



Strengthening Support Vector Machine Capability through Harris Hawks Optimization for Arabic Text Categorization

Hosam Ali AlZawam * hosamzwaml@gmail.com

Department of Computer Science, Faculty of Education, Gharyan University Gharyan, Libya

Tarik Milod Alarbi Ahmad, tarik.ahmad@gu.edu.ly, <https://orcid.org/0009-0003-3955-5695>

Department of Computer Science, Faculty of Science, Gharyan University Gharyan, Libya

Received: 10/8/2026 Revised: 15/8/2026 Accepted: 05/9/2026 Publication: 27/9/2026

Abstract:

The automated categorization of Arabic text remains a persistently difficult problem, largely because the language combines a productive root-and-pattern morphology, pervasive orthographic variation, and a scarcity of large, carefully annotated corpora—factors that together inflate the feature space and obscure the signal available to a learner. Within this setting, the Support Vector Machine (SVM) continues to serve as a strong discriminative baseline, yet its effectiveness hinges critically on how well its penalty term, kernel family, and kernel coefficient are configured, and on the dimensionality of the representation it must operate over. This study introduces a hybrid framework, termed SVM-HHO, in which Harris Hawks Optimization (HHO), a swarm-intelligence metaheuristic modelled on the cooperative besiege-and-pounce behavior of Harris' hawks—is entrusted with a single, unified search that simultaneously calibrates the SVM hyperparameters and isolates a compact, highly discriminative subset of Term Frequency–Inverse Document Frequency (TF-IDF) features. The framework was assessed on a balanced, six-domain corpus of 6,000 Modern Standard Arabic news articles spanning Sports, Politics, Economy, Health, Technology, and Culture. Under this evaluation, SVM-HHO attained an accuracy of 99.5% together with precision, recall, and F1-score of 99.4%, surpassing by clear margins the untuned baselines examined—standard SVM (92.3%), Random Forest (94.2%), K-Nearest Neighbors (90.5%), and Naïve Bayes (88.3%)—while its accuracy also stands favourably beside a previously published SVM-PSO model (99.2% accuracy) reported on a related multi-domain Arabic corpus. Taken together, these findings indicate that the adaptive interplay between exploration and exploitation that characterizes HHO yields a particularly well-matched joint configuration of parameters and features, producing an accurate yet computationally frugal classifier for Arabic Natural Language Processing.

Keywords: Arabic Text Categorization; Support Vector Machine (SVM); Harris Hawks Optimization (HHO); Feature Selection; Metaheuristic Optimization; Natural Language Processing (NLP).

1. Introduction

With several hundred million speakers, Arabic is among the most widely used languages globally, yet it poses distinctive obstacles for computational processing. Its writing system is governed by a derivational root-and-pattern structure, permits diacritics to be omitted, draws no distinction between upper- and lower-case letters, and exhibits considerable lexical divergence across dialects and regions. Such properties enlarge the dimensionality of the resulting representations and give rise to lexical and morphological ambiguities that conventional, language-agnostic text-processing pipelines are ill-equipped to resolve [1]. Document classification—the assignment of one or more predefined categories to a body of

text—forms the backbone of numerous real-world applications, from filtering and routing news streams to opinion mining, topic detection, and information retrieval [2]. Of the diverse learners applied to this task, the Support Vector Machine (SVM) has long been treated as a benchmark method, owing to its principled foundation in structural risk minimization and its ability to draw effective decision boundaries in high-dimensional spaces [3]. Its empirical performance, however, is far from automatic: it depends heavily on how the penalty parameter, the choice of kernel, and the kernel coefficient are set, and an unsuitable configuration tends to leave the model either underfitted or poorly generalizing to unseen documents. A second, equally decisive factor lies in the representation of the text itself. Encoding documents through term-weighting schemes such as Term Frequency–Inverse Document Frequency (TF-IDF) [4] produces feature vectors that are both highly sparse and very high-dimensional, in which a substantial proportion of the coordinates contribute little to class discrimination. Preserving these uninformative dimensions not only raises the computational burden of training but also amplifies the tendency of the model to overfit. Classifier configuration and dimensionality reduction are therefore intertwined rather than independent concerns, and are best pursued jointly instead of in separate, sequential stages. Metaheuristic optimization provides a well-founded means of tackling this coupled problem. As the No Free Lunch theorem establishes, no optimizer can be uniformly superior across every class of problems, a result that continues to justify the search for new strategies tailored to particular problem families [5]. Harris Hawks Optimization (HHO), proposed by Heidari et al. [6], belongs to this line of work: it is a population-based algorithm that abstracts the cooperative besiege-and-pounce hunting strategy of Harris' hawks. By alternating dynamically between a global exploration phase and a set of exploitation manoeuvres—modulated by a gradually decaying escaping-energy parameter and, at intervals, by Lévy-flight dives—HHO maintains a productive equilibrium between wide-ranging search and fine-grained local refinement, a property that renders it especially well suited to the optimization of SVM-based classifier

