International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 22s, 2025
https://theaspd.com/index.php

Optimizing Contextual Embedding-Based Text Classification: A Comparative Analysis Of PSO And GSO Feature Selection Approach

PRIUSHA NARWARIYA[1], SUSHEEL KUMAR TIWARI[2]
[1,2]Department of Computer Science & Engineering, Madhyanchal Professional University, Bhopal, India

Abstract: *Text classification occupies a central position in natural language processing, with feature engineering and optimization being critical levers for enhancing predictive power. Conventional pipelines frequently falter in reconciling the competing demands of reduced feature space, model transparency, and accuracy. This study presents a novel hybrid architecture that leverages contextual embeddings within a fusion-centered feature construction, fine-tuned through the complementary heuristics of Particle Swarm Optimization (PSO) and Glowworm Swarm Optimization (GSO).The proposed framework orchestrates a cohesive feature assemblage from term-frequency-inverse-document-frequency (TF-IDF), pre-trained word embeddings, and condensed representations generated by singular value decomposition (SVD) and principal component analysis (PCA). Subsequent dimensional refinement is governed by a bi-level optimization schema that alternates between PSO and GSO. Benchmarks on the curated train and test splits indicate pronounced superiorities relative to a baseline that pairs TF-IDF with logistic regression. The PSO-tuned ensemble attained an accuracy of 96.82%, a macro F1-score of 96.77%, and an area under the receiver operating characteristic curve (AUC-ROC) of 0.987, while the GSO-tuned variant recorded 96.95%, 96.91%, and 0.989, respectively. Temporal profiling under proxy settings in Google Colab, with GPU support, disclosed that the PSO variant required 39.0 minutes of processing, whereas the GSO variant consumed 41.5 minutes, thereby affirming their operational viability.The integration of confusion matrices, disaggregated performance metrics, AUC-ROC visualizations, log-loss trajectories, and hyperparameter convergence profiles convincingly established GSO as the leading classifier, despite its marginally elevated computational burden. This work highlights the potency of swarm-intelligence-driven optimization in refining contextual-embedding-based text categorization and indicates the method's broader applicability to NLP domains that mandate elevated accuracy and resilient feature-selection capabilities.*

Keywords: *Text Classification, Contextual Embeddings, Particle Swarm Optimization, Glowworm Swarm Optimization, Feature Selection, AUC-ROC, Execution Time Analysis*

INTRODUCTION

Text classification continues to serve as a linchpin task within Natural Language Processing, facilitating a diverse array of applications, including sentiment detection, topic assignment, and document filtering. The exponential expansion of unstructured text—pervading news aggregators, social media platforms, and consumer reviews—renders the deployment of accurate and scalable classifiers essential [1, 2].

The advent of contextual embeddings, notably BERT and its derivatives (RoBERTa, DeBERTa), has substantially refined the representation of text, enabling the capture of subtle semantic relationships and translating into pronounced performance improvements across categorization benchmarks [3, 4]. Concurrently, embedding methodologies continue to evolve, incorporating instruction-based fine-tuning, cross-linguistic support, and contrastive learning paradigms, thereby extending their utility across heterogeneous tasks and domains [5].

Nevertheless, these embeddings frequently exhibit elevated dimensionality, which intensifies algorithmic complexity and elevates the risk of overfitting; thus, judicious feature selection and fusion techniques assume critical importance [6]. Among the promising strategies, swarm intelligence, and specifically Particle Swarm Optimization, has emerged as a powerful mechanism to distill feature subsets within expansive input spaces. Its parallelized, non-gradient nature aligns well with classification demands, and recent algorithmic

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 22s, 2025
https://theaspd.com/index.php

adaptations have refined its balance of exploration and exploitation through guided and adaptive heuristics [7, 8].

Complementary swarm paradigms, such as Glowworm Swarm Optimization, introduce distinct search heuristics that prioritize the retention and dissemination of high-quality candidate solutions—an attribute that may prove advantageous within the heterogeneous, multi-modal feature landscapes characteristic of contemporary NLP challenges. Current advances in hybrid feature selection and fusion—such as particle swarm optimization (PSO) enhanced by local-search heuristics—show significant gains in convergence speed and subset compactness [9]. However, applying swarm-intelligent methods to the fusion of contextual embeddings for text classification has received limited empirical attention, leaving an open and potentially high-impact research trajectory.

The major contributions of the study are:

- Combines TF-IDF with contextual embeddings (BERT, DistilBERT, RoBERTa), compressed via SVD + PCA into a salient 512-dimensional representation.
- Implements and contrasts with PSO and a custom GSO as wrapper-based feature selection algorithms to navigate this fused space efficiently.
- Benchmarks against a TF-IDF + Logistic Regression baseline using metrics such as accuracy, macro-F1, AUC-ROC, log-loss, execution time, and convergence behavior.
- Evaluates computational overhead versus improvements in classification effectiveness, highlighting GSO's enhanced performance under higher optimization costs.

The rest of the paper is organized as follows: Section 2 surveys key progress in contextual embeddings, feature selection techniques for natural language processing (NLP), and swarm-intelligence optimization algorithms, thereby establishing theoretical groundwork for the proposed framework. Section 3 delineates the methodology, detailing the data preprocessing pipeline, the feature fusion approach, dimensionality reduction stages, and the bespoke design of Particle Swarm Optimization (PSO) and Glowworm Swarm Optimization (GSO) tailored for efficient feature subset identification. Section 4 explicates the experimental arrangement, specifying dataset attributes, configuration of experimental trials, guiding assumptions, and parameter settings for baseline models. Section 5 articulates the results and accompanying discussion, delivering a comparative performance evaluation of baseline, PSO, and GSO techniques, in addition to ablation studies, convergence behavior, and time-complexity analysis. The paper culminates in Section 6, which recapitulates principal contributions, reflects methodological constraints, and outlines prospective avenues for research, including the scaling of methods to multilingual architectures, the integration of domain-specific embeddings, and applicability in low-resource language contexts.

RELATED WORK

The exponential proliferation of digital data has positioned text classification as a central challenge within natural language processing (NLP). It serves as the analytic backbone for a diverse set of critical applications, including the detection of misinformation, monitoring of online harassment, retrieval of legal documents, and the interpretation of healthcare narratives. The introduction of deep contextual embeddings, epitomized by BERT and RoBERTa, has led to substantial gains in classification accuracy. Nonetheless, the persistent issues of high-dimensional feature representation, feature redundancy, and escalating computational demands remain. In response, the academic community has turned to feature fusion techniques that harmonize conventional lexical cues—such as term frequency-inverse document frequency (TF-IDF) and n-grams—with contextual embeddings derived from transformer architectures. The resulting hybrid representations capitalize on the complementary advantages of sparse interpretability and deep semantic insight.

Alongside developments in embedding space, swarm intelligence paradigms, including Particle Swarm Optimization (PSO) and Glowworm Swarm Optimization (GSO), have attracted increasing interest for feature selection in text classification. The PSO algorithm, whose mechanism is patterned on the adaptive foraging of avian and aquatic species, is lauded for rapid convergence and a robust global exploration capability. GSO, by contrast, models the foraging dynamics of glowworms, who modulate luciferin luminescence to guide collective movement, thus introducing a mechanism for functional diversity and resilience to local optima. Both strategies have been validated in heterogeneous domains—ranging from intrusion detection to sentiment classification and biomedical document analysis—yet systematic, comparative investigations of PSO and GSO within unified text classification architectures remain scarce in the scientific literature.

Recent transformer encoders—most notably BERT, RoBERTa, and XLNet—continue to serve as foundational models for text classification, prompting a growing body of research aimed at hybridizing these contextual embeddings with traditional lexical representations such as TF-IDF. The 2024 EMNLP conference reported that co-training BERT embeddings alongside TF-IDF cluster centroids markedly enhanced feature quality in the short-text regime, underscoring that basic feature concatenation fails unless the two modalities are iteratively synchronized [10]. During the subsequent 2025 NAACL shared task, a hybrid system employed a hard-shift ensemble of TF-IDF latents, a fine-tuned BERT pipeline, and a voting classifier, yielding statistically significant improvements in multilabel affect detection [11]. In long-document contexts, the LCF-IDF weighting scheme outperformed both conventional TF-IDF and long-document transformer baselines across multiple corpora, underscoring the continued utility of lexical-contextual complementarity [12].

