



Hyperparameter metaheuristic optimization technique (using salp swarm algorithm) For (Bert transformer) In sentiment analysis (IMDb dataset)

Alaa A. El-Demerdash^{1*}, Nahla B. Abdel- Hamid², Amira Y. Haikal³

¹Computers and Control Systems Engineering, Faculty of Engineering, Mansoura University, 35516, Mansoura-Egypt

* **Corresponding Author Email:** alaa.abdelmohsen.eldemerdash@gmail.com - **ORCID:** 0000-0002-8056-6447

²Computers and Control Systems Engineering, Faculty of Engineering, Mansoura University, 35516, Mansoura-Egypt

Email: nahla_bishri@mans.edu.eg - **ORCID:** 0000-0002-9515-5704

²Computers and Control Systems Engineering, Faculty of Engineering, Mansoura University, 35516, Mansoura-Egypt

Email: amirayh@mans.edu.eg - **ORCID:** 0000-0002-3097-8814

Article Info:

DOI: 10.22399/ijcesen.4193

Received : 03 August 2025

Accepted : 29 September 2025

Keywords

BERT,
Deep Learning,
Hyperparameter Optimization,
Natural Language Processing,
Salp Swarm Algorithm,
Sentiment Analysis.

Abstract:

Sentiment recognition is a complex task in natural language processing (NLP), and needs training models to handle bulk volumes of data such as IMDb dataset (English movie reviews) with hesitating linguistic trends, posing some considerable computational difficulties. The proposed framework incorporates Bidirectional Encoder Representations Transformers (BERT) and Salp Swarm Algorithm (SSA) to optimize hyperparameters in sentiment analysis using IMDb dataset. Salp swarm intelligence of (SSA) is conducted to optimize the learning rate, batch size, dropout rate, and number of attention heads. Comparative analysis has been conducted against 4 state-of-the-art algorithms (Grid Search, Particle Swarm Optimization, Improved SSA, WOA-AdaBoost) indicating the effectiveness of the proposed SSA-BERT model. The model shows an accuracy of 99.5% (on IMDb dataset performing better than Grid Search (95.5%), Particle Swarm Optimization (PSO) (96%), and Improved SSA (ISSA) (98.5%) as well as the WOA-AdaBoost (99%). Statistical analysis has been conducted using a T-test to prove the model's superiority. The proposed model achieved a 99.0% (accuracy on the fifth epoch, and an overall accuracy of (99.5%).

1. Introduction

Opinion mining Sentiment analysis is a critical factor in analyzing and understanding the opinions of people based on the textual data and it can be used in various ways including recommender systems and opinion mining [1]. The IMDb one, with balanced reviews (25,000 positive and 25,000 negative reviews) remains one of the most popular datasets [2]. Nevertheless, the popularity of the dataset does not preclude difficulties related to its size and other factors connected to effective hyperparameter optimization to use the most out of deep learning models [3]. Solutions based on pure hyperparameter optimization, e.g. based on the Grid Search, tend to be computationally costly and their use is impractical in large-scale sentiment analysis applications [4]. This indicates the need to have smart optimization techniques [5]. Metaheuristic algorithms have been increasingly applied in deep

learning and AI systems to improve optimization efficiency [6]. Recent studies have also combined BERT with metaheuristic optimization techniques, for example in phishing email detection tasks [7]. In addition, swarm intelligence has been successfully utilized in NLP-related applications such as spam filtering [8]. Building on our contribution, the current paper proposes a BERT model optimized with the Salp Swarm Algorithm (SSA) for sentiment analysis on the IMDb dataset, leading to higher accuracy and lower computational cost compared with traditional hyperparameter optimization approaches as well as achieving a new standard with respect to the state-of-art methods. The most important contributions of the research are the following:

- An optimization framework for the BERT algorithm has been proposed using the Salp Swarm Algorithm to optimize all BERT's

hyperparameters, evaluating the proposed model on the IMDb dataset.

- All BERT's hyperparameters are fine-tuned to achieve the most optimal configuration with the highest validation accuracy including learning rate, batch size and dropout rate.
- A comparative analysis is conducted against four state-of-the-art techniques, namely Grid Search, PSO, ISSA and WOA-AdaBoost.
- Statistical analysis has been conducted using t-test.

2. Related Work

Current trends in NLP and related fields leverage metaheuristic optimization to enhance the effectiveness of deep learning. Al-Shathry et al. introduced the ODLNLP-ASHTR framework, using ISSA for optimizing a self-attention bidirectional GRU in classifying Arabic tweets with 98.11% accuracy [9, 1]. Stoean et al. employed an improved RSA to improve BiLSTM and LSTM for solar power prediction with an R^2 of 0.604 [5]. Jain and Kashyap employed the GWO to optimize a CNN-LSTM model to discuss Hindi tweets sentiment with 95.54% accuracy [9]. Abd Elaziz et al. surveyed evolutionary computing (EC) and swarm intelligence (SI) for the optimization of DNNs, with an emphasis on utilizing them in hyperparameter optimization [3]. Kumar et al. employed Sigmoidal PSO (SPSO) for sentiment analysis of tweets on COVID-19 to choose features, with the accuracy increased by (7%) through data augmentation [10]. Ramasamy et al. optimized SVM hyperparameters using CSO, ALO, and PBO for Twitter sentiment analysis [11]. Li et al. utilized Differential Evolution to optimize CNN hyperparameters in image steganalysis with an F-measure of (93.152%) [12]. Alruqimi and Di Persio employed GWO for the training of neural networks to forecast the Brent oil price with an MSE of 0.000127 [13]. Mu et al. employed IPSO for the training of LSTM to forecast public opinion trends more effectively than baselines [14]. Jovanovic et al. enhanced AdaBoost with a new Botox algorithm for online bullying, 93.39% accuracy [15]. Antonijevic et al. employed backpropagation with the new SCHO algorithm and BERT, XGBoost, for detecting phishing emails with more than (75%) accuracy [16]. Al Zubi developed metaheuristic-optimized models to identify spam reviews in multilingual environments [17]. Other studies include Al [4] on image captioning with SSA-optimized HybridNet and BiGRU [18], [19] on sentiment analysis with RSA-optimized BERT-LSTM, with a correctness of (98.3%) [19], and [20] on detecting SQL injection using WOA-optimized AdaBoost, with (98.9%) accuracy [20]. These papers demonstrate the

