A Study on Text Feature Selection Using Ant Colony and Grey Wolf Optimization

1st Joan Angelina Widians
Doctoral Program dept. of Computer Science and Electronics
Universitas Gadjah Mada
Dept. of Informatics, Faculty of Engineering
Universitas Mulawarman
Samarinda, Indonesia
angelwidians@unmul.ac.id

2nd Retantyo Wardoyo
Dept. of Computer Science and Electronics, Faculty of Mathematics and Natural Sciences
Universitas Gadjah Mada
Yogyakarta, Indonesia
rw@ugm.ac.id

3rd Sri Hartati
Dept. of Computer Science and Electronics, Faculty of Mathematics and Natural Sciences
Universitas Gadjah Mada
Yogyakarta, Indonesia
shartati@ugm.ac.id

***Abstract*—Text classification (TC) is widely used for organizing digital documents. The issues in TC are numerous characteristics and high elements dimension. Many pattern classification issues require feature selection (FS), which is pertinent. FS removes unneeded and redundant data from the dataset. The Ant Colony Optimization (ACO) and Grey Wolf Optimizer (GWO) for FS are the main topics of our thorough assessment of the literature on the Swarm Intelligence (SI) algorithm. Furthermore, it illustrates how the hybrid SI technique is used in FS across various sectors. The hybrid SI technique uses applicable data from various FS methods to find feature subsets with smaller sizes and better classification performance than those found by regular FS algorithms.**

Keywords—Text classification, Feature selection, Swarm Intelligence, Ant Colony Optimization, Grey Wolf Optimizer.

I. INTRODUCTION

Machine learning (ML) has grown in importance and popularity as the popular intelligent approach for learning from written documents. Text Classification (TC) is one of the most popular methods for classifying documents automatically. The performance of ML issues is high-dimensional data, particularly multiple characteristics. An essential step in TC is feature selection (FS), representing a subset of features. Algorithms must carry out the essential FS process for ML [1] [2]. Finding the optimum component without stressing the label's colossal group instance is a challenge for feature optimization that FS techniques may be capable of addressing in domains with many high-dimensional features [3].

Swarm Intelligence (SI) algorithms are used to address FS [4][5][6]. The solution to this issue is SI-based searching, which uses randomized searching to discover suitable sources globally [7][8][9]. The authors analyzed the Ant Colony Optimization (ACO) and Grey Wolf Optimizer (GWO) algorithms for the text FS problem in great detail. This research suggests an in-depth analysis of the FS process using SI methods. The associated works are described in Section 2. The FS problem's SI algorithm is in Section 3. Section 4 shows the hybrid SI for FS and its variations. The paper is concluded in the final portion.

KEMENDIKBUD-RISTEK Indonesia funds this research (1900/UN1/DITLIT/Dit-Lit/PT.01.03/2022).

II. RESEARCH METHOD

A. Swarm Intelligence

Fast expansion SI, which has existed since the 2010s. Recently, various techniques for addressing optimization issues suggested by natural mechanisms have been explored [10][11][12]. Computer programs known as SI algorithms are built on population-based algorithms initially found in nature. The primary strategies for the SI algorithm's self-organization are listed in Table 1 [6]. High-dimensional datasets have been successfully analyzed using SI procedures. Consequently, the SI framework is in Fig. 1 [8].

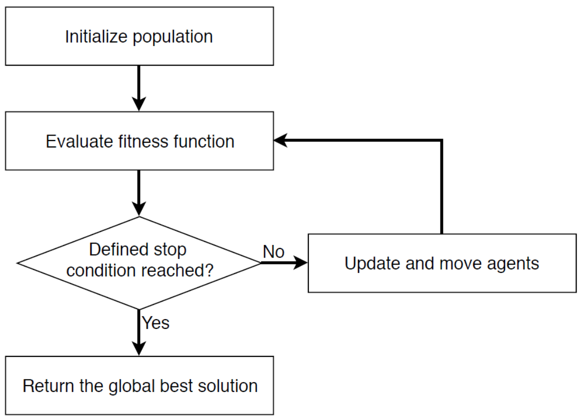

Fig. 1. Swarm Intelligence Framework

The SI framework includes initializing the population, establishing the stop condition, assessing the fitness function, updating and relocating agents, and returning the best response. Prior to initialization, determine the values of the algorithmic parameters. The first step in the SI algorithm is population initialization. The primary objective of the stop condition is algorithmic termination. The search agents are evaluated at the third stage of the SI framework, which also analyzes the fitness function. The fitness function may be one-dimensional, such as step condition, or simple, like classification accuracy. The SI algorithm's agents move and update based on a theoretical basis.

TABLE I. STRATEGIES FOR SWARM INTELLIGENCE ALGORITHMS

Strategies	Functional Description
Agent	Generate the subtask following the work assignment.
Collaboration	Exchange data directly or indirectly among the agents to establish intelligent behavior that is utilized to inform population decisions.
Exploration	Using the search specified space, find the best global solution.
Exploitation	In the allocated area for the exploration stage, expand local search efforts.
Fitness Function	Accomplish the evaluation procedures involved with each prospective solution.