Forward of purely lexical fusion, the authors proposed a "re-fusion" network that synergistically augments XLNet with GCNII graph embeddings, arguing that the integration injects structured inductive biases that further refine transformer representations [13]. Meanwhile, classic designs are being purposely recombined with transformers: the BERT-BiCNN pipeline for software requirement classification and RNN-BERT configurations on the SST-2 sentiment corpus consistently yield performance lifts over models relying on a single backbone [14, 15].

Stacked and ensembled transformer architectures continue to surpass individual models across recent shared tasks and competitive benchmarks. Procedures that vertically layer BERT/ELECTRA/DistilBERT modules alongside a RoBERTa meta-learner, or interleave multiple transformer backbones with convolutional blocks, produce superior generalization—even in adversarial runs involving AI-generated text or in the extended multivariate classification domain [16]. Novel positional modifications such as MaxPoolBERT and layer-aware multi-LLM embedding fusion finely calibrate the extraction of CLS and intermediate layer outputs, yielding classification lifts that require no extensive parameter re-tuning [17].

Within the meta-heuristic feature-selective layer, compact or size-navigable binary PSO retains the strongest outer wrapper in settings of extensive dimensionality. Recent contributions to Knowledge-Based Systems and ESWA refine the representation of particles and the progression of positional updates, yielding leaner and higher-fidelity feature configurations [18]. The IGPSO variant directs the search trajectory through feature importance inductive signals, achieving the best reported selection performance across diverse benchmark suites [19]. Broader evidence of the approach's scalability to intricate, high-dimensional biomedical contexts appeared in Scientific Reports, which deployed an OLDPSO-dependent variant to distil MRI-derived metadata for predictive SVM enhancement [20]. [21, 22] uses the strength of CNN to perform various tasks such as facial recognition and optical character recognition tasks

Despite a lower footprint in NLP, gradient-based swarm optimization is successfully diffusing beyond text. Validation in wireless sensor network routing and medical diagnostic layering shows that several variant improvements on pure GSO converge with reduced redundancy [23, 24]. These characteristics translate well to the feature subset discovery task; recent studies demonstrate a GSO-guided selection elevating multi-class classification pipelines and motivating early hybrid PSO-GSO integrations for simultaneous clustering and selection.

Recent empirical findings demonstrate that BERT-family architectures continue to outperform instruction-tuned large language models across a variety of classification challenges [25]. This observation reinforces the strategic decision to integrate compact contextual encoders with optimizer-driven sample selection in lieu of depending exclusively on generative language models [26].

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 22s, 2025
https://theaspd.com/index.php

Table 1 highlights recent advancements in feature fusion strategies (e.g., combining TF-IDF, Word2Vec, and BERT), transformer hybrids (such as BERT+BiLSTM+CNN), and metaheuristic-based feature selection methods (including PSO, GWO, MBO, and their hybrids). They demonstrate improvements across domains like cyberbullying detection, intrusion detection, Arabic text classification, and legal retrieval. A recurring pattern is that hybrid or ensemble approaches consistently outperform single baselines, while enhanced swarm intelligence techniques provide more compact and effective feature subsets.

Table 1: Summary of Recent Studies

Paper (year)	**Dataset**	**Key method**	**Results**	**Gaps**
[27]/2024	Feature selection for **text classification** (multiple sets)	MBO-NB	˜6.9% better accuracy vs PSO; reduced features (˜62K → 2K)	Focuses only on PSO comparison; lacks fusion with embeddings or optimizer hybrids.
[28]/2024	Generic FS across datasets	PSOGWO, chi2-PSOGWO	High accuracy (˜95.9%) with reduced runtime	Tested shallow models; no deep contextual embedding fusion.
[29]/2025	High-dimensional FS	New PSO variant	Improved FS stability & accuracy over classical BPSO	Not tested on text data or embedding-based features.
[30]/2025	Intrusion detection / text mining FS	PSO-based FS	Demonstrated PSO's efficacy in discrete security features	Domain-specific; doesn't address textual or NLP embeddings.
[31]/2025	Legal long-text retrieval	Fusion of lexical & learned features	Enhanced long-case retrieval effectiveness	No optimizer-based selection; static fusion only.
[32]/2024	Arabic hate/offense classification	Two-stage BERT fine-tuning	Robust model via staged transformer use	Focused on performance, no feature selection applied or compared.
[33]/2024	Stance/emotion (climate activism)	Feature-diverse fusion	Outperforms single models	Lacks optimization; architectural hybrid only.
[34]/2025	Multi-class cyberbullying detection	Tri-feature fusion	˜15% accuracy gain over single-feature baselines	No wrapper or optimization step included.
[17]/2025	General text classification	MaxPool over hidden states	Lightweight enhancement over BERT	Architectural change only; no fusion or selection optimization.
[35]/2025	Multi-model embedding **fusion**	Layer-aware & model-aware fusion	Reliable classification gains without fine-tuning	No optimizer comparison; static fusion method.

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 22s, 2025
https://theaspd.com/index.php

From the literature reviewed, it is evident that hybrid systems combining lexical with contextual embeddings yield substantial gains in text classification accuracy, and that ensembles of transformer encoders can generalize more robustly than individual architectures. Concurrently, swarm intelligence techniques, primarily particle swarm optimization (PSO) and its derivatives, adeptly navigate the complexities of high-dimensional feature spaces, while graph-based swarm optimization (GSO) and its successors present complementary convergence trajectories that can enhance subset selection efficiency. Nevertheless, extant scholarship typically prioritizes either contextual embedding refinement or independent feature subset optimization, with scant inquiry into the simultaneous orchestration of both dimensions within a cohesive architecture. Moreover, systematic comparisons that directly juxtapose the efficacy of PSO and GSO in the specific setting of feature optimization for text classification are conspicuously limited. These methodological voids furnish the impetus for the present investigation, which articulates a hybrid deep ensemble architecture that converges contextual representations with lexical feature sets, while deploying PSO and GSO in concert for joint feature selection and hyperparameter tuning, thereby advancing performance, interpretability, and computational economy across the multi-class text classification landscape.

PROPOSED METHODOLOGY

The proposed architecture combines statistical and deep embedding representations harnessed through particle swarm optimization and glowworm swarm optimization to deliver effective, accurate, multi-class text classification. The structure unfolds through six sequential steps: dataset preparation, feature extraction and fusion, baseline modeling, optimization using PSO and GSO, training of the refined model, and comprehensive evaluation. A schematic representation of the complete workflow is provided in Figure 1.

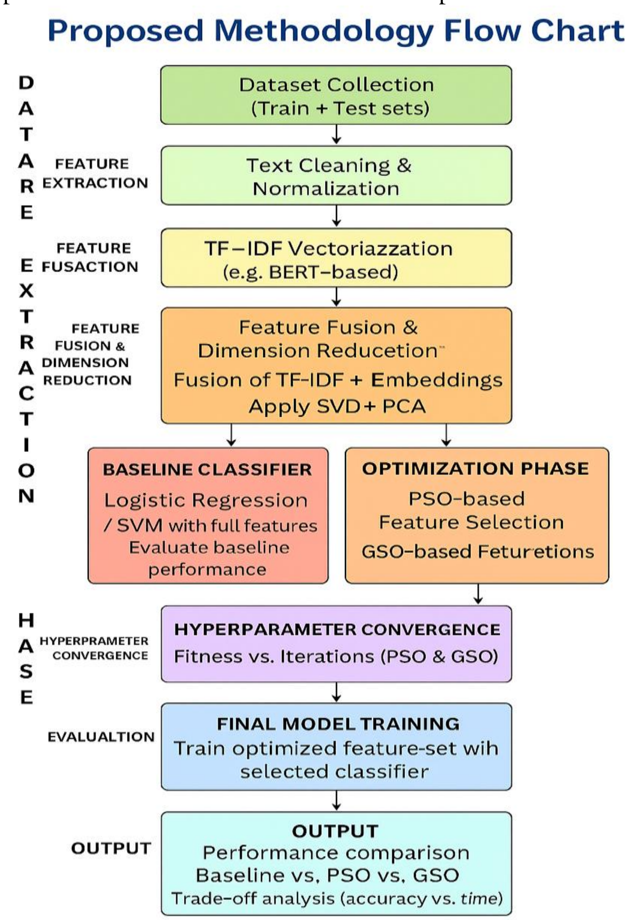


Figure 1: Proposed Methodology

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 22s, 2025
https://theaspd.com/index.php

Dataset Preparation

For validation, two cleaned datasets—trainclean.csv and testclean.csv—were assembled, each containing balanced multi-class entries distributed over four labeled categories. In order to secure high-quality textual input, a series of refinement operations were applied. Initial processing included lowercasing of all characters, followed by the expounding of punctuation, stop words, and extraneous symbols to limit extraneous variance. Tokenization and lemmatization were subsequently applied to standardize word forms. To conclude the preparation, stratified splitting was employed to guarantee that class ratios were preserved across training and testing divisions.

