



OPEN Sentiment classification via improved feature selection using Boolean operator-based particle swarm optimization

Harish Dutt Sharma¹, Raja Rao Budaraju², Neeraj Kumar³, Vijay Kumar³✉, Nemi Chandra Rathore⁴, Gaddam Ravindra Babu⁵ & Arvind Dhaka⁶✉

Sentiment analysis plays a vital role in understanding user opinions across digital platforms. However, accurate classification in high-dimensional text data remains a significant challenge, primarily due to irrelevant and redundant features. This paper introduces a novel Boolean Operator-based Particle Swarm Optimization (BOPSO) algorithm that enhances the feature selection process for sentiment classification. Unlike traditional PSO, the proposed BOPSO integrates Boolean logic operators (Adder, Subtractor, XOR) into the velocity and position update equations to natively handle binary feature inclusion decisions, improving both exploration and exploitation in the search space. The model is evaluated on nine benchmark sentiment datasets, using five filter-based objective functions: Chi-Square, Correlation, Gain Ratio, Information Gain, and Symmetrical Uncertainty. Classification performance is assessed using Naïve Bayes, SVM, and ANN classifiers. The results demonstrate that BOPSO achieves an average accuracy improvement of 1.8% to 4.5% over state-of-the-art optimization techniques, including DE, GWO, ABC, and CS. Specifically, BOPSO with ANN achieves up to 100% accuracy on the laptop dataset and consistently outperforms others in precision, recall, and F1-score. This study confirms that BOPSO not only reduces feature dimensionality effectively but also improves sentiment classification performance significantly across domains.

Keywords Boolean operated PSO, Sentiment classification, Boolean operators

The exponential increase in digital documents in the modern era has enabled information mining with the aid of the most recent IT-focused methodologies feasible. In particular, text mining is essential for obtaining important information from these papers. Sentiment analysis, which is another term for opinion mining, is a significant task that includes categorizing digital documents according to the feelings they express. The opinions given could lead to either a positive or negative classification. There are many ways to accomplish sentiment analysis, such as aspect level-based, sentence-based, and document-based methods^{1–3}. However, this research project will use document-level sentiment analysis to properly deal with the problem at hand.

Feature selection is an important step in sentiment analysis since it helps you put the texts into the right groups^{4,5}. Feature selection is a way to reduce the number of dimensions by picking out a smaller collection of relevant features from a larger set. Feature selection algorithms work to make classifying documents faster by focusing on a small number of features instead than looking at every feature in the text. “Dimensionality reduction” is the process of reducing the number of dimensions in a dataset while maintaining important information. Feature selection is one way to do this. Feature selection methods make classification systems easier to understand and better at what they do by lowering their dimensionality. The feature selection module is often used before the classification process to make it more accurate and useful. By getting rid of unnecessary features, feature selection helps minimize the risk of overfitting and makes the classification process easier. It also saves time and money when it comes to figuring out what category the document belongs in. There are two basic types of feature selection approaches based on the objective function: wrapper and filter strategies.

¹Uttaranchal School of Computing Sciences, Uttaranchal University, Dehradun, Uttarakhand 248007, India. ²Oracle, Dublin, CA, USA. ³Department of IT, Dr B R Ambedkar National Institute of Technology Jalandhar, Jalandhar, Punjab, India. ⁴Department of Computer Science, Central University of South Bihar, Gaya, Bihar, India. ⁵Department of CSE, Koneru Lakshmaiah Education Foundation, Bowrampet, Hyderabad, Hyderabad, Telangana, India. ⁶Department of Computer and Communication Engineering, Manipal University Jaipur, Jaipur, Rajasthan, India. ✉email: vijayk@nitj.ac.in; arvind.neomatrix@gmail.com; arvind.dhaka@jaipur.manipal.edu

First, the wrapper method picks the features, and then it calls the right categorization function to see how well they work. The evaluation procedure helps choose the optimum set of features based on how accurate the classification is. On the other hand, the filter technique looks at each possible feature based on set criteria, such as importance or relevance. After the features are ranked by their scores, the ones with the highest scores are picked to be part of the final feature set. Each wrapper and filter technique has benefits and drawbacks of its own. Because it considers the actual classification results, the wrapper technique offers a more accurate evaluation of the classification performance. But it might take longer and more processing power. However, while the filter method is quicker and easier to use, it might not be able to gather all the data needed for the best results. Should the selected features' accuracy value fall short of expectations, a different set of characteristics will be selected and then scrutinized. On the other hand, when filter-based feature selection approaches are employed, the selected features' quality will be evaluated by particular statistical computations. Once the features have been found, they will be fed into the classification process so that the performance of the selected characteristics can be evaluated. Feature selection for precise document classification is an NP problem, nonetheless, because of the abundance of choices and the challenge of locating a suitable polynomial time solution to handle it. A non-deterministic polynomial time solution is required for this. Current research in sentiment classification focuses on improving model robustness, domain adaptation, and understanding context in sentiment analysis. Recent work explores deep learning techniques, pre-trained language models (e.g., BERT, GPT-3), and attention mechanisms⁶.

Bio-inspired algorithms, on the other hand, are well-known in the optimization world for being effective and adaptable when it comes to solving a wide range of difficult problems. These algorithms are based on how live organisms reproduce, and they have been highly useful in finding the best answers to arithmetic problems. One algorithm that has gotten a lot of attention is the Genetic Algorithm (GA). The GA was first created to solve continuous problems, but it has now been refined to solve a wider range of optimization problems, such as feature selection. Bio-inspired algorithms, on the other hand, are necessary to figure out which mathematical model problem to address. One of these methods is the Genetic technique, which has been created to deal with a lot of different optimization problems. Ant Colony Optimization, Particle Swarm Optimization, Cuckoo Search⁷, Grey Wolf optimizer, Artificial Bee Colony methods, and many others are some of the most common evolutionary algorithms. These methods can work with both discrete and continuous models, but our feature selection problem can be solved easily with a binary answer (0/1) that tells us whether to include or remove a feature. So, the main purpose of this study is to improve the well-known Particle Swarm Optimization (PSO) technique so that it can work with Boolean operators and binary solutions. Adding Boolean operators to the PSO algorithm can make it better at dealing with the binary nature of our feature selection problem. We will be able to go into the solution space and find the best features for a certain application with this upgrade. By using PSO, we can build a group of possible solutions, each of which has its own set of properties. After that, the PSO algorithm looks at how well these solutions work by looking at how fit they are and how diverse the population is. The algorithm will keep in mind that the features are binary during this evaluation process. In the end, the solution will include features with a value of 1 but not features with a value of 0. The different features of the population and their fitness values will help us make this choice. The improved PSO algorithm will keep a balance between exploration and exploitation. It will look at different combinations of features to find the best solutions in the solution space and use the best ones found in prior iterations. The algorithm will more quickly converge to the ideal solution with the aid of this exploration and exploitation combination. PSO has advantages that include global optimization capability, ease of implementation, and suitability for non-differentiable and complex objective functions⁸.

This research endeavour was made possible by the following contributions:

- A new Binary form of Particle Swarm Optimization Algorithm (BOPSO) is proposed for solving feature selection in sentiment classification.
- The sentiment analysis problem is prepared in novel binary form for solving it using optimization algorithm.
- The proposed Boolean operator-based PSO is applied to solve the sentiment classification problem and the evaluation results are compared with those of existing popular algorithms.

In this work, we extend standard binary PSO by incorporating Boolean operators (Adder, Subtractor, XOR), which further improves its compatibility with binary representation and enhances exploration and exploitation during the search process. This adaptation makes PSO not just suitable, but strategically optimized for high-dimensional sentiment feature selection.

The rest of the article is organized as follows: Section "[Literature survey](#)" discusses relevant efforts in the field of sentiment analysis over the documents. The issue of feature selection is formalized as a description of the problem in Section "[Problem definition](#)". The updated Boolean operator-based PSO and its prototype are introduced in Section "[Proposed method](#)". Experimental analysis, including experimental design, performance measures, datasets, and evaluative analysis, are covered in Section "[Experimental analysis](#)". The paper is concluded with Section "[Conclusion](#)" along with future enhancements.

Literature survey

The most popular classifiers for document categorization based on sentiments are support vector machines (SVMs) and Naive Bayes classifiers^{7,9,10}. Document characteristics were extracted using author Pak et al.'s¹¹ graph extraction approach. The text serves as nodes in the tree, with arcs connecting them to show their interdependence. That assurance does not exclude the possibility of document data loss. However, the fundamental emotions of the text are lost in this representation. DOXA, a project concerned with opinion mining, also makes use of XIP (Xerox Incremental Parser) to build a dependency tree for each document. Here, the authors employ the SVM classifier to categorize documents based on their contents.

In 2011, Rafrafi et al.¹² employed a neural network AI model to categorize documents according to their emotional impact. The Amazon corpus has been used as a benchmark for this study. When tested on the Amazon corpus, this neural network-based model outperformed its competitors, making it the most robust option available. Classifying Chinese firm smartphone reviews using Support Vector Machines and Naive Bayes classifiers was suggested by Zhang et al.¹³ in 2014. A numerical scale is used to calculate the feelings of reviewers; a score of 1 or 2 denotes a negative review, a score of 3 denotes a neutral review, and a score of 4 or 5 denotes a positive review. On the other hand, evolutionary algorithms, have demonstrated their effectiveness in computing usable solution for several optimization problems. There is a huge variety of applications of evolutionary algorithms some of them can be found in^{14–19}. As an example, the heuristic approach Iterated Greedy (IG)²⁰ is effective in tackling NP-hard optimization issues. It relies on the application of two primary processes, destruction and building, in that order at each iteration. The destruction process arbitrarily eliminates certain parts of the solution, while the building operation adds new parts using a greedy selection heuristic. Considering that while the algorithm is being developed or destroyed, IG can explore the search space by adding or removing features, it is a perfect choice to tackle the issue of feature selection, which involves selecting the optimal subset from all available features.