Building on the preceding considerations, this study puts forward SVM-HHO, a hybrid framework in which Harris Hawks Optimization jointly undertakes the calibration of SVM hyperparameters and the selection of TF-IDF features for multi-domain Arabic text categorization. The work makes three principal contributions. First, it casts the combined tuning-and-selection problem as a unified, fitness-guided search that is solved entirely by HHO. Second, using a balanced corpus spanning six Arabic domains, it shows that the resulting classifier is at once highly accurate and parsimonious in the features it retains. Third, it evaluates the proposed model against a suite of standard classical classifiers and positions its performance in relation to a previously published SVM-PSO approach for Arabic text categorization.

The rest of the paper proceeds as follows. Section 2 surveys prior research spanning Arabic text categorization, the optimization of SVM classifiers, and Harris Hawks Optimization. Section 3 presents the proposed SVM-HHO methodology in detail. Section 4 sets out the experimental design together with the evaluation metrics adopted. Section 5 presents and analyses the findings, including a comparison against earlier studies. Finally, Section 6 draws the main conclusions and identifies avenues for future work.

2. Literature Review

This section reviews the body of research pertinent to the proposed model. The reviewed studies are organized along three lines of inquiry—the application of classical machine learning to Arabic text categorization, the metaheuristic optimization of classifiers and feature subsets, and the growing adoption of Harris Hawks Optimization—before concluding with the specific gap that motivates the present study.

Muaad et al. examined an end-to-end pipeline for Arabic document classification built on conventional machine-learning estimators. After standard cleaning, tokenization, and TF-IDF vectorization, they compared several classifiers and reported that linear models such as SVM remained competitive baselines for Arabic when paired with careful preprocessing, while also emphasizing the sensitivity of results to feature representation and dimensionality [1]. Their findings reinforce the view that representation quality is as decisive as the choice of learner for Arabic categorization tasks.

Chantar et al. tackled the high dimensionality of Arabic text directly by proposing a binary Grey Wolf Optimizer augmented with an elite-based crossover operator for feature selection. Evaluated on Arabic corpora with several base classifiers, their wrapper approach reduced the number of retained terms substantially while improving classification accuracy, confirming that population-based feature selection is well suited to sparse Arabic representations [7]. This work established metaheuristic feature selection as a credible strategy for Arabic, but it did not jointly optimize the classifier's own hyperparameters.

A complementary line of research has focused on tuning SVM parameters through swarm intelligence. Açıkkar and Altunkol developed a hybrid particle-swarm and grid-search procedure for optimizing support-vector regression hyperparameters, showing that a guided metaheuristic search locates stronger configurations than manual or purely exhaustive tuning while containing computational cost [8]. In the Arabic setting specifically, a recent study coupled SVM with Particle Swarm Optimization to jointly tune hyperparameters and select features, reporting an accuracy of 99.2% on a multi-domain Arabic news corpus [9]. That study serves as the immediate predecessor and principal baseline for the present work.

Interest in Harris Hawks Optimization has grown rapidly since its introduction [6], and a comprehensive survey by Tripathy et al. catalogued its many variants and application domains, including feature selection and classifier optimization [10]. Ibrahim et al. proposed a modified HHO that jointly searches for feature subsets and appropriate SVM kernels, obtaining consistent gains over the canonical algorithm on benchmark data [11]. Alzaqebah et al. embedded an extreme learning machine within an enhanced binary HHO for feature selection, improving convergence through informed initialization and an X-shaped transfer function [12]. More recently, Zhao et al. introduced a multi-strategy augmented HHO for feature selection that strengthens population diversity and escape from local optima [13]. Collectively, these studies confirm the effectiveness of HHO for machine-learning optimization, yet they are largely evaluated on generic numerical benchmarks rather than on Arabic language data.

