can be defined by

$$S_i = \sum_{j=1, j\neq i}^{N} s\left(d_{ij}\right)\hat{d}_{ij}, \quad (20)$$

where $d_{ij} = |x_j - x_i|$ is the distance among the ith and jth grasshoppers and $\hat{d}_{ij} = \left(x_j - x_i\right)/\left(d_{ij}\right)$ is the unit vector from the ith to the jth grasshoppers. The function s denotes the social force which is defined by

$$s(r) = f e^{(-r/l)} - e^{-r}, \quad (21)$$

where f is the intensity of attraction and l is the attractive length scale. During the food searching process, the grasshopper generates 3 distinct kinds of regions with respect to social interaction called comfort, repulsive and attractive regions [19]. If the distance is more among the grasshoppers, then the function "s" can not employ robust force. This problem can be resolved by defining the G element as follows.

$$G_i = -g\hat{e}_g, \quad (22)$$

where g is the gravitational constant and $\hat{e}_g$ represents a unity vector. Besides, the A element can be computed using Eq. (23):

$$A_i = u\hat{e}_w, \quad (23)$$

where u is the constant drift and $\hat{e}_w$ is a unity vector in the way of wind. The substitution of , G, and A in Eq. (19), it is obtained as

$$X_i = \sum_{j=1, j\neq i}^{N} s\left(|x_j - x_i|\right)\frac{x_j - x_i}{d_{ij}} - g\hat{e}_g + u\hat{e}_w, \quad (24)$$

where N denotes the grasshopper count. The modified equation can be employed for resolving the optimization issue which is given below:

$$X_i^d = c\left(\sum_{j=1, j\neq i}^{N} c\frac{ub_d - lb_d}{2} s\left(|x_j^d - x_i^d|\right)\frac{x_j - x_i}{d_{ij}}\right) + \hat{T}_d, \quad (25)$$

where ub_d and lb_d are the upper and lower bounds in the Dth dimension, $\hat{T}_d$ denotes the target value, and c is the falling co-efficient for shrinking the comfort, repulsive, and attracting zones. Fig. 3 demonstrates the flowchart of GOA technique.

It is considered that the way of wind is normally towards the target. The nymph moves on rolling into ground to identify the food whereas the adult moves on jumping in the air, generating exploration as well as exploitation. They can be effectively balanced by reducing the value of variable c in (26) equivalent to the iteration count which is defined below.

$$c = c_{\max} - l\left[\frac{c_{\max} - c_{\min}}{L}\right], \quad (26)$$

where $c_{\max}$ and $c_{\min}$ denotes the maximum and minimum values, l specifies the present round, and L denotes the higher iteration count.

3. PERFORMANCE VALIDATION

This section inspects the performance of the GOA-BiLSTM model on the applied four datasets such as canon dataset, iPod dataset, DVD dataset, and Nokia dataset.

Table 1 and Figs. 4-5 perform a brief comparative results analysis of the GOA-BiLSTM model on the applied Canon dataset. The figure has shown that the CSK model has failed to demonstrate effective performance by accomplishing a reduced sensitivity of 84.18%, specificity of 51.53%, accuracy of 77.51%, F-score of 85.62%, and kappa of 34.07%. Eventually, the SVM model has showcased slightly improved outcomes by offering a sensitivity of 85.36%, specificity of 54.36%, accuracy of 80.34%, F-score of 87.92%, and kappa of 35.43%. Simultaneously, the NN model has showcased moderate outcome over the earlier techniques by obtaining a sensitivity of 86.71%, specificity of 58.71%, accuracy of 81.91%, F-score of 88.82%, and kappa of 41.64%. Concurrently, the PSO algorithm has reached a somewhat improved results with the sensitivity of 86.70%, specificity of 64.10%, accuracy of 82.54%, F-score of 89.02%, and kappa of 46.63%.

Moreover, the ACO algorithm has led to the reasonable performance with a sensitivity of 98.18%, specificity of 90.07%, accuracy of 96.38%, F-score of 97.68%, and kappa of 89.38%. Furthermore, the ACO-K algorithm has reached a competitive sensitivity of 98.59%, specificity of 93.47%, accuracy of 97.48%, F-score of 98.39%, and kappa of 92.55%. But the presented GOA-BiLSTM model has demonstrated better results with the sensitivity of 99.92%, specificity of 98.76%, accuracy of 99.57%, F-score of 99.46%, and kappa of 98.89%.

