

Received March 19, 2022, accepted April 5, 2022, date of publication April 8, 2022, date of current version April 18, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3165814

Wrapper and Hybrid Feature Selection Methods Using Metaheuristic Algorithms for English Text Classification: A Systematic Review

OSAMAH MOHAMMED ALYASIRI^{1,2}, YU-N CHEAH¹, AMMAR KAMAL ABASI^{1,3},
AND OMAR MUSTAFA AL-JANABI¹

¹School of Computer Sciences, Universiti Sains Malaysia, Penang 11800, Malaysia

²Karbala Technical Institute, Al-Furat Al-Awsat Technical University, Karbala 56001, Iraq

³Machine Learning Department, Mohamed Bin Zayed University of Artificial Intelligence (MBZUAI), Abu Dhabi, United Arab Emirates

Corresponding author: Yu-N Cheah (yncheah@usm.my)

This work was supported by the Fundamental Research Grant Scheme (FRGS), Ministry of Higher Education Malaysia, under Grant 203/PKOMP/6711796.

ABSTRACT Feature selection (FS) constitutes a series of processes used to decide which relevant features/attributes to include and which irrelevant features to exclude for predictive modeling. It is a crucial task that aids machine learning classifiers in reducing error rates, computation time, overfitting, and improving classification accuracy. It has demonstrated its efficacy in myriads of domains, ranging from its use for text classification (TC), text mining, and image recognition. While there are many traditional FS methods, recent research efforts have been devoted to applying metaheuristic algorithms as FS techniques for the TC task. However, there are few literature reviews concerning TC. Therefore, a comprehensive overview was systematically studied by exploring available studies of different metaheuristic algorithms used for FS to improve TC. This paper will contribute to the body of existing knowledge by answering four research questions (RQs): 1) What are the different approaches of FS that apply metaheuristic algorithms to improve TC? 2) Does applying metaheuristic algorithms for TC lead to better accuracy than the typical FS methods? 3) How effective are the modified, hybridized metaheuristic algorithms for text FS problems?, and 4) What are the gaps in the current studies and their future directions? These RQs led to a study of recent works on metaheuristic-based FS methods, their contributions, and limitations. Hence, a final list of thirty-seven (37) related articles was extracted and investigated to align with our RQs to generate new knowledge in the domain of study. Most of the conducted papers focused on addressing the TC in tandem with metaheuristic algorithms based on the wrapper and hybrid FS approaches. Future research should focus on using a hybrid-based FS approach as it intuitively handles complex optimization problems and potentially provide new research opportunities in this rapidly developing field.

INDEX TERMS Feature selection, text classification, metaheuristic optimization algorithms, wrapper-based approach, hybrid-based approach.

I. INTRODUCTION

The huge quantity of digital data on the Internet, such as emails, social media sites, and libraries, are increasingly generated every second. The generated data could be in the form of numbers, text, audio, video, graphs, and others. Knowledge of different sorts from several domains such as financial, medical, statistical, logical, amongst others, can be extracted from such data to gain insights and make predictions.

The associate editor coordinating the review of this manuscript and approving it for publication was Huaqing Li¹.

A significant part of the data accessible/available today is stored as text [1], [2]. Data in the form of text, otherwise called textual data, constitute a large portion of the generated data. For example, around a billion messages and tweets are posted on Facebook and Twitter pages monthly. Moreover, more than a million articles were edited on Wikipedia in 2020.

The mining of textual data employs several techniques (from the field of statistics to artificial intelligence) to categorize texts (e.g., news filtering, topic identification, and document routing) [3], [4] from the domain that characterizes the problem that is to be solved. Classifying textual data

is also known as topic classification, text categorization, or text classification (TC). In this regard, many businesses, institutions, and people have text data that are highly unstructured. This makes it extremely difficult for them to analyze, understand, and reduce data complexity to easy and quick access to information and sort it on a huge scale. To solve this problem, businesses and others leverage TC with many techniques and methods that have been proposed for classifying text documents automatically, such as machine learning approaches and others [5], because of its scalability and real-time analysis of unstructured text. This, in turn, saves time, automates business processes, and usually helps make informed business decisions.

Therefore, there is a need to focus on the problem of organizing and managing the phenomenal growth of unstructured textual data. TC is one of the supervised learning methods; it is the primary goal to systematically organize the given documents into their relevant categories/classes (e.g., topics) [5], [6]. For example, a document with the words “viruses”, “pandemic”, “lockdown”, “death”, “vaccination” is assigned to the “Covid-19” class label. The common approach for TC consists of various essential steps:

- **Preprocessing the textual data** using several techniques such as tokenization, stopword removal/filtering, stemming, and cleansing, amongst others [7].
- **Feature Extraction/ document representation** to represent text documents such as bag of words (BOW) model, vector space model (VSM), and term frequency-inverse document frequency (TF-IDF) [7], [8].
- **Dimensionality reductions** such as Feature selection (FS) to reduce the high-dimensional of features space [5].
- **Data Mining/Pattern Discovery** involves building, developing, and training the models using machine or deep learning algorithms [5].
- **Evaluation/Interpretation** involves analyzing the outcome/results of the model, such as (accuracy (A), precision (P), recall (R), time complexity (T), and F-measure (F)) [9].

Among the above essential steps, feature extraction is the main focus of this work, which is conducted using different text representation formats such as the bag of words (BOW) model (set of words/terms or features), which is represented as the feature-vector with its associated frequency. However, the problem of having a huge feature-vector size that may comprise tens or hundreds of thousands of features is that it highly affects the performance of TC accuracy caused by the high-dimensional data [10], [11]. To resolve this problem, the Feature selection (FS) methods (in the dimensional reduction step) are advised to select the significant/valuable features. This, in turn, performs decent text representation and minimize computational time by reducing the overfitting and the error rate of the classifier to achieve a precise classification [12].

Three primary FS methods are used for TC: the filter-based approach, the wrapper-based approach, and the embedded-based approach [13]–[15]. The filter-based approach performs a statistical analysis over the feature space by ranking

each feature of the dataset based on some univariate [e.g., Information Gain (IG)] or multivariate (e.g., Correlation-based Feature Selection (CFS)) methods, then selecting top-N features having the highest-ranking features [16], [17]. Yet, the filter-based methods are restricted to some limitations [18], [19]. However, the wrapper-based approach evaluates the usefulness of features based on the used classifier performance. It is computationally more expensive when compared to the filter approach due to the repeated learning steps and cross-validation [16]. Although it is computationally expensive, it creates an interaction between the search feature subset and the classification algorithm, leading to better feature subset selection. However, using an optimization algorithm as FS in a wrapper-based approach could lead to better results than the filter-based approach. Because the optimization algorithm uses an objective function to evaluate the consistent features, taking into account the classification accuracy and the number of selected features. The examples of the wrapper-based approach are recursive feature elimination, sequential (forward /backward) feature selection (SFS) algorithms, and genetic algorithm (GA) [20], [21]. Besides, apart from the wrapper and embedded-based approaches that used the classification method in their mechanism to measure the performance of TC, the wrapper-based uses feature dependencies. In contrast, the embedded-based approach uses less computational cost features. Furthermore, the embedded-based approach incorporates the FS method into the classifier’s training process (learning process) using no search algorithm such as a metaheuristic algorithm. The decision tree such as Classification and Regression Trees (CART) has a built-in mechanism (embedded-based) to implement variable selection [13], [15], [22], [23]. Nevertheless, recently the researchers proposed a two-stage approach combining a less expensive filter-based approach to rank the features and an expensive wrapper-based approach to eliminate further irrelevant features called the hybrid-based FS approach. It differs from the other approaches by taking advantage of filter and wrapper. At first, it applies the filter method to select a feature subset with the highest-ranking score, followed by the wrapper method to further optimize the selected subsets using the optimization algorithm [24], [25].

Most of the traditional approaches, such as the filter-based approach, have several flaws, such as their failure to provide adequate performance in FS because they neglect feature interactions. Some features, for example, may be redundant and irrelevant on their own but extremely successful when combined with others, and the top-scoring features may be redundant. Most filter methods, in fact, analyze features independently and apply their methods directly on original datasets with high dimensions [26], [27]. The challenges mentioned above have mandated researchers to seek other methods of getting better performance and results during the FS process. The research pursuit had led to the introduction of metaheuristic-based FS methods for TC.

Metaheuristic refers to general ideas, techniques, or approaches that are applicable to many problems [28].

Metaheuristics are estimate algorithms in which each of the algorithms has a different historical background (for instance: Evolutionary-based techniques, Swarm-based techniques, Physics-based techniques, Human-based techniques, etc.). It can also be seen as a set of algorithmic concepts utilized for defining heuristic techniques that can be applied to diverse optimization problems with little modifications to adapt them to specific problems [29].

Since the FS problems are known as non-deterministic polynomial-time hard (NP-hard) problems [30], metaheuristic algorithms have been successfully used in solving such problems, as it involves finding the approximate optimal solution by relying on the core factors; an obtained information and experience throughout the search process represented as memory to guide the iterative generation process for more prominent solution spaces, and that however improves the results of the TC problem [31].

Several review studies have been published in the literature that provides useful information on FS in TC [5], [6], [14], [32]. However, none of these review studies and the ones that are accessible focus on studying the FS utilizing metaheuristics techniques in TC. All of the available ones are either cursory analyses or offer a small collection of work in the field. Notwithstanding, this review considered having research questions formulated to direct the researchers to particular research gaps, methods, and challenges that have not been addressed in the current field of study.

This systematic literature review (SLR) provides a complete analysis and synthesis of 6 years of research on the FS utilizing optimization methods that contribute to developing a solid foundation for future studies. Hence, the contribution of this SLR is to:

- Examine the exploited FS approaches (wrapper- and hybrid-based) that are based on metaheuristics to improve the English TC.
- Assess the effect of the metaheuristic algorithms as an FS method on the TC accuracy compared to the typical FS methods.
- Assess the effect of the modification or hybridization of the metaheuristic algorithms on the text FS problem.
- Identify the gaps in the current studies and their future directions.
- Serve as a hands-on guide for discovering the appropriate modelling technique for English TC problems.

Consequently, this study is meant to answer the research questions (RQs) in the following section and act as a referential guide. Therefore, the structure of the paper is delineated as follows: The introduction in Section I is followed by the review methodology in Section II that provides the details on how the papers in the study were selected. Then, the findings from the reporting stage are discussed in Section III to identify the existing literature based on the metaheuristic-based methods that answered the research questions. Section IV presents the conclusion of the study and its limitations.

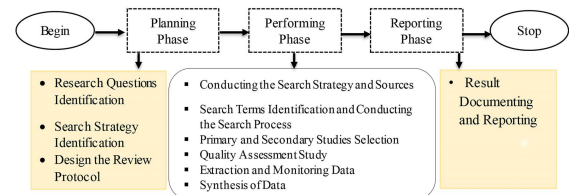


FIGURE 1. Phases of the systematic literature review.

II. REVIEW METHODOLOGY

A systematic literature review (SLR) was carried out to identify the current literature relevant to the Feature selection (FS) based on metaheuristic optimization methods for the text classification (TC). The review approach was performed based on the procedures for carrying out an SLR in computer science and software engineering research by [33]-[35]. The SLR involves three major phases: planning, performing, and reporting, as presented in Figure 1.

A. PHASE 1: PLANNING

The activities related to the planning phase were identifying the necessity for the review, stating the RQs, the search strategy identification, and the review protocol design, which could be utilized to obtain the RQs and the review methodology that would be used to conduct the review [34]. This SLR answered the four research questions:

RQ1: What are the different FS approaches applied to metaheuristic algorithms to improve TC?

RQ2: Does applying metaheuristic algorithms for TC lead to better accuracy than typical FS methods?

RQ3: How effective are the modified, hybridized metaheuristic algorithms for text feature selection problems?

RQ4: What are the gaps in the current studies and their future directions?

The search strategy is the first mapping work that can help determine the right research step. A search strategy begins by identifying important key terms and their alternatives and synonyms. Consequently, choosing the key terms that are closely related to the work will give good results to retrieve the relevant research papers [36]. Therefore, SLR is done based on predefined search strategies that focus on identifying the studies that are relevant to the SLR research questions. The strategy aims to identify the primary studies such as resources and search key terms to be used in the SLR, as shown in phase 2. The most common search strategy approach is to break down the research question into individual terms. Then a list of abbreviations, synonyms, and alternative spellings is drawn up. More terms can be acquired by considering subject headings utilized in journals and databases. The performing phase highlights the search keywords and study sources used in this SLR.

B. PHASE 2: PERFORMING

The performing phase comprises of conducting the search strategy and sources, primary studies selection, quality assessment study, extraction, monitoring, and synthesis of

data. In this review, the SLR method was applied to evaluate all the available research associated with the predefined research questions in the planning phase.

1) SOURCE

Seven (7) database sources were utilized as the primary sources to identify the available studies which apply to the highlighted RQ. The sources are as follows: IEEE, Scopus, Research Gate, Springer, Google Scholar, Science Direct, and Taylor & Francis. These databases enable the discovery of published materials in the form of journals, bulletins, conference proceedings, book chapters, symposiums, gray literature, and workshops. Our justification for choosing the seven sources stemmed from the originality and reputability of the sources. Furthermore, we established that other databases referred to these databases as the main sources for the existing studies on FS.