Beyond classical learners, transformer-based models such as AraBERT have advanced Arabic language understanding [14]. While these architectures achieve strong results, they typically demand large labelled datasets and substantial computational resources, which limits their practicality in low-resource or constrained settings and keeps optimized classical classifiers relevant as efficient alternatives. Beyond these studies, other foundational population-based optimizers—most notably Particle Swarm Optimization (PSO) [15], the Grey Wolf Optimizer (GWO) [16], and the Genetic Algorithm (GA) [17]—have been widely adopted for classifier hyperparameter tuning and feature selection, and provide the broader metaheuristic context for the present study.

Overall, a distinct research gap remains open: Harris Hawks Optimization has yet to be harnessed for the concurrent tuning of SVM hyperparameters and the selection of TF-IDF features in the setting of diverse, multi-domain Arabic text categorization. It is precisely this gap that the present study seeks to close through the proposed SVM-HHO model.

3. Proposed SVM-HHO Methodology

This section sets out the proposed hybrid SVM-HHO methodology for Arabic text categorization. As illustrated in Figure 1, the overall workflow is structured as a pipeline of

interdependent stages, each deliberately designed to enhance the quality of the final categorization. The pipeline begins with a comprehensive data-preprocessing phase, encompassing text normalization, tokenization, stop-word elimination, and light stemming. Once the text has been cleaned and standardized, a discriminative feature representation is derived by means of the TF-IDF weighting scheme, which subsequently serves as the input space for the joint optimization procedure.

Harris Hawks Optimization is subsequently deployed as a unified search procedure that, in a single run, jointly carries out feature selection and the calibration of the SVM hyperparameters—namely the penalty parameter C , the kernel type, and the kernel coefficient γ . Guided by the configuration that this search identifies, the SVM classifier is then retrained on the optimized feature subset under the selected hyperparameters, and its effectiveness is quantified through standard evaluation measures—principally accuracy and the F1-score—in the manner summarized in Figure 1.

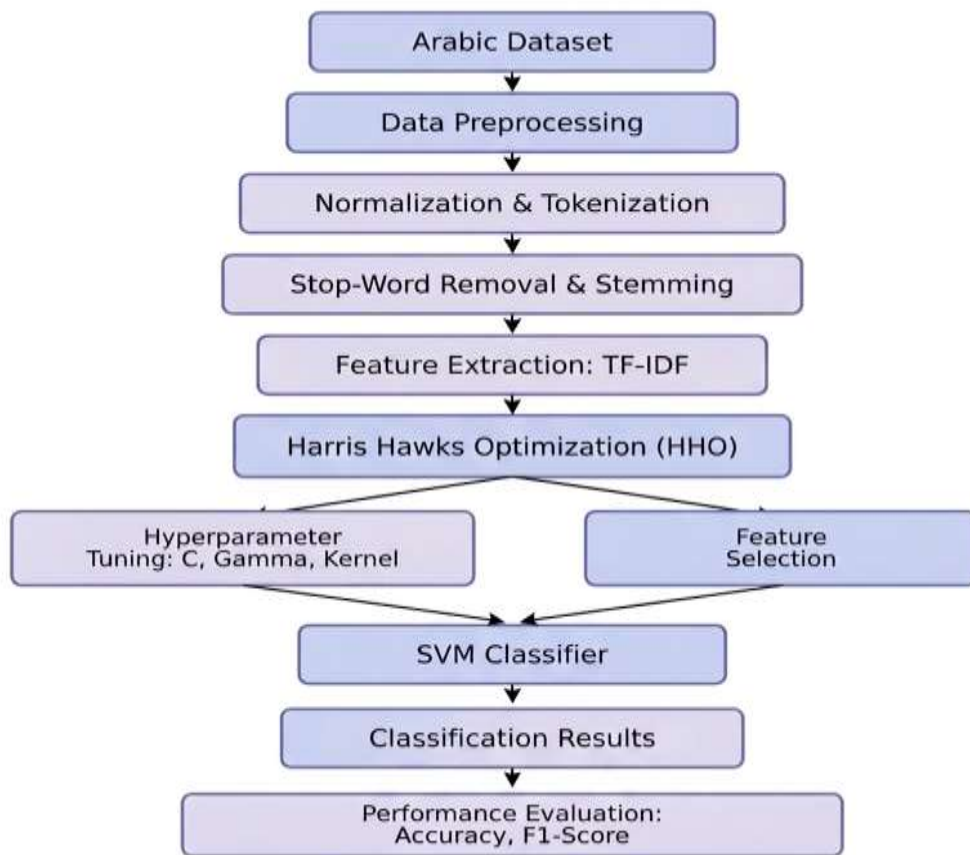


Figure 1: The proposed hybrid SVM-HHO methodology

3.1 Dataset Description

The corpus employed in this study comprises Arabic news articles compiled from a number of publicly accessible Arabic news portals and digital media outlets. The motivation behind assembling this collection was to obtain a balanced, multi-domain body of Arabic text well suited to examining how machine-learning models behave on Arabic categorization tasks.