Table 1: Comparison of Proposed GOA-BiLSTM with Existing Methods for Canon Dataset

Sl. No	Dataset	Classifier	Sensitivity	Specificity	Accuracy	F-score	Kappa
1	Canon	GOA-BiLSTM	99.92	98.76	99.57	99.46	98.89
2		ACO-K	98.59	93.47	97.48	98.39	92.55
3		ACO	98.18	90.07	96.38	97.68	89.38
4		PSO	86.70	64.10	82.54	89.02	46.63
5		CSK	84.18	51.53	77.51	85.62	34.07
6		SVM	85.36	54.36	80.34	87.92	35.43
7		NN	86.71	58.71	81.91	88.82	41.64

Table 2 and Figs. 6-7 accomplish a brief comparative outcomes analysis of the GOA-BiLSTM method on the applied iPod dataset. The figure portrayed that the PSO manner has failed to showcase effective performance by accomplishing a reduced sensitivity of 80.65%, specificity of 93.89%, accuracy of 91.44%, F-score of 77.76% and kappa of 72.47%. Likewise, the SVM method has exhibited somewhat higher outcomes by offering a sensitivity of 81.99%, specificity of 85.42%, accuracy of 84.83%, F-score of 64.88%, and kappa of 55.74%.

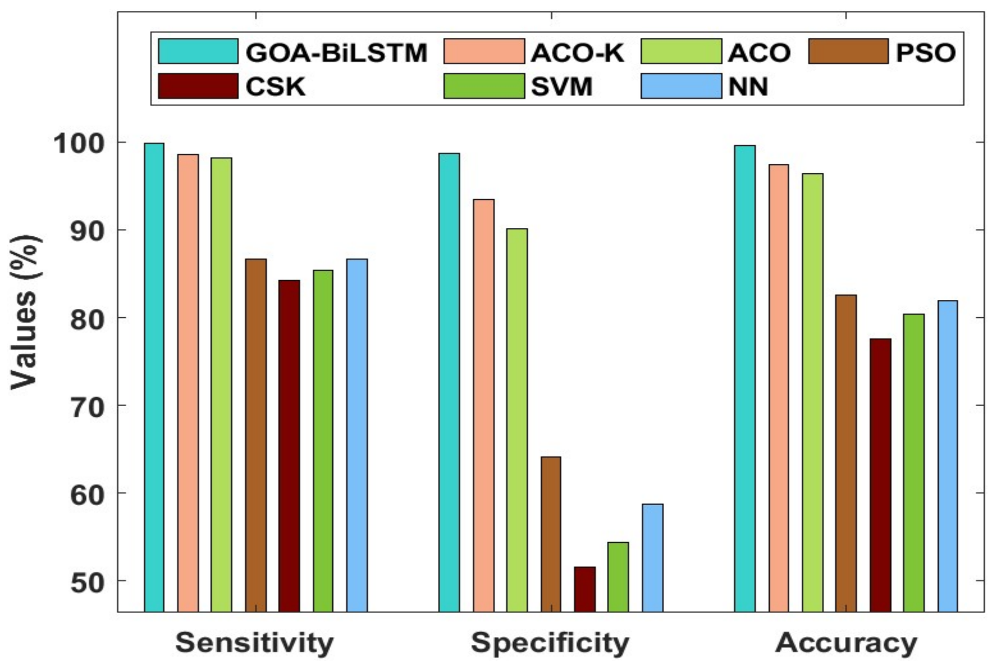

Figure 4: Result analysis of GOA-BiLSTM model under Canon Dataset-I

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

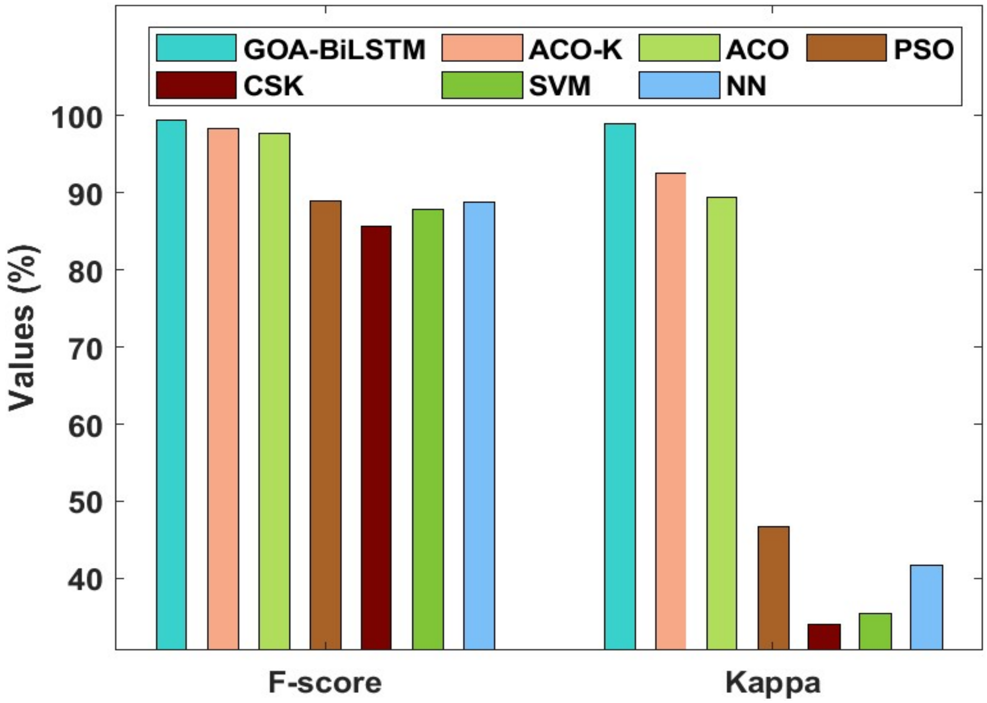

Figure 5: Result Analysis of GOA-BiLSTM Model under Canon Dataset-II

