SSRG International Journal of Electronics and Communication Engineering
ISSN: 2348-8549/ https://doi.org/10.14445/23488549/IJECE-V11I12P105

Volume 11 Issue 12, 44-55, December 2024

Original Article

Modeling of Snow Ablation Optimization Algorithm with Deep Learning Approach for Sentiment Classification on Social Media Corpus

K. Manikandan[1], V. Ganesh[2]

[1]*Department of Computer and Information Science, Annamalai University, Tamilnadu, India.*
[2]*Department of Computer Science, Government Arts College(A), Kumbakonam, Tamilnadu, India.*

Corresponding Author : skmaniswamimalai@gmail.com

Received: 10 October 2024 Revised: 14 November 2024 Accepted: 05 December 2024 Published: 30 December 2024

Abstract - *Recently, Sentiment Analysis (SA) has been a tedious process in natural language processing (NLP), particularly for Social Media (SM) text, which tends to be brief, noisy, and informal. SA is a way of extracting data about an entity and automatically detecting the subjectivity of that entity. The SA aims to determine whether text created by the user conveys optimistic, adverse, or impartial feelings. The goal is to automatically classify sentiments towards specific features like products, topics, or movies, utilizing Deep Learning (DL) as an advanced technique to meet the growing demand for accurate SA. Therefore, this study proposes a new Snow Ablation Optimization with a Deep Learning for Sentiment Detection and Classification (SAODL-SDC) approach on the SM corpus. The presented SAODL-SDC approach primarily intends to recognize the class of opinions in SM data. In the SAODL-SDC technique, a multi-faceted approach begins with data preprocessing and bag of words (BoWs) feature extraction. The SAODL-SDC technique employs a Convolutional Long Short-Term Memory Autoencoder (CLSTM-AE) technique for sentiment detection. The hyperparameter tuning process using SAO is utilized to improve the effectualness of the CLSTM-AE technique. The SAODL-SDC technique is examined under Sentiment140 and the Airline's datasets. The performance validation of the SAODL-SDC approach portrayed superior accuracy values of 94.28% and 97.00% over existing techniques.*

Keywords - *Sentiment analysis, Snow Ablation Optimization, Bag of words, Social media, Deep learning.*

1. Introduction

SA or opinion mining is the research of an individual's emotions, thoughts, and attitudes from written languages [1]. Currently, individual consumers have revealed great attention in thoughts about services and products on the web, and this data mainly disturbs consumer decisions. Owing to the massive volume of data and thoughts being formed, distributed, and moved daily through the internet and other SM platforms, SA has become significant for emerging opinion mining methods [2]. SA intends to specify the opinions stated in a specific text and to find positive and negative views, which are chiefly significant for decision-making selections and strategies. SA denotes the organization of the opinions, feelings, and subjective text. SA delivers the comprehension data linked to public views as it analyses dissimilar reviews and tweets. It is a confirmed device to forecast numerous necessary actions like box office execution of movies and standard polls [3]. Public opinions assess definite entities such as products, individuals, or positions and may originate on dissimilar sites such as Yelp and Amazon [4]. The sentiments can be categorized into positive, negative, or neutral. The foremost intention of SA is to define the representative way of consumer reviews mechanically. The demand for SA is higher owing to the upsurge in analyzing and constructing hidden data from SM sites using unstructured data [5].

As Machine Learning (ML) models have been severely enhanced in the past few years, a superior method has been utilized to improve the accuracy of SA predictions [6]. ML is sturdily linked to arithmetical analysis, which concentrates on generating forecasts using digital computers. The study of mathematical development completes the application to modify the group of Artificial Intelligence (AI). Data processing plays a central part in research regions and the area of research between AI [7]. While considering its use across numerous companies and research issues, AI is also measured as analytical systematic. Over the past few years, DL has rented more information from the brains of humans, statistics, and functional mathematics. Recently, with the increasing volume of data and the fast growth of computers, there has been an assortment of big data problems [8]. Then, the practicality of DL is slowly enhanced in speech and image detection and so on. Moreover, DL is also employed in NLP, where text SA is a significant application region. The rise of SM has made it crucial to analyze public sentiment, as online opinions significantly influence consumer decisions [9]. With vast amounts of user-generated content, automated SA is required to extract insights effectually. By integrating DL techniques with optimization algorithms like SAO, sentiment classification models can be improved for more accurate and real-time analysis, enhancing decision-making processes [10].

This study proposes a new Snow Ablation Optimization with a Deep Learning for Sentiment Detection and Classification (SAODL-SDC) approach on the SM corpus. The presented SAODL-SDC approach primarily intends to recognize the class of opinions in SM data. In the SAODL-SDC technique, a multi-faceted approach begins with data preprocessing and bag of words (BoWs) feature extraction. The SAODL-SDC technique employs a convolutional long short-term memory autoencoder (CLSTM-AE) technique for sentiment detection. The hyperparameter tuning process using SAO is utilized to improve the effectualness of the CLSTM-AE technique. The SAODL-SDC technique is examined under Sentiment140 and the Airline's datasets. The major contribution of the SAODL-SDC approach is listed below.

- The SAODL-SDC model incorporates data preprocessing with BoWs for feature extraction, ensuring effectual handling of text data. This methodology allows the model to concentrate on crucial terms while mitigating dimensionality. It contributes to improved feature representation, enhancing the performance of sentiment detection techniques.
- The CLSTM-AE technique captures local and sequential text data dependencies for more precise sentiment detection. This hybrid model integrates convolutional layers with LSTM units to improve feature extraction and understanding of context. It also enhances SA by effectively learning hierarchical representations of text.
- The SAO model is utilized for hyperparameter tuning to optimize the performance of the CLSTM-AE model. This technique fine-tunes the model's parameters, ensuring improved convergence and higher accuracy in sentiment detection. SAO enhances the method's capability to generalize and improve prediction outcomes by systematically refining hyperparameters.
- The SAODL-SDC method presents a robust solution for improving sentiment detection accuracy, specifically in complex or noisy datasets. It confirms more precise sentiment classification by incorporating data preprocessing, feature extraction, and advanced model optimization. This integrated strategy enhances the model's resilience and performance across varying types of input data.
- The novelty of the SAODL-SDC approach is in integrating the CLSTM-AE with SAO for hyperparameter tuning. This unique incorporation improves SA accuracy by effectively addressing complex, noisy data and fine-tuning model parameters. The method enhances performance and robustness in sentiment classification tasks by using advanced DL and optimisation techniques.

2. Literature Works

Ashok et al. [11] proposed a complete study of emotion on Twitter, using various sorts of innovative DL and Neural Network (NN) techniques, namely Convolutional NNs (CNNs) and Recurrent NNs (RNNs). Furthermore, the method discovers the efficacy of Hybrid Ensemble techniques in improving SA accuracy and enhanced time. The HCCRNN method utilizes a classy DL method for SA on Twitter data. In [12], a novel DL-based technique was developed for SA. This technique executes emotion identification on vectorized analyses utilizing dual models of Word2Vec. Bellam and Prasanna [13] projected a new DL-based technique, which uses CNN and restricted Boltzmann machine (RBM) models, which comprises a complete pipeline that includes feature extraction, data preprocessing, and identification utilizing a mixture of CNN and RBM. Alsayat [14] used an altered DL technique with an innovative word embedding model and generated an LSTM technique. Similarly, an ensemble technique is projected. The influences of this research work are dual. (1) a strong structure based on word embedding and an LSTM method is used. (2) a hybrid ensemble method is employed for SA.