2) SEARCH TERMS IDENTIFICATION AND CONDUCTING THE SEARCH PROCESS

The selection of the relevant studies is represented by the inclusion and exclusion criteria in FS for TC using the metaheuristic algorithm and their related concepts. Several search terms were derived from the pre-identified RQ: “feature selection”, “text classification”, “metaheuristic”, “optimization”, “swarm intelligence”, “evolutionary algorithms”, “filter”, “wrapper”, “embedded”, and “hybrid” were used. For the advanced search, this study adopted a general approach of breaking down the RQ into individual terms followed by executing advanced search strings using Booleans that is “ORs” and “ANDs” [33], [34], as follows: (“Dimension reduction”) AND (“feature selection” OR “text feature selection”) AND (“text classification” OR “text categorization” OR “document categorization” OR “document classification” OR “topic classification”) AND (“metaheuristic” OR “meta-heuristic” OR “optimization” OR “swarm intelligence” OR “evolutionary algorithms”) AND (“Machine learning algorithms”).

3) INCLUSION AND EXCLUSION CRITERIA

To ensure that the selected studies are relevant and within the scope of the study objective, inclusion and exclusion criteria are a must for an SLR [33],[34]. The goal of this SLR is to highlight the current FS methods that relied on the metaheuristic-based methods (i.e., adaptation, modification, and hybridization) for TC. The inclusion and exclusion criteria applied to choose the relevant studies are as shown in Table 1.

4) QUALITY ASSESSMENT

Concerning the issue of quality assessment, the questions to assess the scope were used to validate the criteria. The “quality assessment” can be viewed as a critical step to assess the quality of the selected literature. Quality assessment includes questions aimed at assessing the scope that the reviewed articles have addressed bias and internal and

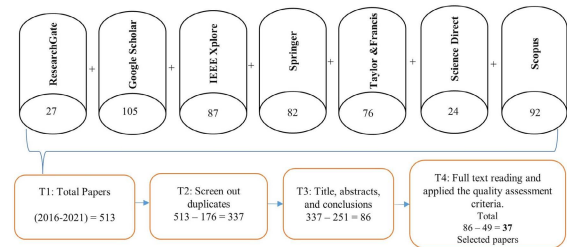


FIGURE 2. The used databases for the process of selecting studies and their results.

external validity [35]. The five (5) questions of quality assessment were answered as there are only three options: Yes = 1; Partially = 0.5; and No = 0, as presented in Table 2.

C. PHASE 3: REPORTING

In the reporting phase, the findings were stated in the following results section.

III. RESULTS

A. FINDINGS OF THE SELECTED STUDIES

The findings of this paper answer our specific research questions that guided this SLR. The selection of the studies was made by searching on a study source, then screening and filtering were performed in four iterations (T). In the first iteration, it extracted 513 relevant papers from the digital search as possible sources ($n = 513$). In the second iteration, duplicate articles were excluded using the Mendeley software ($n = 337$). In the third iteration, the studies were scanned and filtered by title, abstracts, and conclusions and excluded the articles unrelated to our domain’s scope ($n = 86$). In the last fourth iteration, the articles were scanned by reading the full text and applying the exclusion criteria, filtering the quality assessment stage results for all papers; only 37 articles were accepted and identified as final sources for the data synthesis ($n = 37$). The search process, results, and the process of paper selection are shown in Figure 2.

The number of selected papers published per year has increased considerably. Generally, the average quality score appears to increase from 2016 to 2018. Whereas its growth in the years 2019 to 2020 indicates that more researchers developed an interest in the study but decreased again in 2021, indicating that the area of TC using metaheuristics algorithm as an FS needs deep research. This paper strives to help continue the usage of metaheuristic algorithms in the future search, as it highlights the current research gaps and future directions in the findings section. The distribution of publication changes in the years (2016-2021) is presented in Figure 3.

Thirty-seven (37) articles were selected as high-quality that could be used to answer the research questions in this SLR. As shown in Table 3 below, twenty-eight (28) articles were rated (76%) as very good quality and nine (9) articles (24%) were rated as good. Other poor-quality articles were not considered in the results since they might not be making any impact.

TABLE 1. Inclusion and exclusion criteria.

Inclusion criteria	Exclusion criteria
<ul style="list-style-type: none"> · Published between the years 2016 and 2021 · Articles are written in the English language · FS methods solely used metaheuristic algorithms for TC · Benchmark datasets that are in English language only · Accessible in full form · Papers indexed in SCOPUS and ISI only 	<ul style="list-style-type: none"> · Not related to the above-stated research questions in this SLR · Articles that are conceptual (theoretical) have no experiments and results · Studies that used traditional FS methods (filter-based method) without using metaheuristic methods (wrapper-based method) for TC · Repetitive studies

TABLE 2. Quality assessment criteria.

No.	Item	Answer
Q1	Is there a vivid description of the objectives of the study?	Yes/Partially/No
Q2	Does the existing work adequately describe the research method?	Yes/Partially/No
Q3	Does the existing work describe the characteristics of datasets?	Yes/Partially/No
Q4	Do the relevant papers detail and clarify the model architecture?	Yes/Partially/No
Q5	Did the study rely on standard measures to verify the results?	Yes/Partially/No

TABLE 3. Quality assessment score.

Quality Scale	Very Poor (<1)	Poor (1 - <2)	Good (2 - <3)	Very Good (3-4)	Total
Number of papers	0	0	9	28	37
Percentage (%)	0	0	24	76	100

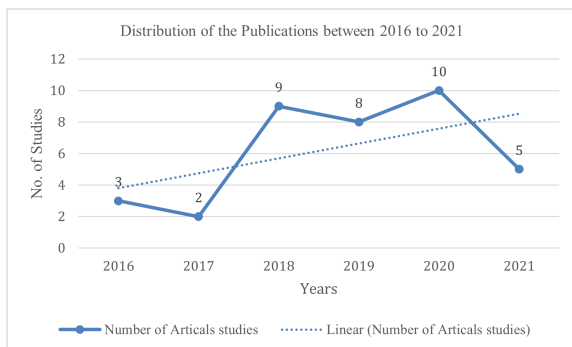


FIGURE 3. Distribution of the publications (2016-2021).

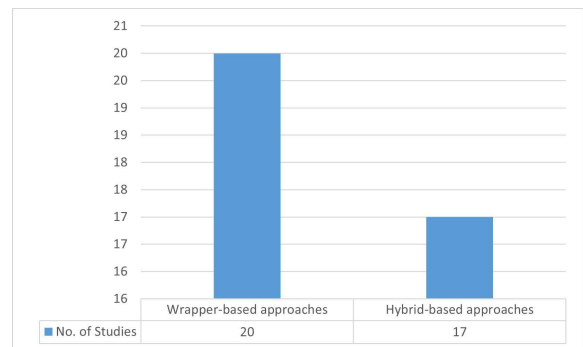


FIGURE 4. The statistical distribution of the selected studies in two FS paradigms.

Based on the quality assessment, questions (from Q1 to Q5) are predefined in the review methodology. The summary of the quality assessment for the 37 papers (P1 to P37) selected for review in this SLR is presented in Table 4.

In Table 4, “P” represents “the reviewed paper.” The review shows that FS is of considerable importance in data analysis, pattern classification, data mining, and machine learning applications. Thus, in many pattern classification problems, a good FS technique can reduce the cost of feature measurement, which automatically increases the classifier efficiency and classification accuracy. Hence, FS can be a preprocessing tool of great significance before solving the categorization problems.

B. DISCUSSION OF FINDINGS BASED ON THE RESEARCH QUESTIONS

This subsection discussed the answers to the stated Research Questions:

RQ1: What are the different FS approaches applied to metaheuristic algorithms to improve TC?

Answer 1: Various investigations were conducted to identify the appropriate FS approaches that used the metaheuristic optimization methods. Generally, the FS approach follows one of these paradigms: Filter, Wrapper, Embedded,

and Hybrid -approach. To meet the search terms of this SLR, two paradigms were found in the conducted studies: The Wrapper approach and the Hybrid approach (aka filter-wrapper) as statistically distributed in Figure 4. Each of the FS approaches (i.e., Wrapper and Hybrid approaches) either adapted, modified or hybridized metaheuristic algorithm(s) in their implementation as described below:

- **Adapt:** Is to adapt a new metaheuristic algorithm by enabling an intelligent behavior to solve sophisticated high-dimensional text data. The reason to adapt the metaheuristic algorithms is to tackle the text FS problem and reduce the high-dimensional datasets that are not plausible to be solved using traditional FS methods to help the classifier obtain better results.

- **Modify/improve:** Is to mitigate some of the core issues on the metaheuristic algorithms that suffer from multiplicity, stacking in local optimal space, and other issues. For example, improved performance in obtaining faster convergence and robust global search efficiency could be achieved by balancing exploitation and exploration in the algorithms search space, which in return improves the classification performance [1], [48], [50], [66].

- **Hybridize:** Is to combine the best operators from two metaheuristic algorithms to create a new, improved one.

TABLE 4. Quality assessment results.

ID/Q	P1 [37]	P2 [20]	P3 [2]	P4 [1]	P5 [38]	P6 [39]	P7 [40]	P8 [41]	P9 [42]	P10 [43]	P11 [44]	P12 [45]	P13 [46]
Q1	1	1	1	1	0.5	1	1	1	1	0.5	0.5	1	1
Q2	0.5	5	1	1	0.5	1	0.5	1	0.5	0.5	0.5	1	1
Q3	0	1	0.5	0	0.5	1	0.5	1	0.5	0	0.5	1	1
Q4	0.5	0.5	1	1	1	1	0.5	1	0.5	1	0.5	1	1
Q5	0.5	1	1	1	0.5	1	0.5	1	0.5	1	1	1	0.5
Total	2.5	4	4.5	4	3	5	3	5	3	3	3	5	4.5
ID/Q	P14 [47]	P15 [48]	P16 [49]	P17 [50]	P18 [10]	P19 [51]	P20 [52]	P21 [53]	P22 [27]	P23 [54]	P24 [11]	P25 [55]	P26 [56]
Q1	1	1	0.5	0.5	1	1	0.5	0.5	1	1	1	0.5	0.5
Q2	1	1	1	1	1	1	0.5	1	1	1	1	1	1
Q3	1	1	1	1	0.5	1	1	0.5	1	1	1	0	1
Q4	1	1	1	1	1	1	1	0.5	1	1	1	0.5	0.5
Q5	1	0.5	1	1	0.5	1	0.5	0.5	0.5	0.5	1	0.5	0.5
Total	5	4.5	4.5	4.5	4	5	3.5	3	4.5	4.5	5	2.5	3.5
ID/Q	P27 [57]	P28 [58]	P29 [59]	P30 [13]	P31 [60]	P32 [61]	P33 [62]	P34 [63]	P35 [64]	P36 [65]	P37 [66]		
Q1	0.5	1	1	1	0.5	0.5	0.5	1	0.5	1	1		
Q2	1	1	1	1	1	0.5	1	1	1	1	1		
Q3	0	0.5	0.5	1	1	1	1	1	1	1	1		
Q4	0.5	0.5	1	1	0.5	1	1	1	1	1	1		
Q5	0.5	1	1	1	1	1	1	1	0.5	1	1		
Total	2.5	3.5	4.5	5	4	4	4.5	5	4	5	5		

It is for creating a better algorithm that optimizes the solution (feature subsets) to enhance the quality of the initial candidate solutions using the local search strategy. Thus, an improved algorithm aids in avoiding local optima trapping, avoiding premature convergence, efficiently and effectively exploring the search space, and making excellent decisions [10], [51], [52].

It is important to note that: there is a difference between the used terminologies, hybrid and hybridize. The hybrid-based approach is the FS method, while hybridize refers to the hybridization between two metaheuristic algorithms.

The approaches and their techniques mentioned above are used to find the optimal feature subsets and reduce the dimensionality of the text data by selecting the smallest subset of features to improve the performance of the classifier, which is to yield a high classification accuracy. Consequently, the error rate will be decreased. The following subsections provide a thorough explanation of the approaches and their techniques.

1) WRAPPER-BASED FS APPROACHES

Twenty (20) papers out of thirty-seven (37) papers propose metaheuristic optimization algorithms or a combination of algorithms (i.e., adapted, modified, and hybridized algorithms) used for FS, which is known as the wrapper method [1], [2], [11], [20], [37]–[52]. It typically uses an evaluation function to select the feature subset in tandem with the classifier algorithm process. Wrapper-based FS categories as statistically shown in Figure 5 according to the conducted number of studies.

a: ADAPTED METAHEURISTIC METHODS

A suitable metaheuristic method was adapted in fifteen studies (15) to accomplish the FS problem in TC, relying on its mechanism of finding a new subset of the relevant features to tackle the massive number of features in the original text data. Bidi, and Elberichi [37], and Kumbhar et al. [20] used Genetic Algorithm (GA) as an FS method to improve the

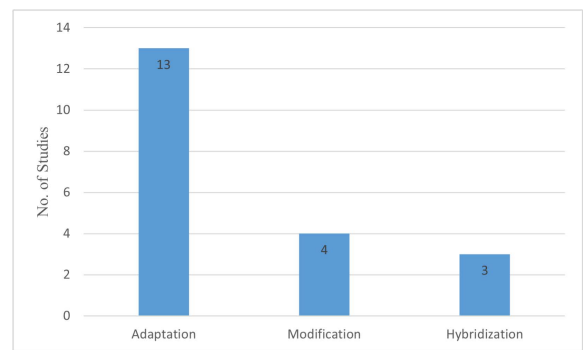


FIGURE 5. Wrapper-based FS categories using of metaheuristic techniques.

text classification. In [37], a text representation method (i.e., bag of words (BOW), N-gram, stemming, and conceptual representation) are used prior to the selection of the optimal number of the features subsets, and, in contrary, in [20], they did not use any text representation methods. Using the GA method, both researchers used a classification method based on the optimally selected number of feature subsets. In [37], Naive Bayes (NB), K-Nearest Neighbors (KNN), and Support Vector Machines (SVMs) were replicated as classifiers to evaluate and classify the candidate subset of features. Besides, in [20], the text classification performance had increased using the optimally selected features in tandem with the Fuzzy classifier as a classification method.