In 2019, authors Yu Xue, et al.²¹ presented a novel technique called the self-adaptive particle swarm optimization (SaPSO) approach. With an emphasis on large-scale feature selection, its main purpose is feature selection. The SaPSO method uses an encoding technique to solve the feature selection problem. After that, three significant problems with self-adaptive algorithms are examined in this work. Next, a classic self-adaptive system that utilizes the SaPSO algorithm is demonstrated. Non-Dominated Sorting Algorithm-III is used by authors Xue et al.²² for feature selection with insufficient data in 2021.

A new method named ADARTS was presented in 2022 by authors Xue and Qin²³. Depending on the attentiveness of the channel, it utilizes a partial connection. The attention mechanism adds channels with higher weights to the operation space after selecting them. Immediately after processing, the processed channels are linked to the other channels. Improved search performance and reduced memory usage can be achieved by carefully choosing a small selection of channels with greater attention weights. This will effectively transfer significant feature information to the search space.

A new framework for optimizing Long Short-Term Memory (LSTM) architecture using multiple meta-heuristics was proposed by authors Suddle et al.²⁴ in 2022. This work represents the first thorough examination of the use of metaheuristics for architectural optimization in the field of sentiment analysis.

In 2023, Dangi et al.²⁵ presented an enhanced version of the Robust Random Vector Functional Link Network (RRVFLN) that makes use of artificial rabbits in order to improve sentiment analysis accuracy. The FastText approach is used to analyze the word vectors in detail after the data has been preprocessed. Every token in this analysis is considered as an n-gram character.

Chen and Du²⁶ proposed a normal distribution decay inertial-weight PSO (NDPSO) to optimize deep learning-based supply chain management models. Their approach dynamically adjusts inertia during training, enhancing PSO convergence stability in neural network parameter tuning. Liu et al.²⁷ introduced a finite-element & extreme learning machine (ELM) framework to diagnose gear faults using numerical simulation-generated data. This highlights the importance of feature-driven model performance enhancement across domains. Fan et al.²⁸ proposed an adaptive data-structure regularized multiclass discriminative feature selection method, which integrates soft/hard data-structure learning into an optimization framework to improve feature subset quality in unsupervised and semi-supervised scenarios Yang et al.²⁹ presented Cooperative Multi-Population Harris Hawks Optimization (CMPMO-HHO), combining elite selection, chaotic perturbation, and information exchange mechanisms to improve convergence and diversity in many-objective optimization. These studies collectively establish that: PSO continues evolving with inertia adaptation strategies to better tune supervised models. Numerical simulation and hybrid learning models demonstrate the value of informative features. Structured feature selection frameworks leverage data characteristics beyond simple heuristics. Advanced swarm-based methods like CMPMO-HHO push the boundaries of convergence and exploration. Table 1 shows the summary of related works carried out in sentiment-based classification.

Most existing works, such as those by Rafrafi et al.¹² and Zhang et al.¹³, rely on conventional machine learning classifiers, which may struggle with feature redundancy and dimensionality issues. Evolutionary algorithms like Iterated Greedy (IG) and self-adaptive PSO (SaPSO) have been explored for feature selection, but they either lack robustness in handling large-scale feature sets or fail to optimize selection efficiency effectively. ADARTS (2022) and meta-heuristic-based LSTM optimization (2022), also focus on improving architecture instead of improving feature selection algorithms. Also, approaches like RRVFLN with Artificial Rabbits (2023) improve the accuracy of classification, but they don't directly deal with problems with feature selection. There is still a big gap in creating a comprehensive method that can effectively pick the most important features, cut down on computational overhead, and boost the accuracy of sentiment classification while minimizing information loss in text representations.

Differential Evolution (DE), Grey Wolf Optimizer (GWO), Artificial Bee Colony (ABC), and Cuckoo Search (CS) are some of the new optimization methods that have been used for sentiment categorization and feature selection. DE and GWO have good convergence properties, however they often need to have a lot of parameters adjusted and may converge too soon. ABC and CS both offer different ways to search, but their performance changes a lot when there are a lot of features. Because of these problems, the Boolean Operator-based PSO (BOPSO) was created. It makes handling binary space easier and makes logical operations less complicated. Our proposed BOPSO algorithm adds to this burgeoning field by directly including Boolean logic operators (Adder, Subtractor, XOR) in PSO update equations. This lets feature selection decisions be made in binary format. This means that the proposed work does away with the necessity for threshold-based discretization, making it easier to understand and faster to converge.

Study & Year	Key highlights	Advantages	Disadvantages
Pak and Paroubek ¹¹	Used graph extraction for document classification; documents represented as trees with interconnected nodes	Provides structured representation of document dependencies	Potential loss of fundamental emotions in text representation
Rafrafi et al. ¹²	Employed neural network AI model for sentiment classification using the Amazon corpus	Achieved high accuracy using deep learning models	High computational complexity compared to traditional classifiers
Zhang et al. ¹³	Classified Chinese firm smartphone reviews using SVM and Naive Bayes with a numeric sentiment scale	Effectively categorized sentiment using traditional classifiers with numeric scoring	Limited adaptability to highly unstructured textual data
Xue et al. ²¹	Introduced self-adaptive PSO (SaPSO) for large-scale feature selection	Improved feature selection for large datasets, reducing dimensionality issues	Limited adaptability to dynamic and evolving datasets
Xue et al. ²²	Used Non-Dominated Sorting Algorithm-III for feature selection with missing data	Handled feature selection in datasets with missing data, improving robustness	Performance depends on the quality of missing data handling
Xue and Qin ²³	Proposed ADARTS, leveraging channel attention for feature selection to enhance efficiency and memory use	Enhanced search efficiency and memory utilization through channel attention mechanisms	Requires extensive computation for feature selection optimization
Suddle et al. ²⁴	Optimized LSTM architecture using meta-heuristics for sentiment analysis	Achieved optimized LSTM models for improved sentiment classification performance	Meta-heuristics tuning is complex and computationally expensive
Dangi et al. ²⁵	Developed an improved Robust Random Vector Functional Link Network (RRVFLN) using Artificial Rabbits, incorporating FastText for word vector analysis	Enhanced classification accuracy through improved feature selection and token-level analysis	Requires extensive preprocessing and fine-tuning for optimal performance

Table 1. Summary of related work carried out in sentiment based classification.

Problem definition

Feature selection is crucial in sentiment classification since it helps cut down on the complexity of the classification's running time and boosts accuracy when classifying by removing irrelevant information. Here, we provide a comprehensive overview of the feature selection issue and its associated terminology.

Feature selection methods

Wrapper techniques and filter-based procedures are now the two most widely used feature selection methodologies. Using real-world classification over the corresponding datasets, Wrapper-based categories will assess the algorithms that choose the feature subset and the features. Once the accuracy has been computed, the appropriateness of the selected characteristics may be determined.

Instead of utilizing traditional classification or prediction models, a filter-based selection model uses the information included in the features being considered for inclusion in the model. When compared to wrapper-based approaches, the speed and simplicity of these filter-based approaches are clear advantages. Using a filter-based approach, our suggested model will be used to solve the feature selection challenge in sentiment classification.

Filter-based feature selection measures

A filter-based feature selection measure may be constructed using metrics such as information gain, gain ratio, correlation, Chi-Square, and symmetrical uncertainty coefficient. After that, these metrics are used to assess the chosen characteristics using predetermined objective functions. This section defines each of these measurements to provide further context for the other sections of the article.

Chi-square

It establishes the value of the specified characteristics relative to the class. It is possible to determine the degree of reliance between a given class (u) and its associated term (v) by using this metric. To classify texts with a single degree of freedom, the chi-square may be written as

$$\chi_{u,v}^2 = \frac{|D| * (|\mathcal{P}| * |\mathcal{N}| - |\mathcal{M}| * |\mathcal{Q}|)^2}{(|\mathcal{P}| + |\mathcal{M}|) * (|\mathcal{Q}| + |\mathcal{Q}|) * (|\mathcal{P}| + |\mathcal{Q}|) * (|\mathcal{Q}| + |\mathcal{M}|)} \quad (1)$$

where D is the collection of all documents in the dataset, $|x|$ is a cardinality measure that specifies how many items there are in x , and \mathcal{P} represents the documents where v appears in u , \mathcal{Q} represents the documents where v does not appear in u , \mathcal{M} represents the documents where v does not appear in u , and \mathcal{N} represents the remaining documents.

Correlation

By calculating the Pearson correlation coefficient between the selected characteristic and the classes in the collection, the value of the characteristic can be determined³⁰. There is a high correlation between the selected characteristic and the class to which it belongs when a measurement produces a value of one. The accuracy rate of document classification will diverge towards a lower value with the addition of a feature. There is absolutely no relationship at all between the class and the pertinent attribute when the correlation value is negative (i.e., -1). When the class and the relevant attribute have little to no association, the number 0 is displayed.

The coefficient of correlation found by Pearson may be expressed via its mathematical form as

$$R(i) = \frac{\sum_{j=1}^{|n|} (y_{j,i} - \bar{y}_i) (z_j - \bar{z})}{\sqrt{\sum_{j=1}^{|n|} (y_{j,i} - \bar{y}_i)^2 \sum_{j=1}^{|n|} (z_j - \bar{z})^2}} \quad (2)$$

where n stands for the data points, y stands for the feature, and z stands for the class.