The authors [15] projected novel DL and graph-based techniques. Twitter is employed for research, and language specialists eliminate and interpret tweets to progress the dataset. The LSTM-GRU technique categorizes hate content into six classes. The Girvan–Newman model is also utilized. Abid et al. [16] introduced a framework in which the Contextual Representation (CR) was developed based on the textual structure. In CR, advanced word representation methods, like FastText (sub-word data) and GloVe (global vectors), jointly yield word representation upon the input series utilizing a weight device. Next, an exclusive method is projected: a 3-parallel layer dilated convolution system with global mean pooling. Halawani et al. [17] propose a model employing the Harris Hawks Optimizer with DL (ASASM-HHODL) technique. This technique mainly procedures the raw text of SM into a beneficial layout. Also, the approach utilizes skip-gram and fastText-assisted word embedding to inspect the lessening of linguistic handling dependence on preprocessing. Furthermore, attention-based Bi-LSTM (Abi-LSTM) is used to identify sentiments. Besides, the hyperparameter of the Abi-LSTM technique is tuned using the HHO method. Rudiyanto and Setiawan [18] analyze Twitter users' sentiments using DL models, comprising CNN, TF-IDF for feature extraction, and Word2Vec for feature expansion while optimizing the model's performance with Particle Swarm Optimization (PSO).

Mu et al. [19] present a novel DIBTBL methodology by integrating an extended sentiment dictionary, BERT embeddings, TextCNN-BiLSTM for feature extraction, and an improved BWO model for parameter optimization, followed by MLP for sentiment classification. Jain and Kashyap [20] analyze Indian Twitter users' opinions on COVID-19 using Hindi tweets, feature extraction with NLP, Grey Wolf Optimization (GWO) for feature selection, and a CNN-LSTM hybrid model for sentiment classification. Vatambeti et al. [21] utilize a DL approach, combining CNN and Bi-LSTM models with feature extraction via a pseudo-inverse AE and optimization through Elephant Herd Optimization (EHO).

Biswas, Daniel, and Neogi [22] introduce the Spider Monkey Optimization with a Stacked RNN (SMO-SRNN) model by using SRNN for classification and SMO for fine-tuning its hyperparameters. Al-Onazi et al. [23] propose the Modified Seagull Optimization with DL-assisted Affect Classification on Arabic Tweets (MSGODL-ACAT) technique, using preprocessing, GloVe word embedding, and Deep Belief Network (DBN) and MSGO-based classifying and tuning process. Rasappan et al. [24] propose an Enhanced Golden Jackal Optimization-assisted LSTM (EGJO-LSTM) comprising data collection, preprocessing, feature selection with LF-MICF and IGWO, and classification into negative, positive, or neutral sentiments.

While various SA techniques show promise, many face challenges such as increased computational complexity, difficulty handling noisy or dynamic data, and need help capturing nuanced sentiments. Optimization methods like PSO, HHO, and SMO may improve performance but can lead to longer training times or overfitting, especially with large or imbalanced datasets. A critical research gap is the need for more efficient and scalable SA models to handle noisy, context-dependent, and dynamic SM data. While hybrid DL models and optimization techniques show promise, more exploration is needed to balance accuracy, computational efficiency, and adaptability to rapidly evolving online content. Further investigation is required into models that can seamlessly integrate feature extraction, sentiment classification, and real-time updates.

3. The Proposed Method

This paper introduces a novel SAODL-SDC technique on an SM corpus. The technique primarily recognizes the class of opinions that exist in SM data. Figure 1 represents the entire flow of the SAODL-SDC method.

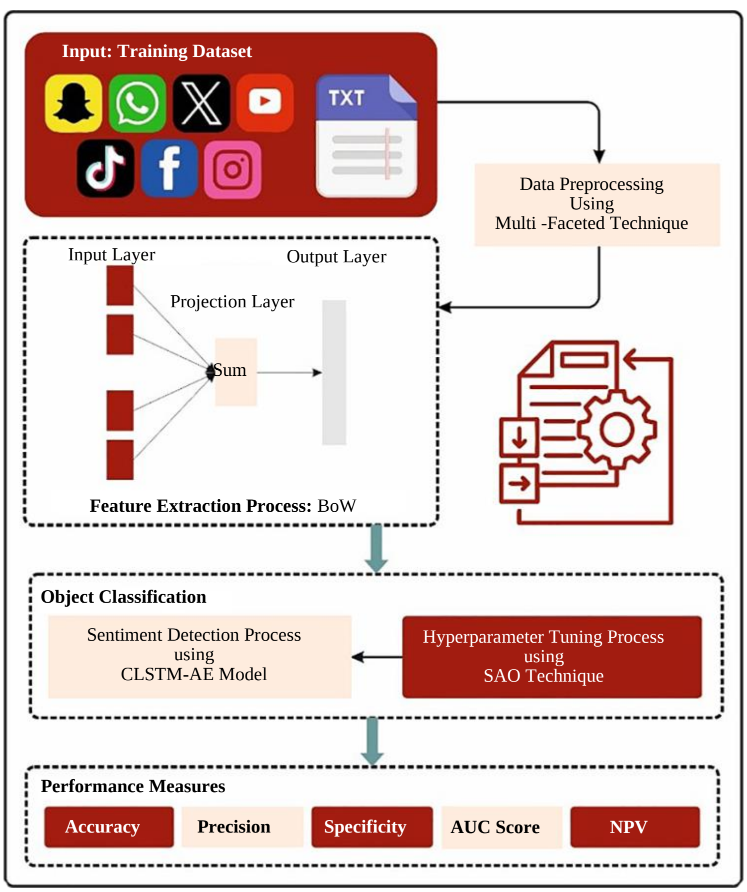

Fig. 1 Structure of SAODL-SDC technique

3.1. Data Preprocessing

Text data preprocessing is the fundamental step in creating a good ML model for classification [25]. Preprocessing is used to convert the raw information into understandable form, and it is also a complex procedure that requires many steps. In this work, the following step is performed in preprocessing. Initially, all the words in the dataset are transformed into lowercase. In general, text messages frequently have different capitalizations that do not significantly affect the final model. As a result, all the words are decreased to lowercase for simplicity. Tokenization was applied in the next step. A word tokenizer is used to split all the lowercase text messages into a collection of words. For example, the tokenizer breaks the message "alright i have a new goal now" into a collection of seven words {' alright', 'i', 'have', 'new', 'goal', 'now'}. Therefore, it generates the data frame of individual words for further processing. After tokenization, all the words with a length lesser than two are eliminated. The alphanumeric words, viz., words mixing digits and letters, are also eliminated from the dataset as punctuations, numbers, and symbols did not contribute to the learning. Then, the stop word was eliminated. Like 'or', 'and', 'this', 'that', etc., frequently used stop words. This can be removed from the textual data, and stop words do not contribute to the knowledge source. Lemmatization is the last step of preprocessing. Lemmatization decreases a term to its base form by eliminating an inflectional ending called a lemma. For instance, a lemmatization minimizes the words car and cars to the car. Lemmatizer exploits morphological analysis of words and vocabulary and returns the dictionary or base form of the word.

3.2. Feature Extraction

BoW is a simple and popular feature extractor method utilized in NLP [26]. It defines the occurrence of all textual content. It generates a matrix of occurrence for a document or sentence, which ignores word order and grammar. Then, this frequency ("occurrence") of words is utilized as a feature for learning. The fundamental concept of using Bow is that the same document has the same content. This approach has a considerable disadvantage for its intuitive clarity and simplicity. The BoW encoder utilizes a corpus or word set and signifies the assumed text with the corpus dimension vector. If the word exists in the text, then the equivalent component is the word frequency. If the one-hot vector encrypts an individual word, then the feature space would contain a length comparable to the cardinality of the dictionary. This increases together with the increasing number of words in the dictionary.