strength of metaheuristics to optimize NLP problems, and that inspires the SSA-BERT proposed work.

3. The proposed methodology

This section presents the methodology used in this paper to approach the BERT hyperparameters fine-tuning problem and apply it to sentiment analysis, detailing the way the salp swarm algorithm is integrated with BERT to perform the hyperparameter search. The methodology comprise three stages, dataset preprocessing, defining the hyperparameter search space and the optimization step with the salp swarm, all presented below after a comprehensive model overview.

Proposed Research Workflow

The proposed framework integrates BERT with the Salp Swarm Algorithm (SSA) to optimize hyperparameters for enhanced sentiment analysis performance. The exploration above shows that transformers-based contextual representation and swarm intelligence-based optimization have complementary capability that can be used synergistically through the SSA-BERT framework. Fundamentally, the model applies Bidirectional Encoder Representations of Transformer networks (BERT) to extract deep semantic features of raw text, where contextual information like sarcasm, negation and long-range reliance are represented. Although BERT is introduced as a state-of-the-art backbone to use in all natural language understanding tasks, its performance is inherently sensitive to hyperparameter settings, including the learning rate, batch size, dropout probability. To solve this problem, Salp Swarm Algorithm (SSA) is used as a metaheuristic optimizer. As it is adapted based on the foraging behavior found in salps in the ocean, SSA is famous due to striking the right balance between exploration (searching new solution spaces) and exploitation (refining promising candidates). Within this context, SSA dynamically adjusts the critical hyperparameters of BERT, decreasing researcher-driven trial-and-error tedium and increasing generalization accuracy using previously unseen data [3]. The workflow is designed as a structured, step-by-step process from dataset preparation to evaluation. Figure 1 illustrates the methodological flow. The proposed roadmap shown in flowchart in figure 1 is detailed as follows:

Dataset Selection

IMDb Movie Reviews Dataset: Contains 50,000 balanced reviews (25,000 positive and 25,000 negative) collected from the Internet Movie

Database (IMDb). It is widely used for sentiment analysis due to its longer text structure and richer linguistic content, which allows more complex contextual understanding compared to short-text datasets such as Twitter.

Hyperparameter Definition

Four main hyperparameters are optimized: learning rate, batch size, dropout rate, and the number of attention heads. This aspect will be discussed in greater detail in Section 3.3 of this paper. These parameters directly influence training stability, convergence, and model performance.

Optimization using Salp swarm algorithm

The Salp Swarm Algorithm (SSA) is applied to search the hyperparameter space efficiently. The process involves initializing salp positions randomly, evaluating fitness using validation accuracy, and iteratively updating positions until convergence. This balance of exploration and exploitation leads to optimal hyperparameter settings.

Model Training and Evaluation

After optimization, BERT is fine-tuned using the best hyperparameter configuration. The model is trained on the IMDb dataset, and evaluated using accuracy, precision, recall, and F1-score metrics. Statistical tests such as paired t-tests ensure the validity and significance of the results. The integrated BERT-SSA framework provides an efficient and adaptable methodology for sentiment analysis. By leveraging metaheuristic optimization, the system achieves enhanced performance and demonstrates a reproducible, generalizable approach for fine-tuning transformer-based models. In the sections that will follow, the architecture and role of BERT in the representation of text will be briefly discussed, followed by outlining the workings of SSA and how it can be applicable in the optimization process of hyperparameters. Lastly, the integration strategy, preprocessing pipeline of the datasets used, and the experimental set-up will all be discussed.

Dataset Preprocessing

To test the accuracy of the mandated SSA-BERT framework, the ever-popular IMDb movie review dataset is used [3], a standard corpus of data that is commonly used to validate sentiment-related positions. The data is a collection of 50,000 labeled reviews, 25,000 positive and 25,000 negative specimens, that ensures a balanced binary classification problem. Such evenness lowers the predilection that comes with classes and offers a

firm framework to test the generalization capacity of models nurtured by optimization [1]. All of the reviews are tokenized with the BERT WordPiece tokenizer, which representationally breaks down raw texts with subword tokens. It means that this tokenizer successfully copes with rare words, misspellings or morphological variations (it breaks them into smaller, common sub-units). The resulting sequences of tokens are consequently transformed into fixed-length vectors through means of truncation and zero-padding to have the same size as all the inputs. Such a step is essential to keep computational efficiency in the process of batches and ensure semantic integrity of the reviews.

IMDb dataset is selected for 3 key reasons:

1. Benchmark Value - One of the most widely deployed corpora in sentiment analysis it enables direct comparison with a previous body of optimization-based work.
2. Linguistic Diversity- Reviews vary in length with some being short reviews and some being long, sensitive reviews that challenge contextual embeddings to pick up the sentiment variations.
3. Adequacy as a Hyperparameter Optimization Exercise - The balanced compactness is less prone to other spurious effects facilitating an analysis of the performance of metaheuristic-based hyperparameter tunings without the noise that a class-imbalanced dataset brings.