At the same time, the CSK technique has demonstrated moderate results over the earlier techniques by obtaining a sensitivity of 82.17%, specificity of 92.57%, accuracy of 90.83%, F-score of 75%, and kappa of 69.44%. Similarly, the NN method has achieved a somewhat increased outcome with a sensitivity of 83.27%, specificity of 85.68%, accuracy of 85.27%, F-score of 65.90%, and kappa of 57.02%. Moreover, the ACO technique has led to reasonable performance with a sensitivity of 91.15%, specificity of 99.35%, accuracy of 97.51%, F-score of 94.28%, and kappa of 92.69%. Also, the ACO-K algorithm has attained a competitive sensitivity of 94.91%, specificity of 99.50%, accuracy of 98.50%, F-score of 96.50%, and kappa of 95.55%. Finally, the projected GOA-BiLSTM technique has outperformed efficient outcomes with the sensitivity of 98.96%, specificity of 99.98%, accuracy of 99.71%, F-score of 99.02%, and kappa of 98.92%.

Table 2 Comparison of proposed GOA-BiLSTM with existing methods for iPod dataset

Sl. No	Dataset	Classifier	Sensitivity	Specificity	Accuracy	F-score	Kappa
1	iPod	GOA-BiLSTM	98.96	99.98	99.71	99.02	98.92
2		ACO-K	94.91	99.50	98.50	96.50	95.55
3		ACO	91.15	99.35	97.51	94.28	92.69
4		PSO	80.65	93.89	91.44	77.76	72.47
5		CSK	82.17	92.57	90.83	75	69.44
6		SVM	81.99	85.42	84.83	64.88	55.74
7		NN	83.27	85.68	85.27	65.90	57.02

Journal of Theoretical and Applied Information Technology
31st January 2022. Vol.100. No 2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