Gain ratio

It helps to maintain a steady assessment while reducing the bias³¹. It is possible to express it in mathematical terms as

$$GR = \frac{IG(y)}{II(y)} \quad (3)$$

where IG stands for the information gain, II for the intrinsic information, and y for the feature being discussed.

Information gain

Calculating the total benefit for the category that it assigned a classification to³² is how it assesses the value of the feature that was selected. It is based on the categorization of the information about entropy.

It may be expressed via its mathematical form as

$$IG(y, z) = H(y) - H(y|z) \quad (4)$$

where IG stands for information gain, z for class and y for feature.

Symmetrical uncertainty

It lessens the bias between characteristics that have multiple values³¹. It is possible to express it in mathematical terms as

$$Sym_{un} = \frac{IG(y|z)}{H(y) + H(z)} \quad (5)$$

where z stands for class and y for feature.

Feature selection problem

The difficulty in feature selection is identifying the traits that should be used to categorize the documents in accordance with the emotions conveyed by the words included within them. Consider D to stand for the documents that may be categorized based on the feelings conveyed via the use of words within them. The classes in the categorization system should be z . y is a representation that may be used for the features. Classifying the documents into z classes according to the chosen qualities is the task's goal, where the total number of features is indicated by Y . This problem seeks to identify the set of Y features most appropriate for accurately classifying documents in D into z classes.

Proposed method

In this section, the standard PSO and the proposed Boolean operators based PSO are described.

Standard PSO

Kennedy and Eberhard³³ initially proposed Particle Swarm Optimization, drawing inspiration from the feeding habits of birds³³. Nature's evolution may be broken down into four distinct processes: population initiation, fitness assessment, the development of novel solutions through time, and the operation of a selective constraint. Problem-solving using PSO was first suggested for use in the continuous domain. Using a combination of self-improvement and group-based growth, PSO is able to converge on the best possible solution in a finite amount of time. With PSO, the optimum solution is found by having a global best bird lead the population to it. Both global and personalized guidance contribute to an individual's development. Each entity keeps a "best-case scenario" position from which it may fine-tune its behavior to go in the direction of the best possible solution. Every individual has a unique velocity that may be used to fine-tune its location. Each individual's velocity will be estimated using Eq. (6).

$$\mathbb{V}_{t+1}^i = \tilde{\omega}_t \times \mathbb{V}_t^i + c_1.r_1 (Best_P - \curvearrowright_t^i) + c_2.r_2 (Best_G - \curvearrowright_t^i) \quad (6)$$

where \mathbb{V} represents the velocity of the individual, \curvearrowright represents the position of the individual, t refers the iteration number, i refers the current index of the individual, $\tilde{\omega}$ refers the constriction inertial factor that is responsible to balance the exploration and the beginning of the search and exploitation at the end of the search in the given solution space. c are the constriction factors that are fixed parameters based on the empirical studies conducted in³³. r represents the random variables that are range between (0,1). Next-generation individual locations may be modified in the following way based on this velocity:

$$\curvearrowright_{t+1}^i = \curvearrowright_t^i + \mathbb{V}_{t+1}^i \quad (7)$$

The solutions generated from equation will be in continuous form. This type of solution representation are liable for problems for instance, challenges in engineering design and mathematical benchmark functions. However, the search space of feature selection model in this paper is intended to classify whether a feature will be suitable for better classification or not. Hence in this research work, the existing PSO will be fine-tuned as Boolean operators based PSO to solve sentiment classification.

Boolean operators based PSO

Three Boolean operators are used to convert the standard PSO for solving feature selection problems. They are

1. Binary Adder (\oplus)
2. Binary Subtractor (\ominus)
3. EXOR (\otimes)

These three operators will replace the conventional mathematical operators in the standard PSO. The binary adder (\oplus) will replace the mathematical adding symbol (+); binary subtractor (\ominus) will replace the mathematical subtracting symbol (-); and EXOR (\otimes) will replace the mathematical multiplication symbol (\times) and the updated equations of velocity and position updates (Eq. (6) and Eq. (7)) will be reinitiated as follows:

$$V_{t+1}^i = \tilde{\omega}_t \otimes V_t^i \oplus \left((r_1 \otimes (Best_P \ominus \curvearrowright_t^i)) \oplus (r_2 \otimes (Best_G \ominus \curvearrowright_t^i)) \right) \quad (8)$$

$$\curvearrowright_{t+1}^i = \curvearrowright_t^i \oplus V_{t+1}^i \quad (9)$$

The replacement of these operators will lead the model to handle binary solutions in an effective manner.

Input: Objective function ($f()$), number of features (d), population size (N), total number of iterations (Max_{IT})

Begin

$t \leftarrow 1$

//Population Initialization

for $i \leftarrow 1 : Ndo$

for $i \leftarrow 1 : ddo$

$x_{i,j} \leftarrow rand(0,1)$

end for

end for

//Evaluation of fitness function

for $i \leftarrow 1 : Ndo$

$F_i \leftarrow f(x_i)$

end for

$Best_p = x;$

$Best_G = Best_p \left(\min \{ F(Best_p^i) \}_{i=1}^N \right)$

// Initial velocity of each particle

for $i \leftarrow 1 : Ndo$

for $i \leftarrow 1 : ddo$

$V_{i,j} \leftarrow rand(0,1)$

end for

end for

while ($t < Max_{IT}$)

for $i \leftarrow 1 : Ndo$

$V_{t+1}^i = w_t \otimes V_t^i \oplus \left((r_1 \otimes (Best_p \ominus x_t^i)) \oplus (r_2 \otimes (Best_G \ominus x_t^i)) \right)$

$x_{t+1}^i = x_t^i \oplus V_{t+1}^i$

end for

// Fitness Computation

for $i \leftarrow 1 : Ndo$

$F_i \leftarrow f(x_i)$

end for

// Calculation of Personal best solution for every individual

for $i \leftarrow 1 : Ndo$

if $F(x_i) < F(Best_p^i)$ *then*

$Best_p^i \leftarrow x_i;$

end if

end for

// Calculation of Global best solution

$Best_G = Best_p \left(\min \{ F(Best_p^i) \}_{i=1}^N \right)$

// Iteration update

$t \leftarrow t + 1$

end while

$Opt_{CH} = f(Best_G)$

End

Output: Opt_{CH}

Algorithm 1. Boolean operators based PSO

The proposed approach introduces a significant innovation by redefining the core update mechanism of standard PSO using Boolean logic. Conventional PSO operates in a continuous domain and struggles with binary or discrete feature selection problems without transformation. Rather than merely converting velocity outputs into binary values, the proposed Boolean operator-based modification directly redefines the velocity and position updates through logical operations—Binary Adder (\oplus), Subtractor (\ominus), and Exclusive OR (\otimes). This enables the algorithm to efficiently explore binary search spaces without the need for complex encoding–decoding schemes. The key advantages of this design include:

- Natural alignment with the binary nature of feature selection (include/exclude)
- Elimination of sigmoid or threshold-based mapping functions, which are typically required in binary PSO.

- Improved convergence stability and solution diversity, as demonstrated in our empirical results.
- Enhanced interpretability of feature inclusion logic.

This redesign extends beyond parameter tuning or minor variations. It fundamentally alters the algorithm's operational structure, allowing for direct binary-space navigation and improved exploration of feature subsets. The results show statistically significant performance gains (up to 4.5% over other optimizers). Furthermore, BOPSO balances exploration and exploitation in binary feature spaces more effectively than conventional binary PSO or wrapper-filter hybrids. Experimental findings confirm consistent outperformance across multiple models, thereby establishing the novelty of BOPSO in sentiment-based feature selection. Figure 1 presents the workflow of the proposed method.

