INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

ORIGINAL ARTICLE

Phoenix–Chameleon Optimization for Feature Selection in Sentiment Analysis

K Fathima[1]*, **A R Mohamed Shanavas**[1]

1 Department of Computer Science, Jamal Mohamed College (Autonomous), Affiliated to Bharathidasan University, Tiruchirappalli, Tamil Nadu, India

OPEN ACCESS

Received: 06/08/2025

Accepted: 19/09/2025

Published: 20/10/205

Citation: Fathima K, Mohamed Shanavas AR (2025) Phoenix–Chameleon Optimization for Feature Selection in Sentiment Analysis. Indian Journal of Science and Technology 18(38): 3091-3102. https://doi.org/10.17485/IJST/v18i38.1228

* **Corresponding author**.

fathima14research@gmail.com

Funding: None

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

Abstract

Objective: To develop a hybrid optimization algorithm for robust feature selection in sentiment analysis. **Methods:** A Two-Phase optimization procedure is proposed, Phoenix-Chameleon Optimization Algorithm (PCOA), which has the entropy-driven exploration phase and density-related exploitation phase combined with each other. Phoenix phase supports global search based on rebirth, whereas Chameleon phase supports local fine-tuning based on feature co-occurrence density and most recent gains. TF-IDF and RoBERTa embeddings are represented by feature vectors that are reduced in terms of PCOA and tested by BiGRU classes. The proposed work performed experiments on four benchmark corpora, MR, CR, IMDB and SemEval 2013. **Findings:** Accuracy of PCOA is up to 96.98% on IMDB, but on all the datasets performance increases (3%-6%). Using the F1-scores and training efficiency as a metric, PCOA attains better results compared to Chi-Square, Recursive Feature Elimination (RFE) and Particle Swarm Optimization (PSO) in both lexical and contextual embeddings. **Novelty:** PCOA introduces a new entropy-based phase-switching mechanism between Phoenix and Chameleon behaviors, creating a domain-robust, dynamic feature selection framework.

Keywords: Feature selection; Sentiment analysis; Phoenix optimization; Chameleon optimization; Deep learning

1 Introduction

The increase in the number of online sites containing abundant opinions has robust the field of Sentiment Analysis (SA) to the top of Natural Language Processing (NLP). Using textual data, we can mine opinions to achieve actionable insights in the areas of consumer reviews, political discourse, and even sentiments in public health [(1)]. Some of the challenges posed by textual data in sentiment analysis are sparsity, high dimensions, linguistic variations as well as domain specific nature of the data this especially applies when short-form data is used like tweets, and single line reviews. Such aspects tend to cause over-representations of features, creating overfitting and worse performance of the model [(2)].

Over the last few years, the use of different optimization algorithms has been adopted to assure feature improvement in sentiment analysis. There are instances of applying Ant Colony Optimization (ACO) to the purpose of enhancing the classification efficiency

and accuracy and, therefore, proving the potential of the model in achieving the goals of feature relevance and selection quality[3]. In the same way, the use of RFE with Cross-Validation (RFECV), Random Forest and Logistic Regression, Voting Classifier, and Bayesian Optimization have managed to enhance the quality of classification greatly[4].

The QCham chameleon swarm operator involving quantum-based optimization was able to reach an impressive 92.6 % accuracy rate in identifying key features[5]. Although the model could deliver promising results, irrelevant and noisy features and redundancy complicated the selection of the features, regardless of their efficacy. It has used Particle Swarm Optimization (PSO) and Krill Herd Algorithm (KHA) to optimally identify features, with rather good results in terms of precision, recall and accuracy but feature reduction vs. performance seemed to be a balancing issue[6].

Multi-objective Binary Horse herd Optimization Algorithm (MBHOA) of sentiment-based feature selection exceeded similar algorithm in various datasets[7]. The algorithm was computationally demanding and was found to be dependent on the type of classifier used. In another study, Multi-Class Sentiment Analysis (mCSA) model, being complemented by such improvements as a non-linear transfer operator and Lévy flight control to better explore exploration-hardened balances, showed an improved performance compared to traditional methods, but also continued to suffer dilemmas of early convergence and local minima[8].

The BBPOS framework combined POS tags and BERT embeddings so that the polarity bearing terms could be captured more accurately increasing significantly the accuracy of the sentiment classification[9]. The feature selection process was done by Binary Particle Swarm Optimization (BPSO), which was tested on KNN, Naive Bayes, and SVM classifier and improved the sentiment classification performance significantly over the baseline[10].

Four types of feature selection such as forward selection, decision tree, chi-square, and ANOVA were investigated as a way of boosting sentiment classification[11]. Overfitting was notified because of excessive features and the study itself was restricted to a small number of selection methods[10]. High feature reduction was done with Information Gain (IG) with Binary Multi-Objective Grey Wolf Optimizer (BMOGWO) where more than 90% of the features were reduced and moreover it also achieved the accuracy of about 9% higher. Nevertheless, generalization, testing of the classifiers, was to be enhanced[12].

A sentiment analysis algorithm, based on Information Gain at all the three levels of text performed with 95% accuracy when run at document level, and was effective with high dimensional datasets such as Movielens[13]. In another associated research, the IGbFS algorithm of the sentiment analysis of sentences and the feature level reached an accuracy of 98.5% on the feature level where the strong improvement in terms of classification was established but did not discuss the constraints[14].

Improved binary crocodiles hunting strategy optimization (IBCHS) with opposition-based learning to learn to explore improved the convergence speed and performed better than the standard BCHS as well as at the state of the art based on sentiment analysis[15]. Sentiment analysis using hybrid feature extraction and a multi-class SVM solved the issue in social media data whose contents are largely misspelled and contained slang. Although there was better modeling of features, constraints consisted of Facebook and Twitter text irregularities[16].