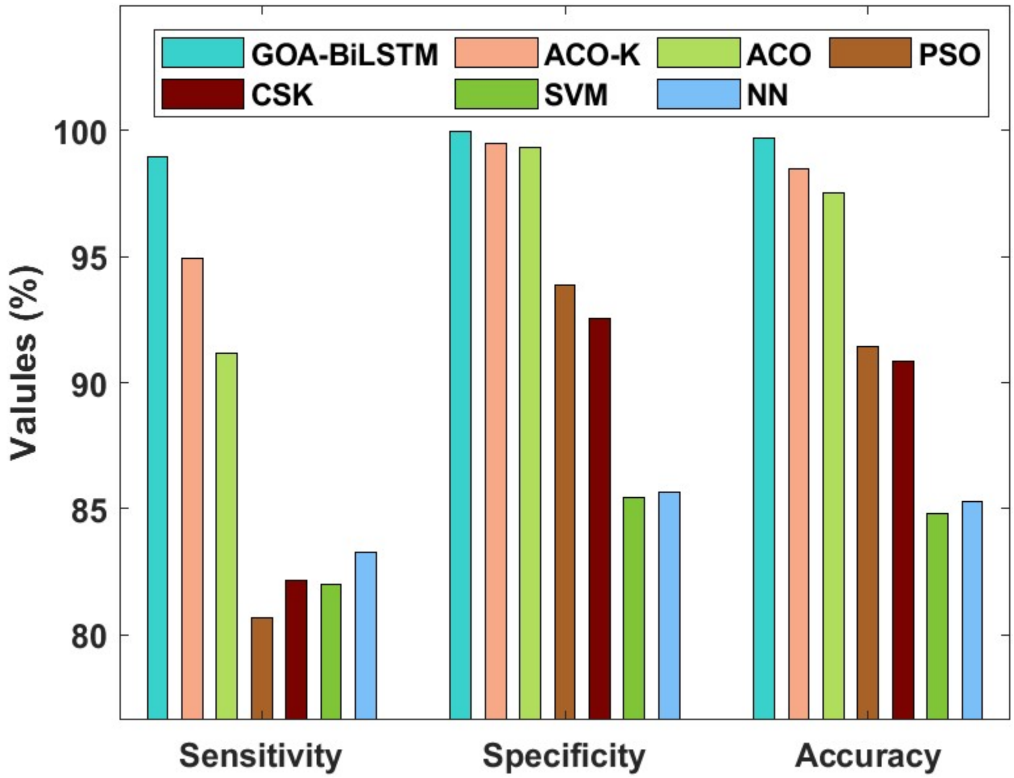

Figure 6: Result Analysis of GOA-BiLSTM Model under iPod Dataset-I

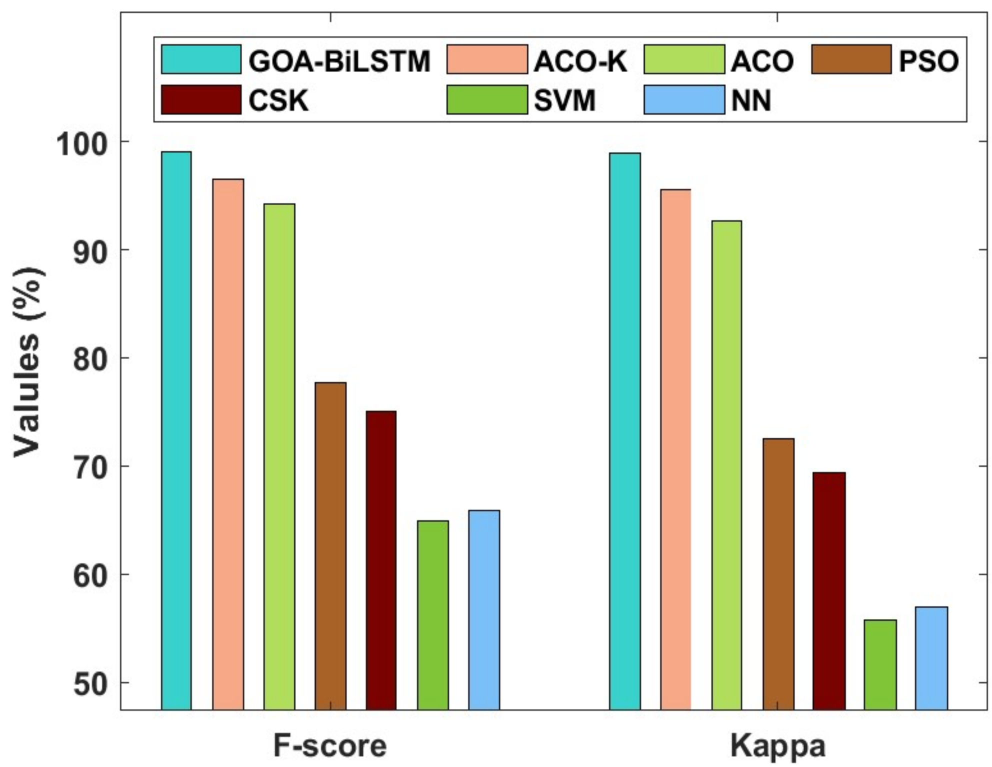

Figure 7: Result Analysis of GOA-BiLSTM Model under iPod Dataset-II

Table 3 and Figs. 8-9 execute a brief comparative results analysis of the GOA-BiLSTM technique on the applied DVD dataset. The figures exhibited that the NN method has failed to showcase effective performance by accomplishing a lesser sensitivity of 93.93%, specificity of 71.61%, accuracy of 87.84%, F-score of 91.82%, and kappa of 68.15%. The CSK model has outperformed with slightly increased result by offering a sensitivity of 96.27%, specificity of 72.85%, accuracy of 90.10%, F-score of 93.48%, and kappa of 73.05%.