3.3. Sentiment Detection using CLSTM-AE

The SAODL-SDC technique involves the design of a CLSTM-AE network for sentiment detection [27]. This technique is chosen because it captures local patterns through convolutional layers and long-range dependencies through LSTM units. This hybrid architecture effectively handles complex, sequential data, making it ideal for SA in SM texts. CLSTM-AE improves feature extraction, reduces overfitting, and offers enhanced accuracy in detecting sentiments from noisy and unstructured data compared to other models. CAE is a common AE variant where a convolution layer substitutes a fully connected (FC) layer. CAE has the benefits of both the convolutional layer and the unsupervised pre-training AE's capability. In comparison to the traditional network of AE, the CAE holds a convolution layer in the encoding and a de-convolution layer in its place of an FC layer in the decoding. The developed Convolutional AE (CAE) contains convolution, pooling, and de-convolution layers. The encoder has a convolutional and pooling layer, and the decoder uses a de-convolution layer. Encrypting the outcome of the convolutional process with a max-pooling layer authorizes high-layer representation, which is constant to smaller input versions and decreases the computation price of the projected model. The convolution and de-convolution layer are monitored by an activation function (AF) signified below:

$$h^k = \sigma\left(\sum_{l\in L} x^l \otimes w^k + b^k\right) \quad (1)$$

Here, σ denotes the function of activation, h^k represents the latent illustration of the present layer's kth feature mapping, $\otimes$ indicates a 2-D convolutional process, x^l refers to the l^{th} feature mapping of the cluster of L attained from the preceding layer and w^k and b^k denote the weight and bias of the present layer's kth feature mapping, respectively.

The convolution and de-convolution layers execute valid and complete convolution, respectively. For instance, if the feature map (x^l) and the filter size is $p \times p$ and $q \times q$, then when implementing the valid convolution, the size turns into $(p - q + 1) \times (p - q + 1)$, and after implementing the full convolution, the size turns into $(p + q - 1) \times (p + q - 1)$.

A max-pooling layer merges factors relying on the pooling kernel dimension by using the highest action within an input; it builds a condensed-size output.

Time-based features in time sequence sensor data are highly significant and demonstrate human drive. Recently, RNNs depending on LSTM have attained remarkable performance in dissimilar areas. The LSTM structure is answerable for removing time-based factors from physical signals owing to its sequential features and prolonged dependencies.

The CAE output and the compacted features are inputs for determining the hidden time-based connections during the time-frames. At time-frame t, x^t denotes the given signal, and h^t denotes the hidden layer (HL). At $t-1$, C^{t-1} refers to the cell memory state. b^f, b^i, b^c, b^o and w^f, w_i, w^c, and w^o denote the biases and weights subsequently. $tanh$ and σ represent the AF. Initially, the LSTM evaluates the prior data from C^{t-1} by utilizing a forget gate, which is given below:

$$f^t = \sigma(w^f[h^{t-1}, x^t] + b^f) \tag{2}$$

Whereas f^t are both $zero$ and one to depict the complete block and data transfer, correspondingly. The LSTM utilizes a dual-stage process to compute and retain future data. The initial portion normalizes the parameters to be used through the below-mentioned calculation:

$$i^t = \sigma(w^i[h^{t-1}, x^t] + b^i) \tag{3}$$

The 2nd step defines an optimum value of the state $\tilde{C}^t$ by using the below-given calculation:

$$\tilde{C}^t = tanh(w^c[h^{t-1}, x^t] + b^c) \tag{4}$$

In 3rd step, the present state C^t is described by using the below calculation:

$$C^t = f^t * C^{t-1} + i^t * \tilde{C}^t \tag{5}$$

The filtered type of compressed cell state $tanh\ (C^t)$ is the HL output h^t. The data is retained using the sigmoid AF o^t and h^t, are depicted below:

$$o^t = \sigma(w^o[h^{t-1}, x^t] + b^o) \tag{6}$$

$$h^t = o^t * tanb(C^t) \tag{7}$$

FL layers are employed to track higher-level representation. In this research paper, the outputs of LSTM are served into dual HL, and a softmax is used for the previous classification stage. Fig. 2 defines the structure of CLSTM.

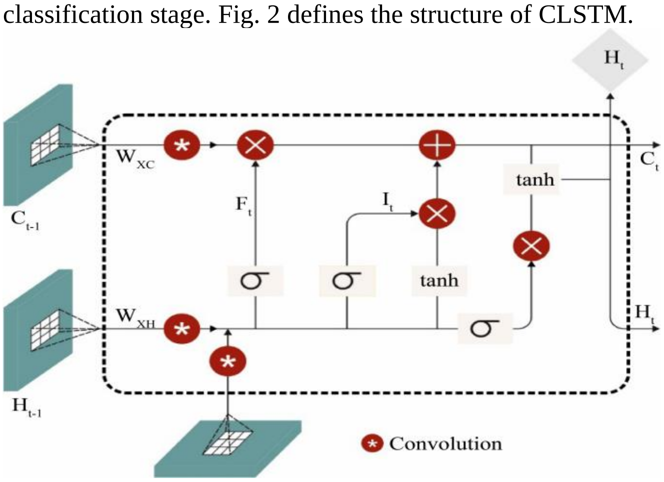

Fig. 2 Structure of CLSTM

3.4. SAO-based Parameter Tuning

Finally, the hyperparameter tuning process using SAO is utilized to improve the efficiency of the CLSTM-AE method [28]. This method is chosen for parameter tuning because it can effectively explore large search spaces and optimize complex models without getting trapped in local minima. Its robustness in handling noisy or imbalanced datasets confirms more accurate hyperparameter adjustments. Compared to other optimization techniques, SAO presents an enhanced convergence speed and improves the overall performance of DL models. The SAO technique was recently developed in 2023 based on three physical processes: sublimation, melting, and evaporation. The SAO model includes the exploitation, exploration, initialization, and 2-population mechanism.

3.4.1. Initialization Stage

The dimension, size, and boundary limit can initialize the population. Based on Equation (8), the initial population is randomly produced. Generally, the whole population can be set by an N-by-dim matrix, where N refers to the population size, and dim is the problem space size.

$$X = 1b + rand \times (ub - lb) \tag{8}$$

In Equation (8), X is the initialized population, lb and ub are the lower and the upper bounds, respectively, and $rand$ is a randomly produced value within [0,1].

3.4.2. Exploration Phase

The water vapour is moved in space without instructions and explored by uneven motion during the exploration stage once the liquid water or snow is converted to water vapour through evaporation or sublimation. In the SAO method, using the irregularity and randomness of Brownian movement makes it easy for individuals to explore potential and valuable regions during the exploration. For typical Brownian movement, the technique exploits the density probability function of the uniform distribution with one variance and 0 means to obtain the stepsize of particle movement. The Brownian movement can be mathematically formulated below:

$$f_B(x; 0,1) = \frac{1}{\sqrt{2\pi}} \times \exp\left(-\frac{X^2}{2}\right) \tag{9}$$

The location updating equation can be given as:

$$X_{i_{new}} = Elite\_poo1(k) + RB_i(j) \times \left(r_1 \times \left(B(j) - X_i(j)\right) + (1 - r_1) \times \left(X(j) - X_i(j)\right)\right) \tag{10}$$

In Equation (10), X_{inew} refers to the location of the individual after the update; $RB_i(j)$ is the random number vector in Brownian movement; r_1 is the random integer within [0, 1]; $B(j)$ is the optimum outcome for the existing populace; $X_i(j)$ refers to the existing individual location; $\overline{X}(j)$ is the centre of mass location of the whole population; k refers to the random value within [1,4], and $Elite\_pool(k)$ is a randomly selected individual from $Elite\_poo1$.