Other studies [2], [1], and [38] investigated the use of metaheuristic algorithms on subsets of text data from three benchmark datasets. In [2], Invasive Weed Optimization (IWO) presented as an FS method to be compared with Particle Swarm Optimization (PSO) and GA for optimally selecting the significant number of feature subsets. The selected subset of features using different metaheuristic methods was evaluated by the NB classifier to test the accuracy of TC using different configurations. In [1], different Term Frequency (TF) methods (i.e., TF, NORMTF, LOGTF, ITF, SPARCK) were used to weight all significant feature in all vectors. For all the

weighted features, the Flower Pollination Algorithm (FPA) method was employed to further select the optimal subsets of features. It was then evaluated and classified using the Ada-boost algorithm. While, in [38], the Crow Search Algorithm (CSA) was advised as an FS method, and the KNN classifier was used to evaluate the selected subsets of features to perform TC. Likewise, Artificial Bee Colony (ABC) had been adapted as an FS method in [39] and [40], and several classification methods (e.g., SVM) were tested for the evaluation and classification.

Contrary to the method presented in [39], where the number of selected classes was eight classes, in [41] and [42], the TC techniques presented were validated using Ten classes from the Reuters dataset. Additionally, in [41] and [42], the Ant Colony Optimization (ACO) and Firefly Algorithm (FFA) were the FS methods, respectively. In [41], the ACO and Artificial Neural Network (ANN) (ACO-ANN) had the capability to congregate promptly since it has effective searchability in the search space problem. Thus, it allows the efficient determination of the minimal feature subset. At the same time, the ANN was used to create a practical model that can be used from a given set of new inputs to predict the optimal output set or classifies the best subset of features from all subset features and predicts the solution. However, in [42], FFA and KNN were implemented using the same number of classes (Ten classes). While in [43], PSO was used in conjunction with a feature weighting and a parameter-tuned KNN classifier (different K-value). Others replicated the standard Binary Gray Wolf Optimizer (BGWO) and KNN classifier on newly extracted data for TC, and the model showed better accuracy using BGWO and a selected subset of features [44].

Ensemble approaches, on the other hand, merged many machine learning algorithms into a single predictive model to reduce variance (bagging), bias (boosting), or enhance prediction accuracy (stacking). As a result, by integrating the output of different weak learning classifiers, the accuracy can be improved. Therefore, Khurana *et al.* [45] propose a novel approach (BBO-bagging). They used a combination of optimization algorithms (Biogeography-Based Optimization (BBO), GA, and PSO) as an FS technique with the ensemble classifier to get optimal performance of TC. They used ten text datasets and one real-time dataset to train and test the retrieved features on six classifiers: NB, KNN, and SVM, Random Forest (RF), Decision Tree (DT), and ensemble (Bagging).

Nonetheless, a common way to calculate two or more contradictory objective functions simultaneously, researchers suggested multi-objective optimization methods (example of the multi-objective optimization methods are the conflicting functions of high-quality feature subset and reasonable running time). In [46], a multi-objective optimization method named Multi-Objective Relative Discriminative Criterion (MORDC), first objective employed RDC to computes the relevancy of the features to the target class, wherein the other objective computes the redundancy of a feature with

those with other selected features in the solution by applying the GA. The DT, Multinomial Naive Bayes (MNB), and Multilayer Perceptron (MLP) classification methods were used to assess the performance of the proposed method.

b: MODIFIED/IMPROVED METAHEURISTIC METHODS

Advanced and persistent techniques to modify/improve the working mechanism of the metaheuristic methods are being proposed to address the FS problem from different perspectives such as convergence, optimal solution, and algorithm efficiency. Janani and Vijayarani [47] proposed a greedy search algorithm to modify the global optimal in ABC to filter the irrelevant features and select the optimal feature subsets called the Optimization Technique for Feature Selection (OTFS). For the selected features, a machine learning method was developed based on the Probabilistic Neural Network (PNN). The PNN was further modified to improve the model's hidden layer (pattern layer) by advising an orthogonal matrix that chose the demonstrative features from the training documents. The developed machine learning method is named an automatic text classification (MLearn-ATC) model for classifying the different benchmark and real datasets based on the relevant selected features using OTFS. Thiyagarajan and Shanthi [48], proposed a crossover mechanism alongside the Artificial Fish Swarm Algorithm (AFSA) to alleviate the problem of multiplicity, which resulted in a new FS method called the Modified Artificial Fish Swarm Algorithm (MAFSA). Ada-boost, SVM, and the NB classifiers were used to evaluate the feature subsets selected using AFSA and perform the classification. In [49], the standard PSO algorithm has been adapted and modified by adding a weighting mechanism called Self-Inertia Weight Adaptive Particle Swarm Optimization (SIW-APSO) to enhance the performance of TC. SIW-APSO has a fast convergence phenomenon that yielded high search competency and a better selection of features. In addition, the KNN technique is used to classify text. M. Mahmoudi and F. S. Gharehchopogh [50], proposed a new mechanism to solve the FS problem of the Shuffled Frog-Leaping Algorithm (SFLA) stuck in the local optima. A combination of the best and worst search space solutions are combined to handle the stuck in the local optima. The DT classifier that used to classify and evaluate the selected subset of features.

c: HYBRIDIZE METAHEURISTIC METHODS (COMBINING BETWEEN TWO METAHEURISTIC ALGORITHMS)

Three studies proposed to hybridize metaheuristic algorithms to get the relevant and optimum feature subset from the original dataset for the FS problem. First, Maruthupandi and Devi [10] hybridized an ABC and Bacterial Foraging Optimization (BFO), known as ABC-BFO, for selecting the most significant feature subset for the prediction. To perform a multi-label classification, the ANN was used. Srilakshmi *et al.* [51], proposed a classification method based on three processing steps: VSM for feature extraction followed by the FS process that is performed using

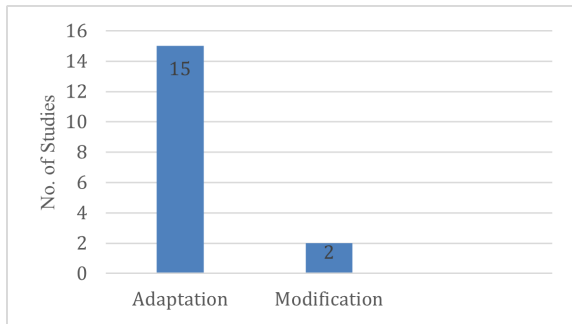


FIGURE 6. Hybrid-based FS categories using of metaheuristic techniques.

a hybridization method, and then a classifier is used for the classification. The proposed FS hybridizes the grasshopper optimization algorithm and the Crow Search Algorithm (GCOA). Thus, the optimally selected subset of features using GCOA is evaluated, and TC is achieved using a Deep Belief Network (DBN). In [52], the metaheuristic algorithms ACO and GA named ACOGA are hybridized to be used as an FS method based on a KNN classifier. A comparison of related studies that have employed the wrapper-based FS using the metaheuristics method for TC is given in Table 5.

2) HYBRID-BASED FS APPROACHES

The optimization algorithms for FS in the wrapper-based FS approaches suffer from high computational resources when identifying the optimal feature subset (due to their randomized mechanism). To resolve these issues, many researchers merged intelligent optimization algorithms with traditional FS methods as a hybrid method. Firstly, it employs a filter method to prune the high dimension in the data and create a subset of features using traditional FS methods (e.g., IG). Secondly, the wrapper method refines the selected subsets by using the metaheuristic methods (e.g., GA). Seventeen out of thirty-seven studies were conducted to propose a hybrid approach, and their distribution is as shown in Figure 6.

a: ADAPTED METAHEURISTIC METHODS

In [53], traditional FS methods IG and CHI were firstly presented for the pre-select feature subsets. Then, a small world optimization algorithm (SWA) was used to refine the selected features further and produce the most effective feature subsets by filtering out the unwanted features to limit the search space of SWA, and that relatively saved the consumed time for the problem-solving. KNN and SVM, however, were used for TC.

In [27], a two-stage FS method was presented involving traditional FS methods (Correlation (CO), Information Gain (IG), Gain Ratio (GR), and Symmetrical Uncertainty (SU)) and followed by a PSO algorithm as a filter and wrapper, respectively. The optimally selected subset of features using the stages was further evaluated, and text documents were classified using the NB classifier. Following the presented FS method in [27], T. Londt improved the PSO performance using the multi-objective function [54]. Likewise, Thirumoorthy and Muneeswaran [11] presented a new hybrid FS

method called the Normalized Difference Measure (NDM) as a filter-based method and a Binary Jaya Optimization (BJO) algorithm called NDM-BJO as a wrapper-based method to reduce the high dimensionality of feature space. At the same time, the NB and SVM classifiers were used to evaluate the nominated optimal feature subsets for the TC problem. Similar to [27], [11], a two-stage FS method were presented in [13], the first stage constituted a filter-based local FS method utilizing four different kinds of univariate methods (i.e., Chi-Square (CHI), Deviation From Poisson Distribution (DP), Discriminative Features Selection (DFSS), and Relative Discrimination Criterion (RDC)). Contrary to [27], [11], this method utilized three different feature set construction methods that are to employ globalization policies (i.e., Feature Set Construction Using Maximum Globalization Policy (MAX), Feature Set Construction By Using Weighted Averaging Globalization Policy (AVG), and Feature Set Construction By Including Equal Number of Features For Each Class (EQ)), also used two kinds of dimension reduction methods as feature transformation methods in the second stage (i.e., Principal Component Analysis (PCA) and Latent Semantic Indexing (LSI)). Then, GA is used as a wrapper-based FS method. Finally, the learning models were built using SVM to evaluate and classify each feature subset.

Kyaw and Limsiraratana, presented several models to solve the multi-dimensional FS problem on news document classification (BBC news) [55]–[59]. Their main goal was to improve news categorization performance while maintaining a fair level of complexity by reducing the number of selected features from a multi-dimensional text feature set. First, in [55], the filter approach PCA and Best First Search (BFS) were used to pick out selected text features throughout the feature reduction phase. Then, the Wolf Intelligence-based Optimization of Multi-dimensional Feature Selection system (WI-OMFS) was used to optimize the selected feature subset using PCA and BFS to obtain the optimal subset of features in the wrapper approach. Finally, the learning models were built using NB, SVM, and J48 to evaluate and classify each feature subset. In [56], however, a comparative study was conducted to compare the performance of using different filter-based methods (CFS, BFS, and IG) to perform the pre-selection of feature subsets and to be an input feature to the wrapper ACO and ABC based FS method(s). Then, the J48 decision tree was applied to evaluate each subset selected by the proposed method and classify the text documents. Similarly, in [57], a CFS was proposed as a filter-based approach to select a feature subset, and then an Evolutionary algorithm (EA) and GA were used to optimize the initially selected feature subsets by the CFS method. Consequently, the proposed methods were evaluated and tested using NB and SVM classifiers. In [58], the CFS and PCA as a filter-based method were implemented for the pre-select feature subset, and three nature-based inspired metaheuristic algorithms (Cuckoo Optimization (CO), Firefly optimization (FFA), and Bat optimization (BO)) were exploited as the wrapper FS approach to provide an optimal

TABLE 5. Comparison of related studies that have employed the wrapper-based FS using metaheuristics method for TC.