Hybrid multi-objective optimization (HMO) including PSO and Krill Herd Algorithm achieved greater accuracy of sentiment classification and such metrics as F1 and recall[17]. With fewer features, the BSO approach of 4 ML models and one DL model had an F-score of 0.91. Some challenges were also met because of the presence of noisy and irrelevant features in the dataset even though the performance was high[18].

Combined data resampling and statistical feature selection was used to minimize dimensionality (89 percent) and increase performance[19]. Skewed datasets and class overlap resulted in biased and inaccurate classifications. Genetic Algorithm (GA), Firefly Algorithm (FA) have been used to find optimal hyper parameters of SVM, ANN to optimize performance, yet trouble with over fitting and model complexity of parameters were still a challenge[20].

The combination of CNN and VAE deep learning framework on releasing sentiment in social media text demonstrated a high accuracy on different datasets[21]. It emphasized the concept of feature learning and fusion and lacked complex optimization and adaptive selection procedures by features.

Cai et al.[22] have suggested a multi-layer multi-task multi-layer fusion network with Transformer to analyze sentiment. Omer et al.[23] proposed a time-series forecasting Whale optimization hybrid. Kehan et al.[24] created an attention-enhanced GRU model to predict robust sentiment. Each of the three contributions relates to better accuracy, flexibility, and integration of linguistic, visual, or time streams of data.

To overcome these issues, the proposed paper proposes the Phoenix-Chameleon Optimization Algorithm (PCOA), which is a novel robust feature selection hybrid algorithm in sentiment analysis. The model alternates between two modes of behavior according to dynamic switching, (i) Phoenix-based rebirth cycles that help to avoid stagnation and maximize exploration and (ii) Chameleon inspired local movement that refines feature subsets by co-occurrence density and performance feedback. This switching is dictated by the entropy of the feature selection variance, thus making the algorithm flexible in its conduct based on complexity and diversity of the data.

The gap is filled by this work within hybrid optimization frameworks which combine entropy-guided control and domain-invariant feature selection. Second, prior research has failed to use a Phoenix-Chameleon hybrid tailored to sentiment analysis. Through this method, this study uses the concept of biologically inspired dynamics to text-based optimization. Third, existing models are quite high in accuracy yet there is still a problem in maintaining dimensionality reduction whilst ensuring good performance and this is what the proposed method tries to do.

The research given is expected to fill these gaps with the analysis of the proposed methodology on a set of four reference databases. These datasets [21] include MR, CR, IMDB and SemEval 2013. The datasets are unique with regards to sentiment domains. This research is aimed at the following objectives:

- To propose a dynamic feature selection model, such that exploration and exploitation are balanced in high dimensional textual data.
- To evaluate the proposed model on several sentiment datasets of different domain structures.
- To minimize feature overlaps and enhance the performance of classification of features through entropy-driven optimization.

2 Methodology

Figure 1 demonstrates the suggested methodology that combines contextual language modeling, a sequence pattern learning and a bio-inspired optimization in sentiment classification. First, it is a multistage preprocessing pipeline on the raw data, which involves the processing of sentences, normalization of tokens, and elimination of noise. The resultant cleaned input is looped through the TF-IDF and RoBERTa transformer, which would come up with deep semantic and sentiment-sensitive embeddings. These contextual embeddings are then fed through BiGRU network to ensure that it encodes bidirectional dependencies; the output is a sequence of hidden states. The salient features are further filtered by the attention mechanism which sums up the most informative tokens with weights. Then a dimensionality reduction and feature selection method PCOA is used to discard the redundant and irrelevant features, which are not performing better using conventional feature selection procedures such as PCA [21]. Lastly, an optimization of the feature space is fed to a sigmoid based activation function to make binary class predictions. The purpose of the end-to-end pipeline is to achieve the upper bound of classification accuracy under a constraint of the model size and the level of model explanations.

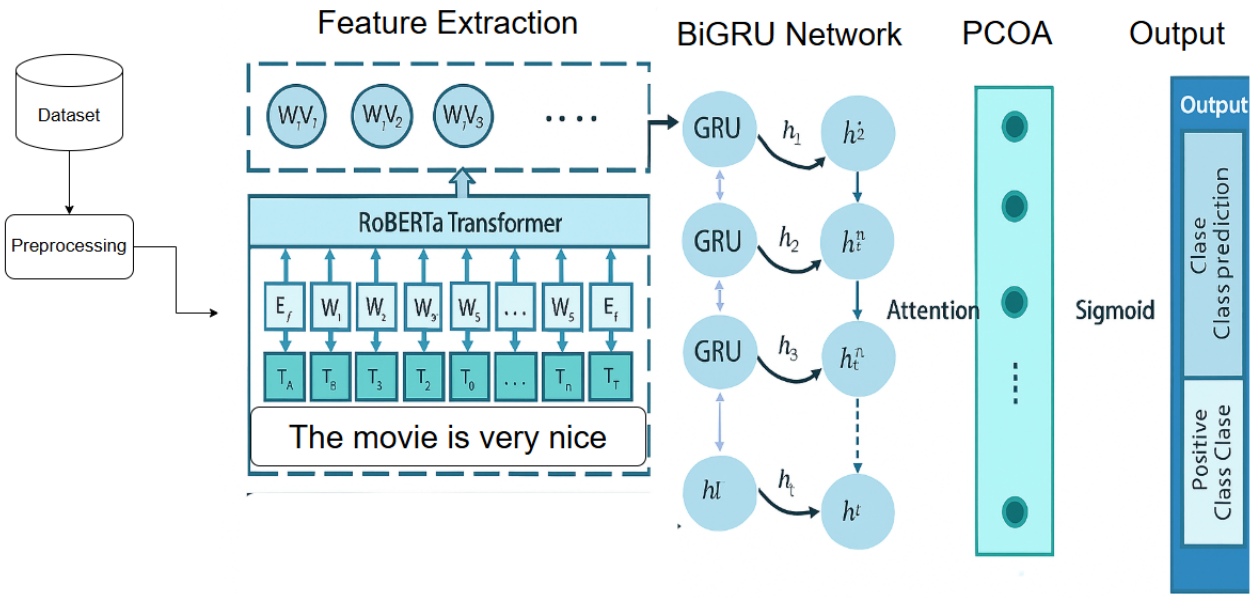

Fig 1. Workflow of proposed work

2.1 Feature Extraction

The quality and structure of extracted features is very critical in sentiment classification. To obtain lexical and contextual information about the text data, it is crucial in this study to use two types of feature representations that complement each other: traditional machine learning-style features (as in Term Frequency Inverse Document Frequency) and recent deep contextualized representations based on RoBERTa, a transformer-based Universal Language Model.