The $\overline{X}(j)$ and $Elite\_poo1(k)$ is mathematically stated as:

$$\overline{X}(t) = \frac{1}{N}\sum_{i=1}^{N} X_i\,(t) \tag{11}$$

$$Elite\_pool(t) \in [B(t), X_{second}(t), X_{third}(t), Z_c(t)] \quad (12)$$

Where $B(t)$ is the optimum solution for the existing population, $X_{second}(t)$ and $X_{third}(t)$ are individuals with the 2nd and 3rd maximum fitness values, and $Z_c(t)$ denotes the individual average at the maximum fitness value.

The top fitness individual in the population is known as an elite individual, which makes it easier to compute $Z_c(t)$.

$$Z_c(t) = \frac{1}{N_1}\sum_{i=1}^{N_1} X_i(t) \quad (13)$$

In Equation (13), N_1 is the number of elite individuals, equivalent to half of N. $Elite\_pool(k)$ refers to the set having the optimum solution, the 2nd optimum solution, the 3rd optimum solution, and the elite individual mean in the populace. The random number is derived from the set to update the location during exploration.

3.4.3. Exploitation Stage

The snow is changed into water through the melting process during exploitation. The degree-day technique characterizes the snowmelt procedure, and the equations are given below:

$$M = DDF \times (T - t_1) \quad (14)$$

In Eq. (14), M refers to the snowmelt rate, which simulates melting behaviour during the exploitation. The average daily temperature t_1 refers to the base temperature of 0.

$$M = DDF \times T \quad (15)$$

In Eq. (15), DDF is the degree-day feature that intervals within [0.35,0.6]. DDF is used to update the mathematical formula:

$$DDF = 0.35 + 0.25 \times \frac{e^{\frac{FEs}{FEs_{\max}}} - 1}{e - 1} \quad (16)$$

In Equation (16), FEs refers to the number of existing computations, $and\ FEs_{\max}$ shows the maximal number. The equation to calculate the snowmelt rate is given below:

$$M = DDF \times T(t) = \left(0.35 + 0.25 \times \frac{e^{\frac{FEs}{FEs_{\max}}} - 1}{e-1}\right) \times T(t),$$
$$T(t) = e^{\frac{-FEs}{FEs_{\max}}}. \quad (17)$$

The location updating equation can be given as follows:

$$X_{i_{new}} = M \times B(t) + RB_i(t) \times \left(r_2 \times \left(B(t) - X_i(t)\right) + (1 - r_2) \times \left(X(t) - X_i(t)\right)\right) \quad (18)$$

Here, M refers to the snowmelt rate and r_2 is the random integer within $[-1$ and $1]$. This equation suggests that non-optimal individuals are more likely to achieve better outcomes by moving towards the current best location.

3.4.4. Dual-Population Mechanism

In the SAO method, a novel dual-population mechanism is used. In summary, the location updating formula is given below:

$$X_{i_{new}} = \begin{cases} Elite\_pool(t) + RB_i(t) \times (r_1 \times \left(B(t) - X_i(t)\right) \\ +(1 - r_1) \times (\overline{X}(t) - X_i(t)))\, i \in index1 \\ M \times B(t) + RB_i(t) \times (r_2 \times \left(B(t) - X_i(t)\right) \\ +(1 - r_2) \times (\overline{X}(t) - X_i(t)))\, i \in index2 \end{cases} \quad (19)$$

In Eq. (19), $index1$ is the set of randomly chosen N_a individuals from the entire population, and $index2$ is the residual individuals.

The SAO utilizes a fitness function (FF) to enhance accomplishment. It defines a positive numeral to denote the improved outcome. Also, the minimizer of the classifier error rate is dignified as FF, as supposed in Eq. (20).

$$fitness(x_i) = ClassifierErrorRate(x_i) = \frac{No.of\ misclassified\ samples}{Total\ no.of\ samples} \times 100 \quad (20)$$

4. Performance Validation

The SAODL-SDC method's result analysis is tested using the Sentiment140 [29] and the Airlines [30] dataset. The Sentiment140 dataset contains 2000 samples, and the Airlines dataset comprises 2000 samples, as depicted in Table 1.

Table 1. Dataset description

Sentiment140 Dataset	
Classes	**Sample Numbers**
Negative	1000
Positive	1000
Total Samples	**2000**
Airlines Dataset	
Classes	**Sample Numbers**
Negative	1000
Positive	1000
Total Samples	**2000**

Figure 3 illustrates the outputs of the SAODL-SDC methodology under Sentiment140. Figures 3a-3b portray the SAODL-SDC methodology's confusion matrices on 70:30 of TRAS/TESS. The output indicated that the SAODL-SDC method identified and classified the two classes in detail.

Equally, Figure 3c depicts the PR of the SAODL-SDC method. The output showed that each class's SAODL-SDC technique has attained maximum PR performance. Lastly, Figure 3d exhibits the ROC of the SAODL-SDC method. Thus, the SAODL-SDC approach attains superior outputs with the greatest ROC under diverse classes.

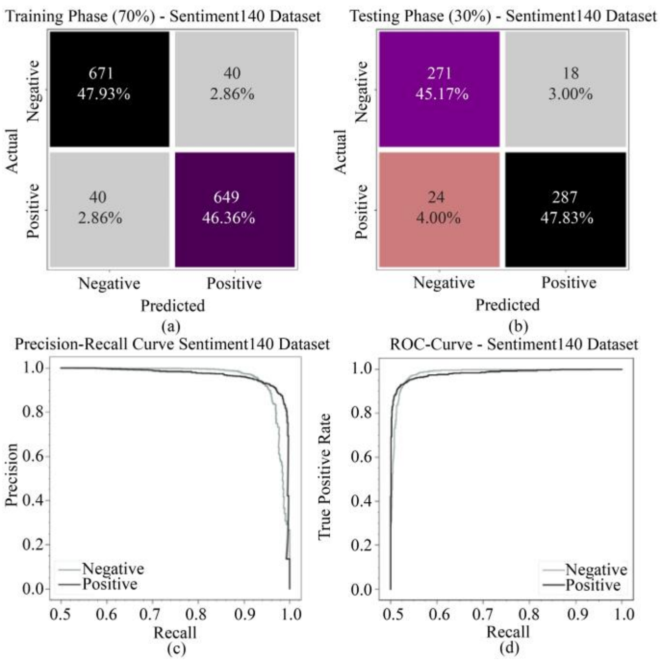

Fig. 3 Sentiment140 dataset (a-b) confusion matrices and (c-d) PR and ROC curves

Table 2 and Figure 4 highlight the outputs of the SAODL-SDC approach under Sentiment140. The outcomes represented that the SAODL-SDC approach correctly identified negative and positive samples. On 70%TRAS, the SAODL-SDC model offers an average $accu_y$ of 94.28%, $prec_n$ of 94.28%, $reca_l$ of 94.28%, F_{score} of 94.28%, and MCC of 88.57%. Additionally, on 30%TESS, the SAODL-SDC model provides an average $accu_y$ of 93.03%, $prec_n$ of 92.98%, $reca_l$ of 93.03%, F_{score} of 93.00%, and MCC of 86.01%.