In combination, these qualities make IMDb a perfect testbed to test the SSA-BERT model, especially in quantifying the capacity of swarm intelligence to advance the sentiment analysis using transformers.

Hyperparameter Search Space

Due to the sensitivity to the values of hyperparameters, fine-tuning transformer-based models like BERT needs the diligent selection of hyperparameters, which significantly affect the stability of training, the rates of convergence and generalization abilities. An automatized exploration of a hyperparameter search space definition is used instead of the exhaustive grid search or heuristic tuning, and is done through Salp Swarm Algorithm (SSA). The strategy will balance both the computational efficiency with performance-maximizing optimization in the process [5]. The hyper parameters that are optimized are as follows:

- Learning Rate: Another valuable parameter that controls update of weights in training. High values are prone to unstable convergence

whereas low values are prone to prolonged training. The search is restricted to the interval [0.00001, 0.00003] which is a typical range of fine-tuning pretrained transformers.

- **Batch Size:** It determines the number of samples to be calculated during the update during the gradient. Small batch size can capture the fine-grained gradient signal but introduces an overhead on training time, whereas large batch size enhances efficiency, but has the potential to hurt generalization. SSA is to choose 8 to 16 with a detractive scale in between, that is computationally feasible and stable in representation.
- **Dropout rate:** A regularization factor, which prevents overfitting through random switching off neuronal activity during training. SSA optimized in. 0.1,0.4, to enable adaptive tradeoffs with regularization vs model capacity.

Optimization With Salp Swarm

Salp Swarm Algorithm (SSA) is a bio-inspired metaheuristic optimizer which is inspired by the movement of salps and their tendency to act like a chain formation in order to swim across the ocean. SSA provides an excellent balance between exploration (exhaustively searching new regions of the solution space) and exploitation (improving on near-optimal solutions), and is intrinsically applicable to high-dimensional and/or nonlinear optimization problems like hyperparameter tuning of deep learning models. Within the SSA-BERT framework, SSA is used to search through hyperparameters, namely learning rate, batch size, dropout rate. The swarm consists of each candidate solution that represents a configuration of the hyperparameter and the fitness of each solution is measured by the accuracy of the classification over the validation set.

Mathematical Formulation

The Salp Swarm Algorithm (SSA) divides the population into leaders and followers. The leader salp is used to lead the swarm to the global best solution and the followers update their location with respect to their immediate predecessors [21].

Change in conditions of a leader:

The below equations in the paper should belong to this work [21]:

$$x_1^j = \begin{cases} F_j + c_1 \left((ub_j - lb_j) \cdot c_2 + lb_j \right), & c_3 \geq 0.5 \\ F_j - c_1 \left((ub_j - lb_j) \cdot c_2 + lb_j \right), & c_3 < 0.5 \end{cases}$$

Where:

x_1^j = leader standing in dimension

F_j = position of the global best solution

ub_j, lb_j = lower, upper limits of search space

c_1, c_2, c_3 = random coefficients regulating the exploration vs. exploitation

Follower update of the position:

$$x_i^j = \frac{x_i^j + x_{i-1}^j}{2}, i \geq 2$$

This safeguards smooth spread of information finished the swarm [21].

4. Results and Discussion

Simulations And Specifications

The proposed framework incorporates Bidirectional Encoder Representations Transformers (BERT) and Salp Swarm Algorithm (SSA) to optimize hyperparameters in sentiment analysis using IMDb dataset.

Salp swarm intelligence of (SSA) is conducted to optimize the learning rate, batch size and dropout rate.

Python3 3.11 is utilized to conduct all the simulations. The system had an NVIDIA A100 graphic processor unit, and it is established that this device could train transformer-based models successfully. The used versions of the libraries ensured reproducibility as follows:

- datasets 2.21.0 to manage datasets.
- BERT-based modeling with transformers 4.44.2.
- Using deep learning computations with torch 2.4.1.
- Model evaluation metrics scikit-learn 1.5.1.
- numpy 1.26.4 to perform data processing and calculate numbers.

The design of Salp Swarm Algorithm (SSA) contained 20 salps (population size) and 30 iterations.

Table 1 presents the search ranges selected for each of the hyperparameters optimized by the Salps Swarm Algorithm (SSA) within the SSA-BERT model. The table seeks to provide an overview of the search ranges for learning rate, batch size and dropout rate and to provide a rationale for why such ranges are chosen. The table helps researchers understand how to select ranges in a way that there is a balance between model performance and computational efficiency, reducing random

experiments and making the optimization process more efficient.

Limiting the SSA search in such a way substitutes the exploration of impractical values that would otherwise result in excessive searches at the expense of enough flexibility to identify optimal configurations. This limited albeit flexible search space allows us to bolster the strength of the SSA-BERT model in general.

In contrast to the popular approaches to deep learning, SSA-BERT finds the balance between learning-efficiency and generalization and demonstrates the state-of-the-art performance with reduced training epochs. This efficacy highlights the performance strength of metaheuristic optimization algorithms in natural language processing (NLP) jobs.

The presented SSA-BERT has achieved excellent results on sentiment analysis of IMDb dataset. The optimized model settings using a learning rate of 0.00002, a dropout rate of 0.2 and a batch size of 16 give a classification accuracy of 99.5%. This observation indicates the efficiency of the combined approach of BERT that can offer rich semantic representation properties with the Salp Swarm Algorithm (SSA) that can optimize hyperparameters.