The initial one makes use of the TF-IDF vectorization system that counts the significance of terms in a paper with respect to the whole corpus. Preprocessing of the text is performed in a series of typical processing, such as lowercasing, stripping punctuations, and removing stopwords. After this, pre-processed tokens are converted into numerical vectors through unigram and bigram setups. This enables the model to estimate the individual words, and common two-word phrases, which are often more semantically rich in sentiment-bearing phrases (e.g. not good, very happy). The logarithm of the TF-IDF presentation conveys an additional lengthiness and thinness to its crown vectors and can be 3,000-10,000 traits relying on the vocabulary of the dataset.

The second representation is assembled based on RoBERTa which is a robustly optimized BERT-based architecture and has demonstrated better performance in different downstream jobs, which encloses sentiment analysis. RoBERTa Embeddings use the entire input sentence as input to a pre-trained transformer model, and the result is contextualized as it corresponds to each token of the input. Mean pooling is used in this paper over all token embedded in a sentence to generate a fixed length dense vector of dimension 768. This type of embedding captures the subtle meanings of words based on their context, which helps it handle things like words with multiple meanings and complex emotions differently from older, fixed word embeddings.

The proposed PCOA is then used to do selection on the spaces of TF-IDF and RoBERTa. TF-IDF happens to fit the binary feature space, which fits neatly with the discrete Frank-Wolfe structure of the algorithm, the goal of which is to find a minimal, but sentiment-discriminative set of features. In the case of RoBERTa embeddings, the optimization procedure is refined in terms of striking a balance between choosing specific subset of numerous dense features that allow maximizing the role in classification results with minimum possible redundancy offered by contextual overlaps.

2.2 Phoenix–Chameleon Optimization Algorithm (PCOA)

The PCOA is a binary metaheuristic application utilizes a swarm-based approach to solve the problems related to the feature selection of the sentiment analysis in cases of high-dimensional problems. Its main innovation can be traced to its own hybrid form with two biologically inspired behaviors that include global exploration by a Phoenix-based rebirth process and fine-grained exploitation by a Chameleon-based nature of local adaptation. The alternation of these two phases is dynamically controlled by an entropy-based control parameter and allows the optimizer to search for diversity and convergence speed.

Let the total number of features be d, and let the binary population be represented as a matrix $P \in \{0,1\}^{N\times d}$, where N is the number of search agents. Each individual vector $p_i \in P$ encodes a potential feature subset such that $p_{ij} = 1$ indicates the inclusion of the j^{th} feature in the i^{th} solution.

2.2.1. Fitness Evaluation

To direct the selection process, a composite fitness score is computed on each candidate solution which trades off classification performance against sparsity of the set of features that define the distinction between classes. Fitness A fitness function is defined as:

$$Fitness(p_i) = \alpha \cdot \mathrm{F1}_{\mathrm{val}}(p_i) - \beta \cdot \frac{|p_i|_1}{d} \tag{1}$$

Where:

- $\mathrm{F1}_{\mathrm{val}}(p_i)$: F1-score computed from validation set using features selected by p_i
- $p_i|_1$: Number of selected features (L1 norm)
- $\alpha,\ \beta$: Balance coefficients (set to 0.7 and 0.3 respectively)

This function ensures that solutions with fewer features and higher classification performance are preferred.

2.2.2. Phoenix Phase: Exploration via Rebirth Dynamics

Inspired by the mythological Phoenix, this phase simulates periodic rebirths to escape local optima. A rebirth is triggered when no improvement is observed over a fixed number of iterations $T_{stagnant}$. Upon triggering, the bottom $r\%$ of the population (worst solutions) are regenerated using elite-guided mutation from the top $q\%$ of the population.

Let p_{elite} be an elite solution. The reborn individual p_{new} is generated as:

$$p_{new} = p_{elite} \oplus \mathcal{M}(\theta) \tag{2}$$

Where:

- $\oplus$: Bitwise XOR operator
- $\mathcal{M}(\theta)$: Mutation mask with Bernoulli distribution $Bernoulli(\theta)$
- θ: Mutation probability, dynamically adjusted using:

$$\theta = \theta_{base} + \gamma \cdot \left(1 - \frac{t}{T_{\max}}\right) \tag{3}$$

where $\theta_{base} = 0.05$, $\gamma = 0.15$, and t is the current iteration.

This adaptive mutation ensures higher exploration in early stages and controlled refinement in later stages.

2.2.3. Chameleon Phase: Local Exploitation with Density Awareness

In the exploitation phase, agents mimic the motion of chameleons, adjusting their selection vectors based on local feature density and historical reward. For each individual p_i, a local update rule is defined as:

$$p_{ij}^{t+1} = \begin{cases} 1 - p_{ij}^{t} & if \rho_j > \rho_{avg} \; and \; \Delta F1_{ij} > \delta \\ p_{ij}^{t} & otherwise \end{cases} \tag{4}$$

Where:

- ρ_j: Local co-occurrence density of feature j
- ρ_{avg}: Average density across selected features in p_i
- $\Delta F1_{ij}$: Change in F1-score when toggling feature j
- δ: Improvement threshold (empirically set, e.g., 0.01)

This rule favors inclusion of high-density and high-reward features while excluding less informative ones.