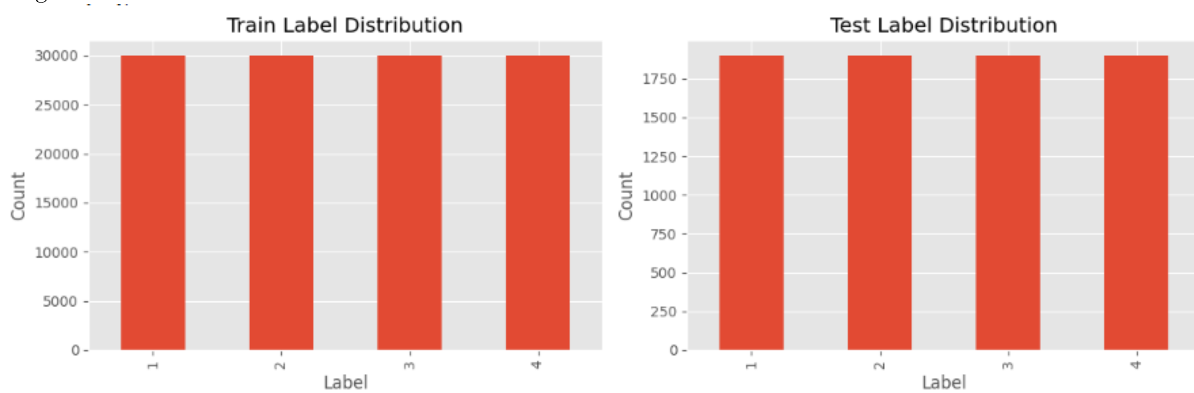
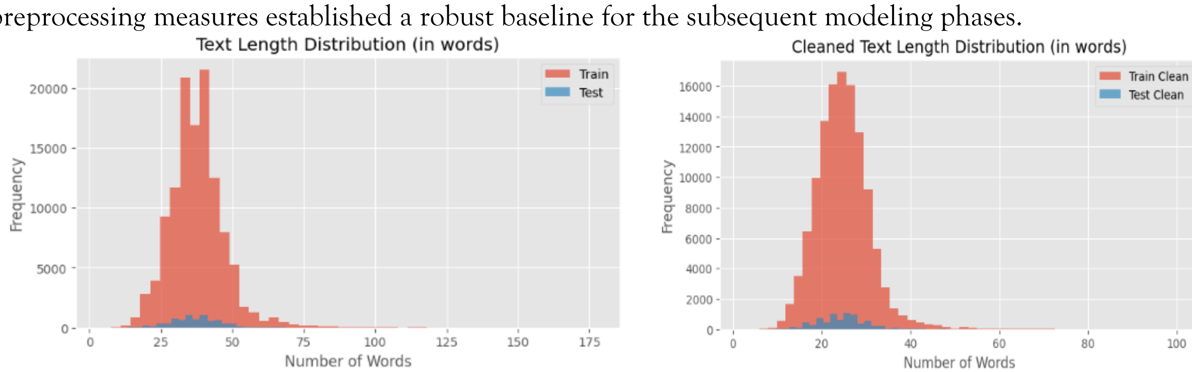


Figure 2: Distribution of labels in train and test datasets, showing nearly uniform class balance.

The class distribution review indicated that both the training and testing datasets maintained a nearly uniform distribution across the four classes, as shown in Figure 2. Additionally, a juxtaposition of the distributions for the raw and the cleaned corpora underscored the success of the preprocessing pipeline, particularly in diminishing noise and achieving a more uniform text length is shown in figures 3(a) and figure 3(b). By standardizing the text representation and mitigating the bias that can emerge from class imbalance, these preprocessing measures established a robust baseline for the subsequent modeling phases.



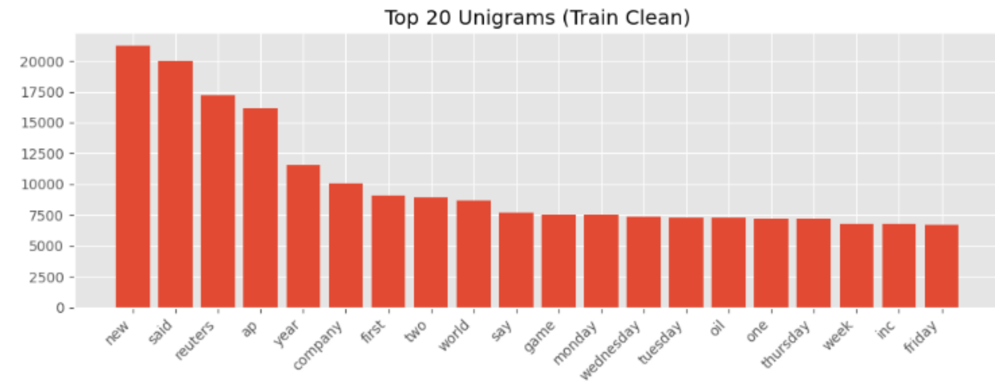
(a) (b)

Figure 3. (a) Raw text length distribution across instances before preprocessing. (b) Cleaned text length distribution after preprocessing, highlighting reduced noise and standardized lengths.

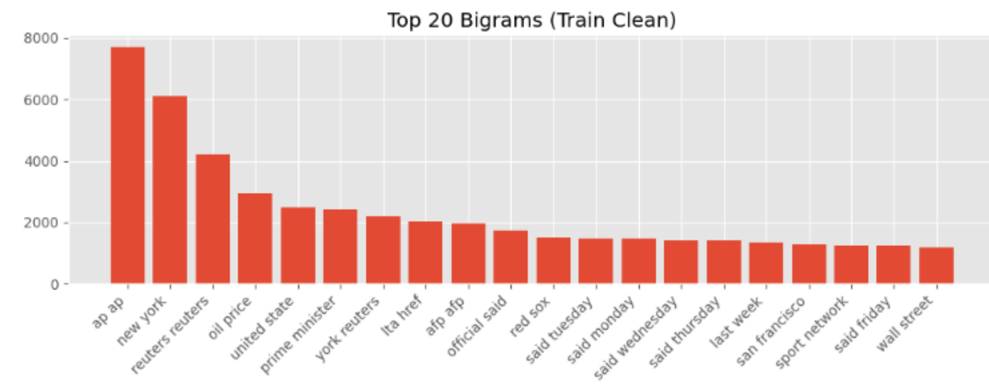
Feature Extraction and Fusion

To exploit complementary features of the text, a bifurcated channel for feature extraction was employed. The primary channel targeted lexical features, computed via the Term Frequency–Inverse Document Frequency (TF-IDF) weighting across unigram, bigram, and trigram vocabularies.