ID Reference	Dataset(s)	Text Representation (s) Models	Feature Selection Method (s)	Classification Algorithm (s)	Evaluation Metrics
Adapt new metaheuristic methods					
P1 [37]	Reuters-21578, 20 Newsgroup	BOW, N-gram, Stemming, Conceptual, and TF-IDF	GA	NB, KNN, and SVM	F-Measure
P2 [20]	Reuters-21578 (135 doc.) and 20 Newsgroup (114 doc.)	Term-Vector Model or VSM, Document-Term Matrix and TF-IDF	GA	Fuzzy Classifier	Feature Reduction Rate (FRR), Precision, Recall, and F-measure
P3 [2]	Reuters-21578, WebKb, and cade 12	Term frequency (TF) and VSM	IWO	NB	Accuracy, Precision, Recall, F-measure, AUC, and Error rate
P4 [1]	Reuters-21578, WebKb, and Cade 12	TF, NORMT, LOGTF, ITF and SPARCK	FPA	Ada-Boost	Accuracy, Precision, Recall, and F-measure
P5 [38]	Reuters-21578, WebKb, and Cade 12	Term frequency (TF) and VSM	CSA	KNN	Accuracy, Precision, Recall, and F-measure
P6 [39]	Reuters-21578, 20Newsgroup and Real datasets	BOW and TF-IDF	ABCFS	SVM, and improved SVM	Accuracy, Precision, Recall, and F-measure
P7 [40]	Reuters-21578 (top eight classes for R5 sub.)	VSM and TF-IDF	ABC	SVM, NB, and KNN	Accuracy, Precision, Recall, and F-measure
P8 [41]	Reuters-21578 (top 10 classes)	NA	ACO	ANN	Precision, Recall, Macro-F1, Micro-F1, and Accuracy
P9 [42]	Reuters-21578 (top 10 classes)	BOW	FFA	KNN	Precision, Recall, and F-measure
P10 [43]	re0, re1, fbis, oh5, and wap	Uncertainty operator weighted after selected subset of feature by PSO	PSO	KNN	Precision, Recall, F-measure, and Accuracy
P11 [44]	AstArc	BOW and TF-IDF	GWO	KNN	F-measure, Accuracy, and AUC
P12 [45]	10 text datasets from Reuters and OHSUMED. Real-time dataset: Airlines dataset of 539,384 records News and Scientific WebKB, 20 Newsgroups, and Reuters-21578 (eight categories)	BOW	BBO	NB, KNN, SVM, RF, DT and ensemble classifier	Precision, Recall, and F-measure
P13 [46]	WebKB, 20 Newsgroups, and Reuters-21578 (eight categories)	BOW, and TF-IDF	GA	DT, NB, and MLP	Recall, Precision, F-measure, Micro-F1, and Macro-F1
Modified/Improved metaheuristic methods					
P14 [47]	Reuters-21578, 20 Newsgroup, and real data set	Latent Semantic Analysis (LSA) and TF-IDF	OTFS	MLearn-ATC and PNN	Precision, Recall, F-measure, and Accuracy
P15 [48]	OHSUMED	BOW	MAFSA	SVM, NB, and Ada-boost	Precision
P16 [49]	Reuters-21578 (Top ten classes)	BOW, and TF-IDF	SIW-APSO	KNN	Recall, Precision, and F-measure
P17 [50]	Reuters-21578, WebKb, Cade 12, and 20 Newsgroup	VSM, and Cosine Distance Method	SFLA	DT	Precision, Recall, F-measure, AUC, Accuracy, and Error Rate
Hybridize/Combining two metaheuristic methods					
P18 [10]	Reuters Dataset (Top ten classes)	BOW, VSM and TF-IDF	HABBFO	ANN	Precision, Recall, and Hamming loss
P19 [51]	Reuter and 20 Newsgroups	VSM, TF-IDF, and Entropy features	GOA and CSA	DBN	Precision, Recall, and Accuracy
P20 [52]	Reuters-21578	BOW, and TF-IDF	ACOGA	KNN	Precision, and Recall

feature subset that to be fed into two classifiers (J48 and SVM). Further, the presented filter-based methods (CFS) in [58] were replicated in [59] with PCA using different search policies (ACO, ABC, EA, Flower Optimization algorithm (FO), Rhinoceros Optimization algorithm (RO), and Wolf Optimization Algorithm (WO)). The selected feature subsets using the wrapper-based methods were evaluated using the J48 classifier. Finally, researchers in [60] and [61] implemented the IG as a filter-based method to select the top-ranking features. To do so, two-hundred (200) feature subsets are selected by IG to be fed into the Gray Wolf Optimizer (GWO) to select the optimal subset of feature,

and top-ranking features to be fed into the Imperialist Competitive Algorithm (ICA) utilizing NB and KNN as classification methods, respectively. In [63], a hybrid method using global Point-wise Mutual Information (PMI) is proposed in two phases. First, the ranking-based filter approach is implemented for FS by applying the IG method. While the second phase comprises two other subphases, the global PMI- based FS method implemented to select a subset of features based on a class-dependent assumption for computing the correlation between pairs of features and then using Gravitational Search Algorithm (BGSA) is a wrapper method to find the best subset of features. Finally, the 1-NN

TABLE 6. Comparison of related studies that employed Hybrid-based FS using the metaheuristics method for TC.

ID Reference	Dataset(s)	Text Representation (s) Models	Feature Selection Method (s)	Classification Algorithm(s)	Evaluation Metrics
Adapt metaheuristic methods					
P21 [53]	Reuters 21578 Classic Corpus, (R8)	BOW and VSM	Filter: IG and CHI Wrapper: SWA	KNN and SVM	Accuracy
P22 [27]	Reuters-21578 (R8 dataset)	BOW, VSM, and TF-IDF	Filter: CO, IG, GR, and SU Wrapper: PSO	NB	Accuracy
P23 [54]	Reuters-21578 (R8 dataset)	BOW, VSM, and TF-IDF	Filter: CO, IG, GR, and SU Wrapper: Multi-objective PSO	NB	Accuracy
P24 [11]	WebKB, SMS, BBC News and 10 Newsgroup datasets	NA	Filter: NDM Wrapper: BJO	NB, and SVM	Accuracy
P25 [55]	BBC News	BOW, N-gram and TF-IDF	Filter: PCA Wrapper: WI-OMFS	NB, SVM, and J48	Precision, Recall, F-measure, and Accuracy
P26 [56]	BBC News	BOW, N-gram and TF-IDF	Filter: CFS and IG Wrapper: ACO and ABC	J48 decision tree	Accuracy, Error Rate, and Time complexity
P27 [57]	BBC News	BOW, N-gram and TF-IDF	Filter: CFS Wrapper: CSE, EA, and GA	NB, and SVM	Accuracy, Error Rate, and Time complexity
P28 [58]	BBC news	BOW and TF-IDF	Filter: CFS and RS Wrapper: CO, FO, and BO	J48 and SVM	Accuracy
P29 [59]	BBC News	BOW and TF-IDF	Filter: CFS and PCA Wrapper: ACO, ABC, EA, FO, RO, and WO	J48 decision tree	Accuracy and Time complexity
P30 [13]	Reuters (top-10 classes) and Ohsumed dataset	BOW and TF-IDF	Filter: - CHI2, DP, DFFS, and RDC.-PCA and LSI Wrapper: GA	SVM	Macro-F1, and Micro-F1
P31 [60]	Oh0, Oh5, Oh10, Oh15, Re0, La1s, La2s, FBIS, and Tr41		Filter: IG Wrapper: GWO	NB	Precision, Recall, and F-measure
P32 [61]	Reuters-21578	VSM, and TF-IDF	Filter: IG Wrapper: ICA	KNN	Precision, Recall, F-measure, and F1-micro
P33 [62]	National Public Radio (NPR) news, and 4-subject-data	BOW	Filter: DFTF Wrapper: PRO	NB and SVM	Accuracy, Precision, Recall, and F-measure
P34 [63]	TecTc100 and Reuters-21578	BOW, and VSM	Filter: - IG- Global PMI-based FS (GFM, MGFM, MGFWM, CDCG FWM) Wrapper: BGSA	1-NN	Accuracy and AUC
P35 [64]	Yahoo! dataset		Memetic Search: Filter: LFD Wrapper: EDA	NB	Accuracy and Hamming loss
Modified/Improved metaheuristic methods					
P36 [65]	(TREC2007, 20 Newsgroups (top 10 categories)), (Reuters21578, WebKB (top 4 categories))	BOW, VSM and EWS1, EWS2	Filter: CMFS Wrapper: IABCM	FSVM and NB	F-measure and Time complexity
P37 [66]	Oh0, Oh5, Oh10, Oh15, Re0, La1s, La2s, FBIS, and Tr41	BOW	Filter: IG Wrapper: ISCA	NB	Precision, Recall, F-measure, Macro-F1, and Micro-F1

classifier is used to assess the performance of the proposed method.

Notwithstanding, the majority of the hybridized FS approaches started the FS task with a filter-based method succeeded by a wrapper-based method. Whilst, in [64], a memetic FS method that hybridizes an evolutionary feature wrapper succeed by the filter for multi-label TC problem. Firstly, a promising region in the search space is located using an evolutionary estimation of distribution algorithm (EDA) as a wrapper, then an effective score function called Label Frequency Difference (LFD) is advised as a filter for feature selection. However, the same classifier method was used as in [60]. Similarly, in [62], a unique hybridization mechanism is proposed by means of three stages; filter-based method, wrapper-based method, and filter-based method. In the first

stage, Document-frequency Term-frequency (DFTF) is proposed as a filter-based method to select a significant subset of features. Secondly, a wrapper-based method named binary Poor and Rich optimization algorithm (HPRO) is used with DFTF in the first stage to select optimal feature subset using NB classifier. Lastly, the second filter-based method is used to select significant features using Term frequency reordering of document level (TRDL).

b: MODIFIED/IMPROVED METAHEURISTIC METHODS

Wang *et al.* [65], introduced a novel method for the TC task based on an Open Directory Project (ODP) to consider the semantic relation of features and their relevancy to the text document classes. Firstly, the redundant information is filtered out using the conceptualization of equivalence word

set (EWS1) that is developed based on the rich semantic data using ODP. Secondly, a Comprehensive Measurement Feature Selection (CMFS) is proposed to select the Optimal Feature Subset (OFS) and decrease the time taken for the execution, followed by the ABC-based FS (ABFS) to select the optimal features further. Lastly, the proposed model was verified using two classifiers: Fuzzy Support Vector Machine (FSVM) and NB.

Lastly, Belazzoug *et al.* [66], presented an improved Sine Cosine Algorithm called ISCA, which allows the exploration of more search space for FS. Then, a filter-based (IG) was used to get the top highest scores ranked features (informative feature), thus reducing the size of high dimensionality and decreasing the execution time. The informative selected features were used as an input in the wrapper algorithm (ISCA). To validate the efficiency of their work, the NB algorithm was applied for the classification task.

To conclude, the modified/improved metaheuristic methods alongside the FS approach are promisingly affecting the TC accuracy because it tackles the weakness of the traditional FS method via the robustness of metaheuristic algorithms which solves the serious complex optimization problem. Table 6 summarizes the conducted steps of the TC problem based on hybrid approaches and datasets used for evaluation purposes.

RQ2: Does applying metaheuristic algorithms for TC lead to better accuracy than the typical FS methods?

Answer 2: From the related studies, it has been observed that the objective function of metaheuristic methods in TC achieves three things: 1) maximizes the accuracy, 2) minimizes the error rate, and 3) improves time efficiency in the classifiers, thus producing high performance in terms of TC accuracy.

Using GA as an FS method [37], [20] had positively affected the text classification performance. In [37], it can be noticed that the increasing percentage of training data increased the accuracy. However, the investigated Conceptual Representation method has produced the highest accuracy amongst other representation methods with NB, KNN, and SVM. Meanwhile, the SVM classifier has the best performance classification with both FS methods (i.e., GA) and no FS method. The conceptual approach obtains comparable or better results than the others in many situations. However, the proposed method is not compared with the traditional FS method. Besides, in [20], when the number of generations (iterations) increased, the performance of TC had increased as well, and the results indicate that a GA-based FS with the Fuzzy classifier gives promising results compared to the traditional Principal Component Analysis (PCA). In [2], the use of IWO as an FS had reduced the high dimensionality of feature space. Therefore, it helped the classifier perform better accuracy and reduced the error rate compared to GA and PSO because IWO is omitting the less important features, increases the calculation speed, and yields the optimum answer in a shorter time. As the text data is inherently noisy, a weighting method (i.e., TF) is used to assign weight

values for the significant features and based on the weighted features, and an FPA algorithm is used to further reduce the number of features by optimally selecting feature subsets. Such a procedure has given a promising TC accuracy using the Ada-boost algorithm. Moreover, the classification using ITF and NORMTF model is the most accurate compared to other models (LOGTF, ITF, SPARCK) [1]. In [38], CSA as an FS method with KNN overcame the standard KNN for TC and produced a better performance than a previously presented IG method as a traditional FS, IG-GA, and IG-PCA using KNN and C4.5 classifiers. Moreover, the proposed mechanism contributed to identifying weights of features in neighboring documents as well, and the CSA also has a good memory in a way that all crows (crows: operators in CSA) preserve the significant features.

The adaption of ABC for FS has shown a different TC accuracy so long as the used dataset differs in the number of classes; in other words, in [39] and [40], the ABC method was applied to select the optimal number of features. In [40], the three widely used classifiers, namely, SVM, NB, and KNN, were compared against ABC. The experimental results show that SVM with ABC outperformed the other two classifiers. However, the proposed method has not been compared with the traditional FS methods like IG to validate the achieved results. On the other hand, in the study by [39], the SVM and the Improved SVM (ISVM) algorithm were used in the classification phase. Based on the performance measures compared with various FS techniques, namely, IG, CHI, and the original SVM algorithm with ABC, the proposed algorithm ABC with the improved SVM classifier offered better results. It reduced the high dimensionality in the dataset by selecting the important features.

Although the used number of classes (in the dataset) were the same in both [41] and [42], there is a significant difference in the TC accuracy because in [41], the ACO-ANN method was used while it was the FFA-KNN in [42]. Moreover, in [41], a comparison was held to compare the presented mechanism's performance with other metaheuristics methods (GA) and traditional FS methods that are IG and chi-square (CHI). In contrast, FFA-KNN outperformed CHI and IG methods [42].

In [43], the experimental results suggest that using a feature weighting strategy in conjunction with a parameter-tuned classifier increases the performance of the classification model. The weighted uncertainty operator in PSO and the tuned parameter in KNN had achieved better accuracy than the standard technique. In [49], the adaption of the weighting scheme (Self-Inertia Weight) into the PSO has overcome the problem of premature convergence in standard PSO. Furthermore, the experimental results showed that the proposed (Self-Inertia Weight) has better text classification accuracy than the typical (IG, CHI) and standard metaheuristic methods (GA, PSO).