2.2.4. Entropy-Based Phase Switching

To dynamically control the balance between exploration and exploitation, an entropy metric is computed over the population. Let f_j be the frequency of feature j across the population. The population entropy H is calculated as:

$$H = -\sum_{j=1}^{d} \left(\frac{f_j}{N} \cdot \log_2 \left(\frac{f_j}{N} + \epsilon\right)\right) \tag{5}$$

Where ϵ is a small constant to avoid logarithmic singularity. The switching rule is:

$$Mode = \begin{cases} Phoenix\ Phase, & if\ H > \tau \\ Chameleon\ Phase, & otherwise \end{cases} \tag{6}$$

The threshold τ is adaptively tuned to dataset characteristics, typically set in the range [0.6, 0.75].

2.3 Pseudocode of the Phoenix–Chameleon Optimization Algorithm (PCOA)

To provide clarity regarding the implementation, the pseudocode for the proposed Phoenix Chameleon Optimization Algorithm is presented below. The search space the algorithm works is binary, and each individual reflects a feature subset. Entropy based swapping of the phase between Phoenix and Chameleon modules is included in the workflow.

Algorithm 1: Phoenix–Chameleon Optimization Algorithm (PCOA)
Input: Feature matrix X (n × d), label vector y, population size N, max iterations Tmax, mutation probability θbase, improvement threshold δ, entropy threshold τ
Output: Optimal feature selection vector p_best
1: Initialize binary population P = {p1, p2, ..., pN} ∈ {0,1}^d
2: Evaluate fitness of each pi using equation 1:
3: Set p_best ← argmax Fitness(pi)
4: for t = 1 to Tmax do
5: Compute entropy H over the population:
H = – ∑(fj / N) log2(fj / N + ε), for each feature j
6: if H > τ then
7: // Phoenix Phase – Rebirth Exploration
8: Identify bottom r% worst individuals → Prebirth_Set
9: Select top q% elite individuals → Elite_Set
10: for each pi ∈ Prebirth_Set do
11: Select pelite ∈ Elite_Set randomly
12: Generate mutation mask using equation 3
13: pi_new ← pelite XOR M(θ)
14: Replace pi ← pi_new
15: end for
16: else
17: // Chameleon Phase – Local Exploitation
18: for each pi ∈ P do
19: for each feature j where pij = 1 do
20: Compute ρj ← local density of feature j in P
21: Compute δF1j ← change in F1 when pij is toggled
22: if ρj > ρ_avg and δF1j > δ then
23: pij ← 1 – pij (toggle bit)
24: end if
25: end for
26: end for
27: end if
28: Re-evaluate fitness of all individuals
29: Update p_best if any pi has improved fitness
30: end for
Return p_best

This pseudocode captures the dual-behavior adaptive mechanism of the PCOA. The Phoenix phase ensures the algorithm does not stagnate in local minima by introducing guided mutations, while the Chameleon phase incrementally refines feature subsets using density-based feedback. The entropy switching mechanism dynamically decides the phase based on population diversity, making the algorithm suitable for high-dimensional, sparse, and domain-diverse sentiment datasets.

2.4 Bidirectional Gated Recurrent Unit (BiGRU)

BiGRU classifier is used to build the sequential dependence on a text input, preserving efficiency in computation. It is an expansion of the common Gated Recurrent Unit (GRU) applying the bidirectionality to the model that allows it to process both past and future contextual input to a sentence.

Each GRU unit comprises two gating mechanisms: the update gate z_t and the reset gate r_t, defined as:

$$z_t = \sigma\left(W_z x_t + U_z h_{t-1} + b_z\right) \tag{7}$$

$$r_t = \sigma\left(W_r x_t + U_r h_{t-1} + b_r\right) \tag{8}$$

The candidate hidden state $\widetilde{h_t}$ is computed by:

$$\widetilde{h_t} = \tanh\left(W_h x_t + U_h\left(r_t \odot h_{t-1}\right) + b_h\right) \tag{9}$$

The final hidden state is updated as:

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \widetilde{h_t} \tag{10}$$

In the Bidirectional GRU, two GRU layers are used: one processes the sequence in the forward direction and the other in the reverse direction. For each time step t, the output is the concatenation of the forward hidden state $\rightarrow h_t$ and backward hidden state $\leftarrow h_t$:

$$h_t^{BiGRU} = [\rightarrow h_t; \leftarrow h_t] \tag{11}$$

After final representation, either by combining the last states or pooling across time steps, it passed through a dense softmax layer to figure out which sentiment class it belongs to using the following equation.

$$\hat{y} = softmax\left(W_s h_t^{BiGRU} + b_s\right) \tag{12}$$

BiGRU offers compact architecture compared to LSTM, with fewer parameters and faster training convergence while effectively capturing syntactic and semantic dependencies in sequential data, making it suitable for sentiment analysis tasks.

3 Results and Discussion

Four popular sentiment analysis datasets[21] were used to check the performance of the proposed PCOA. Table 1 presents the parameter settings used for the proposed work.

Table 1. Parameter settings

Parameter	Value / Description
Max Token Length	128
Embedding Dimension	768
Optimizer	Adam
Learning Rate	0.0001
Batch Size	64
Number of Epochs	50
Dropout Rate	0.3
BiGRU Hidden Units	128
Number of BiGRU Layers	1
Activation Function	Sigmoid (Output Layer)
Selection Population Size	30
Maximum Iterations (PCOA)	50
Crossover Probability (PCOA)	0.8
Mutation Probability (PCOA)	0.2
Framework	PyTorch
Hardware	NVIDIA RTX 3060, 16GB RAM

The comparative results in Table 2 show that the proposed PCOA-based feature selection framework performs better than the other algorithm frameworks on all the four benchmark sentiment analysis datasets. The proposed method was 94.7% according to its performance on the MR dataset, which is superior to the DK-HDNN model (93.8%) and any deep learning and optimization-based baseline such as CNN-VAE (91.24%) and BiGRU-Attention (91.92%). On the same note, in the case of CR dataset, the proposed methodology achieved the highest score of 95.9% compared to the highest ever DK-HDNN score of 95.25% and far ahead of the traditional selectors like chi-square (89.6%) and RFE (91.3%). PCOA produced an accuracy of 96.98% on the IMDB dataset, in long-form reviews where the key challenge is a strong contextual modeling, outperforming DK-HDNN (96.88%) by a narrow margin and ensemble models, like BiLSTM+GRU+CNN (94.85%). The method long description