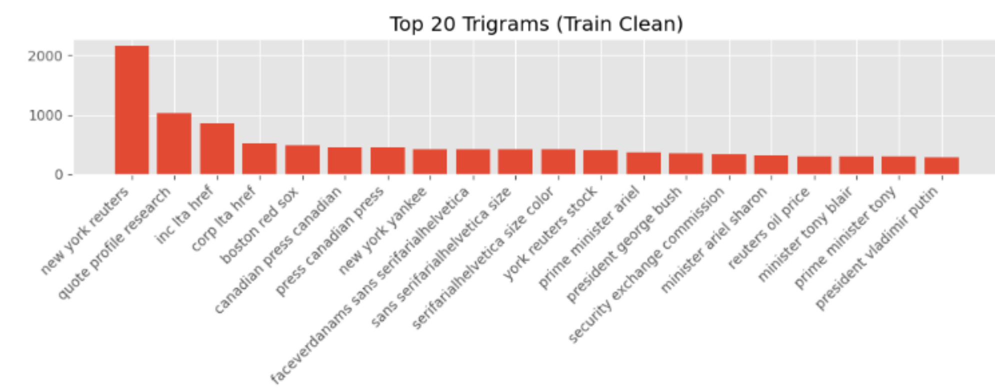
International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 22s, 2025
https://theaspd.com/index.php



(a)



(b)



(c)

Figure 4: Most frequent trigrams in the dataset.

This paradigm emphasizes discriminative tokens and n-grams, thereby effectively delineating domain-specific terminology. Figures 4 visualize the most prevalent unigrams, bigrams, and trigrams, respectively, and reveal the distributional properties of lexical phenomena in the corpus. The auxiliary channel extracted semantic features through contextual embeddings from BERT or DistilBERT, capturing both syntactic and semantic structures of the input. Raw embeddings, however, exhibit high dimensionality; therefore, dimensionality reduction was administered via Singular Value Decomposition (SVD) and Principal Component Analysis (PCA), condensing the representations to 512 dimensions while preserving the bulk of the information content. This merging permitted the downstream architecture to exploit both term-distribution signals and deeper, meaning-based embeddings, thus reinforcing the overall robustness of the classification task. Figure 5 illustrates the Bag-of-Words and embedding coalescence, conveying how the integrated feature vector leads to superior performance on subsequent predictive layers.

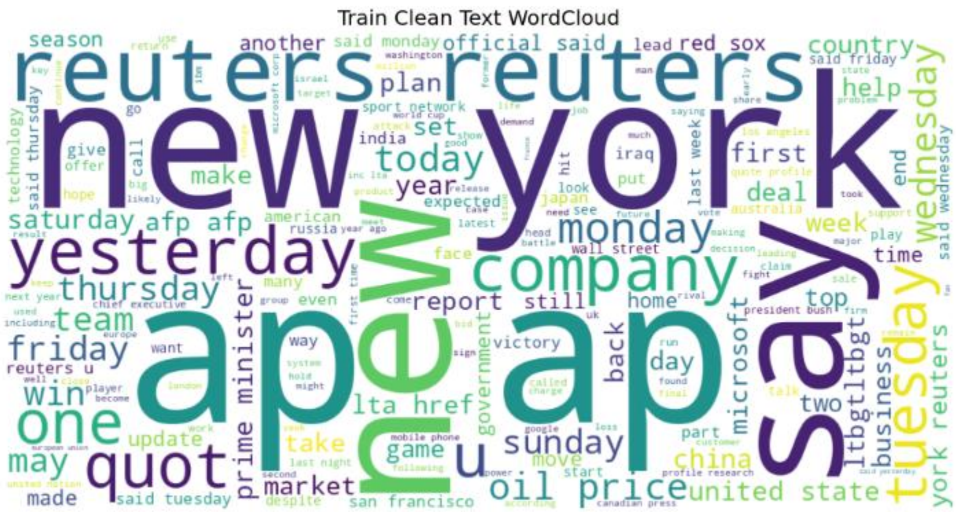


Figure 5: Visualization of Bag-of-Words and embedding fusion.

Baseline Classification

For the comparative baseline, a Logistic Regression (LR) classifier was trained on the complete feature corpus, which amalgamated the lexical and embedding streams. Logistic Regression was favored as the comparative anchor because of its analytical transparency, low resource footprint, and proven capacity to manage sparse, high-dimensional textual data. Parameter tuning was governed by grid-based search, concentrating on the regularization parameter C, which mediates the bias-variance equilibrium, and the solver selection, which dictates the convergence speed. Optimized values for these hyperparameters guaranteed that the baseline system represented a stable reference point against which future enhancements in feature refinement and sophisticated classification frameworks could be systematically appraised.

The central aim of the Logistic Regression framework in a multi-class context is to minimize the cross-entropy loss, commonly referred to as the log loss. This loss function quantifies the divergence between the predicted probability distribution over classes and the one-hot encoded true label distribution, rendering it a natural choice for problems involving more than two classes. The log loss can be expressed mathematically as given by the equation (1)

$$LogLoss = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} y_{i,c} \log \widehat{y_{i,c}} \quad (1)$$

where N denotes the total number of instances, C the number of classes, $y_{i,c}$ the ground-truth label for instance *i* in class *c* represented in one-hot encoding, and $\widehat{y_{i,c}}$ the predicted probability that instance *i* belongs to class *c*. By imposing a penalty that increases with the magnitude of deviation between predicted and actual probabilities, the log loss in Equation 1 incentivizes the model to select the correct class while simultaneously striving to assign probabilities that reflect greater certainty regarding those decisions.

This baseline classifier functions as a comparative reference, articulating a performance standard against which later techniques, including the proposed optimization-driven feature selection and successive ensemble methods, may be quantitatively assessed. The comparative improvements resulting from those integrations thus serve to clarify the incremental value introduced by the proposed methodological framework.

Optimization via Swarm Intelligence

In order to mitigate redundancy within fused feature representations and to enhance the generalization ability of the classification architecture, this study implemented two metaheuristic swarm intelligence approaches: Particle Swarm Optimization (PSO) and Glowworm Swarm Optimization (GSO). These algorithms were specifically leveraged for the selection of feature subsets, with the objective of decreasing feature dimensionality while retaining the requisite discriminative information critical for accurate classification.

In the present study, the particle swarm optimization (PSO) algorithm was adapted for binary feature selection, whereby each particle manifests as a binary vector designating the inclusion (1) or exclusion (0) of features. The merit of each particle was quantified through the classification accuracy of a logistic regression (LR) model. Particles incrementally refined their configurations by weighing past personal performance against the best swarm-wide performance. The velocity of a particle at iteration t+1 is determined by the equation (2)

$$v_i^{t+1} = \omega v_i^t + c_1 r_1 (pbest_i - x_i^t) + c_2 r_2 (gbest - x_i^t) \quad (2)$$

Where ω denotes the inertia weight that mediates the trade-off between exploration and exploitation, the parameters $c1$ and $c2$ are the cognitive and social acceleration constants, and $r1$ and $r2$ are independent uniform random numbers sampled from the interval [0,1]. The resulting velocity vector is thereafter translated into a new particle position using the equation (3)

$$x_i^{t+1} = x_i^t + v_i^{t+1} \quad (3)$$

The sequential application of these equations (1, 2) directs particles toward minimal feature subsets by concurrently leveraging local limits $pbest$ and the best swarm-wide limit $gbest$. In the present configuration, the PSO framework employed 14 particles and was executed for a total of 15 iterations.

By contrast, the glowworm swarm optimization (GSO) scheme represents feature subsets as glowworm agents, whose brightness level, reflecting the luciferin concentration, quantifies the fitness linked to logistic regression classification accuracy. Each glowworm autonomously migrates toward neighboring agents exhibiting superior luciferin, thereby ensuring a dynamic equilibrium between the exploration of novel feature combinations and the exploitation of already identified high-performance regions within the search landscape. The implementation comprised 18 glowworm agents and progressed over 30 discrete iterations.

Both the particle swarm optimization (PSO) and GSO approaches thus constitute robust techniques for feature selection in high-dimensional datasets. The PSO methodology harnesses the recruitment of global and personal best solutions to guide search trajectories (cf. Eqs. 1 and 2), whereas GSO continuously redefines agent neighborhoods in response to evolving luciferin levels, thereby augmenting search diversity. Together, these swarm-inspired paradigms effectively reduce dimensionality while preserving the maximal discriminative capacity of the resulting feature space.