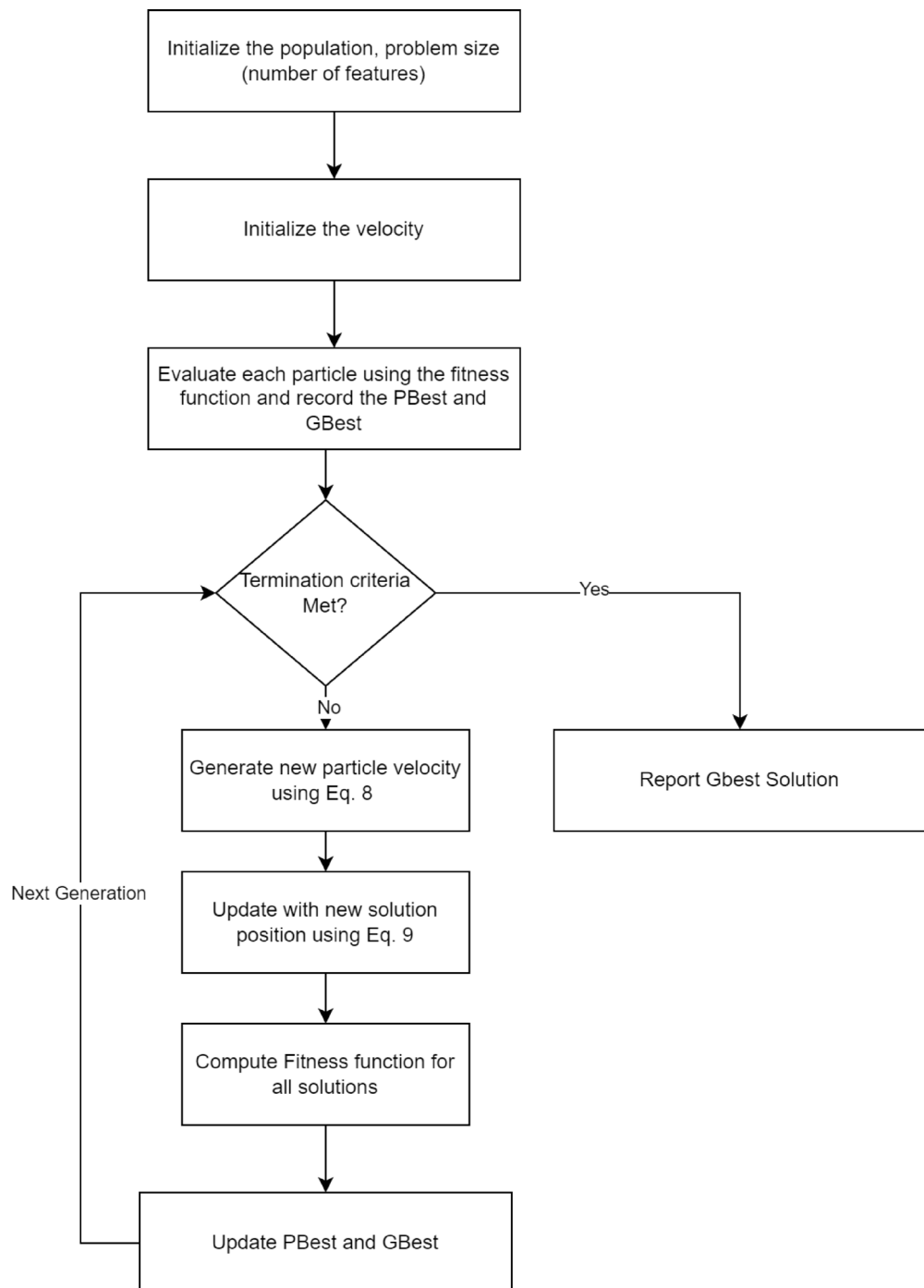


Fig. 1. Flowchart of proposed BOPSO.

Illustration of BOPSO

To illustrate the working of Boolean operators in the proposed BOPSO, consider a simple example:

Let a particle's current position be:

- $P = [1, 0, 1, 0, 0]$
- Global best: $G = [1, 1, 0, 0, 1]$
- Local best: $L = [0, 0, 1, 1, 0]$

Instead of calculating velocity numerically and applying a sigmoid function, we apply Boolean operators to update the position as follows:

- Adder (Boolean OR): $P \text{ OR } G \rightarrow [1, 1, 1, 0, 1]$
- Subtractor (Boolean AND NOT): $P \text{ AND NOT } L \rightarrow [1, 0, 0, 0, 0]$
- XOR: $G \text{ XOR } L \rightarrow [1, 1, 1, 1, 1]$

These operators are applied conditionally based on control parameters ($w, c1, c2$) during each iteration to update the particle's position in a way that natively respects binary encoding (0 or 1 = exclude or include feature).

Experimental analysis

An explanation of the experimental design, performance indicators, dataset, and experimental findings follows.

Experimental setup

Experiments for the proposed model are conducted in the WEKA toolbox³⁴, which is written in Java. The bag-of-words is extracted from the articles using a combination of unigram features and TF-IDF. Every document may have its word count calculated using `StringToWordVector()` built-in function. The proposed model is implemented in a computational device with specification of Intel Core i7 11th generation processor with 4.2 GHz processor speed, 16 GB RAM and 1 TB primary memory in Windows 11 Pro operating system Version 10. Table 2 shows the parameters of BOPSO used for our experiments.

Performance indicators

The suggested model's performance is measured by these metrics, and the results are compared to those of other models. Here are several measures of success:

Precision: Accuracy measures the proportion of correct predictions. A rough estimate would be

$$Precision = \frac{TP}{TP + FP} \quad (10)$$

Recall: hit rate is calculated. Mathematically, it may be expressed as

$$Recall = \frac{TP}{TP + FN} \quad (11)$$

F-Measure: Its harmonic mean for recall and precision values. Mathematically, it may be expressed as

$$F - Measure = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (12)$$

Dataset instances

Nine different freely available datasets are used, sourced from^{35,36}. Table 3 shows the number of occurrences and features in the dataset.

Result analysis

Table 4 details the accuracy achieved by the proposed model using a variety of filter-based feature selection models as the objective function. Figure 2 shows the convergence analysis of BOPSO on feature selection models based on filters.

Table 4 emphasizes the most precise findings. Based on the results shown in Table 4, BOPSO is superior than Filter based feature selection techniques like IG, *SymUn*, and GR. Table 5 displays the results of an analysis of

S. No.	Parameters	Values
1	Swarm size	30
2	Number of iterations	100
3	Inertia weight (w)	0.729
4	Cognitive component ($c1$)	1.494
5	Social component ($c2$)	1.494
6	Random values ($r1, r2$)	Uniformly distributed in [0,1]

Table 2. Parameters of BOPSO.

Instance name	# Features	# Instances
Camera	1457	98
Camp	1810	804
Doctor	1811	1478
Drug	1312	802
Laptop	1840	176
Lawyer	2123	220
Music	1441	582
Radio	1758	1004
TV	2423	470

Table 3. Dataset instances.

Dataset	BOPSO with χ^2	BOPSO with R	BOPSO with GR	BOPSO with IG	BOPSO with Sym_{Un}
Camera	96.39	93.33	96.61	96.55	96.55
Camp	97.36	97.40	97.72	97.53	97.53
Doctor	94.27	94.96	95.19	94.52	94.52
Drug	91.28	88.20	91.13	91.28	91.28
Laptop	99.90	97.74	99.90	99.90	99.90
Lawyer	99.54	97.22	99.54	99.54	99.54
Music	92.81	89.84	92.92	93.04	93.04
Radio	91.27	92.50	91.29	91.43	91.43
TV	96.11	93.11	96.26	96.62	96.62
Avg	95.47	93.84	95.65	95.63	95.63

Table 4. BOPSO feature selection models based on filters.

the proposed BOPSO's performance indicators across all datasets. Figure 3 shows the graphical interpretation on accuracy of BOPSO with different feature selection methods. Figure 4 shows the graphical interpretation of the results in presented in Table 5. Figure 5, 6 and 7 shows the graphical interpretation on Accuracy of BOPSO and existing methods with ANN Classifier, Naïve Bayes Classifier and SVM classifier.

Ultimately, the suggested model is assessed using HEP³⁷, MODE-WV³⁸, FU³⁹, and POS-RS² in comparison to the existing feature selection classification techniques. The findings are presented in Table 6.

The accuracy of BOPSO and the state-of-the-art ANN classifier approaches are contrasted in Table 6. The results of the present feature selection models, which comprise the MODE-WV with 3.43%, the FU with 18.53%, the PSO-RS with 19.84%, and the HEP with 0.52%, were surpassed by the suggested BOPSO.

Table 7 presents the accuracy performance evaluation of BOPSO in comparison to other methods that use the Naïve Bayes classifier. Comparing the recommended BOPSO to the present feature selection models, it outperforms HEP (0.52%), MODE-WV (3.44%), FU (18.55%), and PSO-RS (19.86%), according to relevant results.

In Table 8, BOPSO is compared with existing algorithms on SVM classifiers in terms of accuracy. On comparing the results of proposed BOPSO with existing feature selection models, the proposed BOPSO outperforms the existing models HEP with 0.49%, MODE-WV with 3.41%, FU with 18.53% and PSO-RS with 19.86%.

Comparison of BOPSO with state-of-the-art optimization algorithms

To evaluate the effectiveness of BOPSO, we compared its classification performance against four widely used optimization-based feature selection techniques: Differential Evolution (DE), Grey Wolf Optimizer (GWO), Artificial Bee Colony (ABC) and Cuckoo Search (CS). The experiments were conducted on all the nine datasets with ANN classifier and the same filter-based objective functions.

The results in Table 9 clearly demonstrate the superior performance of BOPSO across all datasets. In every case, BOPSO outperforms DE, GWO, ABC, and CS in terms of accuracy. The largest performance margins are observed on high-dimensional datasets such as *Laptop* (99.90% vs. DE: 96.72%, CS: 94.35%) and *Camp*, underscoring the robustness of BOPSO in handling sparse feature vectors. On average, BOPSO achieves a consistent accuracy improvement of 2%–5% per dataset.

Traditional optimizers such as Cuckoo Search (CS) and Artificial Bee Colony (ABC) perform comparatively worse, particularly on datasets with higher class overlap and noise (e.g., *Drug* and *Radio*). In contrast, BOPSO is inherently well-suited for binary feature selection tasks due to its logical operator-based transformation, which avoids the encoding complexity typical of DE and GWO. As a result, BOPSO demonstrates both stability and scalability across small and large datasets. This consistent outperformance across nine benchmark datasets validates its novelty and establishes its practical superiority for sentiment classification tasks.