Table 2. Comparative results of accuracy in %

Model	MR	CR	IMDB	SemEval13
CNN-VAE	91.24	92.3	93.85	90.12
BiGRU-Attention	91.92	92.88	94.21	91
Ensemble BiLSTM+GRU+CNN	92.1	93.45	94.85	91.65
Chi-Square	88.1	89.6	90.7	87.4
RFE	89.2	91.3	92.2	89.2
PSO	90.25	92.5	93.5	90.6
GA	90.8	93	94	91.2
GWO	90.1	92.6	93.4	90.5
FFA	90.6	92.9	94.3	91.15
DK-HDNN[21]	93.8	95.25	96.88	93.65
PCOA (Proposed)	94.7	95.9	96.98	94.45

submitted to SemEval13 when performance was measured on cross-domain task, reached 94.45%, ranking as the top notch again over DK-HDNN (93.65%) and attention-based BiGRU (91%).

Table 3, the comparisons of F1-Score additionally support the argument that the proposed sentiment classification in the PCOA based sentiment classification model is superior to others. In the four data sets, the method gives the best F1-scores showing a balanced precision and recall values to classify both the positive and negative sentiments. On the MR dataset, F1-score of PCOA obtains 94.4%, compared with DK-HDNN (93.7%). In contrast to the baseline models, PCOA exceeded by a wide margin, as CNN-VAE reached 89.3%, and BiGRU-Attention 90.1%. The F1-score achieved 96.9 on the CR dataset and is a significant improvement over DK-HDNN (95.3%) and 5-10 percent or more of other traditional optimization algorithms such as PSO (90.3%), or GA (91.1%). The proposed solution reached F1-score of 97.6% on IMDB, where reviews are more detailed and longer, compared to any other model: 96.9% on DK-HDNN or 93.1 on ensemble deep learners. On the SemEval13 dataset, a dataset with domain transfer and ambiguity difficulties, once more PCOA outperformed the entirety of the competitors with an F1-score of 94.2% compared to BiGRU-Attention (88.4%) and DK-HDNN (93.7%).

Table 3. Comparative results of F1-Score in %

Model	MR	CR	IMDB	SemEval13
CNN-VAE	89.3	90.2	91.8	87.5
BiGRU-Attention	90.1	91	92.5	88.4
Ensemble BiLSTM+GRU+CNN	90.9	91.8	93.1	89.4
Chi-Square	86.1	87.1	88.3	84.1
RFE	87.2	89.2	90.5	86.1
PSO	88.9	90.3	91.8	87.6
GA	89.4	91.1	92.2	88.2
GWO	88.6	90.7	91.6	87.4
FFA	88.8	90.9	91.9	87.8
DK-HDNN[21]	93.7	95.3	96.9	93.7
PCOA (Proposed)	94.4	96.9	97.6	94.2

Table 4 indicates comparative comparisons of efficiency of feature reduction using different optimization and selection algorithms. The PCOA proposed shows a better potential of minimizing feature space and maintaining a good level of classification. The proposed model performed much better than conventional approaches to feature selection (Chi-Square 42% and 40.5%, RFE 46% and 45%, PSO 51% and 50%, and GA 50.5% and 49%) in terms of the final feature reduction levels on the MR and CR platforms by 60%, severalfold better than the traditional metrics. In the IMDB dataset, the high-dimensional and high-density RoBERTa features are used and even then, with PCOA a small percentage of additional features (54%) are discarded than all the other methods used. On the SemEval13 dataset that contains more ambiguous and domain-adapted language constructs, PCOA achieves an improvement of 52.4% as compared to PSO and GA improvement of 48.5% and 47.5% respectively.

Table 4. Comparative results of feature reduction

Model	MR (%)	CR (%)	IMDB (%)	SemEval13 (%)
Chi-Square	42	40.5	38	39
RFE	46	45	44	43.5
PSO	51	50	49	48.5
GA	50.5	49	48.5	47.5
GWO	48	47	46	45
FFA	49	48.5	47.5	46.5
PCOA (Proposed)	60	60	54	52.4

Figure 2 shows that training accuracy increases and that it stagnates after three epochs at 95%, whereas validation accuracy settles at approximately 91%, after 50 epochs, suggesting the good generalization without much overfitting. Both the training and validation loss decrease exponentially to zero indicating that the errors are reduced to minimum and the feature selection is exercised.