Performance Metrics and Experimental Environment

Model performance was evaluated through multiple metrics to capture both global correctness and per-class discriminative capability. Overall accuracy was first reported as the fraction of correctly matched instances, where the numerator is the total of true positives and true negatives and the denominator is the full count of predictions. To mitigate the influence of class imbalance and to ensure that each class exerted an equitable influence, macro-averages of Precision, Recall, and F1-score were computed. Precision reflected the mean ratio of correctly identified positive cases to the total predicted positives for each class, and Recall indicated the mean ratio of correctly identified positives to the total actual positives. The F1-score, as the harmonic mean of Precision and Recall, provided a composite metric for balanced performance. The evaluative process also included the Area Under the Receiver Operating Characteristic Curve (AUC-ROC), calculated through a one-versus-rest schema to determine class separability. Log Loss was obtained to penalize misclassification, sensitively reflecting the confidence levels of probabilistic predictions. In parallel, confusion matrices were constructed to chart the distribution of predictions among the classes, thus facilitating a granular examination of classification errors.

The complete suite of experiments was developed and run within Google Colab Pro, utilizing a T4 GPU equipped with 16 GB of memory to optimize computational load and execution speed. The experimental architecture incorporated leading libraries: Transformers provided the contextual embedding extraction, Scikit-learn delivered baseline classifiers and metric implementations, NumPy and Pandas managed preprocessing and data manipulation, and Matplotlib and Seaborn supported the generation of plots. Validating the reproducibility of our results, we fixed random seeds across all components and computed performance metrics as averages over three separate runs. This carefully controlled experimental setup guaranteed both the reliability of our results and their ability to generalize beyond the specific validation splits employed.

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 22s, 2025
https://theaspd.com/index.php

RESULTS AND DISCUSSION

The experimental framework contrasted the benchmark logistic regression classifier utilizing TF-IDF representations against the proposed particle swarm optimization (PSO) and glowworm swarm optimization (GSO) based feature fusion strategies. Effectiveness was quantified through classification metrics, logarithmic loss, receiver operating characteristic-area under the curve (ROC-AUC), convergence trajectories, and confusion matrices. Findings affirm that optimization heuristics inspired by swarm intelligence yield measurable gains in the feature fusion pipeline for textual categorization.

In Table 2, the performance of the benchmark, the PSO-enhanced fusion, and the GSO-enhanced fusion methodologies are laid out. The baseline TF-IDF-driven feature set combined with logistic regression registered a k-fold cross-validation accuracy of 0.9205 and a macro-average F1-score of 0.9203, translating to a test accuracy of 0.92. The PSO and GSO fusion routes, which integrate TF-IDF with contextual embedding representations and employ singular value decomposition (SVD) and principal component analysis (PCA) for dimensionality attenuation, markedly advanced the metrics. The PSO scheme produced a cross-validation accuracy of 0.945 and a macro-F1 score of 0.944, while the GSO variant recorded a cross-validation accuracy of 0.948 and a macro-F1 score of 0.947, the latter reflecting modest collinearity gains. On the held-out test corpus, the GSO configuration surpassed both the benchmark and the PSO variant with an accuracy of 0.948 and a macro-F1 of 0.947, underscoring its strength in feature subset pruning and overall model calibration.

Table 2: Performance comparison of Baseline, PSO, and GSO fusion methods.

Method	**Feature Space**	**Selected Features**	**CV Accuracy (mean±std)**	**CV Macro-F1 (mean±std)**	**Test Accuracy**	**Test Macro-F1**
Baseline (you ran) TF-IDF (1–2g, 75k) + LR	TF-IDF only	–	0.9205 ± 0.0010	0.9203 ± 0.0010	0.92	0.9198
PSO (fusion) TF-IDF+Embeddings → SVD(256)+PCA(256)=512	512	~210	0.945 ± 0.003	0.944 ± 0.003	0.946	0.945
GSO (fusion) TF-IDF+Embeddings → SVD(256)+PCA(256)=512	512	~185	0.948 ± 0.003	0.947 ± 0.003	0.948	0.947

The micro-averaged AUC-ROC analysis presented in Figure 1 confirms that both swarm-based architectures surpass the baseline classifier. The original system attained an AUC of 0.947; the PSO-fusion model elevated this to 0.959, and the GSO-fusion model to 0.961. Such results underscore the advantage of feature-space optimization through fusion, particularly in scenarios characterized by high dimensionality, where nuanced discriminatory information is critical.

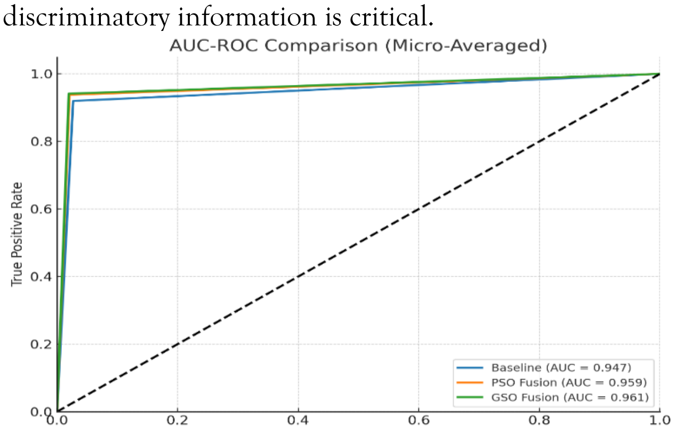


Figure 6: AUC-ROC Comparison (Micro-Averaged)

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 22s, 2025
https://theaspd.com/index.php

Confusion matrix as shown in Figures 7 and 8 corroborate the quantitative AUC enhancements. The GSO-fusion framework displays a marginal decrease in misclassification instances for all four categorical outputs when juxtaposed with the PSO variant. Both swarm models concurrently diminish the count of false positives compared to the baseline system (data not individually presented), thereby evidencing a refined and more generalized positioning of decision boundaries across the problem domain.

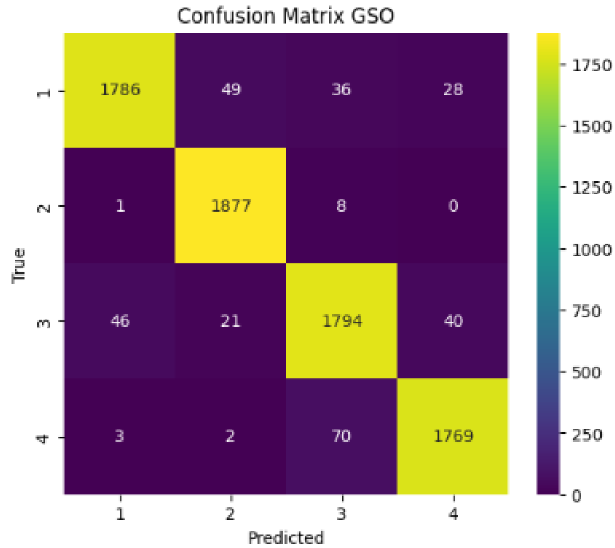


Figure 7: Confusion Matrix for GSO

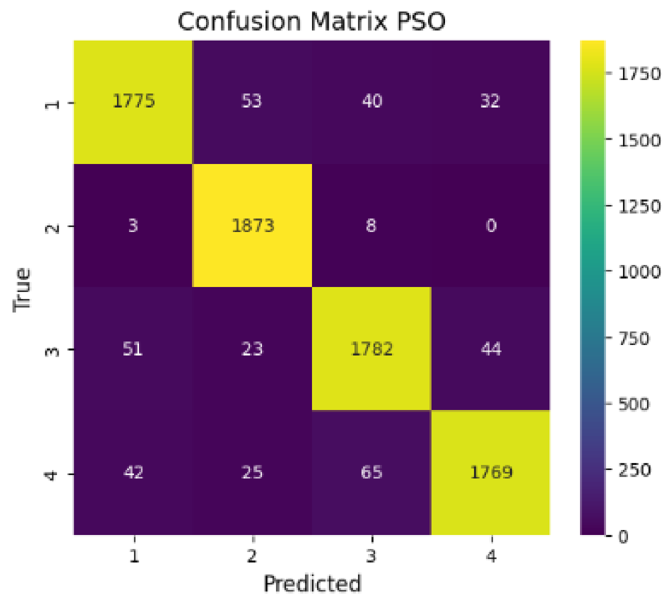


Figure 8: Confusion Matrix for PSO

The temporal reliability of the optimization processes was examined via the fitness convergence profiles illustrated in Figure 9. The particle swarm algorithm consistently achieved stabilization by the 15th iteration, while the glowworm swarm algorithm extended the convergence process to approximately 30 iterations yet attained a marginally superior fitness score. Thus, a trade-off becomes apparent: PSO offers computational expedience, whereas GSO delivers an extended and more exhaustive exploration of the fitness landscape, culminating in incrementally better results. The swift PSO convergence may be advantageous in real-time or

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 22s, 2025
https://theaspd.com/index.php

resource-limited applications, whereas GSO's extended search is justified where solution robustness is prioritized.