B. Text Classification

Text classification (TC), often known as text categorization, defines a document's content and then applies one or more class labels or categories from a specified list to the paper [13][14]. TC can add new patents to existing patent categories and automatically classify webpages or documents based on pre-defined labels, among other valuable applications [15][16]. Indexing, dimension reduction, and automatic categorization are techniques used in TC to determine whether it is categorized [17][18].

C. Feature selection

FS deletes redundant and irrelevant features while selecting a subset of pertinent features. Many scientists conduct thorough experiments to address these problems. Further, efficiently divides these highly dimensional variables and data into essential aspects. The FS technique has four fundamental phases: subset generation, subset evaluation, stopping criterion, and result validations [15]. In a generating technique, feature subsets are processed for evaluations. A random subset of characteristics and features may also be used as a starting point, after which further information may be added, modified, or copied. The evaluation parameter defines if a subset is adequate for analysis. Although an FS technique must be validated, the validation step is not part of the FS process. The FS process is seen in Fig.2 [20].

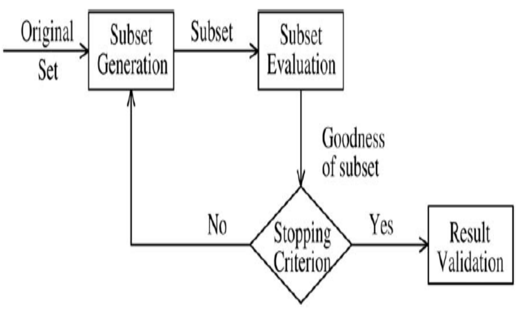

Fig. 2. The Feature Selection System

The four main divisions of FS methods are filter, wrapper, embedding, and hybrid approaches. The filter technique analyzes and selects feature subsets based on available data qualities rather than algorithms [21]. Wrapper approaches: An algorithm (such as a classification algorithm) and the evaluation metrics assess a subset of features (e.g., classification accuracy). Embedded approaches-a subset of the FS process is automatically embedded in an algorithm. Hybrid approaches make use of both wrapper and filter techniques [22].

III. SWARM INTELLIGENCE FOR FEATURE SELECTION

A. Swarm Intelligence for feature selection

SI approaches are the most effective for optimizing the feature subset selection process within a wrapper model method. To evaluate the quality of particular feature subsets, they must continually apply the ML technique, which is very computationally expensive. Improve the FS process within a wrapper model; the SI algorithms seek to identify which subset of all available features pairs with a preset ML algorithm to have the highest prediction performance. The two procedures of selecting a subset of parameters and evaluating the subset using the classification algorithm chosen are maintained in a typical wrapper method until the best outcome (i.e., a sure accuracy) is reached [23].

Wrapper FS techniques reduce the amount of search space available for finding features. The wrapper technique often searches for subsets using the best algorithm before applying a classification algorithm to the subgroup. The wrapper technique offers a learning strategy for feature evaluation. Thus, the attributes are chosen based on how they affect the accuracy improvement. Fig. 3 displays a general structure for a wrapper model of FC for categorization [8].

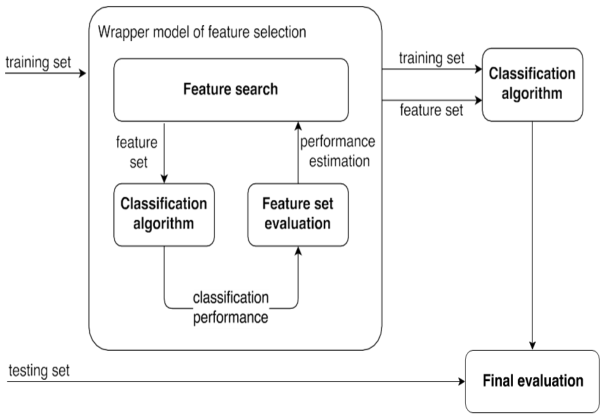

Fig. 3. General structure for wrapper models of feature selection

B. Ant Colony Optimization based feature selection

Dorigo and his associates introduced Ant Colony Optimization (ACO) in the early 1990s. The ACO is a brand-new meta-heuristic for solving challenging combinatorial optimization issues inspired by nature [10][24]. The ACO algorithm was developed first to address the Traveling Salesman Problem (TSP). Since then, it has performed a variety of complex issues, such as the quadratic assignment problem, routing in telecommunication networks, graph coloring problems, scheduling, and others [10].

Although the FS issue may be expressed as a complex diagram, the ACO procedure settles it. Most ACO-based FS methods show features as graph nodes with edges indicating the component that should be chosen next. Ants explore the network for a path that passes by some nodes corresponding to the element. Fig. 4 depicts this configuration. The ant currently resides at node f_1, where it can select which feature to add to its journey (dotted lines). According to the transition rule, it selects feature f_2 next, followed by f_3 and f_4. Upon arrival at f_4, the current subset $\{f_1, f_2, f_3, f_4\}$ is determined to satisfy the traversal stopping criteria [25]. Following its investigation, the subterranean ant generates this element subset as an appropriate option for information pressure. Heuristics or pheromones, in this case, do not control the relationship. As a result, each character may have its own set of heuristics and pheromones [26] [27].