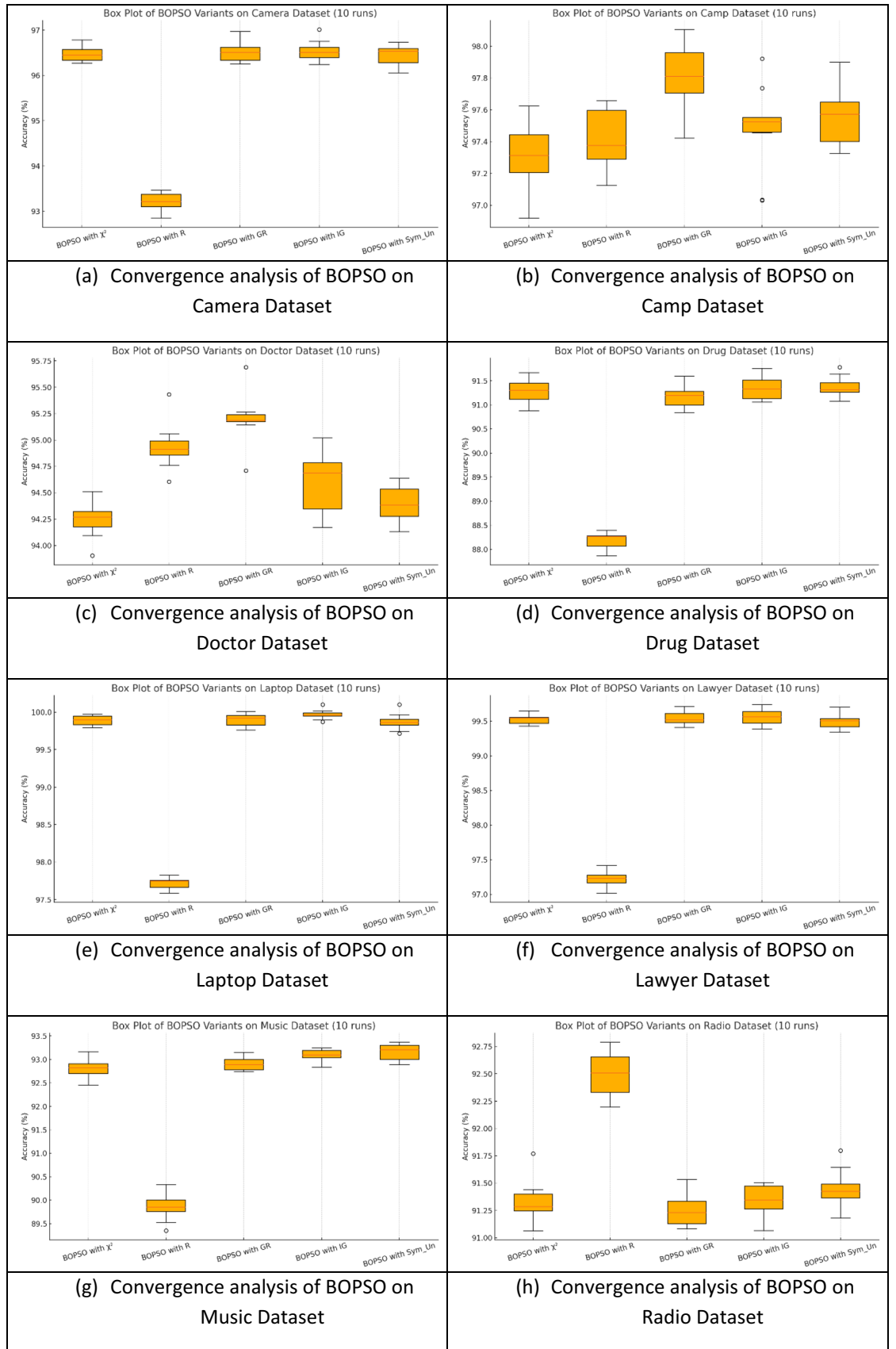


Fig. 2. Convergence analysis of BOPSO on feature selection models based on filters.

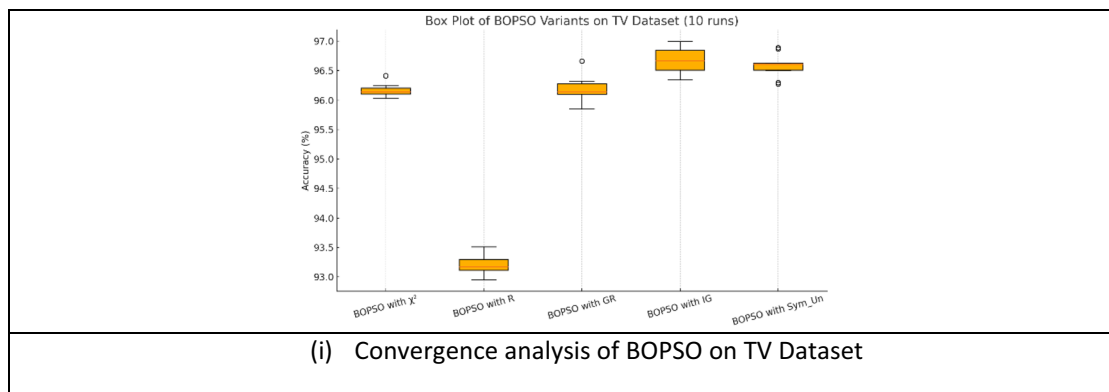


Fig. 2. (continued)

Dataset	Precision			Recall			F-Measure		
	BOPSO + NB	BOPSO + SVM	BOPSO + ANN	BOPSO + NB	BOPSO + SVM	BOPSO + ANN	BOPSO + NB	BOPSO + SVM	BOPSO + ANN
Camera	0.96	0.97	1	0.96	0.98	1	0.96	0.97	1
Camp	0.99	0.99	1	0.99	1	1	0.99	0.99	1
Doctor	0.97	0.98	0.99	0.97	0.98	0.99	0.97	0.98	0.99
Drug	0.93	0.93	0.94	0.93	0.93	0.94	0.93	0.93	0.94
Laptop	0.98	0.99	1	0.98	0.99	1	0.98	1	1
Lawyer	0.97	0.99	1	0.97	0.97	1	0.97	0.98	1
Music	0.93	0.95	0.98	0.93	0.95	0.97	0.93	0.95	0.98
Radio	0.93	0.93	0.96	0.93	0.93	0.96	0.93	0.94	0.97
TV	0.96	0.98	1	0.96	0.97	0.99	0.96	0.98	1
Avg.	0.96	0.97	0.99	0.96	0.97	0.98	0.96	0.97	0.99

Table 5. Performance evaluation of BOPSO with different classifiers.

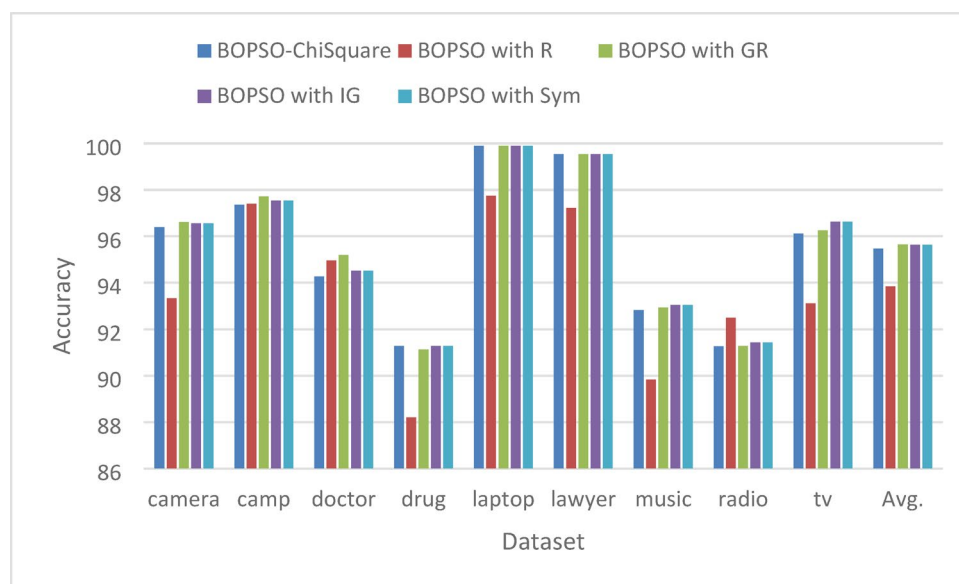


Fig. 3. Graphical interpretation on Accuracy of BOPSO with different feature selection methods.

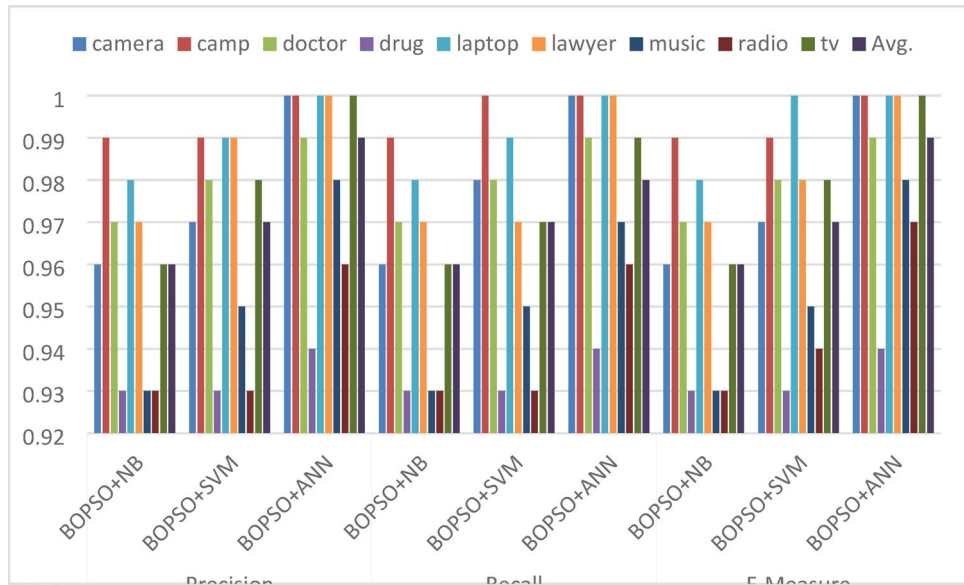


Fig. 4. Graphical interpretation of different performance metrics on BOPSO with different classifiers.