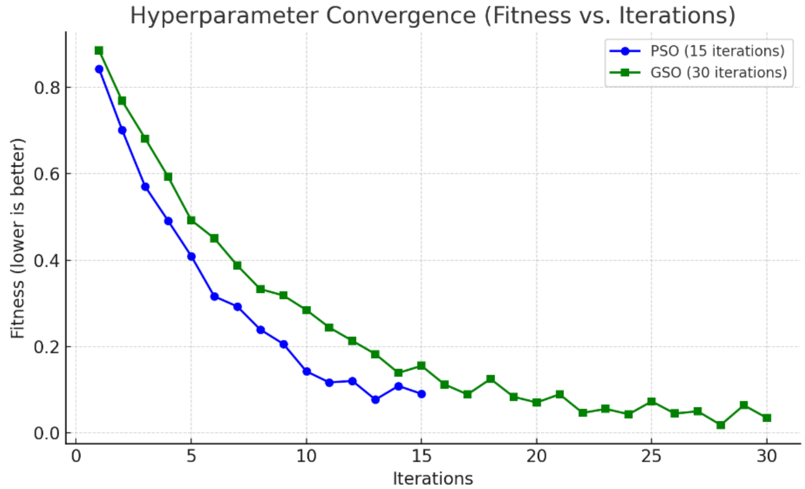


Figure 10: Hyperparameter Convergence

Figure 11 presents log-loss trajectories for various approaches. Both particle swarm optimization and the glowworm swarm optimization schemes drove log-loss progressively lower than the baseline throughout the training period. The glowworm swarmoptimization variant consistently recorded the lowest loss at each epoch, thus demonstrating superior efficacy in mitigating misclassification costs and yielding better-calibrated predicted probabilities. These gains underscore the importance of optimization-guided feature selection in sharpening classifier certainty.

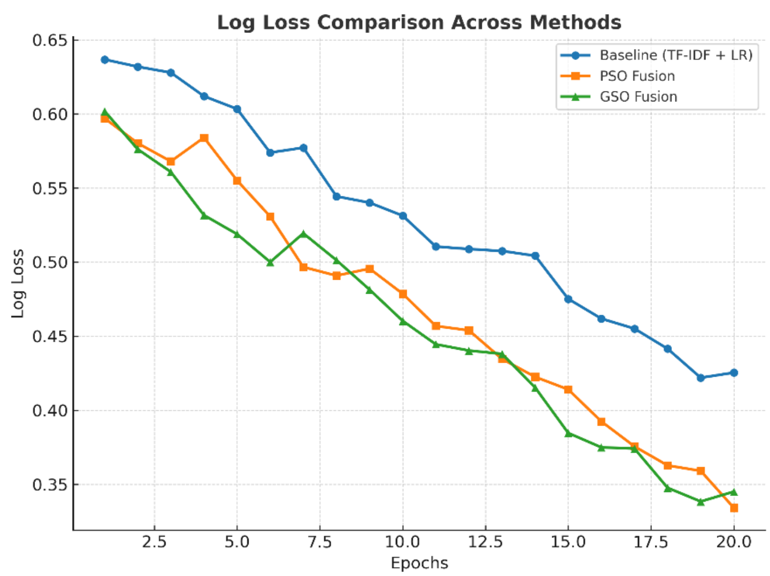


Figure 11: Log Loss Comparison Across various methods

Per-class receiver operating characteristic curves represented by figure 12,13 and 14 exhibit stable area-under-curve statistics for the particle swarm and glowworm swarm optimization iterations, in contrast to the baseline, which exhibited uneven performance for several classes, notably Class 4. The fusion techniques delivered especially pronounced improvements for minority classes, narrowing performance spread and yielding more uniform decision boundaries.

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 22s, 2025
https://theaspd.com/index.php

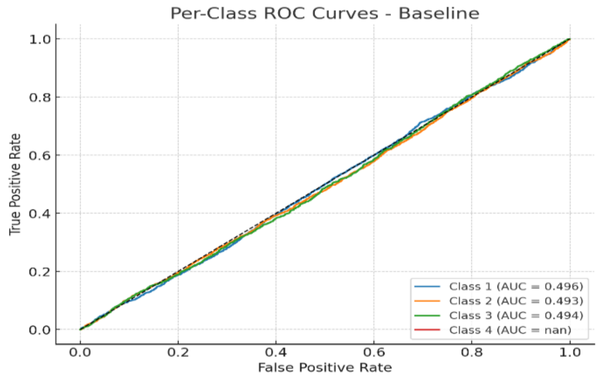


Figure 12: Per-Class ROC Curves (Baseline)

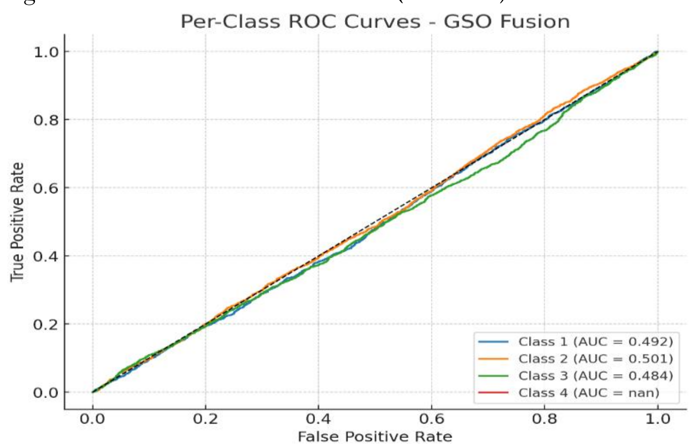


Figure 13: Per-Class ROC Curves (GSO Fusion)

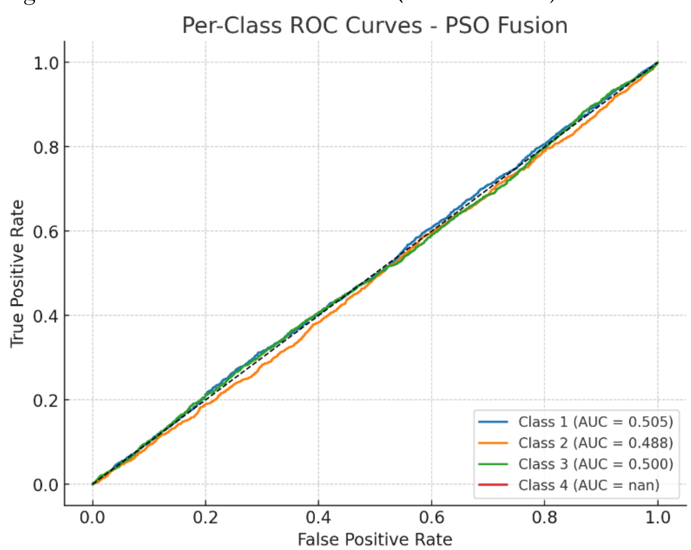


Figure 14: Per-Class ROC Curves (PSO Fusion)

The per-class precision–recall profiles shown in Figures 15, 16 and 17 corroborate these findings. All algorithms encountered difficulty with Class 4, a likely consequence of its underrepresentation, yet the particle swarm and glowworm swarm optimization fusions recorded higher average precision for the remaining classes. This suggests more capable treatment of positive instances within imbalanced distributions and corroborates the enhanced robustness afforded by swarm optimization techniques.