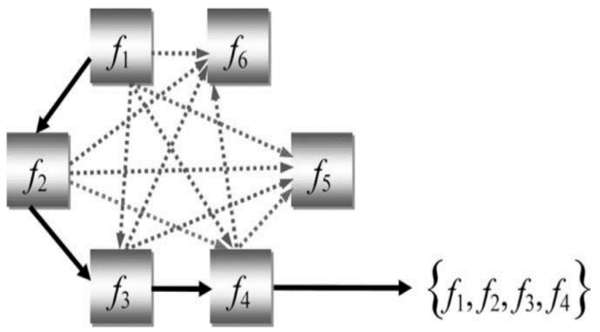

Fig. 4. ACO representation for Feature Selection

Setting all arrays to false forces the ants into an empty memory state before they can search the graph for a solution. Afterwards, a random number is selected to lead each ant's starting node. They provide an exhaustive tour of the entire structure utilizing a choice method based on heuristic data and pheromone trails at each stage. Whenever the halting condition is still unrecognized, the ants will proceed through the nodes in search of the best feature for the subset. The simulated ant calculates the probability that features-*i* will be added to its solution at time-*t* by ant-*k* using a procedure known as the probabilistic transition rule. The probabilistic transition rule comprises heuristic desirability and pheromone levels, as illustrated in Equation (1) [28].

$$p_i^k(t) = \begin{cases} \dfrac{[\tau_i(t)]^\alpha [\eta_i]^\beta}{\sum_{j \in J^k} [\tau_j(t)]^\alpha [\eta_j]^\beta}, & if\ i \in J^k \\ 0 & otherwise \end{cases} \quad (1)$$

J^k is the set of ant k's unvisited features, and η_i is the heuristic desirability of element-*i*.

Afterwards, τ_i (t) is the pheromone value at feature-*i*. While η_j is the heuristic value of feature-*j*, and τ_j (t) is the pheromone value of feature-*j*. At the same time, α, and β are two parameters that determine the relative importance of the pheromone value and heuristic information.

Numerous practical ACO-based FS algorithms have been put out in the literature. ACOFS is a brand-new hybrid ACO algorithm that suggests FS [29]. However, [26] provides an FS method based on ACO that leverages the classifier's performance and the length of the selected feature vector as heuristic data for ACO [30][31]. In ACOFS's hybrid search strategy, the wrapper and filter algorithms are combined [32][33][34][35].

C. *Grey Wolf Optimizer based feature selection*

Grey Wolf Optimizer (GWO) algorithm to address optimization issues. It imitates every step of the grey wolf's hunting procedure, from the search strategy to the stage when the prey is encircled and attacked. The social hierarchy of wolves is divided into four groups, alpha, beta, delta, and omega [36], as seen in Fig. 5.

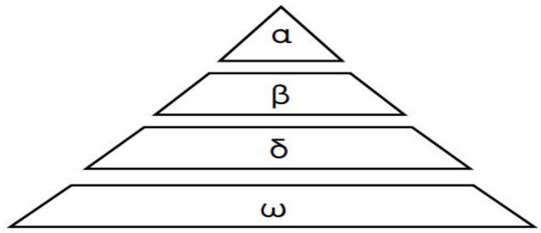

Fig. 5. Hierarchical Structure of Grey Wolves

Alpha (α) wolf is a leader and decision-maker. Beta (β) is alpha's assistant in leading the group. Delta (δ) obeys the orders from alpha and beta. The remaining wolves are called omega (ω), which follow other wolves in their movement. The primary sources of inspiration for GWO are grey wolves' leadership, intelligence, and hunting tendencies in the wild. How grey wolves look for prey is GWO's second source of inspiration. The acts used to complete the hunting phase are chase, encircle, harass, and assault [37][38][39]. Table 2 summarizes some works that use ACO and GWO for FS.

IV. HYBRID SWARM INTELLIGENCE

A. *Hybrid Swarm Intelligence for Various Applications*

Numerous writers have recently developed hybrid techniques to address optimization issues in various application fields. We have investigated a few mixed swarm-based methods that use ACO or GWO as part of the proposed hybridization. Our summary of various recent hybrid swarm-based techniques is presented in Table 3.

B. *Hybrid Swarm Intelligent Based Feature Selection*

Many researchers combine unique characteristics of various SI algorithms to increase overall efficiency. Two ideas from the optimization algorithm are exploitation and exploration. Exploration is a technique for choosing the best location within a search area. Conversely, exploitation leverages a good point in the search space to find a more promising point. A quality strategy incorporating two or more SI procedures balances exploitation with exploration. In other circumstances, researchers used multiple co-evolving swarms, also known as multi-swarms or hybrid swarms, to tackle this multi-objective optimization problem [42]. A mixed GWO and PSO were then presented to balance exploitation and exploration for the FS problem [5]. Table 4 summarizes the hybrid SI for FS that we have developed.