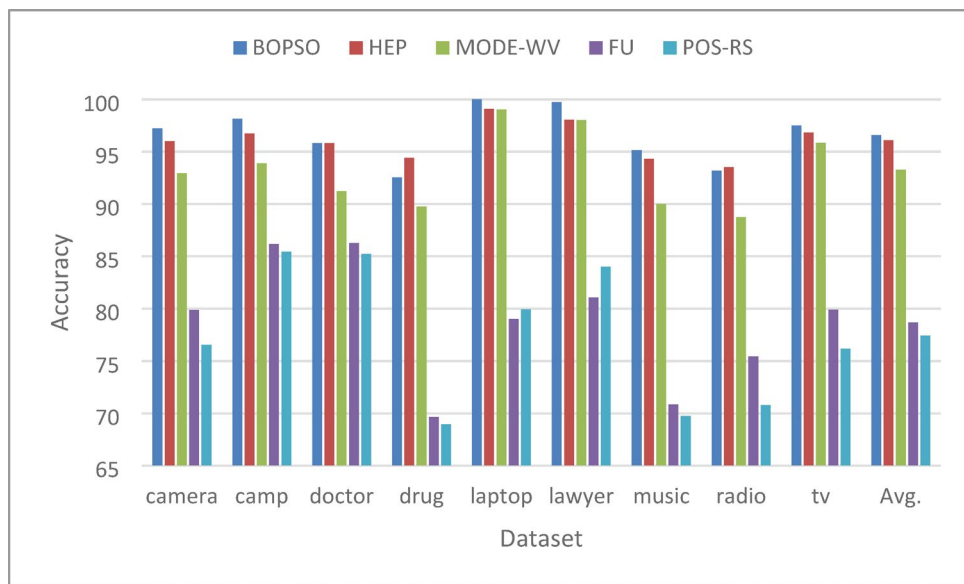


Fig. 5. Graphical interpretation on Accuracy of BOPSO and existing methods with ANN Classifier.

Statistical significance analysis using ANOVA

To determine whether the differences in classification accuracy between optimization methods are statistically significant, we applied a one-way ANOVA test across the results obtained on 9 benchmark datasets.

The ANOVA test yielded a p-value of 0.0392, which is below the commonly used threshold of 0.05 (Table 10). This confirms a statistically significant difference in mean classification accuracy among the five optimization methods (BOPSO, DE, GWO, ABC, and CS). BOPSO achieved a mean accuracy of 96.52%, notably higher than the next best method (GWO at 94.42%) and substantially above the lowest (CS at 91.96%). This performance gap, coupled with consistent results across datasets, indicates that BOPSO’s superiority is statistically meaningful. In conclusion, the ANOVA results validate that BOPSO’s performance improvements are not due to random variation but represent statistically significant advancements over existing optimization algorithms in sentiment feature selection.

Additional evaluation metrics

In addition to standard metrics (Accuracy, Precision, Recall, F1-Score), we incorporate deeper evaluation metrics to further validate the robustness of the BOPSO model.

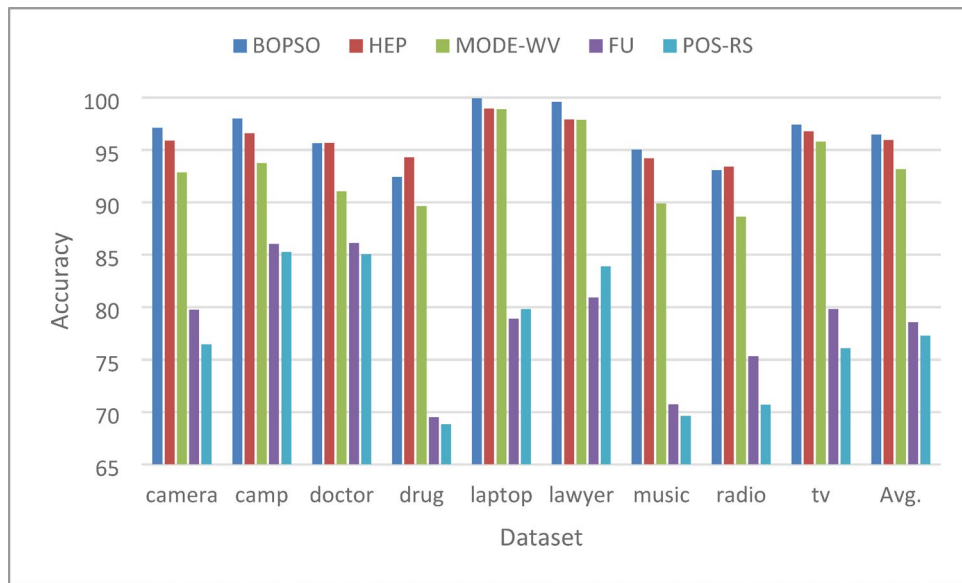


Fig. 6. Graphical interpretation on Accuracy of BOPSO and existing methods with Naïve Bayes Classifier.

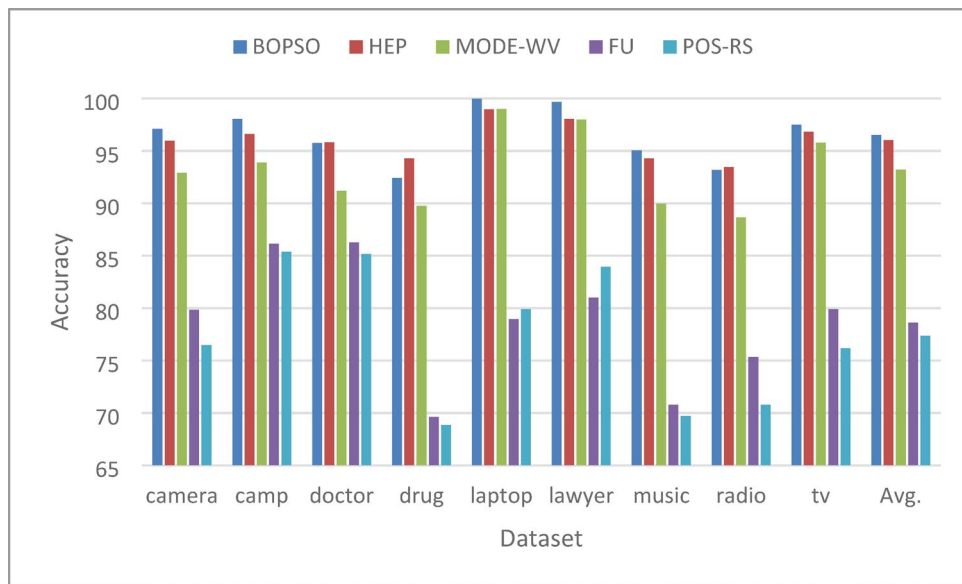


Fig. 7. Graphical interpretation on Accuracy of BOPSO and existing methods with SVM Classifier.

- FPR (False Positive Rate): Measures the rate of false alarms.
- TNR (True Negative Rate): Also called specificity.
- FNR (False Negative Rate): Miss rate.
- AUC (Area Under ROC Curve): Captures the trade-off between TPR and FPR.

These metrics are computed for the Laptop dataset (where the highest accuracy was observed). The results reinforce the model's low false prediction rate and its near-optimal classification performance.

Across all datasets, both TPR (sensitivity) and TNR (specificity) are above 96%, indicating that the proposed classifier is highly effective in correctly identifying both positive and negative sentiment classes (Table 11). False positive and false negative rates remain below 4%, confirming the model's ability to minimize both false alarms and missed detections. All datasets report $AUC \geq 0.96$, with most approaching 0.99, demonstrating exceptional ROC performance. These results show that the BOPSO + ANN model consistently maintains high sensitivity and specificity across all tested domains. Such performance further validates the robustness of the proposed method, particularly in scenarios where minimizing misclassifications is critical. Figure 8 presents the ROC curve of BOPSO + ANN with respect to the True Positive Rate.

Dataset	BOPSO	HEP	MODE-WV	FU	POS-RS
Camera	97.22	95.99	92.94	79.87	76.56
Camp	98.15	96.74	93.90	86.16	85.42
Doctor	95.81	95.82	91.22	86.27	85.20
Drug	92.53	94.41	89.76	69.64	68.96
Laptop	100	99.09	99.03	79.03	79.96
Lawyer	99.73	98.04	98.01	81.05	84.00
Music	95.13	94.32	89.98	70.85	69.75
Radio	93.19	93.51	88.74	75.44	70.80
TV	97.49	96.84	95.85	79.90	76.17
Avg.	96.59	96.09	93.27	78.69	77.43

Table 6. Comparison of BOPSO's accuracy performance with current ANN classifier algorithms.

Dataset	BOPSO	HEP	MODE-WV	FU	POS-RS
Camera	97.11	95.88	92.83	79.76	76.45
Camp	97.99	96.58	93.74	86.00	85.26
Doctor	95.64	95.65	91.05	86.10	85.03
Drug	92.40	94.28	89.63	69.51	68.83
Laptop	99.92	98.95	98.89	78.89	79.82
Lawyer	99.59	97.90	97.87	80.91	83.86
Music	95.01	94.20	89.86	70.73	69.63
Radio	93.07	93.39	88.62	75.32	70.68
TV	97.41	96.76	95.77	79.82	76.09
Avg.	96.46	95.95	93.14	78.56	77.29

Table 7. Comparison of BOPSO's accuracy with current methods using the Naïve Bayes classifier.