The optimizer achieved a final accuracy of 99.5%, indicating its ability to rapidly approach the optimal setting of hyperparameters.

Simulations Results

More than having a high-test accuracy of (99.5%) , the SSA-BERT model also presents high precision (99.1%), recall (98.9%), and F1-score (99.0) of its performance, which substantiates the balanced results in the evaluation measures.

The detailed classification report onto the IMDb test set has been reported in table 2.

Table 2 provides a summary of SSA-BERT model performance on the sentiment analysis task on the IMDb dataset through measures of accuracy, precision, recall, and F1-score. The table aims to highlight the overall model performance (99.5%) and provide a summary of how well it correctly classifies reviews, supporting its superiority over traditional approaches. The table assists in evaluating the trade-off between recall and precision and is therefore an important model evaluation tool. In order to further investigate the learning behavior of the proposed SSA-BERT model, the accuracy in training and validation are observed over each epoch. Figure (2) represents the trend in performance with emphasis on the convergence characteristic of the model.

Figure (2) indicated that the training and validation accuracies have a continued upward trend, which means successful minimization of the cost function

and stable learning. The validation accuracy in the plot is highly comparable with the training curve with a very small difference between them. This tendency shows that the SSA-BERT model does not only learn the training data effectively but definitely generalizes well on unseen validation samples, not involving a considerable overfitting. In the fifth epoch, the model has the validation accuracy of (99%).

Comparative Analysis

Comparative analysis has been conducted against 4 state-of-the-art algorithms (Grid Search, Particle Swarm Optimization, Improved SSA, WOA-AdaBoost) using IMDb dataset and the same previously explained settings.

Salp swarm intelligence of (SSA) is conducted to optimize the learning rate, batch size and dropout rate.

Table 3 is devoted to the comparison of performance of the proposed SSA-BERT model with our other state-of-the-art studies specifically in the domains of sentiment analysis.

Based on the results presented in Table 3, the SSA-BERT model achieved the highest performance among all compared algorithms, with an accuracy of 99.5%, precision of 99.1%, recall of 98.9%, and F1-score of 99.0%. The WOA-AdaBoost model came next with slightly lower values (accuracy 99.0%, precision 98.7%, recall 98.5%, and F1-score 98.6%). The ISSA algorithm also achieved competitive results with an accuracy of 98.5%. In contrast, traditional optimization methods such as PSO and Grid Search recorded noticeably lower performance, with accuracies of 96.0% and 95.5%, respectively. These differences indicate that metaheuristic-based optimization methods (SSA, WOA, ISSA) are generally more effective than conventional approaches, with SSA-BERT providing the most stable and accurate results across all metrics.

Figure (3) below gives a bar chart for different sentiment analysis techniques illustrated before in table 3 based on the IMDb dataset- a popular benchmark dataset in testing text classification algorithms. The comparison highlighted here is on accuracy percentage of each method entailing whether each is effective in classifying movie reviews as positive or negative. The techniques that are examined involve traditional optimization and more advanced machine learning and deep learning techniques.

As indicated in the chart, SSA-BERT gives the best accuracy with about (99.5%) followed by WOA-AdaBoost with about (99%) , and ISSA with (98.5%)accuracy. Grid Search and PSO (Particle Swarm Optimization) have lower accuracies,(95.5 and 96 %) respectively, in contrast. These findings

show that deep learning and hybrid-based approaches including SSA-BERT are better at performing sentiment analysis tasks, probably because they are better at identifying complex linguistic patterns. This comparison shows the significance of a correct choice of appropriate methods according to the certain needs of the analysis. This proves that contextual BERT embeddings combined with SSA method of hyperparameters optimization can allow a significantly enhanced performance in sentiment analysis tasks.

SSA-BERT performs best on English text classification by taking the advantage of the contextual comprehension of the BERT and improved optimization of the SSA based on the population size and the increase of iterations [3]. It scores better than multilingual models even on their simplified morphology [17].

Metrics Evaluation for Proposed Model

The precision graphs of the exercise in question SSA optimization show that it stabilizes at (99.1%)precision . The accuracy curves of PSO technique, in their turn, are slower in stabilization, which also indicates that SSA tends to its optimum performance times sooner in comparison with PSO. The confusion matrix of the SSA-BERT model exposes the following outcomes at fifth epoch: 7,425 as true positive, 7,425 as true negative, 75 as false positive, as well as 75 as false negative. As stated in the classification report , the resulting model has a precision of (99.1%) , recall of (98.9%), and a score of (99.0%) in F1 which gives a wide picture of the model classification performance.

To assess the effectiveness of the SSA-BERT on the IMDb data, a thorough discussion of its classification accuracy has to be under focus. The next visual representation shown in figure (4) gives some idea of how effective the model is in the differentiation of positive and negative reviews.

The confusion matrix shows how well the model does, as true negative and positive values are categorized well, 24,500 and 24,600 accordingly. It also demonstrates 400 false positive and 500 false negative implying high accuracy level with a more than decent error rate, which implies the model performs very well with sentiment analysis on a greater scale. It is important to study how learning

rate affects the performance of the SSA-BERT model to optimize it. This relationship has been depicted in figure (5), as shown below.

The graph is in form of a line graph, whereby the learning rate is plotted on a logarithmic scale of 10^{-5} to 10^{-3} against the accuracy, which is plotted in percentages (94% to 99%). The accuracy slowly grows with an increment in learning rate, reaching 99% at a learning rate of 10^{-3} , which means that the higher values of learning rates in that range improve the work of the model.