Besides, Khurana and Verma [45] proved that the BBO-bagging as an FS method is superior to GA, PSO, and BBO with an ensemble classifier than individual ones

in previous studies. However, not all datasets with the proposed model produced the best performance using all selected classifiers. Some of the classifiers, such as KNN or NB, work better for the high-dimensional datasets, while the SVM performed better with low-dimensional datasets using BBO-bagging. Likewise, in [60], the IG with GWO is better than MFO, SCA, ACO, and GA assessed using most of the benchmark datasets because the behavior of GWO during the search proves that the top three best solutions help to explore and exploit the search space effectively by improving the average population fitness function throughout multiple iterations. Others proved that using IG combined with ICA showed better results than using IG and Mutual information (MI) solely [61]. F. Zarisfi Kermani *et al.* [63], introduced the global PMI-based FS method to improve the quality of the feature subset selected based on correlation criterion. The method considers having a mutual correlation between term and class using class-dependent assumption and using the class-independent assumption that measures the correlation between pair of terms. However, the performance of the proposed model suppressed four single objective FS methods.

For the multi-label TC approach that presented in [64], the proposed memetic search mechanism overcomes the performance of four typical filter FS methods (“maximum discrimination (MD), relevance popularity (RP), variable global feature selection (VGFSS), and normalized difference measure (NDM)”) and three wrapper-based methods (“asynchronously improved particle swarm optimization (AIPSO), enhanced genetic algorithm (EGA), and EGA with class discriminating measure (EGA+CDM)”). However, though the proposed memetic FS problem was promising this case study, it might not be efficient for other text data for two reasons. First, it is biased to solve such a problem; second, the evolutionary wrapper and feature filter were designed separately. Therefore, a new memetic FS method can be designed based on a compact filter and wrapper methods.

The newly proposed hybrid HPRO method in [62] has shown better accuracy compared to various filter-based methods (DFS, NDM, TRDL, and DFTF) and wrapper-based methods such as GA and FFY in terms of classification accuracy using NB classifier.

However, having a modified ABC using the SFS method has affected the performance of the FS process by ignoring the less significant features. The newly modified performance of the FS method is validated by comparing its performance against the widely used optimization methods (ABC, FFA, ACO, and PSO). It is important to mention that the TC performance also produced better accuracy while improving the standard PNN algorithm. Furthermore, the modified PNN outperformed NB, KNN, SVM classification methods [47].

In [48], the experimental results prove that the MAFSA as an FS method had overcome the traditional method AFSA concurrently with the most significant subset of features. Meanwhile, several classifiers were adapted (Ada-boost, SVM, and the NB) for TC.

A few approaches have been proposed to hybridize the metaheuristic algorithms for the FS problem. In [10], the hybridized ABC-BFO approach with ANN outperforms GA with KNN classifier on the top ten classes using the Reuters news dataset. However, ANN was used to train and evaluate the proposed ABC-AFO, while ANN was not used to compare the performance of GA. Similarly, Srilakshmi *et al.* [51] proposed a hybridized GOA and CSA with DBN classifier based on a hybrid weight bounding model. The performance of the newly developed model has been tested against several classification methods (NB, KNN, SVM and Deep Convolutional Neural Network (DCNN) and Stochastic Gradient-CAViaR+ DCNN). The GCOA-DBN has better classification accuracy and time efficiency.

Notwithstanding, multi-objective methods proposed in [46] have shown better results in comparison to the univariate filter algorithms (IG, GI, GR, FS, LS, and RDC), multivariate filter algorithms (mRMR, UFSACO, RRFS), and multi-objective algorithms (MOMI and MECY FS). The improved performance is attributed to the development of an evolutionary optimization technique that considers the relevancy and redundancy of features as goal functions.

On the other hand, a hybrid-based approach was proposed to integrate the filter method IG and CHI to the wrapper method SWA [53]. The approach produced better accuracy than traditional FS methods while using both KNN and SVM. In detail, the pre-selected features using IG and CHI that fed into the SWA based on a different number of features (FN) have promisingly overcome approaches that solely used IG and CHI on both KNN and SVM classifiers. Consequently, their approach concentrated on reducing the dimension of the feature vector, decreasing model complexity, and improving the performance of TC.

The two-stage filter and wrapper method in [27] is tested against traditional FS methods (CO, IG, GR, and SU) acquired higher accuracy. Note that the used subset of features in the proposed method (two-stage) is smaller than what it is in the tested traditional FS methods. However, in [11], a deep analysis has been conducted to compare the proposed NDB-BJO method with the original NDM method and the existing filter-based methods (IG, MI, Document Frequency (DF), Distinguishing Feature Selector (DFS)) alongside the GA and FFA, in which the presented method NDB-BJO outperformed all the stated methods using SVM and NB classifiers. Whilst, in [13], the experimental results showed that DFSS, CHI, or RDC feature selection techniques, rather than DP feature selection methods, are the settings that give the greatest results for two-stage feature selection methods. The AVG and EQ feature set construction approaches appear to be superior to the MAX feature set construction method in most cases. PCA feature transformation produced most of the best outcomes for two-stage feature selection approaches. GA method’s performance is often inferior to PCA and LSI approaches. However, in many cases where transformation-based approaches were successful, GA did not increase performance.

The presented models in [55]–[59] are being compared in terms of computational time and TC accuracy. In [55], the computation time of the presented WI-OMFS-Filter as an FS method fluctuated according to two parameters: number of iterations and population. The performance in terms of computational time peaked using WI-OMFS + J48, and the second-best using the NB, while SVM was the worst. At the same time, WI-OMFS-Filter + J48 achieved the best TC accuracy and the second-best in the SVM, while NB was the worst with regard to the incremented number of iterations and populations. In contrast, the conducted study in [56] is a comparative study to investigate the performance of ACO and ABC as wrapper-based methods according to the number of selected features and for both the iteration and population parameters. The classification accuracy was enhanced in the experiment for the ACO-based FS method by reducing the number of chosen features (NF) as the population size (PS) and iteration number (IN) grew. However, as the NS and IN have grown, so have the computing and hardware costs as shown in Table 7.

In [57], the experimental findings proved that both EA and GA with CFS filter were better than CSE wrapper by selecting the optimal feature subset, where the NB was more promising for TC than SVM. Besides, the EA overcame GA for both CFS and CSE-FS approaches. Similarly, in [58], amongst the three assessed metaheuristic algorithms, CO, BO, and FFA as a wrapper-based approach, and the classifiers SVM and J48, the CO with J48 have achieved the best accuracy on different PS values. Note that the best accuracy achieved using the proposed system (CO + J48) is better than the traditional method (SFS and RS) in terms of the number of the selected global feature subset. In [59], WO and J48 have the best accuracy of the other search policies (ACO, ABC, EA, Flower FO, and RO). Besides, the traditional FS methods GS and CFS optimally selected the least number of features than the search policies.

In [65], the modified ABC (ABCFS) algorithm that is incorporated with the semantic knowledge from ODP achieved better results compared to several typical FS techniques (IG, MI, CMFS, Multi-label FS based on max-dependency and min-redundancy (MDMR), Cumulate conditional mutual information minimization criterion FS (CCM), Global MI-based FS algorithm (GMFS)) using two classifiers (FSVM and NB) and different percentage of the selected number of features.

Finally, in [66], the improved ISCA is statistically outperformed, standard SCA, other SCA methods (i.e., OBL (SCA), Levy (SCA), and Weighted (SCA)), Moth-Flam optimization algorithm (MFO), GA and ACO. Besides, improved ISCA dramatically outperforms the traditional FS methods in terms of TC accuracy as well as the time complexity and the least number of the selected feature. To recap, despite improvements on the FS methods, there are certain weaknesses that can be addressed, as shown in Table 7.

RQ3: How effective are the modified, hybridized meta-heuristic algorithms for text feature selection problems?

Answer3: Janani and Vijayarani [47], mitigated the problem of global optima in the ABC method by incorporating the SFS method that effectively improves the selection of the optimal feature subsets by initializing an empty set. The SFS will add the optimal features sequentially to the initiated set by considering the global objective function. Additionally, they claimed that the proposed algorithm uses the least amount of time and memory to complete the task.

Thiyagarajan and Shanth [48], modified the original AFSA by adding a crossover operation into the Artificial Fish (AF's) search space, in which a blindness search for the food (features) by AF's is directed to generally find better direction and improve the time taken for convergence. The obtained results proved that considering the crossover in the AFSA algorithm will increase the performance of selecting an optimal feature subset and bring down unwanted ones.

M. Asif *et al.* [49], effectively improved the feature selection mechanism based on the proposed weighting scheme, i.e., SIW-APSO. Practically, a group of arbitrary particles is initialized to search for a search space. The initialized particles change their positions by communicating with each other, and better local and global positions are acquired.

M. Mahmoudi and F. S. Gharehchopogh [50], proposed a balancing equation to find the optimal search space solution based on approximating the best and worst solutions to prevent the SFLA from stuck in the local optimum. Maruthupandi and Devi [10] hybridized ABC with BFO to overcome the problem of random search in ABC. The ABC method is hybridized with the BFO algorithm since the initialization of feasible solutions is problematic in some circumstances. The scout bees' random exploration, which initializes the possible solution (feature subset), is difficult to come by in certain cases and its inconvenience solutions. The scout bee component of the ABC algorithm is replaced with the BFO algorithm. By doing so, the hybridization helped select the important features from the text documents that improved the TC accuracy.

Srilakshmi *et al.* [51], suggested GCOA by merging the GOA and CSA algorithms. The CSA is derived from the incentive gained from crows' clever behavior in seeking prey and locating it using memory. Furthermore, the method successfully balances the diversification and intensification stages, and the convergence rate is quite high with relatively little computing time. To address optimization problems, the GOA is based on the swarming behavior of grasshoppers. The GOA is capable of finding the optimum answer to optimization issues while balancing exploitation and exploration. Yet, it suffers from a low convergence rate. Therefore, CSA is integrated with GOA to resolve the shortcomings of the GOA algorithm. The developed GCOA was utilized to optimize the DBN's weights for better classification.

A. Singh and A. Kumar [52], hybridized different ants from ACO and the populace in the GA for the selection

TABLE 7. Comparison of the proposed models for the TC problem and their issues.

ID Ref.	Feature Selection Method (s)	Compared With Typical and Current Method (s)	Numbers of Population (NP) and Generation (s) (NI)	Performance of Proposed method (s)	Issues	
P1 [37]	GA	NA	50 NP and 500 NI	F-Measure: Reuters-21578: KNN with GA: 0.88, NB with GA: 0.86, SVM with GA: 0.94 20 Newsgroup: KNN with GA: 0.93, NB with GA: 0.87, SVM with GA: 0.946	- Not compared with typical FS and optimization methods - Time inefficient - No parameter tuning	
P2 [2]	GA	PCA	NA of NP, and 25, 50, 75, 100 NI	Reuters-21578: P:96.4, R: 96.1, F: 96.1	- Selective usage of classes - Not compared with optimization methods	
P3 [20]	IWO	NA	50 NP and 100 NI	Highest F-measures with (FS=160): Reuters-21578: P: 0.84, R: 0.86, F: 0.85, AUC: 0.89, A: 0.95, Error rate: 0.04 WebKb: P: 0.80, R: 0.89, F: 0.84 Proposed Model with (FS=80) and highest TF weighted: Reuters-21578: (used SPARCK) P: 97.11, F: 89.18, R: 83.02, A: 93.87	20 Newsgroup: AUC: 0.83, A: 0.96 Error rate: 0.03 Cade 12: P: 0.82, R: 0.85, F: 0.83, AUC: 0.80, A: 0.95, Error rate: 0.04	- Not compared with typical FS and optimization methods - Time inefficient - Low performance
P4 [1]	FPA	NA	NA	Proposed Model with (FS=80) and highest TF weighted: Reuters-21578: (used SPARCK) P: 97.11, F: 89.18, R: 83.02, A: 93.87	Cade 12: (used LOGTF) A: 91.58, P: 96.84, F: 89.18, R: 83.78 WebKb:(used SPARCK) P:96.40, F:89.84 R: 82.97, A: 97.45 WebKb: P: 96.34, R: 86.30, F:91.04 A: 96.00 Cade 12: P: 97.34, R: 85.47, F:91.02 A: 96.86	- Not compared with typical FS and optimization methods - No exact figure of iterations and populations
P5 [38]	CSA	IG, IG-GA, and IG-PCA	NA	Perf. Proposed Model (K=3 and FS= 120) Reuters-21578: P: 95.30, R: 82.02, F: 88.16, A: 97.24	WebKb: P: 96.34, R: 86.30, F:91.04 A: 96.00 Cade 12: P: 97.34, R: 85.47, F:91.02 A: 96.86	- No exact figure of iterations and populations - Optimal features selection - Not compared with optimization methods
P6 [39]	ABCFS	IG and CHI	NA	Average Perf. (ABCFS) with ISVM on: Reuters: P: 0.67, R: 0.70, F: 0.67, A: 0.82	20 Newsgroup: P: 0.70, R: 0.72, F: 0.71, A: 0.82 Real dataset: P: 0.84, R: 0.79, F:0.81, A: 0.83	- No exact figure of iterations and populations - Not compared with optimization methods
P7 [40]	ABC	NA	NP= 2 * No. of features and 100 NI	Accuracy: SVM: 95.07, NB: 92.23, KNN: 87.37 Average Precision: SVM: 84.75, NB: 83.04, KNN: 78.91 Average Recall: Averages of Precision:77.34 Recall: 80.14	SVM: 83.74, NB: 81.96, KNN: 77.25 Average F-measure SVM: 96.08, NB: 82.48 KNN: 78.04 Macro-F1: 79.01 Micro-F1: 89.87 Accuracy: 81.35	- Not compared with typical FS and optimization methods - Time inefficient - Low Performance - No parameter tuning
P8 [41]	ACO-ANN	IG, CHI, GA, and ACO	NP and NI 20, 30	Average of : Accuracy: 0.80 Precision: 0.64 Recall: 0.67 F-measure : 0.65	Macro-F1: 79.01 Micro-F1: 89.87 Accuracy: 81.35	- Not use features weighted - Low performance - No parameter tuning
P9 [42]	FFA	IG and CHI	NA	Average of : Accuracy: 0.80 Precision: 0.64 Recall: 0.67 F-measure : 0.65	Macro-F1: 79.01 Micro-F1: 89.87 Accuracy: 81.35	- Not use features weighted - No exact figure of iterations and populations - Low performance - Not compared with optimization methods
P 10 [43]	Weighted PSO	PSO	NA of NP and 100 NI	Average Accuracy of weighted PSO with parameter tuned KNN:	Re0: 66.5, R1: 36.84 Fbis: 64.43, Wap: 48.3, Oh5:61.0	- Not compared with typical FS - Time inefficient
P 11 [44]	GWO	NA	10 NP, and 100 NI	F-measure: 0.96, and AUC: 0.96	Accuracy: 0.97,	- Not compared with typical FS
P12 [45]	BBO- Bagging	GA, PSO and BBO	100 NP and NI	Average of P: 83.87 R: 70.67	A: 85.16 F: 76.71	- Time inefficient - Not use features weighted - Low Performance
P13 [46]	GA	IG, GI, GR, FS, LS, RDC, mRMR, UFSACO, RRFS, MOMI and MECY_FS	100 NP and NI	Average precision of MORDC using (1500) number of features:MNB on : WebKB: 0.84, 20Newsgroups: 0.55, Reuters-21,578: 0.77	MLP on :WebKB: 0.54, 20Newsgroups: 0.45, Reuters-21,578: 0.69 DT on: WebKB: 0.47, 20 Newsgroups: 0.72, Reuters-21,578: 0.62	- No parameter tuning - Low performance - Fails to identify features that are undetected when optimizing because of using a small number of metrics.