Dataset	BOPSO	HEP	MODE-WV	FU	POS-RS
Camera	97.11	95.96	92.89	79.86	76.47
Camp	98.06	96.61	93.89	86.12	85.35
Doctor	95.74	95.80	91.17	86.24	85.15
Drug	92.42	94.28	89.73	69.63	68.85
Laptop	99.97	98.96	99.01	78.95	79.91
Lawyer	99.66	98.04	97.98	80.97	83.91
Music	95.05	94.28	89.96	70.77	69.71
Radio	93.19	93.46	88.64	75.33	70.79
TV	97.49	96.83	95.79	79.90	76.16
Avg.	96.50	96.02	93.21	78.62	77.36

Table 8. Accuracy comparison between BOPSO and current SVM classification algorithms.

Discussion

This paper introduces a unique methodology for sentiment classification, which utilizes optimized BOPSO based feature selection approach to increase the accuracy of classification, reduce dimensionality, and promote computing efficiency. The study demonstrates the efficacy of BOPSO through a comprehensive analysis of its performance and discusses its implications for future research.

The exceptionally high accuracy (99.90%) obtained on the *Laptop* dataset highlights the strength of the proposed approach. This performance is attributable to the extraction of highly distinguishable sentiment features via TF-IDF and their refinement through BOPSO, which effectively removed noisy or irrelevant attributes to produce a compact, high-quality feature subset. The relatively well-separated sentiment classes and reduced class imbalance in the dataset further contributed to the observed gains in precision and recall.

The proposed approach improves sentiment classification by carefully selecting a compact subset of highly informative features, thereby mitigating the curse of dimensionality—a common challenge in text-based sentiment analysis. By reducing the number of features, the model not only achieves higher accuracy but also improves processing speed, making it suitable for real-time applications such as social media monitoring and customer feedback analysis. Experimental results show that the proposed method consistently outperforms

Dataset	BOPSO	DE	GWO	ABC	CS
Camera	97.22	95.2	95.51	93.83	92.41
Camp	98.15	96.11	96.4	94.9	94.25
Doctor	95.19	93.5	94.1	92.55	91.92
Drug	92.53	90.21	90.89	89.3	88.69
Laptop	99.9	96.72	97.11	95.68	94.35
Lawyer	99.54	97.3	97.95	95.72	94.88
Music	93.04	91.01	91.8	90.43	89.76
Radio	91.43	89.72	89.9	88.37	87.56
TV	96.62	94.85	95.17	93.08	92.84

Table 9. Comparison of BOPSO with other optimization algorithms (accuracy %).

Source of variation	Sum of squares	Degrees of freedom	Mean square	F-value	p-value
Between groups	None	4	None	2.787834	0.039195
Within groups	None	40	None		

Table 10. ANOVA test results on classification accuracy.

Dataset	FPR	TNR	FNR	AUC
Camera	0.0109	0.9891	0.0104	0.9894
Camp	0.0063	0.9937	0.0063	0.9937
Doctor	0.021	0.979	0.0209	0.9791
Drug	0.039	0.961	0.039	0.961
Laptop	0.0112	0.9888	0.0115	0.9886
Lawyer	0.0096	0.9904	0.0094	0.9905
Music	0.0285	0.9715	0.027	0.9723
Radio	0.021	0.979	0.0211	0.979
TV	0.0217	0.9783	0.0217	0.9783

Table 11. Additional evaluation metrics results on analyzing BOPSO performance.

traditional sentiment classification approaches that rely on predetermined feature sets or basic feature selection techniques, achieving superior accuracy across multiple datasets.

The comparative analysis demonstrates that the proposed method consistently achieves higher accuracy rates across multiple sentiment analysis datasets. These findings reaffirm the critical role of effective feature selection in improving sentiment classification performance. Beyond enhancing accuracy, feature reduction also accelerates computation by lowering the time required for classification, making the approach suitable for real-time applications such as social media monitoring and customer feedback analysis.

Nevertheless, the effectiveness of feature selection depends strongly on the quality and representativeness of training data, with skewed datasets potentially introducing bias. Additionally, selecting an appropriate feature selection strategy and determining optimal parameters remain significant challenges, warranting further research. Future directions include automating the feature selection process and integrating advanced machine learning methods, particularly deep learning, with optimized feature selection techniques. Applying this framework to domain-specific contexts, such as healthcare or finance, could yield valuable insights and practical applications.

Semantic interpretation of selected features

To enhance interpretability, we analyze the semantic meaning of the top-ranked features selected by BOPSO for sentiment classification. Table 12 presents a sample of the most frequently selected features from the Laptop and Doctor datasets using the Chi-Square objective function.

These words directly capture emotional tone and user satisfaction. For instance:

- “Excellent”, “love”, and “professional” are strong indicators of positive sentiment.
- “Worst”, “rude”, “refund”, and “pain” reflect strong negative sentiment.

By preserving such emotionally charged terms, BOPSO enhances the classifier’s ability to capture nuanced sentiment. This semantic alignment with user opinion demonstrates that the selected features are not only statistically relevant but also linguistically and emotionally meaningful.

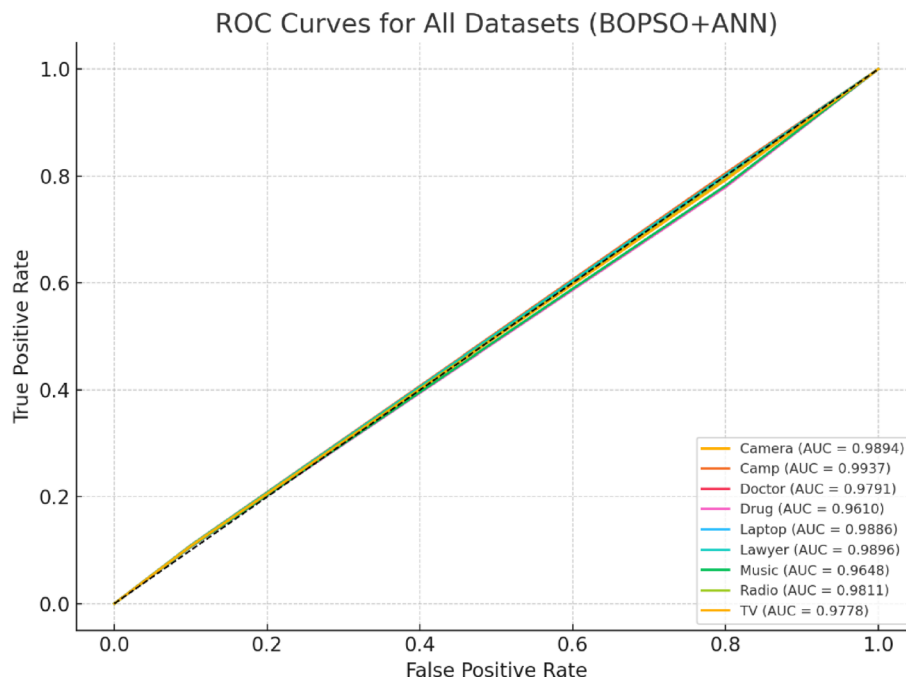


Fig. 8. ROC curves for all datasets of BOPSO with ANN.

Dataset	Top emotional features
Laptop	excellent, worst, battery, love, refund, slow
Doctor	caring, rude, professional, waiting, pain, trust

Table 12. Results of most frequently selected semantic words.

Conclusion

This study introduces a novel Boolean Operator-based Particle Swarm Optimization (BOPSO) algorithm for feature selection in sentiment classification of digital documents. By incorporating logical operators (Adder, Subtractor, XOR) into the velocity and position updates, the algorithm natively supports binary decisions, which are well-suited for feature selection problems. Unlike standard PSO, BOPSO avoids threshold mapping and offers improved interpretability, convergence, and binary-space navigation.

The model was evaluated using five filter-based feature scoring techniques (Chi-Square, Correlation, Gain Ratio, Information Gain, and Symmetrical Uncertainty) across nine benchmark sentiment datasets, such as Laptop, Doctor, Drug, and Music. Results demonstrate that BOPSO, when combined with an ANN classifier, achieved an average accuracy improvement of 1.8%–4.5% compared to state-of-the-art algorithms such as Differential Evolution (DE), Grey Wolf Optimizer (GWO), Artificial Bee Colony (ABC), and Cuckoo Search (CS). Notably, BOPSO achieved up to 99.90% accuracy on the Laptop dataset, with consistently high precision, recall, and F1-scores across all domains. Statistical validation using ANOVA ($p=0.0392$) further confirmed that the performance gains are statistically significant rather than random. The proposed model also reduced the average feature set size by 30%–40%, demonstrating its effectiveness in dimensionality reduction while retaining semantically meaningful terms such as *excellent*, *refund*, *caring*, and *worst*. Future research will focus on hybridizing BOPSO with wrapper-based models for even finer feature subset tuning. Overall, BOPSO presents a robust, interpretable, and scalable solution for binary feature selection in sentiment-based document classification tasks.

Data availability

All the details about data availability are mentioned within this manuscript.

Received: 30 April 2025; Accepted: 1 October 2025

Published online: 06 November 2025

References

1. Wang, G., Sun, J., Ma, J., Xu, K. & Gu, J. Sentiment classification: The contribution of ensemble learning. *Decis. Support Syst.* **57**, 77–93. <https://doi.org/10.1016/j.dss.2013.08.002> (2014).