Table 2. Sentiment recognition output of SAODL-SDC technique on Sentiment140 dataset

Class	$Accu_y$	$Prec_n$	$Reca_l$	F_{Score}	MCC
TRAS (70%)					
Negative	94.37	94.37	94.37	94.37	88.57
Positive	94.19	94.19	94.19	94.19	88.57
Average	94.28	94.28	94.28	94.28	88.57
TESS (30%)					
Negative	93.77	91.86	93.77	92.81	86.01
Positive	92.28	94.10	92.28	93.18	86.01
Average	93.03	92.98	93.03	93.00	86.01

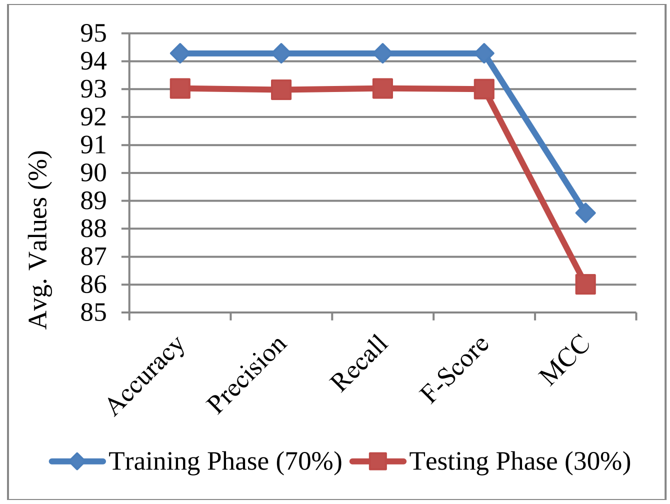

Fig. 4 Average output of SAODL-SDC method on Sentiment140 dataset

Figure 5 shows the accuracy of the Training (TR) and validation (VL) of the SAODL-SDC method under Sentimentown. The accuracy is computed over 0-25 epochs, showing a steady increase in TR/VL accuracy, indicating improved performance with each iteration. The close alignment of TR/VL accuracy suggests minimal overfitting and consistent predictions on unseen data, highlighting the effectiveness of the SAODL-SDC approach.

Fig. 6 portrays the TR/VL loss of the SAODL-SDC technique on the Sentiment140 dataset. The loss is calculated over 0-25 epochs and shows a decreasing trend, indicating the capability of the SAODL-SDC technique to balance data fitting and generalizing issues. The consistent lessening in loss additionally emphasizes the superiority and improved predictions of the method over time.

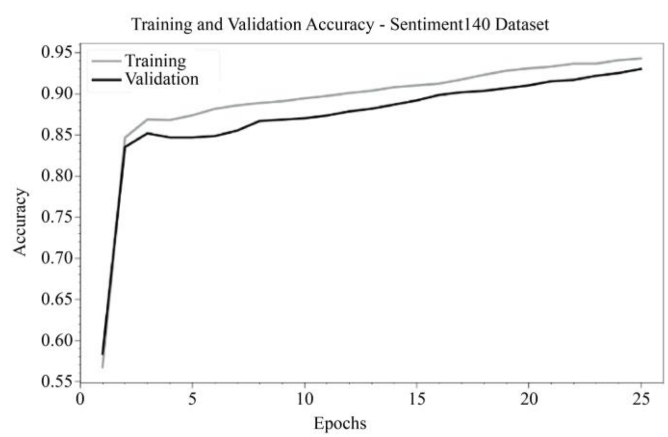

Fig. 5 $Accu_y$ curve of the SAODL-SDC method on the Sentiment140 dataset

Table 3 and Figure 7 illustrate the comparison outcomes of the SAODL-SDC approach under the Sentiment140 dataset [17]. The outputs portray that the TF-RNN and WV-CNN models have shown reduced performance. Simultaneously, the TF-DNN, TF-CNN, WV-DNN, and WV-RNN models have demonstrated reasonable outcomes. Although the ASASM-HHODL technique has managed to reach considerable performance, the SAODL-SDC technique gains better performance with a maximum $accu_y$ of 94.28%, $prec_n$ of 94.28%, $reca_l$ of 94.28%, and F_{score} of 94.28%.

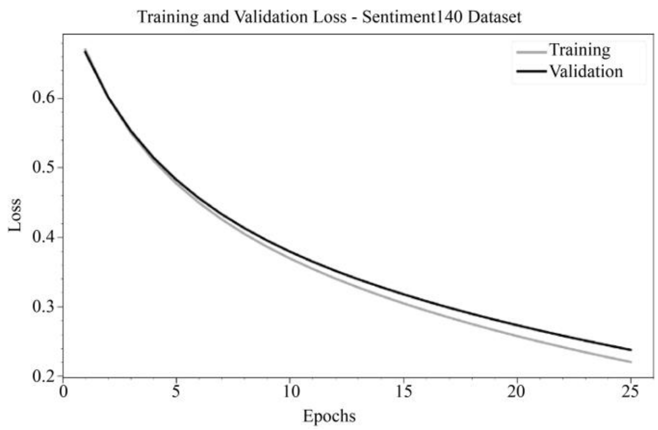

Fig. 6 Loss curve of SAODL-SDC method on Sentiment140 dataset

Table 3. The comparative output of the SAODL-SDC technique with recent models on the Sentiment140 dataset

Sentiment140 Dataset				
Methods	$Accu_y$	$Prec_n$	$Reca_l$	F_{Score}
TF-DNN	78.05	75.10	82.25	77.24
TF-CNN	82.8	79.59	83.48	75.94
TF-RNN	75.05	80.18	81.95	83.32
WV-DNN	81.93	79.95	76.47	83.48
WV-CNN	79.72	82.63	83.77	79.71
WV-RNN	80.98	75.65	78.45	75.98
ASASM-HHODL	84.25	85.83	86.37	86.13
SAODL-SDC	94.28	94.28	94.28	94.28

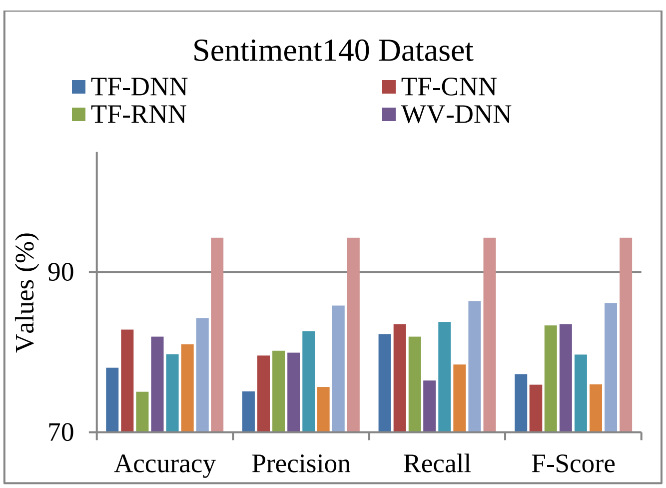

Fig. 7 Comparative outcome of SAODL-SDC technique on Sentiment140 dataset

Table 4 and Figure 8 designate the overall Computational Time (CT) outputs of the SAODL-SDC method under Sentiment 140. The outcomes specify that the TF-RNN and WV-CNN approaches have displayed the worst performance. Simultaneously, the TF-DNN, TF-CNN, WV-DNN, and WV-RNN approaches have established reasonable results. While the ASASM-HHODL method has spread significant performance, the SAODL-SDC model gains enhanced performance with a lesser CT of 1.00s.