The text categorization process is shown in Fig. 6, which typically entails preprocessing, feature selection, classification model creation, and evaluation tasks. In addition, words like lexical items and uncommon words—frequently found in text documents but lacking information that aids in discriminating among text classes—are also eliminated. A select group of highly informative words is kept to represent texts as vectors of features. The feature extraction process takes out via n-gram Term Frequency-Inverse Document Frequency (TF-IDF) computation to obtain the feature vector. The TF-IDF statistic gauges a word's significance to writing within a collection. Each word (or term) in a text is assigned a variable weight, known as term frequency, based on how frequently it appears. Since the maximum value is the number of occurrences, the weight assigned to a word in the document increases as its frequency increases (High TF). If we express raw frequency t as f (t, d), then the simple *tf* scheme in Equation 2 [19]. The IDF technique doesn't consider how frequently a word appears in the text; it solely evaluates terms based on their occurrence in a document. The TF-IDF is commonly employed as a weighting factor in text mining and information retrieval. The extracted features are fed to the FS process to reduce the dimensionality of a feature vector. The formula of TF-IDF in Equations (3) and (4) [22].

$$tf\ (t,d) = f(t,d) \tag{2}$$

$$IDF = log10\left(\frac{D}{dfi}\right) \tag{3}$$

$$W(t,f,d) = tf(t,d) * IDF(t,d) \tag{4}$$

In Equation (3), D is the number of documents containing term-t. While *dfi* is the number of occurrences (frequency) of the word against D. In Equation (4), W is the weight of the d-document against the term-t [44].

FS attempts to improve the accuracy and computational efficiency of text classifiers in text classification by removing irrelevant and noisy features using the suggested approach, hybrid SI. FS with hybrid SI can decrease computer complexity while improving classification accuracy by removing noisy data. Due to FS, the features are at their peak. Develop a text categorization system; documents must first be rendered eligible for classification by being represented as feature vectors. These feature vectors have separate test and training sets. The learning algorithm will be trained on a feature from the training set before creating the classifier. The ML classifier splits the input text into the categories it intends to employ [17].

In the last stage, the best feature selection methods using training data are applied to generate a text classifier. Its classification accuracy is then verified using a new set of test data. The efficacy of the word categorization findings is assessed using the top feature subset. The optimal feature subset's results provide a set of features for categorizing documents. ML classifiers such as Support Vector Machine (SVM), Naive Bayes Classifier, K-Nearest Neighbor (k-NN), and Decision Tree can be used to categorize text [18]. The SVM algorithm is one of the supervised ML techniques for various classification issues. SVM works remarkably well with high-dimensional data. SVM techniques are available in both linear and non-linear variants. The non-linear model does not divide the classes. In the linear variant, classes are split using hyperplanes [45]. Naive Bayes uses the multinomial model for large datasets. Bayesian reasoning and probabilistic inference are utilized to predict the target class. A substantial percentage of the classification is comprised of characteristics. According to the Naive Bayes technique, a feature should be independent of another component given known prior probability and class conditional probability. In contrast, a modified k-NN-based text classification algorithm is proposed and improves the system's performance. Additionally, unlike Naive Bayes, Decision Tree-based text classification does not imply independence among its characteristics. Decision trees perform best as text classifiers when there are few features, considering the allure of building a classifier with various components. In the testing procedure, classification performance is appropriately analyzed [46].

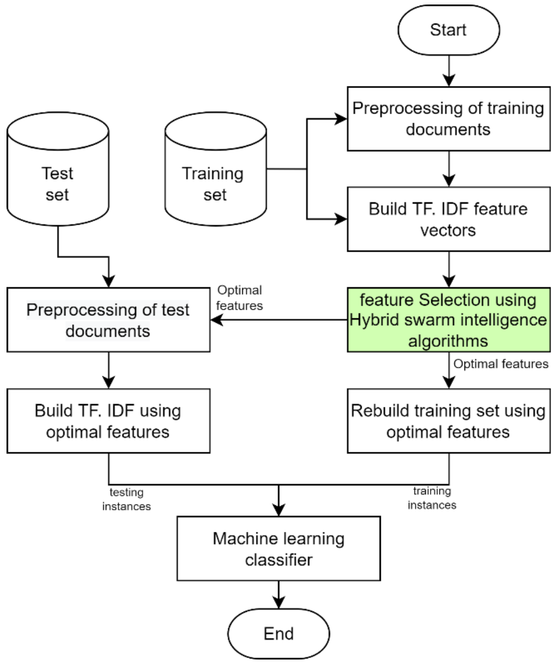

Fig. 6. Block diagram text categorization

V. CONCLUSION

Features selection techniques are classified into four types: filter, wrapper, embedded, and hybrid strategy. This research examines the contemporary feature selection-based wrapper strategy using a combination of two Swarm Intelligence (SI) algorithms. We investigated numerous hybrid swarm-based approaches that require hybridization with ACO or GWO. It seems sensible to use SI approaches to enhance the choice of feature subsets. The hybrid SI technique aims to discover feature subsets with smaller sizes and higher classification performance than those distinct algorithms by incorporating helpful information from various FS algorithms.