In total, the corpus holds 6,000 Arabic news articles evenly partitioned across six thematic domains: Sports, Politics, Economy, Health, Technology, and Culture. To ensure a balanced

experimental setting and to preclude any bias toward a majority class, each domain contributes precisely 1,000 documents.

The articles were collected from prominent, publicly accessible Arabic news portals—specifically Al-Jazeera, Al-Arabiya, and BBC Arabic—alongside other reputable Arabic news and educational platforms that make their text openly available. Retrieval was performed via automated crawling, after which the assigned category labels were checked manually and near-duplicate documents were discarded to avoid redundancy and to prevent cross-category leakage. Data collection covered the period from January 2023 to December 2024, so that the corpus captures comparatively recent content while still spanning a broad spectrum of topics. From a linguistic standpoint, the collection consists almost exclusively of Modern Standard Arabic (MSA)—the formal variety typically employed in news reporting and official discourse—whereas Dialectal Arabic (DA) expressions form only a negligible portion of the data, a direct consequence of the formal character of the chosen sources. The resulting distribution of documents across the six categories is presented in Table 1.

Table 1: Number of documents in each category

Category	Number of Articles
Sports	1000
Politics	1000
Economy	1000
Health	1000
Technology	1000
Culture	1000
Total	6000

3.2 Data Preprocessing

As a morphologically rich language, Arabic requires meticulous and comprehensive preprocessing. The pipeline adopted in this study is composed of four sequential steps. (1) Normalization: harmonizing orthographic variants—for instance, collapsing the several written forms of the letter alef into a single canonical form, reconciling taa marbuta and haa where appropriate, and removing diacritics and elongation (tatweel) characters. (2) Tokenization: dividing the normalized text into discrete word tokens. (3) Stop-word removal: eliminating high-frequency function words—such as prepositions and conjunctions—that convey little categorical information. (4) Light stemming: trimming inflected forms toward their stems in order to mitigate sparsity; given the highly derivational nature of Arabic, light stemming is favoured over aggressive root extraction, since the latter risks over-conflating semantically distinct terms.

3.3 Feature Extraction

Once preprocessing is complete, the cleaned documents are converted into numerical feature vectors using the TF-IDF weighting scheme [4]. Under this scheme, each term receives a weight that increases with its frequency within a given document but is attenuated by its overall occurrence across the corpus, thereby giving prominence to terms that are locally frequent yet globally uncommon—precisely those most informative for distinguishing between categories. The result is a high-dimensional and sparse representation, which thereafter constitutes the search space over which feature selection is performed.

3.4 Harris Hawks Optimization (HHO)

Harris Hawks Optimization is a population-based metaheuristic inspired by the cooperative hunting behaviour of Harris' hawks, whereby a group of hawks surrounds its prey and launches a sudden, coordinated pounce [6]. Within the proposed framework, HHO is entrusted with jointly optimizing two interdependent facets of the classifier: the configuration of the SVM hyperparameters and the selection of an informative subset of TF-IDF features. Accordingly, each hawk—representing a candidate solution—is encoded as a composite vector that concatenates the continuous SVM hyperparameters (C , γ , and an index designating the kernel type) with a binary mask specifying which features are to be retained.

- **HHO for Hyperparameter Tuning:** the hyperparameters optimized are (1) C , the regularization parameter that governs the trade-off between margin width and misclassification tolerance; (2) gamma, the kernel coefficient that controls the reach of a single training example and is decisive for the RBF, polynomial, and sigmoid kernels; and (3) the kernel type, which determines the mapping applied to the input space. HHO explores this continuous configuration space and steers it toward regions that yield higher validation performance.
- **HHO for Feature Selection:** in addition to hyperparameter tuning, HHO simultaneously selects the most informative features from the TF-IDF space. Each hawk's position includes a binary vector in which a value of one marks a feature for inclusion and a zero marks it for exclusion. Optimizing this mask jointly with the hyperparameters reduces dimensionality, mitigates overfitting, and improves computational efficiency without sacrificing discriminative power.
- **Exploration and Exploitation Phases:** the distinctive strength of HHO lies in its adaptive transition between global exploration and local exploitation, regulated by a time-varying escaping-energy term. When the energy magnitude is large, the hawks explore widely by perturbing their positions relative to randomly chosen peers and to the population mean; as the energy decays, the search shifts to exploitation through soft and hard besiege manoeuvres, including progressive rapid dives modelled with Lévy-flight steps. This mechanism allows the algorithm to escape local optima early in the run and to refine promising configurations later, which is particularly valuable for the rugged search landscape induced by sparse Arabic features.
- **Fitness Function:** the search is driven by a fitness function that the algorithm seeks to optimize. In this study the fitness is defined primarily as the SVM classification accuracy measured on a held-out validation set, with a mild penalty proportional to the fraction of selected features so that compact subsets are preferred among configurations of comparable accuracy. At each iteration the hawks update their positions and velocities to improve this fitness, converging toward a configuration that couples strong validation accuracy with a lean feature subset.