Table 4. CT outcome of SAODL-SDC approach with existing method on Sentiment140 dataset

Sentiment140 Dataset	
Methods	**CT (sec)**
TF-DNN	2.20
TF-CNN	3.19
TF-RNN	2.65
WV-DNN	2.81
WV-CNN	3.81
WV-RNN	2.99
ASASM-HHODL	4.19
SAODL-SDC	1.00

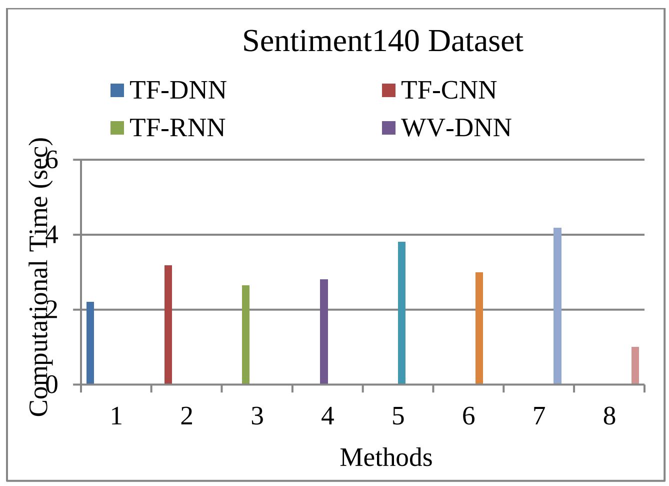

Fig. 8 CT outcome of SAODL-SDC technique on Sentiment140 dataset

Figure 9 establishes the classifier outcomes of the SAODL-SDC method under the Airline dataset. Figs. 9a-9b illustrates the confusion matrices presented by the SAODL-SDC method on 70:30 of TRAS/TESS. The output indicated that the SAODL-SDC approach is renowned and considered perfect for all two classes. Similarly, Fig. 9c defines the PR value of the SAODL-SDC approach. The result showed that the SAODL-SDC approach has attained the most outstanding PR performance in every class. As a final point, Figure 9d establishes the ROC analysis of the SAODL-SDC approach. The figure exposed that the SAODL-SDC approach has resulted in advanced outputs with maximum ROC value under different classes.

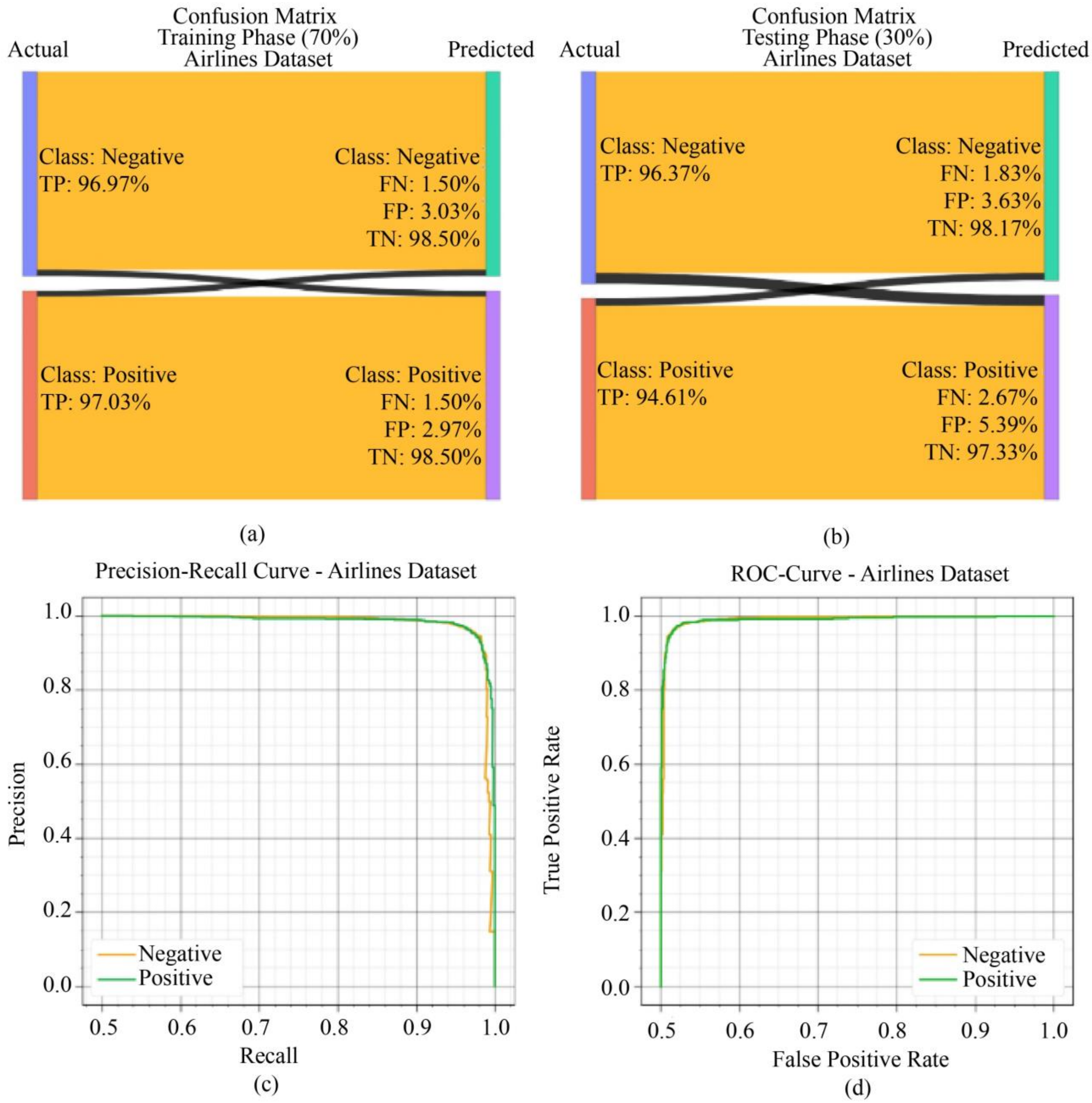

Fig. 9 Airlines dataset (a-b) confusion matrices and (c-d) PR and ROC curves

Table 5. Sentiment recognition output of SAODL-SDC method on Airlines dataset

Class	$Accu_y$	$Prec_n$	$Reca_l$	F_{Score}	MCC
TRAS (70%)					
Negative	96.97	96.97	96.97	96.97	94.00
Positive	97.03	97.03	97.03	97.03	94.00
Average	**97.00**	**97.00**	**97.00**	**97.00**	**94.00**
TESS (30%)					
Negative	94.81	96.37	94.81	95.58	91.01
Positive	96.23	94.61	96.23	95.42	91.01
Average	**95.52**	**95.49**	**95.52**	**95.50**	**91.01**

Table 5 and Figure 10 highlight the complete sentiment recognition outputs of the SAODL-SDC method under Airlines. The outputs denoted that the SAODL-SDC model correctly identified negative and positive samples.

On 70%TRAS, the SAODL-SDC model delivers an average $accu_y$ of 97.00%, $prec_n$ of 97.00%, $reca_l$ of 97.00%, F_{score} of 97.00%, and MCC of 94.00%.

Also, on 30%TESS, the SAODL-SDC approach provides an average $accu_y$ of 95.52%, $prec_n$ of 95.49%, $reca_l$ of 95.52%, F_{score} of 95.50%, and MCC of 91.01%.

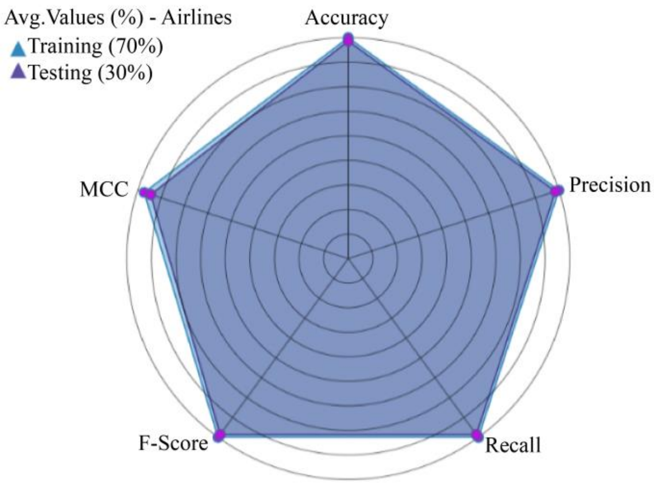

Fig. 10 Average outcome of SAODL-SDC method on Airlines dataset

Figure 11 exhibits the TR/VL accuracy of the SAODL-SDC technique on the Airlines dataset, computed over 0-25 epochs, showing an increasing trend, indicating improved performance with each iteration. Also, the close correlation of TR/VL accuracy values suggests low overfitting and consistent performance, ensuring reliable predictions on unseen data.