Table 3: Comparison of proposed GOA-BiLSTM with existing methods for DVD dataset

Sl. No	Dataset	Classifier	Sensitivity	Specificity	Accuracy	F-score	Kappa
1	DVD	GOA-BiLSTM	99.90	97.12	99.60	99.72	97.90
2		ACO-K	98.23	89.93	96.66	97.94	89.03
3		ACO	97.92	86.50	95.70	97.35	86.03
4		PSO	96.57	79.69	92.61	95.23	78.76
5		CSK	96.27	72.85	90.10	93.48	73.05
6		SVM	96.93	74.88	91.18	94.20	75.85
7		NN	93.93	71.61	87.84	91.82	68.15

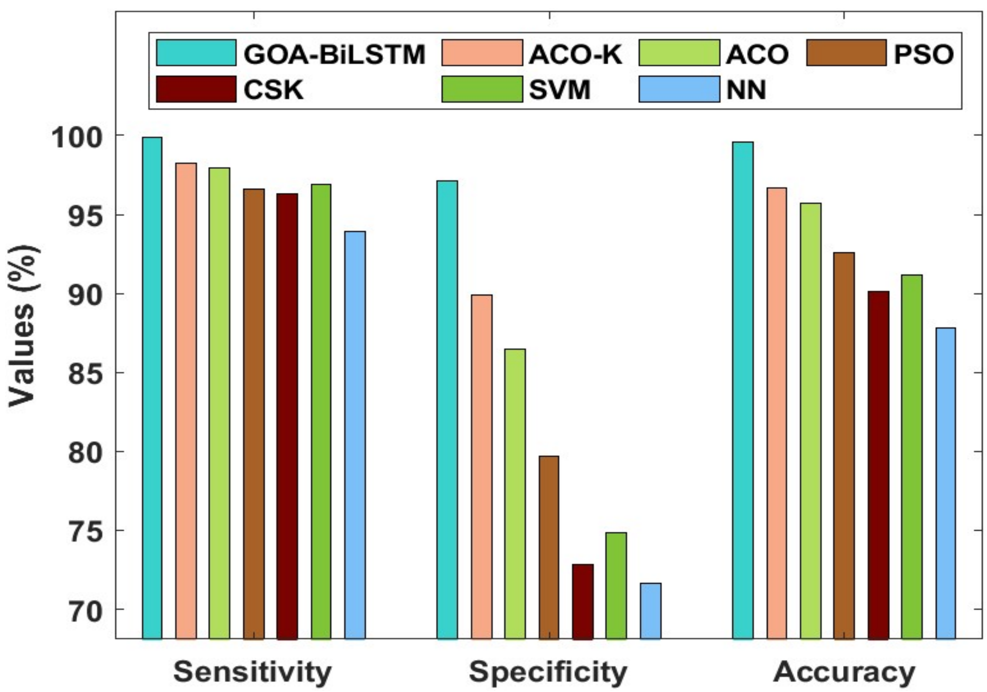

Figure 8: Result Analysis of GOA-BiLSTM Model under DVD Dataset-I

Along with that, the PSO manner has demonstrated moderate outcome over the earlier models by attaining a sensitivity of 96.57%, specificity of 76.69%, accuracy of 92.61%, F-score of 95.23%, and kappa of 78.76%. The SVM technique has achieved a slightly higher outcome with a sensitivity of 96.93%, specificity of 74.88%, accuracy of 91.18%, F-score of 94.20%, and kappa of 75.85%. Furthermore, the ACO method has led to the reasonable performance with a sensitivity of 97.92%, specificity of 86.50%, accuracy of 95.70%, F-score of 97.35%, and kappa of 86.03%. Besides, the ACO-K technique has reached a competitive sensitivity of 98.23%, specificity of 89.93%, accuracy of 96.66%, F-score of 97.94%, and kappa of 89.03%. However, the proposed GOA-BiLSTM algorithm has showcased effectual outcomes with the sensitivity of 99.90%, specificity of 97.12%, accuracy of 99.60%, F-score of 99.72%, and kappa of 97.90%.