TABLE II. SEVERAL WORKS FOR FEATURE SELECTION SUMMARY

Author	Method	FS approach	Domain
[47]	ACO	Wrapper	Protein post-synaptic activity prediction
[26]	ACO	Wrapper	Face recognition system
[48]	ACO	Wrapper	Classification
[25]	ACO	Wrapper	Classification
[30]	ACO	Wrapper	Classification
[31]	ACO	Wrapper	Classification
[49]	ACO	Wrapper	Classification
[50]	ACO	Filter	Classification
[28]	ACO	Hybrid	Classification
[51]	ACO	Wrapper	Text categorization
[52]	ACO	Wrapper	Classification
[4]	ACO	Wrapper	Image classification
[29]	ACO	Hybrid	Classification
[24]	ACO	Wrapper	Classification
[32]	BACO	Wrapper	Reggression
[33]	ACO	Wrapper	Classification
[34]	ACO	Filter	Classification
[35]	ACO	Filter	Classification
[53]	GWO	Wrapper	EMG signals Classification
[54]	GWO	Wrapper	Classification
[21]	ACO	Hybrid	Classification
[13]	Binary GWO	Wrapper	Text Classification
[37]	GWO	Wrapper	Classification
[7]	Improve d BGWO	Wrapper	Classification
[38]	GWO	Wrapper	Classification
[3]	GWO	Wrapper	Text Classification
[39]	IG-GWO	Hybrid	Text Classification
[23]	MOGW	Wrapper	Text Classification

TABLE III. HYBRID SWARM-BASED APPROACHES

Authors	Method	Hybridization principle	Domain
[55]	PSO-ACO	The PSO-ACO algorithm is utilized as the k - means algorithm's initial condition.	Clustering
[56]	ACO-PSO	The ACO pheromone update rules are combined with local exploration and global exploitation to provide the search mechanism of the PSO algorithm.	TSP
[57]	ACO-PSO	According to a comparison of the top solutions that ACO and PSO uncover, the best solution is allocated to the system's overall best explanation.	Turkish energy demand estimation
[58]	ACO-CS	Breast cancer detection using digital mammography using the ACO and CS algorithm.	Image processing
[59]	ACO-PSO	The PSO algorithm evaluates the importance of inter-city pheromones and distance and finds the appropriate parameter values for ACO city selection operations.	TSP
[40]	GWO-PSO	Convex economic load dispatch with GWO-PSO.	Economic Load dispatch
[60]	PSO-GWO	Enhance the exploration in GWO with the PSO exploitation to produce the strengths of both variants.	Convergence performance
[61]	ACO-GWO	GWO based on the ACO for Assembly Sequence Planning (ASP).	ASP
[21]	Binary GA-PSO	To combine the exploration capabilities of PSO and the ability of GA, BGSO can extract data from the acquired feature subsets sufficiently.	Pattern recognition
[41]	GWO-PSO	PSO to improve the GWO method's progress in addressing the ORPD problem in the context of electrical power networks.	Optimal reactive power dispatch (ORPD) problem

TABLE IV. HYBRID SWARM INTELLIGENCE BASED FEATURE SELECTION

Authors	Method	Hybridization principle	Domain
[42]	ACO-GA	ACO-GA hybrid for text FS. The size of the chosen feature subset is used as heuristic data.	Text categorization
[20]	ACO-PSO	ACO-PSO for reduce of the number of selected features	Classification
[62]	GWO-CS	Alpha wolf, Beta wolf, and Delta wolf are the top three solutions that GWO has updated using CS's global-search functionality.	Classification
[14]	ACO-ABC	Combines ACO and ABC to optimize FS	Classification
[63]	PSO-GWO	Incorporating the PSO into the GWO to assign class labels and produce various text document clusters.	Text clustering
[64]	BGWO-PSO	The hybrid BGWO-PSO to find the best feature subset.	Classification
[5]	GWO-PSO	The best subset of the feature is located using the hybrid GWO-high PSO's exploitation and exploration capabilities.	Classification
[65]	GWO-GOA	The hybrid GWO-GOA method processes the text FS by first picking the local optima from the text document and then choosing the best global optima from the local optimum.	Text clustering
[66]	GWO-PSO	Combines BPSO (BPSO) and BGWO (BGWO) to enhance the global search capability when solving the FS problem in high-dimensional datasets.	Classification

ACKNOWLEDGMENT

A grant from KEMENDIKBUD-RISTEK Indonesia aids this research (1900/UN1/DITLIT/Dit-Lit/PT.01.03/2022). The authors thank the Faculty of Engineering, Mulawarman University, for supporting this study.