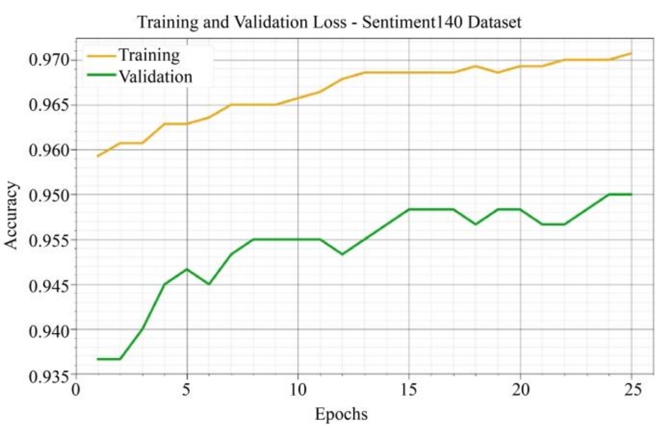

Fig. 11 $Accu_y$ curve of SAODL-SDC method on Airlines dataset

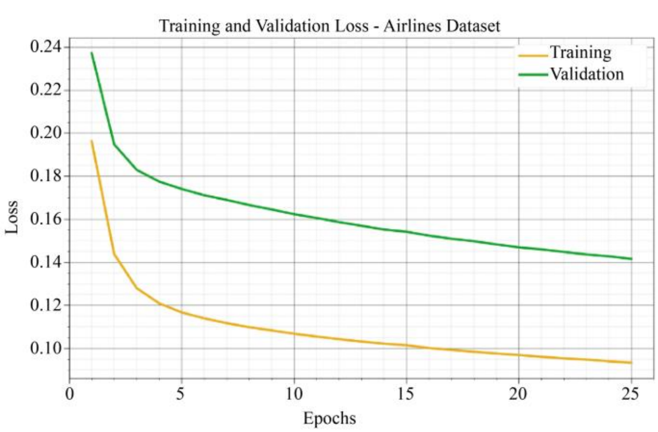

Fig. 12 Loss curve of SAODL-SDC method on Airlines dataset

In Figure 12, the TR/VL loss graph of the SAODL-SDC methodology on the Airlines dataset exhibits a decreasing trend over 0-25 epochs. This indicates the capability of the model to balance data fitting and generalizing issues. The continuous lessening in loss underscores the improved results and refinement of prediction outcomes over time.

Table 6 and Figure 13 show the overall comparison outputs of the SAODL-SDC approach on the Airlines dataset. The results show that the TF-RNN and WV-CNN techniques have shown decreased performance. Meanwhile, the TF-DNN, TF-CNN, WV-DNN, and WV-RNN approaches have demonstrated reasonable results. Whereas the ASASM-HHODL model has managed to grasp considerable performance, the SAODL-SDC method gains better performance with a maximum $accu_y$ of 97.00%, $prec_n$ of 97.00%, $reca_l$ of 97.00%, and F_{score} of 97.00%.

Table 6. The comparative output of the SAODL-SDC technique with recent models on the Airlines dataset

Airlines Dataset				
Methods	$Accu_y$	$Prec_n$	$Reca_l$	F_{Score}
TF-DNN	93.77	89.67	95.08	86.87
TF-CNN	92.38	85.17	93.25	90.00
TF-RNN	93.74	87.87	90.35	86.48
WV-DNN	88.00	94.03	92.93	91.27
WV-CNN	89.83	93.77	94.24	91.96
WV-RNN	90.69	85.72	92.71	93.57
ASASM-HHODL	95.50	95.19	95.36	94.33
SAODL-SDC	97.00	97.00	97.00	97.00

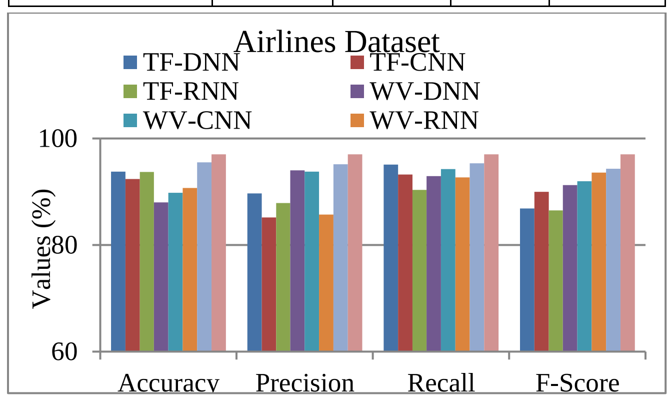

Fig. 13 Comparative outcome of SAODL-SDC technique on Airlines dataset

Table 7 and Figure 14 show the overall CT outputs of the SAODL-SDC approach on the Airlines dataset. The results specify that the TF-RNN and WV-CNN approaches have exposed the worst performance. However, the TF-DNN, TF-CNN, WV-DNN, and WV-RNN methods have established reasonable results. Although the ASASM-HHODL approach has reached considerable performance, the SAODL-SDC approach gains enhanced performance with a smaller CT of 1.72s.

Table 7. CT outcome of SAODL-SDC approach with existing method on Airlines dataset

Airlines Dataset	
Methods	**CT (sec)**
TF-DNN	6.19
TF-CNN	5.39
TF-RNN	3.83
WV-DNN	3.67
WV-CNN	6.53
WV-RNN	4.39
ASASM-HHODL	5.62
SAODL-SDC	1.72

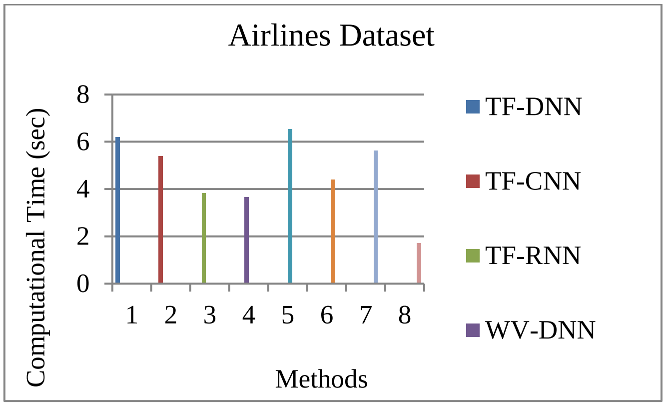

Fig. 14 CT outcome of SAODL-SDC technique on Airlines dataset

Thus, the SAODL-SDC technique has improved the ability to recognize sentiments on SM.

5. Conclusion

This study introduces a new SAODL-SDC technique on SM corpus. The presented SAODL-SDC technique primarily intends to recognize the class of opinions in SM data. In the SAODL-SDC technique, a multi-faceted approach begins with data preprocessing and BoW feature extraction. The SAODL-SDC technique involves the design of a CLSTM-AE network for sentiment detection. The hyperparameter tuning process using SAO is utilized to improve the effectualness of the CLSTM-AE methodology. The SAODL-SDC approach is examined under Sentiment140 and Airlines datasets. The performance validation of the SAODL-SDC approach portrayed superior accuracy values of 94.28% and 97.00% over existing techniques. The limitations of the SAODL-SDC approach include reliance on a specific dataset, which may limit the generalizability of the results across diverse domains or languages.