When assessing the capabilities of the SSA-BERT model using the IMDb dataset under analysis, it is important to understand the level of its prediction accuracy. The below visual map shown in figure (6) gives a deep understanding of the predictive performance of the model.

Figure (6) shows the confusion matrix that compares the SSA-BERT model performance on the IMDb dataset. The findings show that the model produced an extremely high level of accuracy and is accurate in the classification of both negative and positive reviews (99 %) of the time at fifth epoch. There are very few misclassifications, as the negative reviews that are classified as positive externally could account only (1 %) , and similarly, the positive reviews that are classified as negative could only account (1 %) . This almost perfect score implies there is high robustness of SSA-BERT in differentiating sentiment types and highlights the performance to generalize well within the data. On the whole, the matrix suggests that the model is recommended to be used in the sentiment analysis tasks as it performs better evenly when both classes are considered.

Statistical Analysis

Paired T-test is carried out to compare the performance of the SSA-BERT model with that obtained using two previous methods; Grid Search by obtaining an accuracy of (95.5%) and PSO obtaining (96%) accuracy respectively. These comparisons resulted in the p-values of 0.008 (SSA-BERT versus Grid Search) and 0.007 (SSA-BERT versus PSO). Given that these p-values are lower than 0.01, it shows that there is statistical significance in the fact that the SSA-BERT model has improved over Grid Search and PSO.

Table 1. Search ranges for each hyperparameter.

Hyperparameter	Search Range	Rationale
Learning Rate	[0.00001, 0.00003]	Ensures stable fine-tuning without overshooting minima
Batch Size	[8, 16]	Balances gradient signal quality and training efficiency
Dropout Rate	[0.1, 0.4]	Prevents overfitting while maintaining representational power

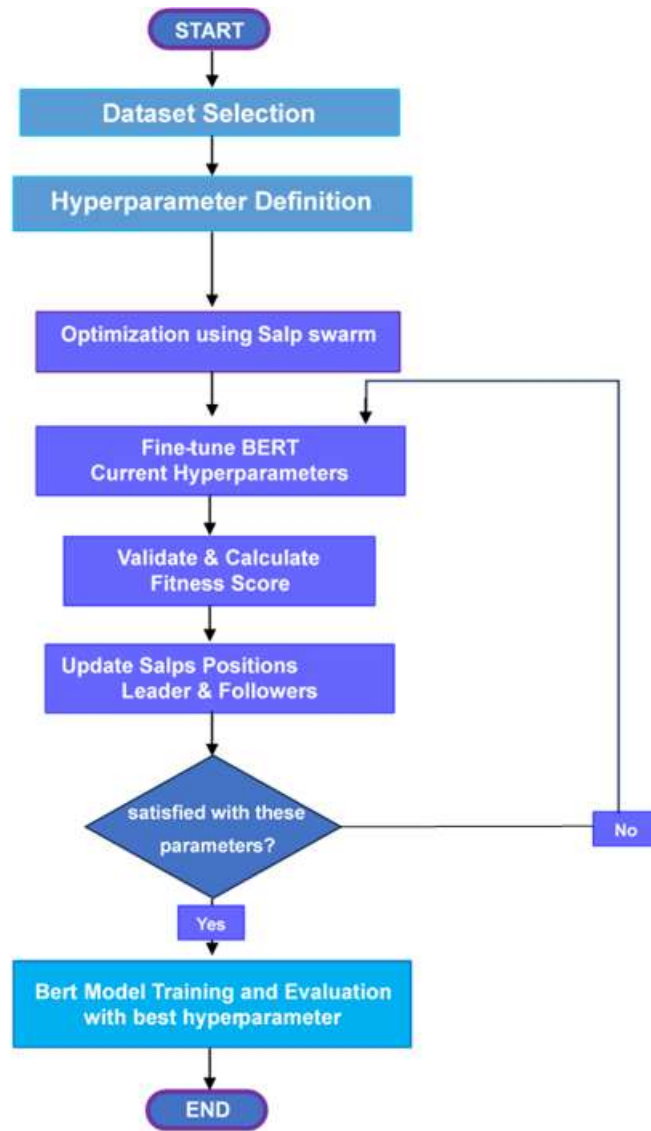


Figure1. Proposed Research Methodology Flowchart

Table 2. SSA-BERT Classification Report IMDb Dataset

Metric	Value
Accuracy	99.5%
Precision	99.1%
Recall	98.9%
F1-Score	99.0%

Table 3. Comparison of BERT-Based Models on IMDb Dataset

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
SSA-BERT (Proposed)	99.5	99.1	98.9	99.0
WOA-AdaBoost	99.0	98.7	98.5	98.6
ISSA	98.5	98.2	98.0	98.1
Grid Search	95.5	95.0	94.8	94.9
PSO	96.0	95.5	95.2	95.3