REFERENCES

[1] V. Bijalwan, P. Kumari, J. Pascual, and V. B. Semwal, "ML approach for text and document mining," *arXiv Prepr. arXiv1406.1580*, 2014.

[2] H. Liu and L. Yu, "Toward integrating feature selection algorithms for classification and clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 4, pp. 491–502, 2005.

[3] I. Al-Jadir and W. A. Mahmoud, "A Grey Wolf Optimizer Feature Selection method and its Effect on the Performance of Document Classification Problem".

[4] B. Chen, L. Chen, and Y. Chen, "Efficient ant colony optimization for image feature selection," *Signal Processing*, vol. 93, no. 6, pp. 1566–1576, 2013.

[5] E.-S. El-Kenawy and M. Eid, "Hybrid gray wolf and particle swarm optimization for feature selection," *Int. J. Innov. Comput. Inf. Control*, vol. 16, no. 3, pp. 831–844, 2020.

[6] K. S. Kyaw and S. Limsiroratana, "Traditional and Swarm Intelligent Based Text Feature Selection for Document Classification," *Proc. - 2019 19th Int. Symp. Commun. Inf. Technol. Isc. 2019*, pp. 226–231, 2019, doi: 10.1109/ISCIT.2019.8905200.

[7] P. Hu, J.-S. Pan, and S.-C. Chu, "Improved binary grey wolf optimizer and its application for feature selection," *Knowledge-Based Syst.*, vol. 195, p. 105746, 2020.

[8] L. Brezočnik, I. Fister, and V. Podgorelec, "Swarm intelligence algorithms for feature selection: a review," *Appl. Sci.*, vol. 8, no. 9, p. 1521, 2018.

[9] K. S. Kyaw, S. Limsiroratana, and T. Sattayaraksa, "A comparative study of meta-heuristic and conventional search in optimization of multi-dimensional feature selection," *Int. J. Appl. Metaheuristic Comput.*, vol. 13, no. 1, pp. 1–34, 2022.

[10] M. Dorigo, M. Birattari, and T. Stutzle, "Ant colony optimization," *IEEE Comput. Intell. Mag.*, vol. 1, no. 4, pp. 28–39, 2006.

[11] S. Suyanto, A. T. Wibowo, S. Al Faraby, S. Saadah, and R. Rismala, "Evolutionary Rao algorithm," *J. Comput. Sci.*, vol. 53, no. April, p. 101368, 2021, doi: 10.1016/j.jocs.2021.101368.

[12] E. Indramaya and S. Suyanto, "Comparative Study of Recent Swarm Algorithms for Continuous Optimization," *Procedia Comput. Sci.*, vol. 179, no. 2019, pp. 685–695, 2021, doi: 10.1016/j.procs.2021.01.056.

[13] 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 Comput. Appl.*, vol. 32, no. 16, pp. 12201–12220, 2020.

[14] P. Shunmugapriya and S. Kanmani, "A hybrid algorithm using ant and bee colony optimization for feature selection and classification (AC-ABC Hybrid)," *Swarm Evol. Comput.*, vol. 36, pp. 27–36, 2017.

[15] X. Deng, Y. Li, J. Weng, and J. Zhang, "Feature selection for text classification: A review," *Multimed. Tools Appl.*, vol. 78, no. 3, pp. 3797–3816, 2019.

[16] B. Agarwal and N. Mittal, "Text classification using ML methods-a survey," in *Proceedings of the Second International Conference on Soft Computing for Problem Solving (SocProS 2012), December 28-30, 2012*, 2014, pp. 701–709.

[17] X. Zhou *et al.*, "A survey on text classification and its applications," *Web Intell.*, vol. 18, no. 3, pp. 205–216, 2020, doi: 10.3233/WEB-200442.

[18] M. Thangaraj and M. Sivakami, "Text classification techniques: A literature review," *Interdiscip. J. Information, Knowledge, Manag.*, vol. 13, p. 117, 2018.

[19] J. Žižka, F. Dařena, and A. Svoboda, *Text mining with ML: principles and techniques*. CRC Press, 2019.

[20] K. Menghour and L. Souici-Meslati, "Hybrid ACO-PSO based approaches for feature selection," *Int J Intell Eng Syst*, vol. 9, no. 3, pp. 65–79, 2016.

[21] M. Ghosh, R. Guha, R. Sarkar, and A. Abraham, "A wrapper-filter feature selection technique based on ant colony optimization," *Neural Comput. Appl.*, vol. 32, no. 12, pp. 7839–7857, 2020.

[22] S. Kannan *et al.*, "Preprocessing techniques for text mining," *Int. J. Comput. Sci. Commun. Networks*, vol. 5, no. 1, pp. 7–16, 2014.

[23] R. Asgarnezhad, S. A. Monadjemi, and M. Soltanaghaei, "An application of MOGW optimization for feature selection in text classification," *J. Supercomput.*, vol. 77, no. 6, pp. 5806–5839, 2021.

[24] E. Saraç and S. A. Özel, "An ant colony optimization based feature selection for web page classification," *Sci. World J.*, vol. 2014, 2014.