Furthermore, the accomplishment of the sentiment detection model can be influenced by noisy, unstructured, or ambiguous data commonly found in SM texts. The study also needs to address the challenge of real-time SA, which is crucial for dynamic applications. Moreover, while the approach effectually classifies sentiments, it may have difficulty detecting subtle or context-dependent sentiments, such as sarcasm or irony. In future work, improving the model's capability to handle diverse languages, integrating multi-modal data (e.g., images, videos), and improving contextual understanding for more accurate sentiment detection could be explored. Additionally, real-time processing and scalability to handle massive datasets will be significant for broader applications.

References

[1] Rezaul Haque et al., “Multi-Class Sentiment Classification on Bengali Social Media Comments Using Machine Learning,” *International Journal of Cognitive Computing in Engineering*, vol. 4, pp. 21-35, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[2] Hameedur Rahman et al., “Multi-Tier Sentiment Analysis of Social Media Text Using Supervised Machine Learning,” *Computers, Materials & Continua*, vol. 74, no. 3, pp. 5527-5543, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[3] Bador Al Sari et al., “Sentiment Analysis for Cruises in Saudi Arabia on Social Media Platforms Using Machine Learning Algorithms,” *Journal of Big Data*, vol. 9, no. 1, pp. 1-28, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[4] Xianyong Li et al., “A Novel Deep Learning-Based Sentiment Analysis Method Enhanced with Emojis in Microblog Social Networks,” *Enterprise Information Systems*, vol. 17, no. 5, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[5] Lal Khan et al., “Deep Sentiment Analysis using CNN-LSTM Architecture of English and Roman Urdu Text Shared in Social Media,” *Applied Sciences*, vol. 12, no. 5, pp. 1-18, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[6] Najla M. Alnaqbi, and Walaa Fouda, “Exploring the Role of ChatGPT and Social Media in Enhancing Student Evaluation of Teaching Styles in Higher Education Using Neutrosophic Sets,” *International Journal of Neutrosophic Science*, vol. 20, no. 4, pp. 181-190, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[7] Sayyida Tabinda Kokab, Sohail Asghar, and Shehneela Naz, “Transformer-Based Deep Learning Models for the Sentiment Analysis of Social Media Data,” *Array*, vol. 14, pp. 1-12, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[8] Nirmal Varghese Babu, and E. Grace Mary Kanaga, “Sentiment Analysis in Social Media Data for Depression Detection Using Artificial Intelligence: A Review,” *SN Computer Science*, vol. 3, pp. 1-20, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[9] S.V. Praveen, and Vajratiya Vajrobol, “Understanding the Perceptions of Healthcare Researchers Regarding ChatGPT: A Study Based on Bidirectional Encoder Representation from Transformers (BERT) Sentiment Analysis and Topic Modeling,” *Annals of Biomedical Engineering*, vol. 51, pp. 1654-1656, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[10] Amjad Iqbal et al., "Sentiment Analysis of Consumer Reviews Using Deep Learning," *Sustainability*, vol. 14, no. 17, pp. 1-19, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[11] Gadde Ashok, N. Ruthvik, and G. Jeyakumar, "Optimizing Sentiment Analysis on Twitter: Leveraging Hybrid Deep Learning Models for Enhanced Efficiency," *International Conference on Distributed Computing and Intelligent Technology*, Cham, Switzerland: Springer Nature Switzerland, pp. 179-192, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[12] Muhammet Sinan Başarslan, and Fatih Kayaalp, "MBi-GRUMCONV: A Novel Multi Bi-GRU and Multi CNN-Based Deep Learning Model for Social Media Sentiment Analysis," *Journal of Cloud Computing*, vol. 12, no. 1, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[13] Thilak Bellam, and P. Lakshmi Prasanna, "Novel DeepLearning based Sentimental Approach to Identifying the Fake News in Social Networking Media Based Smart Application," *Measurement: Sensors*, vol. 33, pp. 1-8, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[14] Ahmed Alsayat, "Improving Sentiment Analysis for Social Media Applications using an Ensemble Deep Learning Language Model," *Arabian Journal for Science, Engineering*, vol. 47, no. 2, pp. 2499-2511, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[15] Mohsan Ali et al., "Social Media Content Classification and Community Detection Using Deep Learning And Graph Analytics," *Technological Forecasting and Social Change*, vol. 188, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[16] Fazeel Abid et al., "Sentiment Analysis in Social Internet of Things Using Contextual Representations and Dilated Convolution Neural Network," *Neural Computing and Applications*, vol. 36, pp. 12357-12370, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[17] Hanan T. Halawani et al., "Automated Sentiment Analysis in Social Media Using Harris Hawks Optimization and Deep Learning Techniques," *Alexandria Engineering Journal*, vol. 80, pp. 433-443, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[18] Regina Anatasya Rudiyanto, and Erwin Budi Setiawan, "Sentiment Analysis Using Convolutional Neural Network (CNN) and Particle Swarm Optimization on Twitter," *Journal of Computer Science and Technology*, vol. 9, no. 2, pp. 188-195, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[19] Guangyu Mu et al., "DIBTBL: A Novel Text Sentiment Analysis Model Based on an Improved Swarm Intelligence Algorithm and Deep Learning," *IEEE Access*, vol. 12, pp. 158669-158684, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[20] Vipin Jain, and Kanchan Lata Kashyap, "Ensemble Hybrid Model for Hindi COVID-19 Text Classification with Metaheuristic Optimization Algorithm," *Multimedia Tools and Applications*, vol. 82, no. 11, pp. 16839-16859, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[21] Ramesh Vatambeti et al., "Twitter Sentiment Analysis on Online Food Services Based on Elephant Herd Optimization with Hybrid Deep Learning Technique," *Cluster Computing*, vol. 27, no. 1, pp. 655-671, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[22] Praloy Biswas, A. Daniel, and Subhrendu Guha Neogi, "Spider Monkey Optimization with Deep Learning-based Hindi Short Text Sentiment Analysis," *Journal of Intelligent Systems & Internet of Things*, vol. 12, no. 1, pp. 97-109, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[23] Badriyya B. Al-Onazi et al., "Modified Seagull Optimization with Deep Learning for Affect Classification in Arabic Tweets," *IEEE Access*, vol. 11, pp. 98958-98968, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[24] Punithavathi Rasappan et al., "Transforming Sentiment Analysis for E-Commerce Product Reviews: Hybrid Deep Learning Model with an Innovative Term Weighting and Feature Selection," *Information Processing & Management*, vol. 61, no. 3, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[25] Surajit Giri et al., "SMS Spam Classification–Simple Deep Learning Models with Higher Accuracy using BUNOW and Glove Word Embedding," *Journal of Applied Science and Engineering*, vol. 26, no. 10, pp. 1501-1511, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[26] Vasily D. Derbentsev et al., "A Comparative Study of Deep Learning Models for Sentiment Analysis of Social Media Texts," *CEUR Workshop Proceedings*, pp. 168-188, 2023. [Google Scholar] [Publisher Link]

[27] Dipanwita Thakur et al., "ConvAE-LSTM: Convolutional Autoencoder Long Short-Term Memory Network for Smartphone-Based Human Activity Recognition," *IEEE Access*, vol. 10, pp. 4137-4156, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[28] Heming Jia et al., "Improved Snow Ablation Optimizer with Heat Transfer and Condensation Strategy for Global Optimization Problem," *Journal of Computational Design and Engineering*, vol. 10, no. 6, pp. 2177-2199, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[29] Sentiment140 Dataset with 1.6 Million Tweets, Kaggle, 2017. [Online]. Available: https://www.kaggle.com/datasets/kazanova/sentiment140

[30] Twitter US Airline Sentiment, Kaggle, 2019. [Online]. Available: https://www.kaggle.com/datasets/crowdflower/twitter-airline-sentiment