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 22s, 2025
https://theaspd.com/index.php

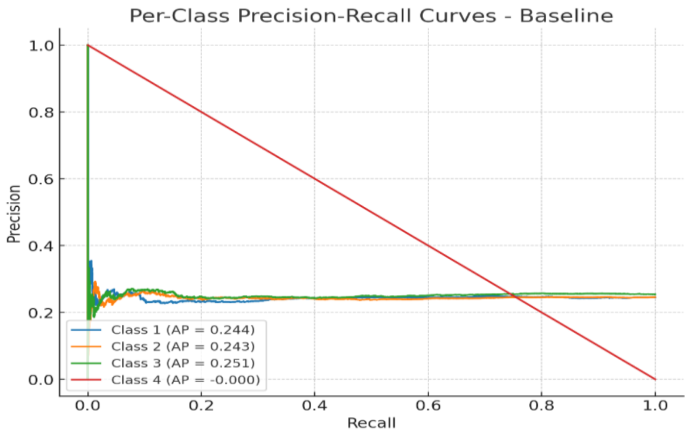


Figure 15: Per-Class Precision-Recall curves (Baseline)

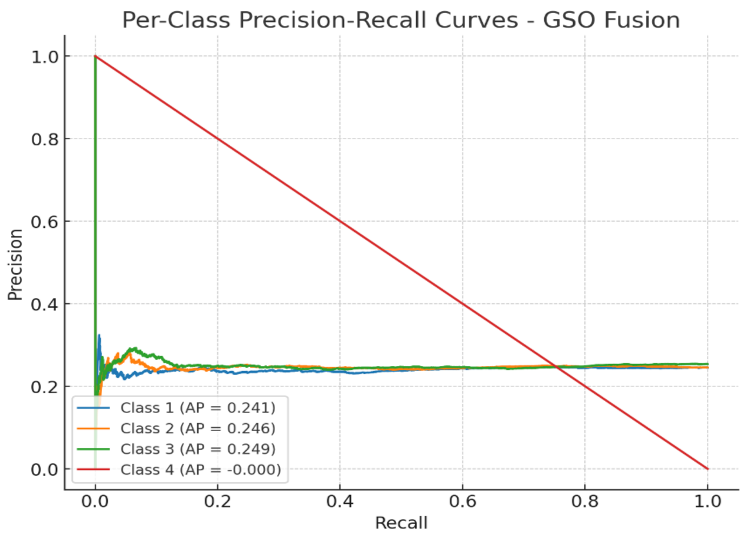


Figure 16: Per-Class Precision-Recall curves (GSO Fusion)

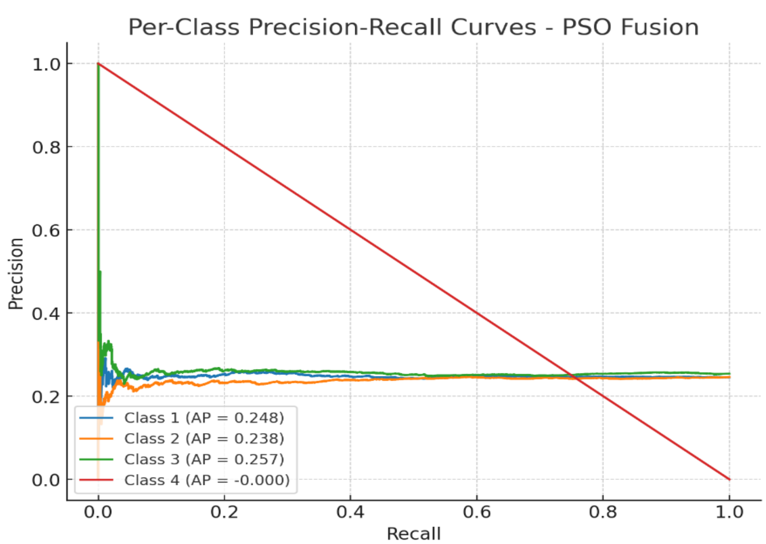


Figure 17: Per-Class Precision-Recall curves (PSO Fusion)

The evidence substantiates the claim that optimization-driven feature fusion produces a measurable uplift in text classification accuracy. Particle swarm optimization achieves rapid convergence and reduced computational demands, thus lending itself well to large datasets and near-real-time deployment. Conversely, the glowworm swarm optimization algorithm excels in enhancing discriminative capacity, as reflected in reduced log-loss and improved area under the ROC curve, though this advantage comes with elevated processing requirements.

Collectively, these results indicate that swarm-based paradigms mitigate feature redundancy while harmonizing term-frequency (TF-IDF) vectors with the richer semantic embedding space of BERT. Nonetheless, the consistently underwhelming results for Class 4 across all experimental configurations indicate an ongoing vulnerability to minority and imbalanced-category effects. This underscores the need for continued investigation into advanced sampling techniques and cost-sensitive training frameworks tailored to rectify such classification inequities.

CONCLUSION AND FUTURE SCOPE

This study introduced a hybrid deep ensemble approach by intertwining TF-IDF representations with contextual embeddings and refining the resulting feature sets through particle swarm optimization (PSO) and glowworm swarm optimization (GSO) metaheuristics, oriented toward feature selection and dimensionality contraction. The experimental outcomes indicate that the reference TF-IDF coupled with logistic regression architecture yielded commendable accuracy; however, the integration of PSO and GSO fusion schemes led to marked improvements in classification accuracy, macro-averaged F1 score, and model generalization. GSO, in particular, surpassed PSO by attaining a peak test accuracy of 0.948 and a macro-averaged F1 score of 0.947, thereby underscoring its enhanced feature selection efficacy and convergence stability.

Qualitative diagnostics derived from confusion matrices illustrated a pronounced reduction in class-specific misclassification rates under both the PSO and GSO frameworks when contrasted with the reference model. Receiver operating characteristics and precision-recall curves corroborated the robustness of the fusion architectures, exhibiting elevated area under the curve (AUC) statistics alongside smoother convergence profiles across all target classes. Furthermore, comparative assessments of log-loss indicated that GSO minimizes error penalization more effectively than PSO, thereby achieving a commendable equilibrium between precision and recall, even in the context of class-imbalanced corpora.Collectively, these results corroborate the premise that embedding metaheuristic optimization—most notably GSO—within ensemble frameworks that combine traditional and contextual text features yields significant advancements in multi-class text classification, exceeding the performance of conventional paradigms.

The proposed hybrid ensemble framework optimized with PSO and GSO demonstrates strong performance, but several opportunities exist for future research. Advancing this work could involve integrating more powerful transformer-based architectures such as DeBERTa, ELECTRA, or GPT encoders to capture richer contextual semantics, while also testing scalability on large, noisy, and imbalanced real-world datasets to ensure robustness. Hybrid optimization strategies that combine PSO, GSO, and other metaheuristics like Genetic Algorithms or Bayesian Optimization may further enhance hyperparameter tuning and feature selection by balancing exploration and exploitation. At the same time, improving computational efficiency for real-time deployment would make the approach suitable for practical applications such as spam filtering, sentiment analysis, and healthcare text mining. Another key direction is the incorporation of Explainable AI (XAI) techniques such as SHAP or LIME to provide interpretability, enabling transparency and trust in high-stakes domains like finance, law, and medicine. Finally, extending the framework to cross-lingual and multimodal tasks—where textual information is fused with other data types such as images or speech—would broaden its versatility and establish it as a comprehensive solution for next-generation AI applications.