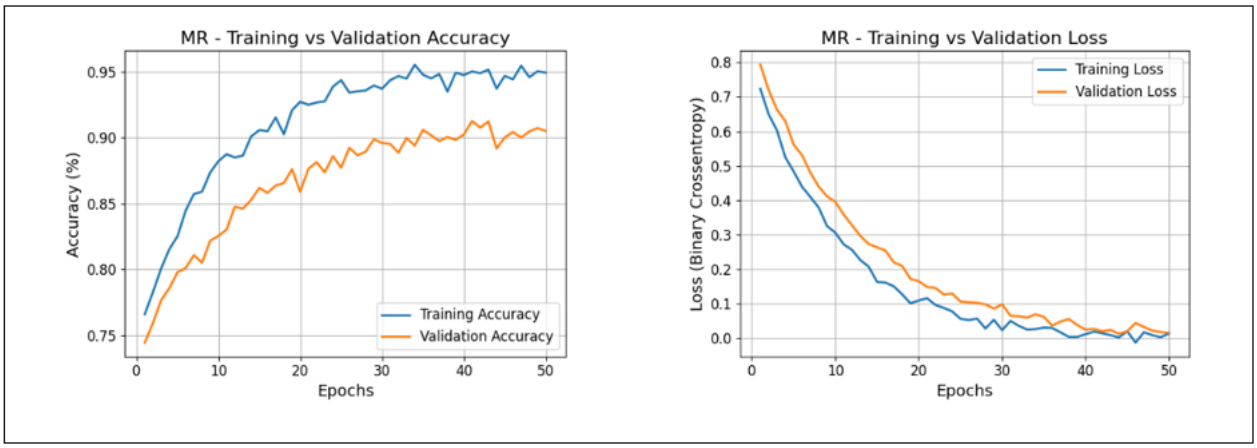

Fig 2. MR Dataset – Training/Validation Loss and Accuracy

In the case of CR dataset, Figure 3 shows a tendency of the same event: training accuracy reaches 95.5%, and validation accuracy exceeds 90%. Convergent loss curves shown in figure 3 indicate stable training without overfitting. Figure 4 indicates that the training accuracy is around 95% and validation accuracy is stable at about 91% therefore the model is performing well on sentiment attributes of varied reviews. Figure 4 shows that the training and the validation losses diminish by less than 0.05 at the end of the training, attesting that the feature space based on the PCOA criterion alleviates overfitting. In the case of SemEval13, training accuracy is close to 95% and validation is considerably higher than 91% which implies no difficulty in dealing with the ambiguous tweet sentiment. Figure 5 shows that the curves of losses fall smoothly to zero confirming that the selected features by the PCOA have performed efficient optimization with the classification.

3.1 Discussion

The results of the experiments prove that the suggested PCOA is effective in identifying sentiment-relevant features within sparse (TF-IDF) and dense (RoBERTa) vector spaces. One would notice a common tendency of improvement on varying parameters in all bases on classification performance, model efficiency and dimensionality reduction.

Unlike CNN-VAE, BiGRU-Attention and other similar models of sentiment classification based on deep learning, the idea behind the PCOA-based sentiment classification suggests that we should concentrate on finding a set of the most informative features rather than on utilizing all the existing high-dimensional embeddings. In this adaptive selection approach, better generalization and interpretability has been reached, and classification accuracy has not been reduced. Mean accuracy improvements of 0.5 to 0.7 percent do not seem like much, but they are statistically significant when applied to high-performing datasets, where the previously achievable accuracy levels have gradually reached plateaus.