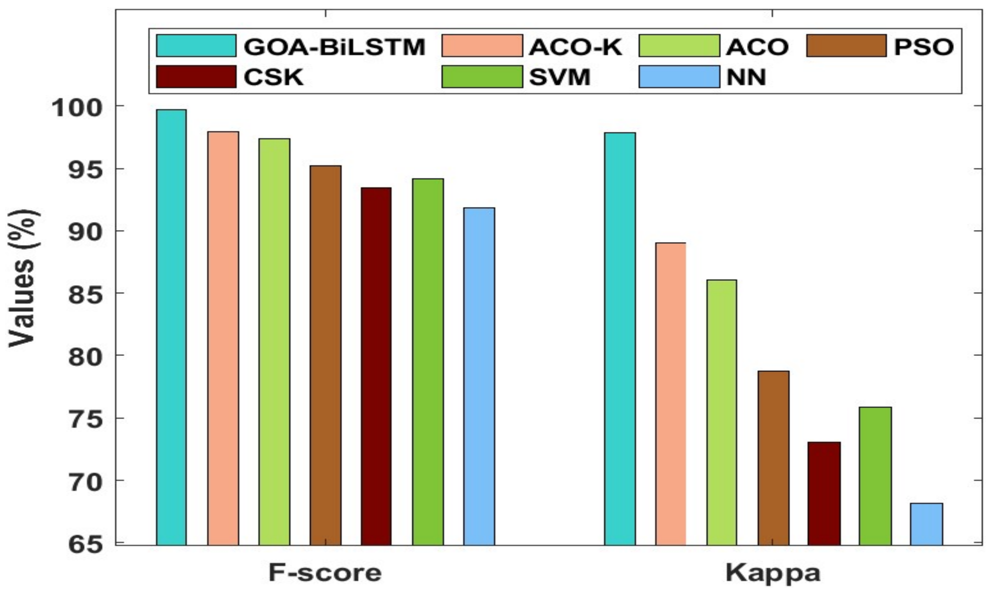

Figure 9: Result Analysis of GOA-BiLSTM Model under DVD Dataset-II

Table 4 and Figs. 10-11 implement a brief comparative outcomes analysis of the GOA-BiLSTM approach on the applied Nokia dataset. The figure demonstrated that the NN technique has failed to exhibit effective performance by accomplishing a minimum sensitivity of 78.16%, specificity of 94.58%, accuracy of 90.59%, F-score of 80.14% and kappa of 73.99%. At the same time, the PSO manner has portrayed slightly enhanced results by offering a sensitivity of 82.70%, specificity of 95.57%, accuracy of 92.64%, F-score of 83.65% and kappa of 78.91%. Next, the SVM algorithm has outperformed moderate outcome over the earlier methods by reaching a sensitivity of 86.56%, specificity of 95.34%, accuracy of 93.33%, F-score of 85.60%, and kappa of 81.27%. The ACO model has obtained a somewhat higher outcome with the sensitivity of 87.32%, specificity of 79.62%, accuracy of 85.20%, F-score of 89.53%, and kappa of 64.36%.

Table 4: Comparison of Proposed GOA-BiLSTM Method with Existing Methods for Nokia Dataset

Sl. No	Dataset	Classifier	Sensitivity	Specificity	Accuracy	F-score	Kappa
1	Nokia	GOA-BiLSTM	97.53	99.42	98.98	98.87	97.42
2		ACO-K	90.71	98.20	96.41	92.36	90.01
3		ACO	87.32	79.62	85.20	89.53	64.36
4		PSO	82.70	95.57	92.64	83.65	78.91
5		CSK	88.88	96.07	94.52	87.5	83.99
6		SVM	86.56	95.34	93.33	85.60	81.27
7		NN	78.16	94.58	90.59	80.14	73.99

Journal of Theoretical and Applied Information Technology
31st January 2022. Vol.100. No 2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