TABLE 7. (Continued.) Comparison of the proposed models for the TC problem and their issues.

ID Ref.	Feature Selection Method (s)	Compared With Typical and current method (s)	Numbers of Population (NP) and Generation (s) (NI)	Performance of Proposed method (s)		Issues
P14 [47]	OTFS	PSO, ACO ABC, and FA	NA	Perf. (OTFS) with MLearn-ATC on: Reuters: P: 0.84, R: 0.83, F: 0.84, A: 0.93	20Newsgroup: P: 0.89, R: 0.82 F: 0.85, A: 0.93 Real dataset: P: 0.89, R: 0.84 F: 0.87, A: 0.96	- Not compared with typical FS - No exact figure of iterations and populations - Time inefficient
P15 [48]	MAFSA	FSA	NA	The average Precision of: MAFSA is 2.27% better than AFSA		- Not use features weighted - No exact figure of iterations and populations - Evaluation metrics
P16 [49]	SIW-APSO	IG, CHI, GA, and PSO	50 NP and 100 NI	Reuters-21578: P: 98.60, R: 96.56, F1: 97.57, Micro F1: 94.43, Macro F1: 84.23		- Time inefficient - Number of selected Feature
P17 [50]	SFLA	Improved SFLA	50 NP and 100 NI	The highest accuracy on: Reuters-21578: (with 240 features): 0.95, WebKb: (with 200 features): 0.95, Cade 12: (300 features): 0.95, 20 Newsgroup: (with 180 features): 0.89		- Time inefficient - Not compared with typical FS - Not compared with optimization methods
P18 [10]	HABBFO-ANN	GA-KNN	NA	Perf. Multilabel ANN Precision: 94.82 Recall: 93.79 Hamming loss: 20.45		- No exact figure of iterations and populations - Not compared with proposed classifier (ANN) - Evaluation metrics
P19 [51]	GOA and CSA	DCNN, and CAViaR+ DCNN	NA	Performance on: Reuter: P: 0.82, R: 0.95, A: 0.94	20 newsgroups: P: 0.95, R: 0.95, A: 0.96	- No exact figure of iterations and populations - Not compared with optimization methods - Evaluation metrics
P20 [52]	ACOGA	ACO, and GA	20 NP and 100 NI	Average Precision: 99.65	Average Recall: 93.17	- Time inefficient - Not compared with typical FS
P21 [53]	Filter: IG and CHI Wrapper: SWA	IG and CHI	0.3, 0.8, and 10 number of nodes with 5 NI	The proposed model achieved the best result with SVM on (FS= 300) Accuracy rate (CHI_FN_SWA):95.57 Accuracy rate (GI_FN_SWA): 95.34		- Not use features weighted - Evaluation metrics - Not compared with optimization methods
P22 [27]	Filter: CO, IG, GR, and SU Wrapper: PSO	CO, IG, GR, and SU	NA	Average Accuracy: CO Size (310): 88.74, IG Size (340): 89.52, GR Size (1020): 87.83, and SU Size (360): 89.34		- No exact figure of iterations and populations - Not compared with optimization methods - Evaluation metrics
P23 [54]	Filter: CO, IG, GR, and SU Wrapper: Multi-objective PSO	2SPSO, OMOPSO-LS, OMOPSO-FS, NSGAII-FS, SPEA2-FS, MOEA/D-FS	40 NP and 100 NI	Best Training and Test Accuracy of OMOPSO-LS respectively: 84.99 and 80.82 on Reuters R8 dataset		- Evaluation metrics - Number of selected feature
P24 [11]	Filter: NDM Wrapper: BJO	DF, IG, MI, DFS, NDM, GA and FFY	50 NP and NI	Accuracy of: WebKB: 88.9, SMS: 97.8, BBC News: 96.7 and 10 Newsgroup: 90.5		- Evaluation metrics - No parameter tunnig
P25 [55]	Filter: PCA Wrapper: WI-OMFS	BFS, GS, and WI-OMFS wrapper	20, 30, 40, 50, 60, 70, 80, 90, 100, 150 NP and NI	The Proposed model achieved the best accuracy with J48: (WI-OMFS filter) Accuracy: 91.63, Precision: 91.60 Recall: 91.60, F-measure: 91.60		- Time inefficient
P26 [56]	Filter: CFS and IG Wrapper: ACO and ABC	BF, GS and Ranker	20, 30 NP and NI	Perf. Proposed model for ACO: Accuracy: 93.11 Error rate: 6.88 T: 4.7 Accuracy of: EA + CFS + SVM: 91.81, Error rate: 0.03, T: 2.84 EA + CFS + NB: 92.21, Error rate: 0.031, T: 0.15	Perf. Proposed model for ABC: Accuracy: 78.18 Error rate:21.82 T: 0.65 Accuracy of: GA + CFS + SVM: 89.78, Error rate: 0.24, T: 2.03 GA+CFS+NB: 90.55, Error rate: 0.42, T: 1.26	- Evaluation metrics - No parameter tunnig
P27 [57]	Filter: CFS Wrapper: EA and GA	EA, GA and CFS	NA			- No exact figure of iterations and populations - Evaluation metrics

TABLE 7. (Continued.) Comparison of the proposed models for the TC problem and their issues.

ID Ref.	Feature Selection Method (s)	Compared With Typical and current method (s)	Numbers of Population (NP) and Generation (s) (NI)	Performance of Proposed method (s)	Issues	
P28 [58]	Filter: PCA and CFS Wrapper: CO, FO, and BO	BFS and RS	NP and NI 20, 30,40,50,60, 70, 80, 90, 100,150, and 200	Best Accuracy for J48 CO: 92.03% FO: 91.81% BO: 89.42%	Best Accuracy for SVMCO: 87.22% FO: 89.60% BO: 88.89%	- Time inefficient - Evaluation metrics
P29 [59]	Filter: PCA and CFS Wrapper: ACO, ABC, EA, FO, RO, and WO	BFS and GS	NP and NI 20, 30,40,50, and 100	Best Accuracy: WO: 90.99 ACO: 87.26 ABC: 39.38		- Time inefficient - Evaluation metrics
P30 [13]	Filter: - CHI2, DP, DFFS, and RDC.-PCA and LSI Wrapper: GA	- CHI2, DP, DFFS, and RDC	100 NP and 20 NI	The best performances of the proposed model mostly include either DFSS, CHI2, or RDC rather than DP, and AVG, EQ better than MAX feature set with using PCA and LSI feature transformation rather than GA on the two datasets based on 500 and 1500 features using different dimension reduction ratio. F-measure: Re0: 70.36, La1s: 77.08, La2s: 78.60, Oh0: 81.67, Oh5: 80.50, Oh10: 73.82, Oh15: 76.43, Fbis: 76.48, Trec41: 87.03		- Time inefficient - Not improved GA and not compared with other optimization methods
P31 [60]	Filter: IG Wrapper: GWO	SCA, GA, ACO, and MFO [66]	30 NP and 50 NI	The performance of the proposed method based on (5%) of selected feature: P: 94.53, R: 86.48, F1: 90.32 Performance of HBPRO with NB on: NPR dataset: P: 0.91, R: 0.89, F1: 0.90, A: 89.26, and 4-subject dataset: P: 0.94, R: 0.93, F1: 0.93, A: 93.16 HBPRO with SVM on NPR dataset: P: 0.90, R: 0.90, F1: 0.90, A: 89.63, and 4-subject dataset: P: 0.92, R: 0.91, F1: 0.91, A: 90.8		- Number of selected Feature - Low Performance
P32 [61]	Filter: IG Wrapper: ICA	IG and MI	15 NP and 200 NI	The experimental results show that combining the ranking-based filter strategy with a subset selection-based filter can provide a final subset of terms with the fewest terms and satisfactory classification performance.		- Low Performance - Not compared with optimization methods
P33 [62]	Filter: DFTF Wrapper: PRO	DFS, NDM, TRDL, DFTF, GA, and FFY	NA	The performance of the proposed method outperforms the seven compared methods in term of accuracy, hamming loss and ranking loss. Highest F1 values with Ratio of selected features: WebKB: 0.74, NB (10%) 0.74 FSVM (8%) 0.90, NB (8%) 0.89 20Newsgroup: FSVM (10%) 0.81 NB (8%) 0.78		- Low Performance - No exact figure of iterations and populations
P34 [63]	Filter: - IG-Global PMI-based FS (MGFM, MGFWM, CDCG FWM) Wrapper: BGSA	IG, GFM, MGFM, MGFWM, CDCGFWM	200 NP and 1000 NI	The performance of the proposed method outperforms the seven compared methods in term of accuracy, hamming loss and ranking loss. Highest F1 values with Ratio of selected features: WebKB: 0.74, NB (10%) 0.74 FSVM (8%) 0.90, NB (8%) 0.89 20Newsgroup: FSVM (10%) 0.81 NB (8%) 0.78		- Time inefficient
P35 [64]	Memetic Search: Filter: LFD Wrapper: EDA	MD, RP, VGFSS, NDM, AIPSO, EGA, and (EGA+CDM)	50 NP and 100 NI	Averages of: Precision: 82,32 Recall: 82,89, F-measure: 82,22		- Its strong bias to the given problem - Time inefficient
P36 [65]	Filter: GMFS Wrapper: IABCM	IG, CMFS MI, MDMR, CCM, and GMFS	20 NP and 4000 NI		Reuters21578: FSVM (10%) 0.74, NB (10%) 0.74 TREC2007: FSVM (10%) 0.97, NB (10%) 0.95	- Time inefficient - Not compared with optimization methods
P37 [66]	Filter: IG Wrapper: ISCA	IG, Correlation, Significance,SCA, OBL (SCA), Levy (SCA), Weighted (SCA), GA, ACO, and MFO	30 NP and 50 NI			- Not use features weighted -Number of selected Feature - Low Performance

of an optimal feature subset. The selection is chosen based on an evaluation measure. All ants, in particular, refresh their pheromone, after which the superior underground insect stores additional pheromone on the nodes of the best course of action, and the best arrangement may be made by GA or ACO. Wang et al. [65], presented a modified ABC algorithm

by integrating semantic information into the FS procedure. Semantic information and filter-based methods are used to maintain high TC accuracy and adequate execution time. The semantic information between two words (features) in a text document is calculated using conditional probability distribution and ODP (knowledge-based), which construct the

Equivalent Word Set (EWS). Then, the constructed EWS was integrated into the enhanced Artificial Bee Colony Memory (ABCM), an advanced metaheuristic algorithm that imitates the artificial bee's memory to benefit from the previous successful experience of the employee bee (onlooker bee).

Finally, Belazzoug *et al.* [66], developed a new ISCA algorithm to overcome the SCA's search space issue. SCA is sometimes stuck in the sub-optimal region due to limitations in the exploration of the search space. Accordingly, the ISCA used a dynamic search to consider the random search to find the solution instead of only relying on the optimal solution. Therefore, determining and improving the metaheuristics performance and their mechanism, such as improving the convergence speed, balancing the global and local search space, and considering the past experience of the algorithms (e.g., employee bees in ABC) help the algorithm to select an optimal feature subset and assess the classifier in their classification.