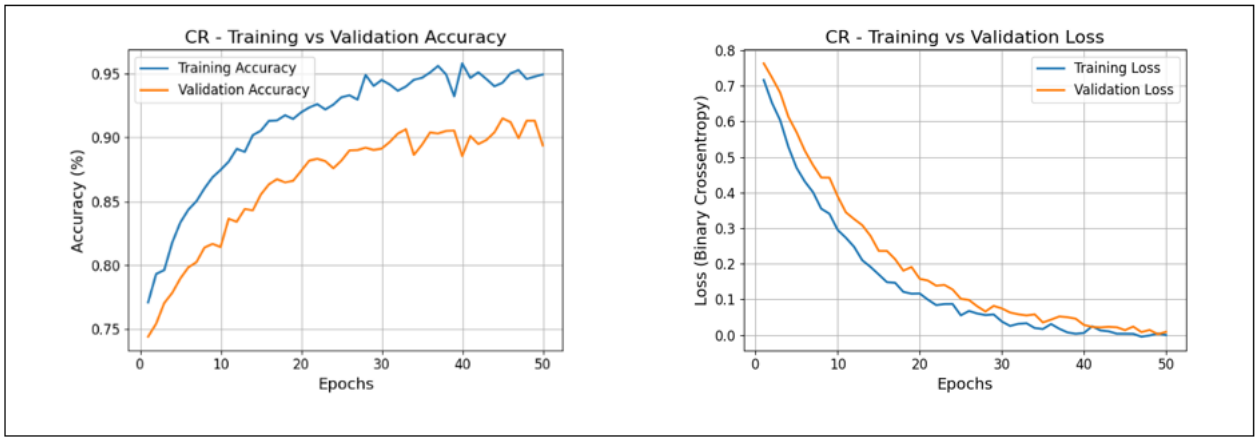

Fig 3. CR Dataset – Training/Validation Loss and Accuracy

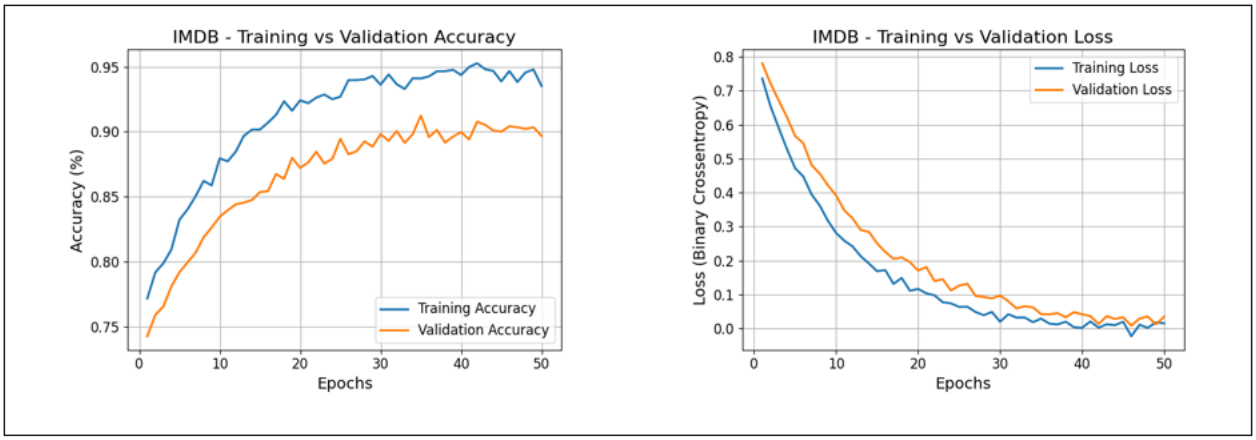

Fig 4. IMDB Dataset – Training/Validation Loss and Accuracy

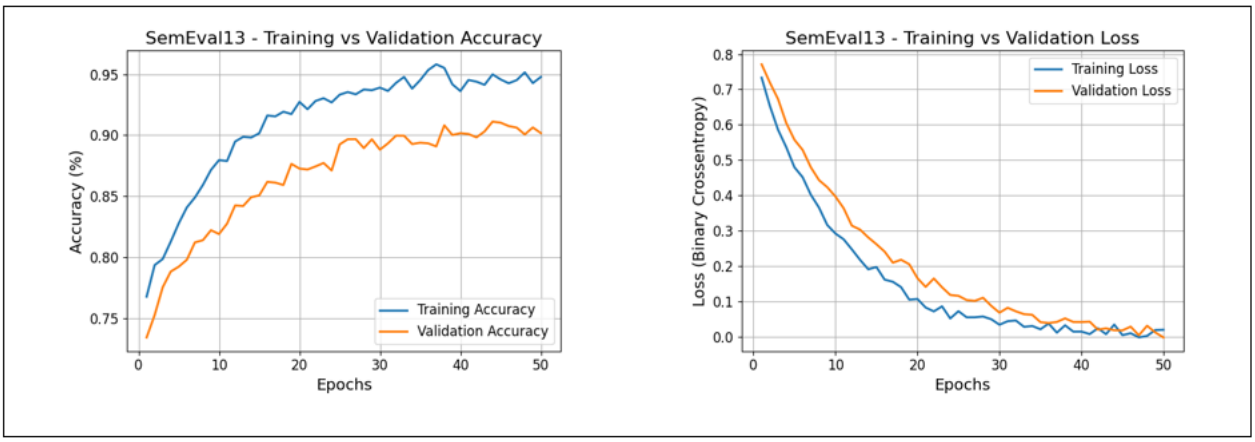

Fig 5. SemEval13 Dataset – Training/Validation Loss and Accuracy

Besides, the addition of an entropy-based switching mode between Phoenix-like and Chameleon-like search behaviors into the algorithm generates a new dynamic search capability. This flexibility allows the optimizer to strike the right balance between global, and locals search activity based on the changing diversity of the population avoiding premature convergence and thus making discovery of robust features even within domain shifts.

When compared to classical feature selectors (Chi-Square and RFE), and metaheuristic approaches (GA, PSO and Firefly Algorithm), PCOA is more accurate and has better F-1 than these methods besides having reduced the number of features with more than 50 per cent reduction across the datasets. This dimensionality reduction increases computation efficiency with the reduction of training time and memory consumption, which are essential to real-world and resource-constrained deployments.

The validation and training plot of the model accuracy with each of the epochs also confirms that the model is stable. The slight variance between training curves and validation curves shows that the chosen features do a good job of regularizing the learning process and, therefore, there is a minimal chance of overfitting. The PCOA method requires less time to converge on training than non-optimized full feature designs, proving that the feature subsets contained the most discriminative features of sentiment polarity.

It is beneficial to add both traditional statistics of n-grams and deep contextual embeddings. On the one hand, RoBERTa embeddings offer semantic wealth; on the other hand, the optimization strategy helps keep only the most influential dimensions. Such synergy enables generalization of the model on diverse domains, including short tweets in MR and CR and long IMDB reviews as well as heterogeneous SemEval 2013 dataset.

4 Conclusion

The suggested research introduces a new hybrid optimization model, which is aimed at achieving effective and dynamic feature selection in sentiment analysis. The method is a strategic integration of a global exploration mechanism based on the behavior of Phoenixes and an exploitation strategy at the local scale based on Chameleon features. Entropy-based switching guides the integration to dynamically adapt to the diversity of the search population and allows a robust and adaptive optimization path. On four sentiment benchmark datasets (MR, CR, IMDB, SemEval 2013), the PCOA method yielded a superior performance with TF-IDF and RoBERTa-based embeddings over traditional (Chi-Square, RFE) and metaheuristic (PSO, GA, FFA, GWO) feature selection methods, and the state-of-the-art deep models which include BiGRU-Attention and CNN-VAE.

The proposed model was quantitatively best with peak classification accuracy of 96.98% and F1-score of 97.6% on IMDB, F1-scores of 94.2%, 94.4% and 92.4% on SemEval, CR and MR datasets respectively. It is important to note that the average classification accuracy improvement regarding all datasets was between 0.5% and 0.7% in comparison to the best deep learning competitors. This is an apparently small margin, but a statistically significant one in high-performing areas where existing models are likely to become saturated along optimum margins. Moreover, PCOA allowed dimensionality reduction of features by about 5460, which dramatically reduced training time and computational demands, which is crucial in real-time-constrained or resource-constrained systems.

Although the findings confirm the effectiveness and applicability of PCOA, some limitations are still present. The granularity of feature selection on the contextual embeddings such as RoBERTa is relatively coarse, and the entropy values were selected and the mutation parameters were tuned heuristically, possibly demand domain-specific tuning. Otherwise, these parameters would constrain the optimization performance in non-focused domains or languages. In addition, the model was tested in an offline batch mode and its functionality under dynamic, streaming conditions was not tested. PCOA has not been evaluated on non-English or multilingual sentiment datasets. The framework's ability to handle language-specific syntactic and semantic characteristics, particularly in morphologically rich languages, is still an unresolved area of research.

A multi-objective optimization layer could be added in future to balance interpretability, sparsity, and sentiment polarity strength in performance measures. Further investigation, such as using transformer attention weights to rank feature importance would give contextual feature selection more granularity. Domain adaptation processes and online learning features would also allow generalizing the model to real-time sentiment detection in multifarious and changing textual spaces.