[25] M. H. Aghdam, N. Ghasem-Aghaee, and M. E. Basiri, "Text feature selection using ant colony optimization," *Expert Syst. Appl.*, vol. 36, no. 3, pp. 6843–6853, 2009.

[26] H. R. Kanan and K. Faez, "An improved feature selection method based on ant colony optimization (ACO) evaluated on face recognition system," *Appl. Math. Comput.*, vol. 205, no. 2, pp. 716–725, 2008.

[27] M. F. Zaiyadi and B. Baharudin, "A proposed hybrid approach for feature selection in text document categorization," *Int. J. Comput. Inf. Eng.*, vol. 4, no. 12, pp. 1799–1803, 2010.

[28] M. M. Kabir, M. Shahjahan, and K. Murase, "A new hybrid ant colony optimization algorithm for feature selection," *Expert Syst. Appl.*, vol. 39, no. 3, pp. 3747–3763, 2012, doi: 10.1016/j.eswa.2011.09.073.

[29] M. Kabir, M. Shahjahan, K. Murase, and H. J. C. Barbosa, "Ant colony optimization toward feature selection," *Ant Colony Optim. Appl. InTech*, pp. 1–43, 2013.

[30] Y. Gómez, R. Bello, A. Nowé, E. Casanovas, and J. Taminau, "Multi-colony ACO and Rough Set Theory to Distributed Feature Selection Problem," in *International Work-Conference on Artificial Neural Networks*, 2009, pp. 458–461.

[31] C.-L. Huang, "ACO-based hybrid classification system with feature subset selection and model parameters optimization," *Neurocomputing*, vol. 73, no. 1–3, pp. 438–448, 2009.

[32] N. Abd-Alsabour, “Binary ant colony optimization for subset problems,” in *Multi-objective Swarm Intelligence*, Springer, 2015, pp. 105–121.

[33] S. Kashef and H. Nezamabadi-pour, “An advanced ACO algorithm for feature subset selection,” *Neurocomputing*, vol. 147, pp. 271–279, 2015.

[34] P. Moradi and M. Rostami, “Integration of graph clustering with ant colony optimization for feature selection,” *Knowledge-Based Syst.*, vol. 84, pp. 144–161, 2015.

[35] S. Tabakhi and P. Moradi, “Relevance–redundancy feature selection based on ant colony optimization,” *Pattern Recognit.*, vol. 48, no. 9, pp. 2798–2811, 2015.

[36] S. Mirjalili, S. M. Mirjalili, and A. Lewis, “Grey wolf optimizer,” *Adv. Eng. Softw.*, vol. 69, pp. 46–61, 2014.

[37] H. Wang, Z. Hu, Z. Yang, and Y. Guo, “A Novel Grey Wolf Optimization Based Combined Feature Selection Method,” in *International Conference on Bio-Inspired Computing: Theories and Applications*, 2019, pp. 569–580.

[38] O. S. Qasim and Z. Y. Algamal, “A gray wolf algorithm for feature and parameter selection of support vector classification,” *Int. J. Comput. Sci. Math.*, vol. 13, no. 1, pp. 93–102, 2021.

[39] O. M. Alyasiri, Y.-N. Cheah, and A. K. Abasi, “Hybrid filter-wrapper text feature selection technique for text classification,” in *2021 International Conference on Communication & Information Technology (ICICT)*, 2021, pp. 80–86.

[40] N. Chopra, G. Kumar, and S. Mehta, “Hybrid GWO-PSO algorithm for solving convex economic load dispatch problem,” *Int J Res Adv Technol*, vol. 4, no. 6, pp. 37–41, 2016.

[41] M. A. M. Shaheen, H. M. Hasanien, and A. Alkuhayli, “A novel hybrid GWO-PSO optimization technique for optimal reactive power dispatch problem solution,” *Ain Shams Eng. J.*, vol. 12, no. 1, pp. 621–630, 2021.

[42] M. E. Basiri and S. Nemati, “A novel hybrid ACO-GA algorithm for text feature selection,” in *2009 IEEE Congress on Evolutionary Computation*, 2009, pp. 2561–2568.

[43] M. Sheikhan and N. Mohammadi, “Neural-based electricity load forecasting using hybrid of GA and ACO for feature selection,” *Neural Comput. Appl.*, vol. 21, no. 8, pp. 1961–1970, 2012.

[44] S. Vijayarani, M. J. Ilamathi, and M. Nithya, “Preprocessing techniques for text mining-an overview,” *Int. J. Comput. Sci. Commun. Networks*, vol. 5, no. 1, pp. 7–16, 2015.

[45] X. Zhou *et al.*, “A survey on text classification and its applications,” in *Web Intelligence*, 2020, vol. 18, no. 3, pp. 205–216.

[46] L. Wang and X. Fu, *Data mining with computational intelligence*. Springer Science & Business Media, 2006.

[47] M. E. Basiri, N. Ghasem-Aghaee, and M. H. Aghdam, “Using ant colony optimization-based selected features for predicting post-synaptic activity in proteins,” in *European Conference on Evolutionary Computation, ML and Data Mining in Bioinformatics*, 2008, pp. 12–23.