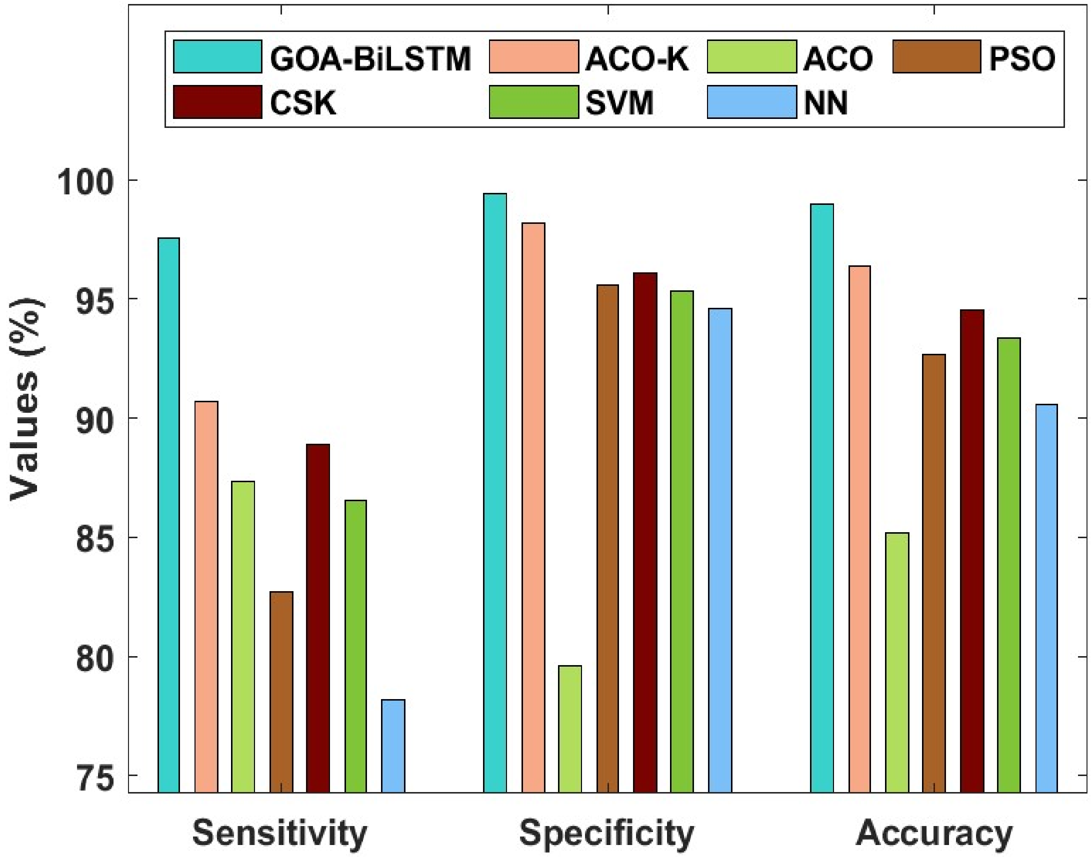

Figure 10: Result Analysis of GOA-BiLSTM model under Nokia Dataset-I

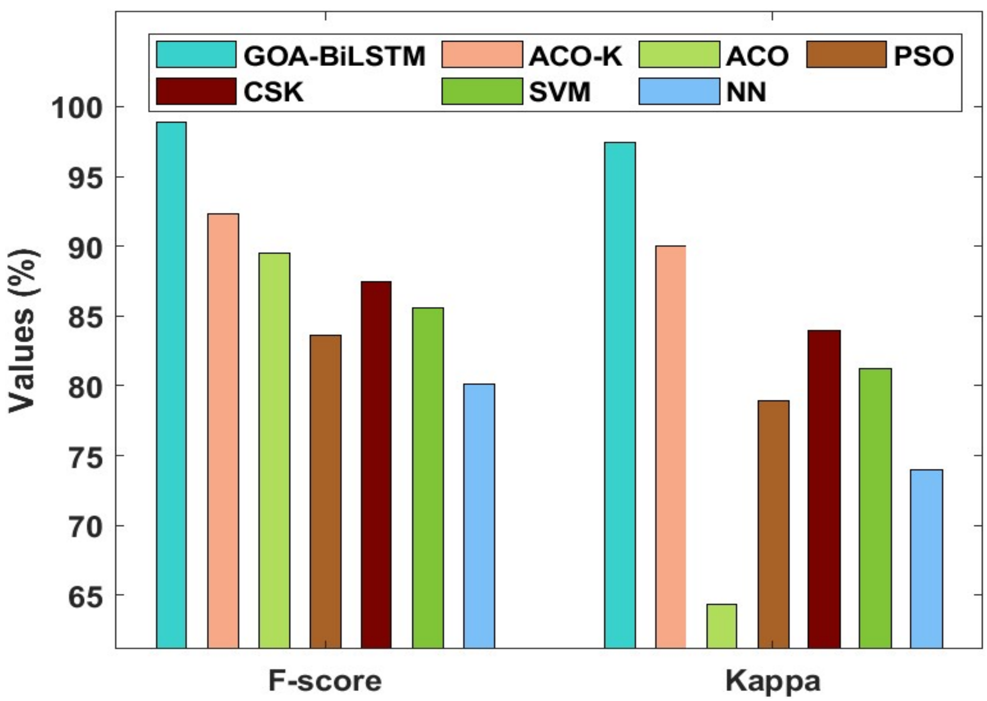

Figure 11: Result Analysis of GOA-BiLSTM Model under Nokia Dataset-II

Additionally, the CSK approach has led to reasonable performance with a sensitivity of 88.88%, specificity of 96.07%, accuracy of 94.52%, F-score of 87.5% and kappa of 83.99%. Likewise, the ACO-K method has achieved a competitive sensitivity of 90.71%, specificity of 98.20%, accuracy of 96.41%, F-score of 92.36% and kappa of 90.01%. Eventually, the proposed GOA-BiLSTM methodology has showcased efficient outcomes with the sensitivity of 97.53%, specificity of 99.42%, accuracy of 98.98%, F-score of 98.87%, and kappa of 97.42%.