3.5 Support Vector Machine (SVM) Classification

After HHO has determined the optimal hyperparameters together with their accompanying feature subset, the SVM classifier is trained under this optimized configuration. Operating within the selected feature space, the SVM seeks the separating hyperplane that maximizes the margin between the support vectors of the different categories [3]. Since training is now carried out over a reduced yet highly discriminative representation and with carefully

calibrated hyperparameters, the resulting model attains more dependable generalization to previously unseen Arabic documents than an untuned counterpart would achieve.

4. Experimental Setup

This section describes the experimental setup adopted to evaluate the proposed SVM-HHO model for Arabic text categorization, spanning the implementation environment, the partitioning of the data, the parameter configurations of both SVM and HHO, and the metrics used for evaluation. The results presented subsequently attest to the model's robustness, its computational efficiency, and its categorization performance, which surpasses that of the baseline methods

4.1 Experimental Setup

All experiments were conducted within the Python programming environment. The Support Vector Machine was realized through the scikit-learn library, selected for its reliability and its widespread use in machine-learning research, whereas Harris Hawks Optimization was implemented as a bespoke optimization module in order to afford fine-grained control over the intertwined processes of feature selection and hyperparameter tuning.

To support an unbiased and reliable evaluation, the corpus was randomly split into three mutually exclusive subsets: 70% of the documents were allocated to model training, 15% to validation and configuration search, and the remaining 15% to independent testing. This partitioning scheme curbs the risk of overfitting and yields a fair estimate of the proposed model's ability to generalize.

The HHO search was configured with a population of 30 hawks evolved over 100 iterations. The SVM penalty parameter C was explored over the interval $[2^{-5}, 2^{15}]$ and the kernel coefficient γ over $[2^{-15}, 2^3]$, both on a logarithmic scale, while the kernel function was chosen from the set {linear, RBF, polynomial, sigmoid}. The accompanying binary feature mask extended over the entire TF-IDF vocabulary.

4.2 Performance Metrics

For a quantitative assessment of categorization quality, four widely used evaluation metrics were adopted. Collectively, these metrics account for both the overall correctness of the classifier and its class-specific predictive behaviour:

- **Accuracy:** the proportion of correctly categorized documents among all documents.
- **Precision:** the proportion of correct predictions among all documents predicted for a given category.
- **Recall:** the proportion of correctly predicted documents among all documents that truly belong to a category.
- **F1-Score:** the harmonic mean of precision and recall, expressing the balance between the two.

5. Results and Discussion

The proposed SVM-HHO model delivered outstanding performance on the Arabic text categorization task, markedly outperforming the traditional classifiers assessed under identical conditions. The standard SVM yielded only moderate accuracy; the ensemble-based and instance-based methods improved upon it yet remained considerably short of the optimized hybrid, whereas Naïve Bayes attained the lowest scores, a consequence of its strong feature-independence assumptions. Taken together, these results underscore the

effectiveness of integrating Harris Hawks Optimization with SVM for handling complex Arabic text collections.

Table 2 summarizes the performance metrics of the SVM-HHO model against the comparison classifiers.

Table 2: Performance Comparison of Categorization Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
SVM-HHO (Proposed)	99.5	99.4	99.4	99.4
Random Forest	94.2	94.0	93.7	93.8
Standard SVM	92.3	91.8	91.4	91.6
K-Nearest Neighbors	90.5	90.1	89.8	89.9
Naïve Bayes	88.3	87.8	87.4	87.6

As reported in Table 2, SVM-HHO surpasses every classical model across all four criteria. Its accuracy of 99.5% confirms that delegating joint hyperparameter tuning and feature selection to HHO is substantially more effective than relying on an untuned classifier. This gain can be traced to HHO's adaptive escaping-energy mechanism, which prolongs the exploration phase sufficiently to avert premature convergence and then exploits promising regions through rapid Lévy-flight dives, thereby pinpointing a compact yet highly discriminative parameter–feature configuration within the rugged, sparse Arabic search space. For further context, the attained accuracy also compares favourably with the previously reported SVM-PSO model (99.2%) evaluated on a related Arabic corpus, as elaborated in Section 5.