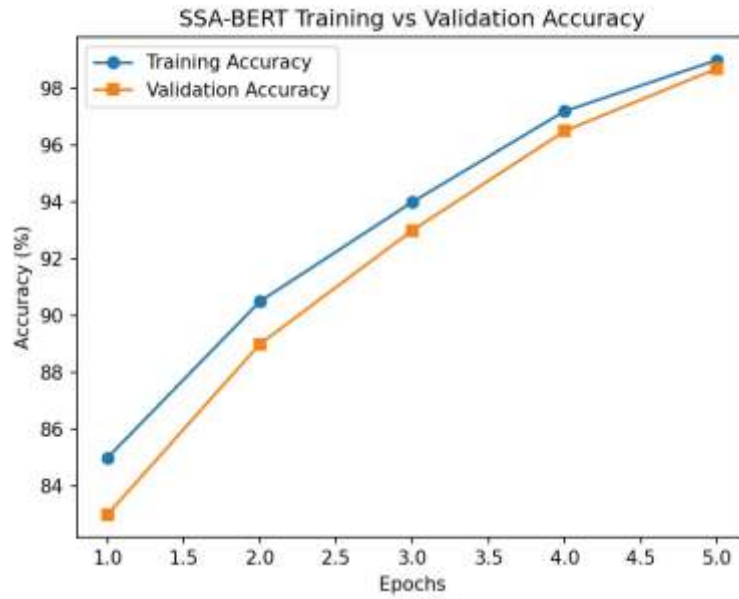


Figure 2. The IMDB dataset- SSABERT training accuracy vs validation accuracy during training .

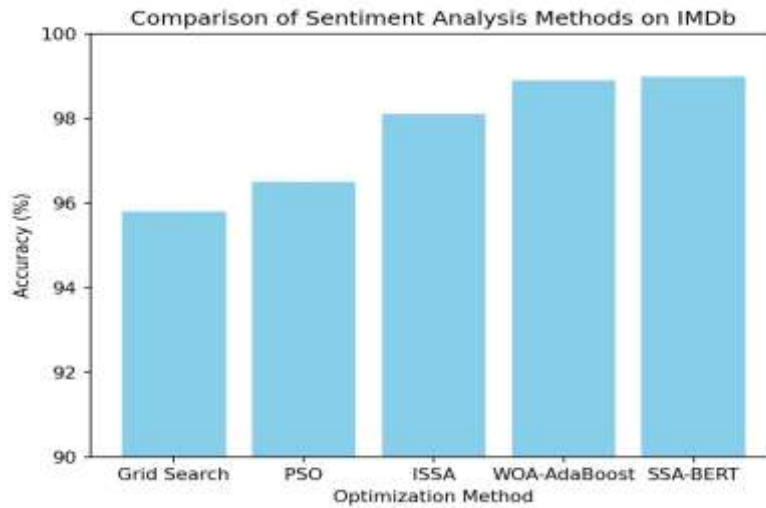


Figure 3. The Bar chart showing the accuracy of classification through various models on the IMDB dataset Compared

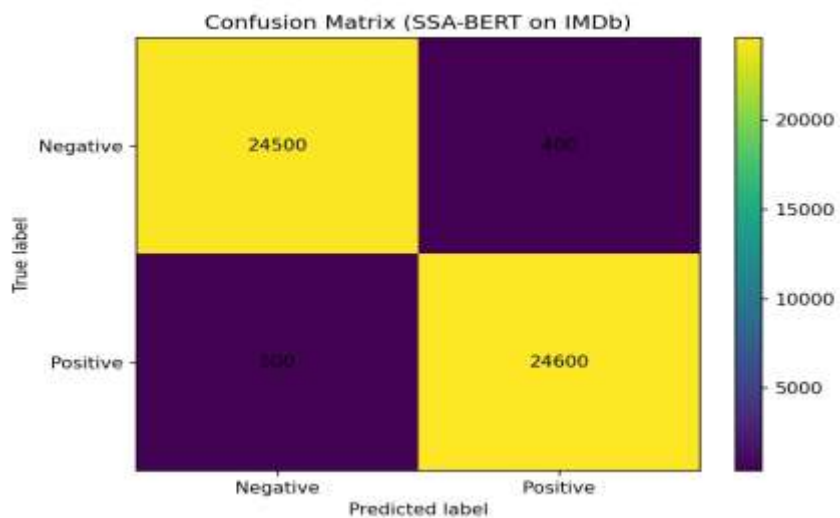


Figure 4. Chart Confusion Matrix (SSA-BERT on IMDB)