4. CONCLUSION

This study has developed a novel SA and classification model by the use of GOA-BiLSTM. Here, the preprocessing is employed to discard the noise that exists in the input data which is applied for enhancing the classifier performance. The GOA-BiLSTM model involves word2vec based feature extraction process to derive a useful set of features. Furthermore, Bi-LSTM based classifier is utilized for the determination of the optimum class labels of the extracted features. Besides, the GOA is exploited for the hyperparameter optimization of the Bi-LSTM model. The GOA-BiLSTM model guarantees an improved outcome through a widespread set of simulations that were carried out on four datasets. The simulation outcome verified the superiority of the GOA-BiLSTM model by accomplishing a higher accuracy of 99.57%, 99. 71%, 99. 06%, and 98.98% on the applied Canon, Nokia DVD, and iPod dataset respectively. As a part of future scope, the classification performance of the GOA-BiLSTM model can be enhanced using advanced deep learning architectures.

REFERENCES

[1] Pak, A.; Paroubek, P. Twitter as a corpus for sentiment analysis and opinion mining. LREc 2010, 10, 1320–1326.
[2] Read, J. Using emoticons to reduce dependency in machine learning techniques for sentiment classification. In Proceedings of the ACL Student Research Workshop, Ann Arbor, Michigan, 27 June 2005; pp. 43–48.
[3] Kennedy, A.; Inkpen, D. Sentiment classification of movie reviews using contextual valence shifters. Comput. Intell. 2006, 22, 110–125.
[4] Wan, X. A comparative study of cross-lingual sentiment classification. In Proceedings of the 2012 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology, Macau, China, 4–7 December 2012; Volume 1, pp. 24–31.
[5] Valakunde, N.; Patwardhan, M. Multi-aspect and multi-class based document sentiment analysis of educational data catering accreditation process. In Proceedings of the 2013 International Conference on Cloud & Ubiquitous Computing & Emerging Technologies, Pune, India, 15–16 November 2013; pp. 188–192.
[6] Gautam, G.; Yadav, D. Sentiment analysis of twitter data using machine learning approaches and semantic analysis. In Proceedings of the 2014 Seventh International Conference on Contemporary Computing (IC3), Noida, India, 7–9 August 2014; pp. 437–442.
[7] Tripathy, A.; Agrawal, A.; Rath, S.K. Classification of sentiment reviews using n-gram machine learning approach. Expert Syst. Appl. 2016, 57, 117–126.
[8] Hasan, A.; Moin, S.; Karim, A.; Shamshirband, S. Machine learning-based sentiment analysis for twitter accounts. Math. Comput. Appl. 2018, 23, 11.
[9] Yassine, M.; Hajj, H. A framework for emotion mining from text in online social networks. In Proceedings of the 2010 IEEE International Conference on Data Mining Workshops, Sydney, Australia, 13 December 2010; pp. 1136–1142.
[10] Denecke, K. Using sentiwordnet for multilingual sentiment analysis. In Proceedings of the 2008 IEEE 24th International Conference on Data Engineering Workshop, Cancun, Mexico, 7–12 April 2008; pp. 507–512.
[11] Jiang, L.; Yu, M.; Zhou, M.; Liu, X.; Zhao, T. Target-dependent twitter sentiment classification. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Portland, Oregon, 19–24 June 2011; Volume 1, pp. 151–160.
[12] Bahrainian, S.A.; Dengel, A. Sentiment analysis using sentiment features. In Proceedings of the 2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT), Atlanta, GA, USA, 17–20 November 2013; Volume 3; pp. 26–29.
[13] Antai, R. Sentiment classification using summaries: A comparative investigation of lexical and statistical approaches. In Proceedings of the 2014 6th Computer Science and Electronic Engineering Conference

(CEEC), Colchester, UK, 25–26 September 2014; pp. 154–159.

[14] Ghiassi, M.; Lee, S. A domain transferable lexicon set for Twitter sentiment analysis using a supervised machine learning approach. Expert Syst. Appl. 2018, 106, 197–216.

[15] Mensikova, A.; Mattmann, C.A. Ensemble sentiment analysis to identify human trafficking in web data. Workshop on Graph Techniques for Adversarial Activity Analytics (GTA 2018), Marina Del Rey, CA, USA, 5–9 February 2018

[16] Chaturvedi, I.; Cambria, E.; Welsch, R.E.; Herrera, F. Distinguishing between facts and opinions for sentiment analysis: Survey and challenges. Inf. Fusion 2018, 44, 65–77.

[17] Ozer, Z., Ozer, I. and Findik, O., 2018. Diacritic restoration of Turkish tweets with word2vec. *Engineering Science and Technology, an International Journal*, *21*(6), pp.1120-1127.

[18] Long, F., Zhou, K. and Ou, W., 2019. Sentiment analysis of text based on bidirectional LSTM with multi-head attention. *IEEE Access*, *7*, pp.141960-141969.

[19] Saxena, A., Shekhawat, S. and Kumar, R., 2018. Application and development of enhanced chaotic grasshopper optimization algorithms. *Modelling and Simulation in Engineering*, *2018*.