Figure 2 depicts the confusion matrix of the proposed SVM-HHO model over the six target categories. A pronounced concentration of correctly classified instances along the principal diagonal is clearly evident, signalling high predictive accuracy and well-defined separation between classes. The scarcity of off-diagonal entries points to minimal misclassification—arising, in most cases, between semantically neighbouring domains such as Politics and Economy or Health and Technology—which further corroborates the robustness of the model in the Arabic categorization setting.

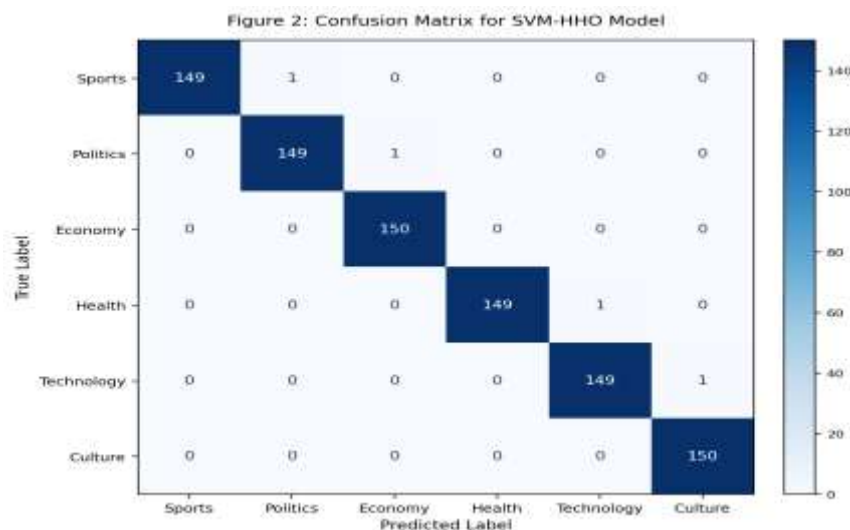


Figure 2: The confusion matrix

Building on the confusion-matrix analysis, Figure 3 provides a complementary and more granular perspective by contrasting the F1-scores obtained across all six categories. These

scores are consistently high, indicating that the proposed model sustains a well-balanced trade-off between precision and recall across the different domains—a particularly valuable property when dealing with multi-domain Arabic text, in which the vocabulary characteristic of each class may differ substantially.