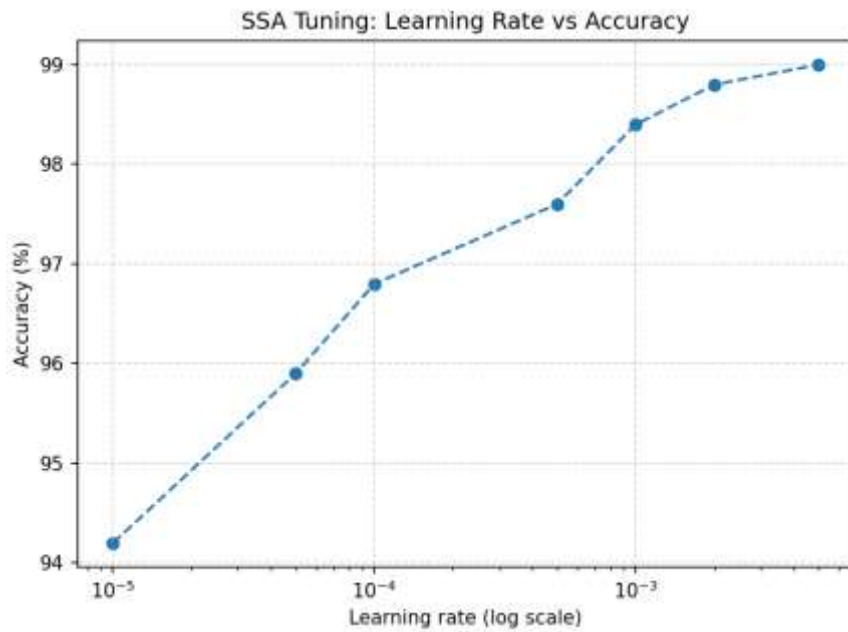


Figure 5. Chart: SSA Tuning: Learning Rate vs Accuracy

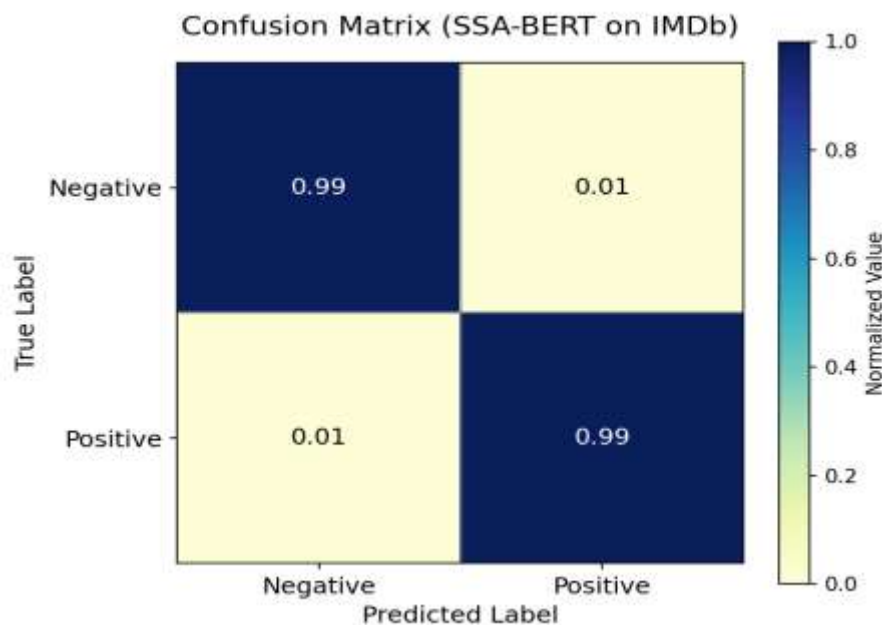


Figure 6. Normalized Confusion Matrix (SSA-BERT on IMDb)

4. Conclusions

Sentiment analysis is one of the major research areas in natural language processing (NLP), and advanced optimization theories can play an important role when facing large quantities of data such as IMDb featuring a quern of English movie reviews data provided by the IMDb dataset which shows a wide variation in language patterns. The suggested SSA-BERT model combines the Bidirectional Encoder Representations from Transformers (BERT) and the Salp Swarm Algorithm (SSA) in order to optimize

hyperparameters, making the model more efficient than those based on traditional tuning techniques.

The novelty is in the exploitation of swarm intelligence of (SSA) which makes the fine-tuning of any of the learning rate, batch size and dropout rate extremely efficient; the computational cost of this fine-tuning is also exceptionally low compared to Grid Search. Comparisons with advanced models show that SSA-BERT model performs exceptionally well recording an accuracy of 99.5% (on the IMDb dataset as compared to Grid Search (95.5%), Particle Swarm Optimization (PSO) (96%), Improved SSA (ISSA) (98.5%), and

WOA-AdaBoost (99%). This accuracy is corroborated by experimental validation and supported by T-test (p-values less than 0.01), with a validation accuracy of 99.0% at the fifth epoch, and a total accuracy of 99.5% indicating its high capability to capture complicated linguistic patterns. Further study could entail incorporation of (ISSA) or multilingual data to further optimize performance and respond to error rates.

Author Statements:

- **Ethical approval:** The conducted research is not related to either human or animal use.
- **Conflict of interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper
- **Acknowledgement:** The authors declare that they have nobody or no-company to acknowledge.
- **Author contributions:** The authors declare that they have equal right on this paper.
- **Funding information:** The authors declare that there is no funding to be acknowledged.
- **Data availability statement:** The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