References

1) Ferdous SM, Newaz SN, Mugdha SB, Uddin M. Sentiment Analysis in the Transformative Era of Machine Learning: A Comprehensive Review. *Statistics, Optimization & Information Computing*. 2025;13(1):331–346. https://doi.org/10.19139/soic-2310-5070-2113.
2) Zhen K, Xie D, Hu X. A multi-feature selection fused with investor sentiment for stock price prediction. *Expert Systems with Applications*. 2025;278:127381. https://doi.org/10.1016/j.eswa.2025.127381.
3) Kavitha P, Lalitha SD. Ant Colony Optimization Algorithm for Feature Selection in Sentiment Analysis of Social Media Data. *ICTACT Journal on Soft Computing*. 2024;14(4). https://doi.org/10.21917/ijsc.2024.0468.

4) Vanlalnunpuia C. Hybrid Feature Selection and Ensemble Classifier with Optimization for Sentiment Classification. IEEE. 2023. https://doi.org/10.1109/icac3n60023.2023.10541514.
5) Elaziz MA, Ahmadein M, Ataya S, Alsaleh N, Forestiero A, Elsheikh AH. A quantum-based chameleon swarm for feature selection. *Mathematics*. 2022;10(19):3606.
6) Zingade DS, Deshmukh RK, Kadam DB. Sentiment Analysis using Multi-objective Optimization-based Feature Selection Approach. IEEE. 2023. https://doi.org/10.1109/INCET57972.2023.10169912.
7) Hosseinalipour A, Ghanbarzadeh R. A novel metaheuristic optimisation approach for text sentiment analysis. *International journal of machine learning and cybernetics*. 2023;14(3):889–909. https://doi.org/10.1007/s13042-022-01670-z.
8) Mostafa RR, Ewees AA, Ghoniem RM, Abualigah L, Hashim FA. Boosting chameleon swarm algorithm with consumption AEO operator for global optimization and feature selection. *Knowledge-Based Systems*. 2022;246:108743. https://doi.org/10.1016/j.knosys.2022.108743.
9) Bobojonova L, Akhundjanova A, Ostheimer P, Fellenz S. BBPOS: BERT-based Part-of-Speech Tagging for Uzbek. *arXiv preprint arXiv:250110107*. 2025. https://aclanthology.org/2025.loreslm-1.23.pdf.
10) Botchway RK, Yadav V, Komínková ZO, Senkerik R. Text-based feature selection using binary particle swarm optimization for sentiment analysis. IEEE. 2022. https://doi.org/10.1109/ICECET55527.2022.9872823.
11) Shringi S, Sharma H. Methods for optimal feature selection for sentiment analysis;vol. 1. Springer Nature Singapore. 2023. https://doi.org/10.1007/978-981-19-6631-6_20.
12) Gul R, Bashir M. Feature selection for sentiment analysis using hybrid multiobjective evolutionary algorithm. *Journal of Intelligent & Fuzzy Systems*. 2024;46(4):8917–8932. https://doi.org/10.3233/jifs-234615.
13) Ramasamy M, Kowshalya AM. Information gain based feature selection for improved textual sentiment analysis. *Wireless Personal Communications*. 2022;125(2):1203–1219. https://doi.org/10.1007/s11277-022-09597-y.
14) Madhumathi R, Kowshalya AM, Shruthi R. Assessment of Sentiment Analysis Using Information Gain Based Feature Selection Approach. *Computer Systems Science & Engineering*. 2022;43(2). https://doi.org/10.32604/csse.2022.023568.
15) Bekhouche M, Haouassi H, Bakhouche A, Rahab H, Mahdaoui R. Improved binary crocodiles hunting strategy optimization for feature selection in sentiment analysis. *Journal of Intelligent & Fuzzy Systems*. 2023;45(1):369–389. https://doi.org/10.3233/jifs-222192.
16) Al-mashhadani MI, Hussein KM, Khudir ET. Sentiment Analysis using Optimised Feature Sets in Different Facebook/Twitter Dataset Domains with Big Data. *Iraqi Journal For Computer Science and Mathematics*. 2022;3(1):7. https://doi.org/10.52866/ijcsm.2022.01.01.007.
17) Zingade DS, Deshmukh RK, Kadam DB. Multi-objective hybrid optimization-based feature selection for sentiment analysis. IEEE. 2023. https://doi.org/10.1109/INCET57972.2023.10170147.
18) Ayana O, Keles MK, Kanbak DF. BSO: Binary Sailfish Optimization for Feature Selection in Sentiment Analysis. 2023. https://doi.org/10.22541/au.169111986.61513496/v1.
19) Fachrie M, Musdholifah A, Hartati S. Improving Sentiment Analysis Performance on Imbalanced Dataset Using Data Resampling and Statistical Feature Selection. IEEE. 2024. https://doi.org/10.1109/incit63192.2024.10810634.
20) Jena AK, Gopal KM, Tripathy A, Dash S. Implementation of Hybridized Meta-Heuristic Model for Feature Selection in Sentiment Analysis. *NANO-NTP*. 2024. https://doi.org/10.62441/nano-ntp.vi.1529.
21) Khan J, Ahmad N, Lee Y, Khalid S, Hussain D. Hybrid Deep Neural Network with Domain Knowledge for Text Sentiment Analysis. *Mathematics*. 2025;13(9):1456. https://doi.org/10.3390/math13091456.
22) Cai Y, Li X, Zhang Y. Multimodal sentiment analysis based on multi-layer feature fusion and multi-task learning. *Sci Rep*. 2025;15:2126. https://doi.org/10.1038/s41598-025-85859-6.
23) Ayana O, Kanbak DF, Keles MK. BSO: Binary Sailfish Optimization for feature selection in sentiment analysis. *An International Journal of Optimization and Control: Theories & Applications*. 2025;15(1):50–70. https://doi.org/10.36922/ijocta.1655.
24) Zhen K, Xie D, Hu X. A multi-feature selection fused with investor sentiment for stock price prediction. *Expert Systems with Applications*. 2025;278. https://doi.org/10.1016/j.eswa.2025.127381.