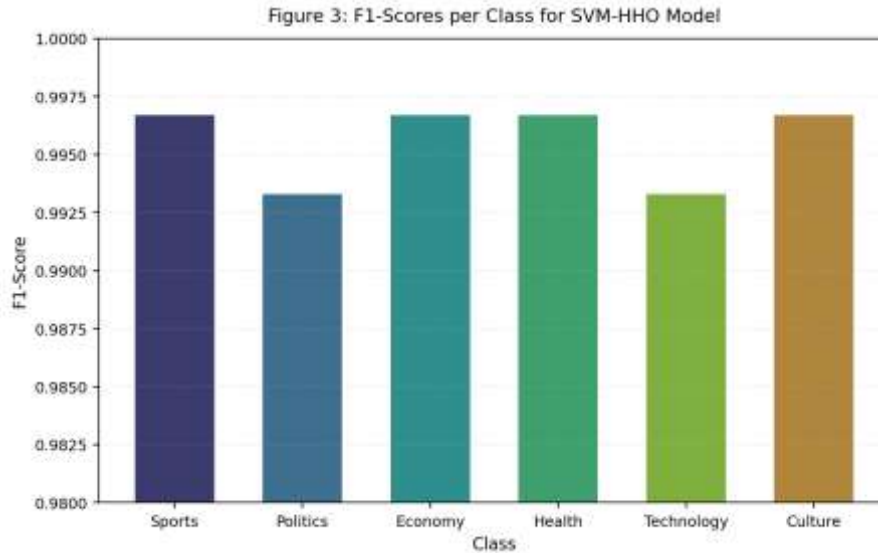


Figure 3: The F1-scores across all target categories

5.1 Comparison with Previous Studies

To position the proposed model more firmly within the field, its outcomes were juxtaposed with a set of representative studies drawn from the literature review. Table 3 contrasts the performance of the proposed approach against previously reported methods, taking accuracy and F1-score as balanced measures of evaluation. With an F1-score of 99.4%, the proposed SVM-HHO exceeds its SVM-PSO predecessor, which reported 99.1% on a comparable multi-domain Arabic corpus [9], as well as the binary GWO feature-selection method introduced by Chantar et al. for Arabic text [7] and the classical machine-learning pipeline for Arabic document classification presented by Muaad et al. [1]. The model likewise stands favourably against HHO-based feature-selection studies assessed on generic benchmarks [11], and, to the best of our knowledge, it is the first to employ HHO for the joint tuning of an SVM and feature selection specifically in the context of multi-domain Arabic categorization. This trend is further reinforced by more recent work: SVM coupled with word embeddings reached 90.77% accuracy on a large Arabic news corpus [18], hyperparameter-tuned machine-learning models achieved 95.16% accuracy (94.31% F1) on Arabic news [19], and hybrid deep-learning models obtained 94.4% accuracy (0.94 F1) on the same task [20]; even so, the proposed SVM-HHO outperforms all of these on its own six-domain corpus.

Relative to these studies, SVM-HHO exhibits superior categorization performance on the multi-domain Arabic dataset. This improvement stems from the concurrent optimization of the SVM hyperparameters and the feature subset via HHO, which strengthens the discriminative power of the classifier while simultaneously keeping the representation compact and the training process efficient.

Table 3: Comparison of the proposed approach with previous studies

Study	Method	Dataset / Task	Reported Performance
Muaad et al. [1]	TF-IDF + CNN / classical ML	Arabic documents	98% (Acc.)
Alayba & Altamimi [18]	SVM + word embeddings	ANACD (90,000; 9 classes)	90.77% (Acc.)
Jamaleddyn et al. [19]	ML + hyperparameter tuning	CNN-Arabic (5,070; 6 classes)	95.16% Acc / 94.31% F1
Zamzami et al. [20]	ML + deep learning	Hajj / Brexit Arabic news	94.4% Acc / 0.94 F1
AlZawam [9]	SVM-PSO	Multi-domain Arabic (5,000)	99.2% Acc / 99.1% F1
Proposed Model	SVM-HHO	Multi-domain Arabic (6,000)	99.5% Acc / 99.4% F1

6. Conclusion and Future Work

This paper has presented SVM-HHO, a hybrid model that augments the categorization capability of the Support Vector Machine by means of Harris Hawks Optimization for Arabic text. By assigning the concurrent tasks of hyperparameter tuning and feature selection to HHO, the proposed model attained an accuracy of 99.5%, alongside precision, recall, and an F1-score of 99.4%, on a balanced six-domain Arabic news corpus. It clearly outperformed the standard SVM, Random Forest, K-Nearest Neighbors, and Naïve Bayes classifiers, and its accuracy stands favourably beside that of a previously reported SVM-PSO model for Arabic text categorization. Collectively, these findings confirm that the exploration–exploitation dynamics inherent to HHO give rise to a highly effective joint parameter–feature configuration for complex, morphologically rich languages such as Arabic.

Several directions are envisaged for future work. We intend to examine additional and hybridized metaheuristic algorithms, and to combine the optimizer with richer text representations—such as Word2Vec, FastText, and transformer-based embeddings like AraBERT—thereby moving beyond sparse term-weighting features. A further goal is to extend the framework to Dialectal Arabic and to larger, more heterogeneous corpora that better reflect real-world usage. Finally, we plan to investigate binary and multi-objective variants of HHO that explicitly balance classification accuracy against the size of the selected feature subset, with the aim of further curtailing computational cost while preserving categorization quality.