[48] T. Deng, C. Yang, Y. Zhang, and X. Wang, “An improved ant colony optimization applied to attributes reduction,” in *Fuzzy Information and Engineering*, Springer, 2009, pp. 1–6.

[49] M. Kabir, M. Shahjahan, and K. Murase, “An efficient feature selection using ant colony optimization algorithm,” in *International Conference on Neural Information Processing*, 2009, pp. 242–252.

[50] Y. Chen, D. Miao, and R. Wang, “A rough set approach to feature selection based on ant colony optimization,” *Pattern Recognit. Lett.*, vol. 31, no. 3, pp. 226–233, 2010.

[51] M. J. Meena, K. R. Chandran, A. Karthik, and A. V. Samuel, “An enhanced ACO algorithm to select features for text categorization and its parallelization,” *Expert Syst. Appl.*, vol. 39, no. 5, pp. 5861–5871, 2012.

[52] P. Shunmugapriya, S. Kanmani, S. Devipriya, J. Archana, and J. Pushpa, “Investigation on the effects of ACO parameters for feature selection and classification,” in *Proc. Springer—the Third Int. Conf. Advances in Communications, Networks and Computing, LNICST*, 2012, pp. 136–145.

[53] J. Too, A. R. Abdullah, N. Mohd Saad, N. Mohd Ali, and W. Tee, “A new competitive binary grey wolf optimizer to solve the feature selection problem in EMG signals classification,” *Computers*, vol. 7, no. 4, p. 58, 2018.

[54] Q. Al-Tashi, H. Rais, and S. Jadid, “Feature selection method based on grey wolf optimization for coronary artery disease classification,” in *International conference of reliable information and communication technology*, 2018, pp. 257–266.

[55] T. Niknam and B. Amiri, “An efficient hybrid approach based on PSO, ACO and k-means for cluster analysis,” *Appl. Soft Comput.*, vol. 10, no. 1, pp. 183–197, 2010.

[56] B. Shuang, J. Chen, and Z. Li, “Study on hybrid PS-ACO algorithm,” *Appl. Intell.*, vol. 34, no. 1, pp. 64–73, 2011.

[57] M. S. Kiran, E. Özceylan, M. Gündüz, and T. Paksoy, “A novel hybrid approach based on Particle Swarm Optimization and Ant Colony Algorithm to forecast energy demand of Turkey,” *Energy Convers. Manag.*, vol. 53, no. 1, pp. 75–83, 2012, doi: 10.1016/j.enconman.2011.08.004.

[58] J. B. Jona and N. Nagaveni, “Ant-cuckoo colony optimization for feature selection in digital mammogram.,” *Pak. J. Biol. Sci.*, vol. 17, no. 2, pp. 266–271, 2014.

[59] M. Mahi, Ö. K. Baykan, and H. Kodaz, “A new hybrid method based on particle swarm optimization, ant colony optimization and 3-opt algorithms for traveling salesman problem,” *Appl. Soft Comput.*, vol. 30, pp. 484–490, 2015.

[60] N. Singh and S. B. Singh, “Hybrid algorithm of particle swarm optimization and grey wolf optimizer for improving convergence performance,” *J. Appl. Math.*, vol. 2017, 2017.

[61] M. F. F. Ab Rashid, “A hybrid Ant-Wolf Algorithm to optimize assembly sequence planning problem,” *Assem. Autom.*, 2017.

[62] H. Xu, X. Liu, and J. Su, “An improved grey wolf optimizer algorithm integrated with cuckoo search,” in *2017 9th IEEE international conference on intelligent data acquisition and advanced computing systems: technology and applications (IDAACS)*, 2017, vol. 1, pp. 490–493.

[63] C. Vidyadhari, N. Sandhya, and P. Premchand, “Particle grey wolf optimizer (pgwo) algorithm and semantic word processing for automatic text clustering,” *Int. J. Uncertainty, Fuzziness Knowledge-Based Syst.*, vol. 27, no. 02, pp. 201–223, 2019.

[64] Q. Al-Tashi, S. J. A. Kadir, H. M. Rais, S. Mirjalili, and H. Alhussian, “Binary optimization using hybrid grey wolf optimization for feature selection,” *Ieee Access*, vol. 7, pp. 39496–39508, 2019.

[65] R. Purushothaman, S. P. Rajagopalan, and G. Dhandapani, “Hybridizing Gray Wolf Optimization (GWO) with Grasshopper Optimization Algorithm (GOA) for text feature selection and clustering,” *Appl. Soft Comput.*, vol. 96, p. 106651, 2020.

[66] J. Yousef, A. Youssef, and A. Keshk, “A Hybrid Swarm Intelligence Based Feature Selection Algorithm for High Dimensional Datasets,” *IJCI. Int. J. Comput. Inf.*, vol. 0, no. 0, pp. 0–0, 2021, doi: 10.21608/ijci.2021.62499.1040.