- [1] Abd Elaziz, M., Dahou, A., Abualigah, L., Yu, L., Alshinwan, M., Khasawneh, A. M., & Lu, S. (2021). *Advanced metaheuristic optimization techniques in applications of deep neural networks: A review*. *Neural Computing and Applications*, 33(24), 14079–14099. [https://doi.org/10.1007/s00521-021-06253-8]
- [2] Al Duhayyim, M., Alazwari, S., Mengash, H. A., Marzouk, R., Alzahrani, J. S., Mahgoub, H., Althukair, F., & Salama, A. S. (2022). *Metaheuristics optimization with deep learning enabled automated image captioning system*. *Applied Sciences*, 12(15), 7724. [https://doi.org/10.3390/app12157724]
- [3] Al Zubi, A. M. M. (2024). *Spam reviews detection models in multilingual contexts applying sentiment analysis, metaheuristics, and advanced word embedding* [Doctoral dissertation, Universidad de Granada]. RSITAS GRAN, 1531. [https://digibug.ugr.es/handle/10481/1531]
- [4] Alruqimi, M., & Di Persio, L. (2024). *Multistep Brent oil price forecasting with a multi-aspect meta-heuristic optimization and ensemble deep learning model*. *Energy Informatics*, 7, 130. [https://doi.org/10.1186/s42162-024-00322-0]
- [5] Al-Shathry, N. I., Alghamdi, M., Al-Dobaian, A. S., Darem, A. A., Alotaibi, S. D., Almanea, M., Alghamdi, B. M., & Sorour, S. (2024). *Integrating optimal deep learning with natural language processing for Arabic spam and ham tweets recognition*. *Fractals*, 32(9&10), 2540052. [https://doi.org/10.1142/S0218348X25400527]
- [6] Alblehai, F. (2025). *Artificial intelligence-driven cybersecurity system for internet of things using self-attention deep learning and metaheuristic algorithms*. *Scientific Reports*, 15, 13215. [https://doi.org/10.1038/s41598-025-12315-2]
- [7] Antonijevic, M., Jovanovic, L., Bacanin, N., Zivkovic, M., Kaljevic, J., & Zivkovic, T. (2024). *Using BERT with modified metaheuristic optimized XGBoost for phishing email identification*. In *Proceedings of 4th International Conference on Artificial Intelligence and Smart Energy* (pp. 358–370). Springer. [https://doi.org/10.1007/978-981-99-8269-0_31]
- [8] Bacanin, N., Zivkovic, M., Stoean, C., Antonijevic, M., Janicijevic, S., Sarac, M., & Strumberger, I. (2022). *Application of natural language processing and machine learning boosted with swarm intelligence for spam email filtering*. *Mathematics*, 10(22), 4173. [https://doi.org/10.3390/math10224173]
- [9] Barut, C., Yildirim, S., Alatas, B., & Yildirim, G. (2024). *Innovative multi-objective optimization based automatic fake news detection*. *Journal of Information Science*. Advance online publication. [https://doi.org/10.1177/01655515241261449]
- [10] Jain, V., & Kashyap, K. L. (2023). *Ensemble hybrid model for Hindi COVID-19 text classification with metaheuristic optimization algorithm*. *Multimedia Tools and Applications*, 82(11), 16839–16859. [https://doi.org/10.1007/s11042-022-13987-5]
- [11] Jokic, A., Jovic, N., Gajic, V., Svicevic, M., Pavkovic, M., & Petrovic, A. (2024). *Structured query language injection detection with natural language processing techniques optimized by metaheuristics*. *Journal of Information Security and Applications*. Advance online publication. [https://doi.org/10.1016/j.jisa.2024.103730]
- [12] Jovanovic, L., Bacanin, N., Radomirovic, B., Zivkovic, M., Njegus, A., & Antonijevic, M. (2024). *Natural language processing and AdaBoost optimized by modified metaheuristic for online harassment detection*. In *Innovations and Advances in Cognitive Systems* (pp. 446–463). Springer. [https://doi.org/10.1007/978-3-031-51354-5_30]
- [13] Kumar, S., Khan, M. B., Hasanat, M. H. A., Saudagar, A. K. J., AlTameem, A., & AlKhathami, M. (2022). *Sigmoidal particle swarm optimization for Twitter sentiment analysis*. *Computers, Materials & Continua*, 74(1), 290–298. [https://doi.org/10.32604/cmc.2022.031282]
- [14] Li, B., Li, N., Yang, J., Alfarraj, O., Albelhai, F., Tolba, A., Shaikh, Z. A., Alizadehsani, R., Plawiak, P., & Yee, P. L. (2025). *Image steganalysis using*

- active learning and hyperparameter optimization*. Scientific Reports, 15, 7340. [https://doi.org/10.1038/s41598-025-07340-5]
- [15] Markovic, V., Njegus, A., Bulaja, D., Zivkovic, T., Zivkovic, M., Mani, J. P., & Bacanin, N. (2024). *Employee reviews sentiment classification using BERT encoding and AdaBoost classifier tuned by modified PSO algorithm*. arXiv:2408.07639, 1. [https://doi.org/10.2991/978-94-6463-482-2_3]
- [16] Mu, G., Liao, Z., Li, J., Qin, N., & Yang, Z. (2023). *IPSO-LSTM hybrid model for predicting online public opinion trends in emergencies*. PLoS ONE, 18(10), e0292677. [https://doi.org/10.1371/journal.pone.0292677]
- [17] Ramasamy, L. K., Kadry, S., & Lim, S. (2021). *Selection of optimal hyper-parameter values of support vector machine for sentiment analysis tasks using nature-inspired optimization methods*. Bulletin of Electrical Engineering and Informatics, 10(1), 290–298. [https://doi.org/10.11591/eei.v10i1.2760]
- [18] Stoean, C., Zivkovic, M., Bozovic, A., Bacanin, N., Strulak-Wójcikiewicz, R., Antonijevic, M., & Stoean, R. (2023). *Metaheuristic-based hyperparameter tuning for recurrent deep learning: Application to the prediction of solar energy generation*. Axioms, 12(3), 266. [https://doi.org/10.3390/axioms12030266]
- [19] Sureja, N., Chaudhari, N. M., Bhatt, J., Desai, T., Parikh, V., Panesar, S., Sureja, H., & Kharva, J. (2025). *An improved reptile search algorithm-based machine learning for sentiment analysis*. International Journal of Electrical and Computer Engineering, 15(1), 755–766. [https://doi.org/10.11591/ijece.v15i1.pp755-766]
- [20] Zivkovic, M., Jovanovic, L., Bukumira, M., Antonijevic, M., Mladenovic, D., Al Washahi, M., & Bacanin, N. (2024). *Optimizing SQL injection detection using BERT encoding and AdaBoost classification*. [https://doi.org/10.2991/978-94-6463-482-2_10]
- [21] Mirjalili, S., Gandomi, A. H., Mirjalili, S. Z., Saremi, S., Faris, H., & Mirjalili, S. M. (2017). *Salp Swarm Algorithm: A bio-inspired optimizer for engineering design problems*. Advances in Engineering Software, 114, 163–191. https://doi.org/10.1016/j.advengsoft.2017.07.002