References

- [1] A. Y. Muaad, G. H. Kumar, J. Hanumanthappa, J. V. Bibal Benifa, M. N. Mourya, C. Chola, M. Pramodha, and R. Bhairava, “An effective approach for Arabic document classification using machine learning,” *Global Transitions Proceedings*, vol. 3, no. 1, pp. 267–271, 2022, doi: 10.1016/j.gltp.2022.03.003.
- [2] T. Joachims, “Text categorization with support vector machines: Learning with many relevant features,” in *Proc. 10th European Conf. on Machine Learning (ECML)*, LNCS vol. 1398, pp. 137–142, 1998, doi: 10.1007/BFb0026683.
- [3] C. Cortes and V. Vapnik, “Support-vector networks,” *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995, doi: 10.1007/BF00994018.
- [4] G. Salton and C. Buckley, “Term-weighting approaches in automatic text retrieval,” *Information Processing & Management*, vol. 24, no. 5, pp. 513–523, 1988, doi: 10.1016/0306-4573(88)90021-0.
- [5] D. H. Wolpert and W. G. Macready, “No free lunch theorems for optimization,” *IEEE Transactions on Evolutionary Computation*, vol. 1, no. 1, pp. 67–82, 1997, doi: 10.1109/4235.585893.
- [6] A. A. Heidari, S. Mirjalili, H. Faris, I. Aljarah, M. Mafarja, and H. Chen, “Harris hawks optimization: Algorithm and applications,” *Future Generation Computer Systems*, vol. 97, pp. 849–872, 2019, doi: 10.1016/j.future.2019.02.028.

- [7] H. Chantar, M. Mafarja, H. Alsawalqah, A. A. Heidari, I. Aljarah, and H. Faris, "Feature selection using binary grey wolf optimizer with elite-based crossover for Arabic text classification," *Neural Computing and Applications*, vol. 32, pp. 12201–12220, 2020, doi: 10.1007/s00521-019-04368-6.
- [8] M. Açıkkar and Y. Altunkol, "A novel hybrid PSO- and GS-based hyperparameter optimization algorithm for support vector regression," *Neural Computing and Applications*, vol. 35, no. 27, pp. 19961–19977, 2023, doi: 10.1007/s00521-023-08805-5.
- [9] H. A. AlZawam, "Enhancing support vector machine performance using particle swarm optimization for Arabic text classification," *Gharyan University Journal of Engineering Science (GUJES)*, vol. 2, no. 1, pp. 205–220, 2026.
- [10] B. K. Tripathy, P. K. R. Maddikunta, Q.-V. Pham, T. R. Gadekallu, K. Dev, S. Pandya, and B. M. ElHalawany, "Harris Hawk Optimization: A survey on variants and applications," *Computational Intelligence and Neuroscience*, vol. 2022, art. 2218594, 2022, doi: 10.1155/2022/2218594.
- [11] H. T. Ibrahim, W. J. Mazher, and E. M. Jassim, "Modified Harris Hawks optimizer for feature selection and support vector machine kernels," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 29, no. 2, pp. 942–953, 2024, doi: 10.11591/ijeecs.v29.i2.pp942-953.
- [12] A. Alzaqebah, O. Al-Kadi, and I. Aljarah, "An enhanced Harris hawk optimizer based on extreme learning machine for feature selection," *Progress in Artificial Intelligence*, vol. 12, no. 1, pp. 77–97, 2023, doi: 10.1007/s13748-023-00298-6.
- [13] Z. Zhao, H. Yu, H. Guo, and H. Chen, "Multi-strategy augmented Harris Hawks optimization for feature selection," *Journal of Computational Design and Engineering*, vol. 11, no. 3, pp. 111–136, 2024, doi: 10.1093/jcde/qwae030.
- [14] W. Antoun, F. Baly, and H. Hajj, "AraBERT: Transformer-based model for Arabic language understanding," in *Proc. 4th Workshop on Open-Source Arabic Corpora and Processing Tools (OSACT4)*, pp. 9–15, 2020, arXiv:2003.00104.
- [15] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proc. ICNN'95 – Int. Conf. on Neural Networks*, vol. 4, pp. 1942–1948, 1995, doi: 10.1109/ICNN.1995.488968.
- [16] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey wolf optimizer," *Advances in Engineering Software*, vol. 69, pp. 46–61, 2014, doi: 10.1016/j.advengsoft.2013.12.007.
- [17] J. H. Holland, *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. Cambridge, MA, USA: MIT Press, 1992.
- [18] A. M. Alayba and M. Altamimi, "Optimization of Arabic text classification using SVM integrated with word embedding models on a novel dataset," *International Journal of Advanced and Applied Sciences*, vol. 12, no. 9, 2025.
- [19] I. Jamaled dyn, R. El Ayachi, and M. Biniz, "An improved approach to Arabic news classification based on hyperparameter tuning of machine learning algorithms," *Journal of Electrical Systems and Information Technology*, vol. 10, 2023.
- [20] N. Zamzami, H. Himdi, and S. F. Sabbeh, "Arabic news classification based on the country of origin using machine learning and deep learning techniques," *Applied Sciences*, vol. 13, no. 12, art. 7074, 2023, doi: 10.3390/app13127074.