RQ4: What are the gaps in the current studies and their future directions?

Answer4: As shown in Figure 3, over the last few years, several studies indicate increasing attention among researchers on investigating TC using the optimization of the FS problem, especially during these three years (2018, 2019, and 2020). Therefore, this section discusses the gaps in the current FS methods after analyzing the information from the selected studies, and recommendations are presented for future research.

- **Selective Classes:** Some of the conducted studies seems choosy in the selected classes (class labels or categories) from the benchmark datasets to assess the performance of their proposed methods for TC problem [10], [19], [40]–[42]. Although these conducted studies are relatively achieved better accuracy using a selective number of classes. Yet, they are unable to select features from a very large search space because the used wrapper approaches herein are not supportive for many classes. Hence, the hybrid approach can be used to resolve the issue of wrapper approaches by relying on filter and wrapper methods.

- **Text representation:** The importance of using newer text representation on text data is to overcome the limitations of traditional text representation models like BOW, as the latter relies on the word frequency in their analysis and ignores the semantics of the words [37], [42]. The investigated experiments in [37] show that the conceptual representation model is a better text representation than BOW. This is because conceptual representations consider the semantics of the words using a lexical knowledge-based dictionary like WordNet. In [65], however, the performance of the IABCM algorithm was improved using a semantic representation from the ODP knowledge-based method. Another direction of study is to consider the feature weights before applying the FS methods as it is proven to improve accuracy [1], [43], [47], [51].

- **Lack of comparison with baseline methods:** Based on the two paradigms of FS methods (wrapper and hybrid), the proposed optimization algorithms were not compared to

standard optimization algorithms or the typical FS methods. The necessity of comparing the metaheuristic methods with typical methods is important because the modern metaheuristics methods resolved the problem of the typical FS methods, e.g., avoiding interaction between features. However, others examined the proposed model using different classification methods with no account of comparing it with different FS methods. Also, traditional or advanced classifiers achieved high performance using feature subsets from optimization or typical FS methods. Yet, some of the proposed models were not validated against different kinds of classifiers to prove the reliability of the FS methods against multiple classifiers. In [45], it has been shown that comparing different classifiers results in different classification accuracy.

- **Wrapper-based approaches:** Most of the metaheuristic methods are adapted to resolve the curse of dimensionality in the data. Yet these adapted methods still suffer from the problem of parameter tuning that leads to high computational time [2], [20], [37], [40], [45]. Moreover, the problem in the adapted methods is being stuck in the local optima of the search space, and the selected features are not appropriately selected. Therefore, the modified [47], [48] and hybridized [10], [50] methods are proposed to address the problem of being stuck in the local optima by balancing exploration and exploitation. Yet these methods (modified and hybridized) still suffer from high time complexity. Subsequently, the hybrid-based approaches provide the recommendation to resolve the problem of computation time by ranking and selecting the significant features using the filter-based method. This helps to reduce the dimensionality of the data and limits the search space of the optimization algorithm in the wrapper-based method to solve the problem of being trapped in the local optima. This also enhances the execution speed.

- **Hybrid (filter-wrapper)-based approaches:** The filter-based methods could be univariate (e.g., IG) or multivariate (e.g., CFS) methods. Either of them is being used as the FS method that technically relies on different factors of various characteristics (e.g., probability distribution). Applying either of the filter-based methods (univariate or multivariate) on a specific dataset will produce a different subset of features than other filter-based methods using the same dataset. Therefore, the applied classification method on each of the obtained subset of features led to a different performance. To cope with this issue, the filter-based approach can be modified by hybridizing it with an ensemble of filter approaches to achieve a better selection of the subset of features. The adapted optimization algorithms in hybrid-based approaches focus on resolving the high computation issue in the wrapper-based method, and the least effort is given to improve the mechanism of optimization algorithms [11], [26], [52], [55]–[59], except for [65] and [66] who developed a unique optimization method (IABCM) that keeps the tracking of the optimal solution using memory. A dynamic search is introduced in ISCA to avoid being trapped in the local optima, respectively. To handle these difficulties, more

work can be done to improve/hybridize the metaheuristics-based method(s) to obtain better performance by overcoming a premature convergence and local search issues (exploitation and exploration) to attain optimal or near-optimal solutions. Meanwhile, improving the statistical FS methods is another research direction to remove redundant features.

- **Single and multi-objective optimization problem:** The majority of the conducted metaheuristic algorithms that proposed to conduct TC problem reliant on the single objective function that suffers from forcing the evolving population to form a particular feature-set due to the use of the single quality function. It is not considering the reconciliation between two or more conflicting processing functions. Therefore, a new direction of research has considered the usage of multi-objective functions like MORDC [46] and multi-objective PSO [54]. Other future research directions may consider the multi-objective function that uses minimizing redundancy, maximizing redundancy, minimizing the number of features, maximizing classifier accuracy, and time efficiency.

- **Metaheuristic algorithms:** Most of the metaheuristic optimization algorithms in the conducted related studies were originated from the Swarm intelligence-based algorithms (SI) [1], [10], [27], [38]–[44], [47]–[51], [54]–[56], [58]–[60], [65], and Evolutionary-based algorithms (EA) [2], [13], [20], [37], [45], [46], [57], [64], and some other studies based on Physics-based algorithms [63], [66], and Human behavior related algorithms [11], [53], [61], [62]. SI and EA are nature-inspired algorithms that share the behaviour of population-based (multiple solutions) stochastic optimization techniques. These algorithms generate a population of solutions that are updated with the number of generations (i.e., iterations) and start their optimization process. However, SI seeks to design intelligent multi-agent systems inspired by the collective actions of social insects such as ants, termites, wasps, bees, and other animal communities, such as flocks of birds or schools of fish, that compete for foods. Where are cooperated by an indirect connection medium and do actions in the decision space. While EA is grounded on the survival of the fittest candidate for a particular climate, they start with a set of solutions that strives to thrive in a particular setting that is defined with a fitness evaluation. The parent population operates in an ecosystem to share its adaptation properties with children who have different processes of evolution, such as genetic crossover and mutation. This approach continues for many generations until the most acceptable solutions for the environment are obtained. Accordingly, having multiple solutions assisted in avoiding local optima, as the particles works cooperatively with a great exploration of search space compared to Physics and Human-based algorithms. For the issue of TC, new algorithms can be adapted, or hybridization between the existing algorithms (e.g., SI) and Physics or Human-based algorithms can be considered as future directions.

- **Unfair assessment:** The majority of the conducted studies used different evaluation metrics to assess the performance of

the models [2], [41], [48], [53], [56], amongst others. However, evaluation metrics have to be generalized or consistent over the proposed models to provide a fair assessment.

In conclusion, future models could be generalized on new test data by considering the following points:

- 1) The contextual analysis of the datasets is an essential factor. An example of this is the use of a different number of classes or features that will lead to a different accuracy using the same technique [1], [20], [38], [40]–[42], [53], [65].
- 2) The text representation (e.g., Conceptual representation) seems to affect the optimally selected features using FS, consequently producing varying accuracy [37], [47].
- 3) Term weighting (e.g., TF-IDF) schemes are used to ignore the less useful features (keywords). Therefore, reducing the burden on the FS in selecting the optimal feature subsets [1], [43], [51], [65].
- 4) To maintain a predictive model, one may consider:
 - a. Comparing different classifiers using the FS method or without FS [2], [20], [27], [45], [48], [55].
 - b. Different metaheuristic algorithms [57], [59], [66].
 - c. Comparing different metaheuristics methods (e.g., GA) with traditional FS methods (e.g., IG) [11], [38], [41], [65], [66].
 - d. Parameter tuning of classifier (e.g., k value in KNN) and metaheuristics methods (e.g., number of iterations (generations) and population size) [38], [39], [47].
 - e. Comparing with the state-of-the-art model(s) on the same benchmark dataset [1], [38], [45].
- 5) Evaluation Metrics (i.e., precision, recall, F-measure, accuracy, time complexity, and error rate) and consider the number of selected features [2], [56], [57], [66].

IV. CONCLUSION AND LIMITATIONS

Text classification (TC) is widely implemented in dealing with the data structure partition in a known area (labeled data), such as email filtering, e-news filtering, topic identification, document routing, and so forth. TC approaches play a key role in tagging text documents into categories/classes based on their content. Features selection (FS) using metaheuristics algorithms can efficiently improve the accuracy of classification, computation demands, and storage, and thus, it has been applied increasingly in TC. This article presents a comprehensive analysis of the current metaheuristics-based FS methods for TC problems. A total of 37 studies was selected from the year 2016 to 2021. This systematic literature review (SLR) provides an essential contribution for TC by shedding light on the previous six years of the existing body of knowledge to comprehensively understand the metaheuristics-based FS methods and their pros and cons in reducing the computational resources to perform TC. Besides, this SLR states the statistical analysis that was carried out to select the relevant studies critically. The selected studies were further analyzed and categorized into

two paradigms (wrapper and hybrid approaches). The wrapper approaches comprise three kinds of techniques (adapted, modified, and hybridized), each with its own characteristics. There is a need for high computational resources for processing wrapper-based approaches that lead to a low classification performance due to the huge search space for the FS. Thus, the hybrid-based approaches have emerged in two techniques (adapted and modified) to cope-up with time inefficiency and the huge search space. Critical research questions were formulated to justify the usage of each approach, its effect on TC accuracy, the comparison with typical and existing models, and finally, the gaps and future directions of each model. Other than those investigated in ongoing research, the hybrid-based approaches require better attention as they are not well explored. To this end, several key points for future models are presented as future work. However, this SLR focuses on the metaheuristic-based FS methods for TC in the English language text, and there is also a need for an SLR that can explore TC models for other text languages such as Arabic, Chinese, Spanish, and so on.