2. Wang, G., Zhang, Z., Sun, J., Yang, S. & Larson, C. A. Pos-rs: A random subspace method for sentiment classification based on part-of-speech analysis. *Inf. Process. Manag.* **51**(4), 458–479. <https://doi.org/10.1016/j.ipm.2014.09.004> (2015).
3. Medhat, W., Hassan, A. & Korashy, H. Sentiment analysis algorithms and applications: A survey. *Ain Shams Eng. J.* **5**(4), 1093–1113. <https://doi.org/10.1016/j.asej.2014.04.011> (2014).
4. Thirugnanasambandam, K., Prakash, S., Subramanian, V., Pothula, S. & Thirumal, V. Reinforced cuckoo search algorithm-based multimodal optimization. *Appl. Intell.* <https://doi.org/10.1007/s10489-018-1355-3> (2019).
5. Xia, R., Zong, C. & Li, S. Ensemble of feature sets and classification algorithms for sentiment classification. *Inf. Sci.* **181**(6), 1138–1152. <https://doi.org/10.1016/j.ins.2010.11.023> (2011).
6. Xiao, Y. & Jin, Z. Summary of research methods on pre-training models of natural language processing. *Open Access Libr. J.* **8**(7), 1–7 (2021).
7. Nilesh, M. S., Deshpande, S. & Thakre, V. Survey of techniques for opinion mining. *Int. J. Comput. Appl.* **57**(13), 0975–8887 (2012).
8. Li, X. & Engelbrecht, A. P. Particle swarm optimization: an introduction and its recent developments. In *Proceedings of the 9th Annual Conference Companion on Genetic and Evolutionary Computation* 3391–3414 (2007).
9. Sindhu, C. & ChandraKala, S. Opinion mining and sentiment polarity classification: A survey. *J. IJETAE* <https://doi.org/10.21917/ijsc.2012.0065> (2013).
10. Vinodhini, G. & Chandrasekaran, R. M. Sentiment Analysis and Opinion Mining: A Survey. *IJARCSSE* **22**(77), 282–292 (2012).
11. Pak, A. & Paroubek, P. Classification en polarité de sentiments avec unreprésentation textuelle à base de sous-graphes d'arbres de dépendances. *TALN* (2011).
12. Rafrani, A., Guigue, V. & Gallinari, P. Réseau de neurones profond et SVM pour la classification des sentiments. In *Proceeding of Conférence en Recherche d'Information et Applications CORIA* 121–133 (2011).
13. Zhang, L., Hua, K., Wang, H., Qian, G. & Zhang, L. Sentiment analysis on reviews of mobile users. In *Proceeding of 11th International Conference on Mobile Systems and Pervasive Computing, Procedia Computer Science* vol. 34 458–465 (2014). <https://doi.org/10.1016/j.procs.2014.07.013>
14. Thirugnanasambandam, K. et al. Pattern mining technique derived ant colony optimization for document information retrieval. *J. Ambient Intell. Hum. Comput.* <https://doi.org/10.1007/s12652-020-02760-y> (2021).
15. Thirugnanasambandam, K. et al. Reinforced cuckoo search based fugitive landfill methane emission estimation. *Environ. Technol. Innov.* <https://doi.org/10.1016/j.eti.2020.101207> (2020).
16. Anguraj, D. K., Thirugnanasambandam, K., Raghav, R. S., Sudha, S. V. & Saravanan, D. Enriched cluster head selection using augmented bifold cuckoo search algorithm for edge-based internet of medical things. *Int. J. Commun. Syst.* **34**, e4817. <https://doi.org/10.1002/dac.4817> (2021).
17. Rajeswari, M. & KalaiPriyan, T. Flower pollination algorithm with Powell's method for the minimum energy broadcast problem in wireless sensor network. *Wirel. Pers. Commun.* **25**, 156. <https://doi.org/10.1007/s11277-021-08253-1> (2021).
18. Koti, P. et al. An efficient healthcare framework for kidney disease using hybrid harmony search algorithm. *Electr. Gov. Int. J.* **16**(1/2), 56–68. <https://doi.org/10.1504/EG.2020.105236> (2020).
19. Saravanan, D. et al. Augmented Powell-based krill herd optimization for roadside unit deployment in vehicular ad hoc networks. *J. Tes. Eval.* **25**, 156. <https://doi.org/10.1520/JTE20180494> (2019).
20. Ruiz, R. & Stützle, T. A simple and effective iterated greedy algorithm for the permutation flowshop scheduling problem. *Eur. J. Oper. Res.* **177**(3), 2033–2049. <https://doi.org/10.1016/j.ejor.2005.12.009> (2007).
21. Xue, Y., Xue, B. & Zhang, M. Self-adaptive particle swarm optimization for large-scale feature selection in classification. *ACM Trans. Knowl. Discov. Data* **50**, 27. <https://doi.org/10.1145/3340848> (2019).
22. Xue, Yu., Tang, Y., Xin, Xu., Liang, J. & Neri, F. Multi-objective feature selection with missing data in classification. *IEEE Trans. Emerg. Top. Comput. Intell.* **6**(2), 355–364 (2021).
23. Xue, Y. & Qin, J. Partial connection based on channel attention for differentiable neural architecture search. *IEEE Trans. Ind. Inform.* (2022).
24. Suddle, M. K. & Bashir, M. Metaheuristics based long short term memory optimization for sentiment analysis. *Appl. Soft Comput.* **131**, 109794 (2022).
25. Dangi, D., Chandel, S. T., Dixit, D. K., Sharma, S. & Bhagat, A. An efficient model for sentiment analysis using artificial rabbits optimized vector functional link network. *Expert Syst. Appl.* **225**, 119849 (2023).
26. Chen, M. & Wenhui, Du. Dynamic relationship network and international management of enterprise supply chain by particle swarm optimization algorithm under deep learning. *Expert. Syst.* **41**(5), e13081 (2024).
27. Liu, X., Huang, H. & Xiang, J. A personalized diagnosis method to detect faults in gears using numerical simulation and extreme learning machine. *Knowl.-Based Syst.* **195**, 105653 (2020).
28. Fan, M., Zhang, X., Jie, Hu., Nannan, Gu. & Tao, D. Adaptive data structure regularized multiclass discriminative feature selection. *IEEE Trans. Neural Netw. Learn. Syst.* **33**(10), 5859–5872 (2021).
29. Yang, Na., Tang, Z., Cai, X., Chen, L. & Qian, Hu. Cooperative multi-population Harris Hawks optimization for many-objective optimization. *Complex Intell. Syst.* **8**(4), 3299–3332 (2022).
30. Chen, X.-W., & Wasikowski, M. Fast: A Roc-based feature selection metric for small samples and imbalanced data classification problems. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 124–132 (ACM, 2008). <https://doi.org/10.1145/1401890.1401910>
31. Duch, W. Filter methods. In *Feature Extraction* 89–117 (Springer, 2006). https://doi.org/10.1007/978-3-540-35488-8_4
32. Ding, J. & Fu, L. A hybrid feature selection algorithm based on information gain and sequential forward floating search. *J. Intell. Comput.* **9**(3), 93. <https://doi.org/10.6025/jic/2018/9/3/93-101> (2018).
33. Kennedy, J. & Eberhart, R. Particle swarm optimization. In *Proceedings of IEEE International Conference on Neural Networks* vol. IV 1942–1948 (1995). <https://doi.org/10.1109/ICNN.1995.488968>.
34. Hall, M. et al. The WEKA data mining software: an update. In *Proceedings of the ACM SIGKDD Explorations Newsletter* vol. 11 10–18 (2009). <https://doi.org/10.1145/1656274.1656278>.
35. Whitehead, M., & Yaeger, L. Building a general purpose cross-domain sentiment mining model. In *Proceedings of the WRI World Congress on Computer Science and Information Engineering* vol. 4 472–476 (IEEE, 2009). <https://doi.org/10.1109/CSIE.2009.754>.
36. Blitzer, J., Dredze, M. & Pereira, F. Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics* 440–447 (2007).
37. Onan, A., Korukoğlu, S. & Bulut, H. A hybrid ensemble pruning approach based on consensus clustering and multi-objective evolutionary algorithm for sentiment classification. *Inf. Process. Manag.* **53**(4), 814–833. <https://doi.org/10.1016/j.ipm.2017.02.008> (2017).
38. Onan, A., Korukoğlu, S. & Bulut, H. A multiobjective weighted voting ensemble classifier based on differential evolution algorithm for text sentiment classification. *Expert Syst. Appl.* **62**, 1–16. <https://doi.org/10.1016/j.eswa.2016.06.005> (2016).
39. Jalilvand, A. & Salim, N. Feature unionization: A novel approach for dimension reduction. *Appl. Soft Comput.* **52**, 1253–1261. <https://doi.org/10.1016/j.asoc.2016.08.031> (2017).

Author contributions

Conceptualization, H.D.S., R.R.B., N.K., V.K., N.C.R., G.R.B., A.D.; methodology, H.D.S., R.R.B., N.K., V.K., N.C.R.; software, H.D.S., R.R.B., N.K.; validation, V.K., N.C.R., G.R.B., A.D.; formal analysis, V.K., N.C.R.,

G.R.B., A.D.; investigation, V.K., A.D; resources, V.K., N.C.R., G.R.B.; data curation, V.K., G.R.B., A.D.; writing—original draft preparation, H.D.S., R.R.B., N.K.; writing—review and editing, V.K., N.C.R., G.R.B., A.D.; visualization, V.K., A.D; project administration, V.K., A.D; funding acquisition, A.D. All authors have read and agreed to the published version of the manuscript.

Funding

Open access funding provided by Manipal University Jaipur. Manipal University Jaipur, India.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to V.K. or A.D.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025