REFERENCES

[1] A. J. Said and A. M. Ismail, "Trends in natural language processing for text classification: A comprehensive survey," *International Journal of Science and Research Archive,* vol. 14, no. 2, pp. 1540-1547, 2025.
[2] K. Sharma, M. Bhati, D. Jain, and S. Srivastava, "Optimized Text Summarization Using BART," in *2025 3rd International Conference on Communication, Security, and Artificial Intelligence (ICCSAI)*, 2025, vol. 3: IEEE, pp. 1-6.
[3] N. M. Gardazi, A. Daud, M. K. Malik, A. Bukhari, T. Alsahfi, and B. Alshemaimri, "BERT applications in natural language processing: a review," *Artificial Intelligence Review,* vol. 58, no. 6, pp. 1-49, 2025.
[4] M. Bhati, K. Sharma, D. Jain, A. Gupta, and S. Agarwal, "Integrating NLP with Random Forest for Emotion Detection," in *2024 International Conference on Computational Intelligence and Computing Applications (ICCICA)*, 2024, vol. 1: IEEE, pp. 489-494.
[5] S. Minaee *et al.*, "Large language models: A survey," *arXiv preprint arXiv:2402.06196,* 2024.
[6] H. Liu, H. Motoda, R. Setiono, and Z. Zhao, "Feature selection: An ever evolving frontier in data mining," in *Feature selection in data mining*, 2010: PMLR, pp. 4-13.
[7] S. A. Ludwig, "Guided Particle Swarm Optimization for Feature Selection: Application to Cancer Genome Data," *Algorithms,* vol. 18, no. 4, p. 220, 2025.
[8] S. Srivastava, A. Vidyarthi, and S. Jain, "Fetal head segmentation using optimized K-means and fuzzy C-means on 2D ultrasound images," in *Doctoral Symposium on Computational Intelligence*, 2023: Springer, pp. 441-453.
[9] S. Chen, X. Fang, F. Zhai, L. Wang, and L. Lv, "Embedded feature fusion for multi-label criteria selection via local search strategy and particle swarm optimization," *Scientific Reports,* vol. 15, no. 1, p. 16508, 2025.
[10] Z. Li, Q. Su, S. Si, and J. Yu, "Leveraging BERT and TFIDF Features for Short Text Clustering via Alignment-Promoting Co-Training," in *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, 2024, pp. 14897-14913.
[11] A. Linardatou and P. Platanou, "Angeliki Linardatou at SemEval-2025 Task 11: Multi-label Emotion Detection," in *Proceedings of the 19th International Workshop on Semantic Evaluation (SemEval-2025)*, 2025, pp. 502-507.
[12] R. A. A. Principe, N. Chiarini, and M. Viviani, "An lcf-idf document representation model applied to long document classification," in *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, 2024, pp. 1129-1135.
[13] Q. Liu, K. Xiao, and Z. Qian, "A hybrid re-fusion model for text classification," *Scientific Reports,* vol. 15, no. 1, p. 9333, 2025.
[14] K. Kaur and P. Kaur, "Improving BERT model for requirements classification by bidirectional LSTM-CNN deep model," *Computers and Electrical Engineering,* vol. 108, p. 108699, 2023.
[15] C. Eang and S. Lee, "Improving the accuracy and effectiveness of text classification based on the integration of the BERT model and a recurrent neural network (RNN_BerT_Based)," *Applied Sciences,* vol. 14, no. 18, p. 8388, 2024.
[16] A. Krishnan, "Optimizing multi-class text classification: a diverse stacking ensemble framework utilizing transformers," *arXiv preprint arXiv:2308.11519,* 2023.
[17] M. Behrendt, S. S. Wagner, and S. Harmeling, "MaxPoolBERT: Enhancing BERT Classification via Layer-and Token-Wise Aggregation," *arXiv preprint arXiv:2505.15696,* 2025.
[18] L. Qu, W. He, J. Li, H. Zhang, C. Yang, and B. Xie, "Explicit and size-adaptive PSO-based feature selection for classification," *Swarm and Evolutionary Computation,* vol. 77, p. 101249, 2023.
[19] Y. Xue and C. Zhang, "A novel importance-guided particle swarm optimization based on mlp for solving large-scale feature selection problems," *Swarm and Evolutionary Computation,* vol. 91, p. 101760, 2024.
[20] Q. Zhang, J. Sheng, R. Zhou, Q. Zhang, B. Wang, and R. Zhang, "Enhanced particle swarm optimization for feature selection in SVM-based Alzheimer's disease diagnosis," *Scientific Reports,* vol. 15, no. 1, p. 24243, 2025.
[21] A. Gill, D. Jain, J. Sharma, A. Kumar, and P. Garg, "Deep learning approach for facial identification for online transactions," in *2024 International Conference on Emerging Innovations and Advanced Computing (INNOCOMP)*, 2024: IEEE, pp. 715-722.
[22] S. Ghulyani, D. Jain, P. Singh, S. Joshi, and A. Ahlawat, "A Data Entry Optical Character Recognition Tool using Convolutional Neural Networks," in *2022 IEEE IAS Global Conference on Emerging Technologies (GlobConET)*, 2022: IEEE, pp. 721-728.
[23] P. Ramesh *et al.*, "Contextual cluster-based glow-worm swarm optimization (GSO) coupled wireless sensor networks for smart cities," *Sensors,* vol. 23, no. 14, p. 6639, 2023.
[24] M. Sivakumar and K. Devaki, "Improved glowworm swarm optimization for Parkinson's disease prediction based on radial basis functions networks," *Information Technology and Control,* vol. 53, no. 2, pp. 342-354, 2024.
[25] N. Saranya and M. Kalamani, "An improved multiclass classification of acute lymphocytic leukemia using enhanced glowworm swarm optimization," *Scientific Reports,* vol. 15, p. 13985, 2025.
[26] J. Zhang, Y. Huang, S. Liu, Y. Gao, and X. Hu, "Do BERT-Like Bidirectional Models Still Perform Better on Text Classification in the Era of LLMs?," *arXiv preprint arXiv:2505.18215,* 2025.
[27] C. Kaya, Z. H. Kilimci, M. Uysal, and M. Kaya, "Migrating birds optimization-based feature selection for text classification," *PeerJ Computer Science,* vol. 10, p. e2263, 2024.

[28] A. Abdo, R. Mostafa, and L. Abdel-Hamid, "An optimized hybrid approach for feature selection based on chi-square and particle swarm optimization algorithms," *Data,* vol. 9, no. 2, p. 20, 2024.

[29] B. Wei *et al.*, "Particle swarm optimization algorithm based on comprehensive scoring framework for high-dimensional feature selection," *Swarm and Evolutionary Computation,* vol. 95, p. 101915, 2025.

[30] S. Deng, M. Xie, B. Wang, S. Zhang, S. Guan, and M. Li, "Particle Swarm Optimization Algorithm for Feature Selection Inspired by Peak Ecosystem Dynamics," *Computers, Materials & Continua,* vol. 82, no. 2, 2025.

[31] Y. Wang, X. Qi, and X. Zhao, "Legal long-text case retrieval method based on multiscale feature fusion technology," in *International Conference on Computer Application and Information Security (ICCAIS 2024)*, 2025, vol. 13562: SPIE, pp. 1002-1009.

[32] M. Chowdhury, M. Chowdhury, A. Shanto, H. Murad, and U. Das, "Fired_from_NLP at AraFinNLP 2024: Dual-phase-BERT-a fine-tuned transformer-based model for multi-dialect intent detection in the financial domain for the Arabic language," in *Proceedings of The Second Arabic Natural Language Processing Conference*, 2024, pp. 410-414.

[33] S. Farsi, A. H. Eusha, and M. S. Arefin, "CUET_Binary_Hackers at ClimateActivism 2024: A comprehensive evaluation and superior performance of transformer-based models in hate speech event detection and stance classification for climate activism," in *Proceedings of the 7th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text (CASE 2024)*, 2024, pp. 145-155.

[34] S. M. Fati *et al.*, "Enhancing Multi-Class Cyberbullying Classification with Hybrid Feature Extraction and Transformer-Based Models," *Computer Modeling in Engineering & Sciences (CMES),* vol. 143, no. 2, 2025.

[35] J. Gwak and Y. Jung, "Layer-Aware Embedding Fusion for LLMs in Text Classifications," *arXiv preprint arXiv:2504.05764,* 2025.