REFERENCES

- [1] H. Majidpour and F. S. Gharehchopogh, "An improved flower pollination algorithm with AdaBoost algorithm for feature selection in text documents classification," *J. Adv. Comput. Res.*, vol. 9, no. 1, pp. 29–40, 2018.
- [2] S. Khanalini and F. S. Gharehchopogh, "A new approach for text documents classification with invasive weed optimization and naive Bayes classifier," *J. Adv. Comput. Eng. Technol.*, vol. 4, no. 3, pp. 167–184, 2018.
- [3] M. M. Mironczuk and J. Protasiewicz, "A recent overview of the state-of-the-art elements of text classification," *Expert Syst. Appl.*, vol. 106, pp. 36–54, Sep. 2018.
- [4] G. Mujtaba, L. Shuib, N. Idris, W. L. Hoo, R. G. Raj, K. Khowaja, K. Shaikh, and H. F. Nweke, "Clinical text classification research trends: Systematic literature review and open issues," *Expert Syst. Appl.*, vol. 116, pp. 494–520, Feb. 2019.
- [5] J. T. Pintas, L. A. F. Fernandes, and A. C. B. Garcia, "Feature selection methods for text classification: A systematic literature review," *Artif. Intell. Rev.*, vol. 54, pp. 6149–6200, Feb. 2021.
- [6] A. Dhar, H. Mukherjee, N. S. Dash, and K. Roy, "Text categorization: Past and present," *Artif. Intell. Rev.*, vol. 54, pp. 3007–3054, Sep. 2020.
- [7] K. Kowsari, K. J. Meimandi, M. Heidarysafa, S. Mendu, L. Barnes, and D. Brown, "Text classification algorithms: A survey," *Information*, vol. 10, no. 4, pp. 1–68, 2019.
- [8] P. D. Turney and P. Pantel, "From frequency to meaning: Vector space models of semantics," *J. Artif. Intell. Res.*, vol. 37, pp. 141–188, Feb. 2010.
- [9] A. I. Kadhim, "Survey on supervised machine learning techniques for automatic text classification," *Artif. Intell. Rev.*, vol. 52, no. 1, pp. 273–292, Jun. 2019.
- [10] J. Maruthupandi and K. V. Devi, "Multi-label text classification using optimised feature sets," *Int. J. Data Mining, Model. Manage.*, vol. 9, no. 3, pp. 237–248, 2017.
- [11] K. Thirumoorthy and K. Muneeswaran, "Optimal feature subset selection using hybrid binary Jaya optimization algorithm for text classification," *Sādhanā-Acad. Proc. Eng. Sci.*, vol. 45, no. 1, pp. 1–13, 2020.
- [12] M. Iqbal, M. M. Abid, M. N. Khalid, and A. Manzoor, "Review of feature selection methods for text classification," *Int. J. Adv. Comput. Res.*, vol. 10, no. 49, pp. 138–152, Jul. 2020.
- [13] A. K. Uysal, "On two-stage feature selection methods for text classification," *IEEE Access*, vol. 6, pp. 43233–43251, 2018.
- [14] P. Kumbhar, "A survey on feature selection techniques and classification algorithms for efficient text classification," *Int. J. Sci. Res.*, vol. 5, no. 5, pp. 1267–1275, 2016.
- [15] J. Cai, J. Luo, S. Wang, and S. Yang, "Feature selection in machine learning: A new perspective," *Neurocomputing*, vol. 300, pp. 70–79, Jul. 2018.
- [16] L. H. J. Tang and S. Alelyani, "Feature selection for classification: A review," *Data Classif. Algorithms Appl.*, pp. 571–605, 2014.
- [17] P. J. Y. Yang, Y. Yiming, and P. O. Jan, "A comparative study on feature selection in text categorization," in *Proc. 14th Int. Conf. Mach. Learn. (ICML)*, 1997, vol. 53, no. 1, pp. 412–420.
- [18] F. P. Shah and V. Patel, "A review on feature selection and feature extraction for text classification," in *Proc. Int. Conf. Wireless Commun., Signal Process. Netw. (WiSPNET)*, Mar. 2016, pp. 2264–2268.
- [19] V. Bolón-Canedo, N. Sánchez-Marroño, and A. Alonso-Betanzos, "A review of feature selection methods on synthetic data," *Knowl. Inf. Syst.*, vol. 34, no. 3, pp. 483–519, 2013.
- [20] P. Kumbhar, M. Mali, and M. Atique, "A genetic-fuzzy approach for automatic text categorization," in *Proc. IEEE 7th Int. Advance Comput. Conf. (IACC)*, Jan. 2017, pp. 572–578.
- [21] S. A. G. Shirazi and M. B. Menhaj, "A new genetic based algorithm for channel assignment problems," in *Proc. 9th Fuzzy Days Comput. Intell. Theory Appl. Int. Conf.*, Dortmund, Germany, Sep. 2006, pp. 85–91.
- [22] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," *Comput. Elect. Eng.*, vol. 40, no. 1, pp. 16–28, Jan. 2014.
- [23] K. P. Singh, N. Basant, and S. Gupta, "An introduction to variable and feature selection," *Anal. Chim. Acta*, vol. 703, no. 2, pp. 152–162, 2003.
- [24] Y. B. Wah, N. Ibrahim, H. A. Hamid, S. Abdul-Rahman, and S. Fong, "Feature selection methods: Case of filter and wrapper approaches for maximising classification accuracy," *Pertanika J. Sci. Technol.*, vol. 26, no. 1, pp. 329–340, 2018.
- [25] G. Naqvi, "A hybrid filter-wrapper approach for feature selection," M.S. thesis, Örebro Univ., Örebro, Sweden, 2012.
- [26] X. Tang, Y. Dai, and Y. Xiang, "Feature selection based on feature interactions with application to text categorization," *Expert Syst. Appl.*, vol. 120, pp. 207–216, Apr. 2019.
- [27] X. Bai, X. Gao, and B. Xue, "Particle swarm optimization based two-stage feature selection in text mining," in *Proc. IEEE Congr. Evol. Comput. (CEC)*, Jul. 2018, pp. 1–8.
- [28] A. E. Ezugwu, A. K. Shukla, R. Nath, A. A. Akinyelu, J. O. Agushaka, H. Chiroma, and P. K. Muhuri, "Metaheuristics: A comprehensive overview and classification along with bibliometric analysis," *Artif. Intell. Rev.*, vol. 54, no. 6, pp. 4237–4316, Aug. 2021.
- [29] P. Agrawal, H. F. Abutarboush, T. Ganesh, and A. W. Mohamed, "Meta-heuristic algorithms on feature selection: A survey of one decade of research (2009–2019)," *IEEE Access*, vol. 9, pp. 26766–26791, 2021.
- [30] D. Mladenović, "Feature selection for dimensionality reduction," in *Proc. Int. Stat. Optim. Perspect. Workshop*. Berlin, Germany: Springer, 2006, pp. 84–102.
- [31] I. R. Vanani and S. Majidian, *Meta-Heuristic Algorithms: A Concentration on the Applications in Text Mining*, 1st ed. New York, NY, USA: Taylor & Francis, 2020.
- [32] X. Deng, Y. Li, J. Weng, and J. Zhang, "Feature selection for text classification: A review," *Multimedia Tools Appl.*, vol. 78, no. 3, pp. 3797–3816, 2019.
- [33] F. W. Neiva and L. S. de Da Silva, "Systematic literature review in computer science—A practical guide," Dept. Comput. Sci., Fed. Univ. Juiz Fora, Juiz de Fora, Brazil, Tech. Rep., Nov. 2016, p. 8, vol. 1.
- [34] B. Kitchenham, "Procedures for performing systematic reviews," *Empirical Softw. Eng. Nat. ICT Aust. Ltd.*, Keele Univ., Keele, U.K., Tech. Rep. TR/SE-0401, 2004.
- [35] B. Kitchenham, "Guidelines for performing systematic literature reviews in software engineering (version 2.3)," Keele Univ., Durham, U.K., Tech. Rep., 2007.
- [36] Z. A. Al-Sai, R. Abdullah, and M. H. Husin, "Critical success factors for big data: A systematic literature review," *IEEE Access*, vol. 8, pp. 118940–118956, 2020.
- [37] N. Bidi and Z. Elberichi, "Feature selection for text classification using genetic algorithms," in *Proc. 8th Int. Conf. Modelling, Identificat. Control (ICMIC)*, Nov. 2016, pp. 1–5.
- [38] A. Allahverdipoor and F. S. Gharehchopogh, "An improved K—Nearest neighbor with crow search algorithm for feature selection in text documents classification," *J. Adv. Comput. Res.*, vol. 9, no. 2, pp. 37–48, 2018.
- [39] J. Balakumar and S. V. Mohan, "Artificial bee colony algorithm for feature selection and improved support vector machine for text classification," *Inf. Discovery Del.*, vol. 47, no. 3, pp. 154–170, Aug. 2019.
- [40] P. Grover and S. Chawla, "Text feature space optimization using artificial bee colony," in *Soft Computing for Problem Solving (Advances in Intelligent Systems and Computing)*, vol. 1057, 2020, pp. 691–703.
- [41] R. J. Manoj, M. D. A. Praveena, and K. Vijayakumar, "An ACO-ANN based feature selection algorithm for big data," *Cluster Comput.*, vol. 22, no. S2, pp. 3953–3960, Mar. 2019.

- [42] R. Janani and S. Vijayarani, "Text classification using K-nearest neighbor algorithm and firefly algorithm for text feature selection," in *Advances in Electrical and Computer Technologies (Lecture Notes in Electrical Engineering)*, vol. 672, T. Sengodan, M. Murugappan, and S. Misra, Eds. Singapore: Springer, 2020.
- [43] A. Khurana and O. P. Verma, "PSO based optimal text classification using tuned k-NN and feature weighting," in *Proc. 4th Int. Conf. Comput. Manage. (ICCM)*, 2018, pp. 137–142.
- [44] M. Aktaş and F. Kılıç, "Binary grey wolf optimizer using archeology and astronomy news for text classification," in *Proc. Int. Conf. Innov. Eng. Appl. (CIEA)*, Muş, Turkey, Jun. 2021, pp. 1–7.
- [45] A. Khurana and O. P. Verma, "Novel approach with nature-inspired and ensemble techniques for optimal text classification," *Multimedia Tools Appl.*, vol. 79, nos. 33–34, pp. 23821–23848, Sep. 2020.
- [46] M. Labani, P. Moradi, and M. Jalili, "A multi-objective genetic algorithm for text feature selection using the relative discriminative criterion," *Expert Syst. Appl.*, vol. 149, Jul. 2020, Art. no. 113276.
- [47] R. Janani and S. Vijayarani, "Automatic text classification using machine learning and optimization algorithms," *Soft Comput.*, vol. 25, no. 2, pp. 1129–1145, Jan. 2021.
- [48] D. Thiyagarajan and N. Shanthi, "A modified multi objective heuristic for effective feature selection in text classification," *Cluster Comput.*, vol. 22, no. S5, pp. 10625–10635, Sep. 2019.
- [49] M. Asif, A. A. Nagra, M. B. Ahmad, and K. Masood, "Feature selection empowered by self-inertia weight adaptive particle swarm optimization for text classification," *Appl. Artif. Intell.*, pp. 1–16, Dec. 2021.
- [50] M. Mahmoudi and F. S. Gharehchopogh, "An improvement of shuffled frog leaping algorithm with a decision tree for feature selection in text document classification," *CSI J. Comput. Sci. Eng.*, vol. 16, no. 1, pp. 60–72, 2018.
- [51] V. Srilakshmi, K. Anuradha, and C. S. Bindu, "Optimized deep belief network and entropy-based hybrid bounding model for incremental text categorization," *Int. J. Web Inf. Syst.*, vol. 16, no. 3, pp. 347–368, Jul. 2020.
- [52] A. Singh and A. Kumar, "Text document classification using a hybrid approach of ACOGA for feature selection," *Int. J. Adv. Intell. Paradigms*, vol. 20, nos. 1–2, pp. 158–170, 2021.
- [53] Y. Lu and Y. Chen, "A text feature selection method based on the small world algorithm," *Proc. Comput. Sci.*, vol. 107, pp. 276–284, Jan. 2017.
- [54] T. Londt, X. Gao, and B. Xue, "Particle swarm optimisation based two-stage feature selection in text mining: A multi-objective approach," *Tech. Rep.*, Dec. 2020, pp. 1–37.
- [55] K. S. Kyaw and S. Limsiroratana, "Towards nature-inspired intelligence search for optimization of multi-dimensional feature selection," in *Proc. 23rd Int. Comput. Sci. Eng. Conf. (ICSEC)*, Oct. 2019, pp. 379–384.
- [56] K. S. Kyaw and S. Limsiroratana, "Traditional and swarm intelligent based text feature selection for document classification," in *Proc. 19th Int. Symp. Commun. Inf. Technol. (ISCIT)*, Sep. 2019, pp. 226–231.
- [57] K. S. Kyaw and S. Limsiroratana, "Case study: Knowledge discovery process using computation intelligence with feature selection approach," in *Proc. 17th Int. Conf. ICT Knowl. Eng. (ICT&KE)*, Nov. 2019, pp. 1–7.
- [58] K. S. Kyaw and S. Limsiroratana, "Optimization of multi-class document classification with computational search policy," *ECTI Trans. Comput. Inf. Technol.*, vol. 14, no. 2, pp. 149–161, 2020.
- [59] K. S. Kyaw and S. Limsiroratana, "Optimization of text feature selection process based on advanced searching for news classification," *Int. J. Swarm Intell. Res.*, vol. 11, no. 4, pp. 1–23, Oct. 2020.
- [60] O. M. Alyasiri, Y.-N. Cheah, and A. K. Abasi, "Hybrid filter-wrapper text feature selection technique for text classification," in *Proc. Int. Conf. Commun. Inf. Technol. (ICICT)*, Jun. 2021, pp. 80–86.
- [61] M. Mojaveriyani, H. Ebrahimpour-Komleh, and S. J. Mousavirad, "IGICA: A hybrid feature selection approach in text categorization," *Int. J. Intell. Syst. Appl.*, vol. 8, no. 3, pp. 42–47, Mar. 2016.
- [62] K. Thirumoorthy and K. Muneeswaran, "Feature selection using hybrid poor and rich optimization algorithm for text classification," *Pattern Recognit. Lett.*, vol. 147, pp. 63–70, Jul. 2021.
- [63] F. Z. Kermani, E. Eslami, and F. Sadeghi, "Global filter-wrapper method based on class-dependent correlation for text classification," *Eng. Appl. Artif. Intell.*, vol. 85, pp. 619–633, Oct. 2019.
- [64] J. Lee, I. Yu, J. Park, and D.-W. Kim, "Memetic feature selection for multilabel text categorization using label frequency difference," *Inf. Sci.*, vol. 485, pp. 263–280, Jun. 2019.
- [65] Y. Wang, L. Feng, and J. Zhu, "Novel artificial bee colony based feature selection method for filtering redundant information," *Int. J. Speech Technol.*, vol. 48, no. 4, pp. 868–885, Apr. 2018.
- [66] M. Belazzoug, M. Touahria, F. Nouioua, and M. Brahimi, "An improved sine cosine algorithm to select features for text categorization," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 32, no. 4, pp. 454–464, May 2020.



OSAMAH MOHAMMED ALYASIRI received the B.Sc. degree in computer science from Mustansiriyah University, in 2009, and the M.Sc. degree in computer science from Dr. Babasaheb Ambedkar Marathwada University, India, in 2013. He is currently pursuing the Ph.D. degree with the School of Computer Sciences, Universiti Sains Malaysia. He is also a Senior Lecturer with Al-Furat Al-Awsat Technical University, Iraq. His research interests include optimization, pattern recognition, text classification, information retrieval, machine learning, and feature selection.



YU-N CHEAH received the B.Comp.Sc. and Ph.D. degrees (Hons.) from Universiti Sains Malaysia, in 1998 and 2002, respectively. He is currently an Associate Professor at the School of Computer Sciences, Universiti Sains Malaysia. His research interests include sentiment analysis, semantic technologies, knowledge management, intelligent systems, and health informatics.



AMMAR KAMAL ABASI received the B.Sc. degree in computer information system from the Jordan University of Science and Technology, the M.Sc. degree in international business from The University of Jordan, and the Ph.D. degree in artificial intelligence and software engineering from Universiti Sains Malaysia. His research interests include evolutionary algorithms, nature inspired computation, artificial intelligence, and their applications to optimization problems.



OMAR MUSTAFA AL-JANABI received the B.Sc. degree from the University of Al-Ma'mun, Iraq, the M.Sc. degree from Acharya Nagarjuna University (ANU), India, and the Ph.D. degree from Universiti Sains Malaysia. His research interests include data science, Bayesian probabilistic models, sentiment analysis, natural language processing, and deep Bayesian models.

...