

A systematic literature review on meta-heuristic based feature selection techniques for text classification

Sarah Abdulkarem Al-shalif¹, Norhalina Senan¹, Faisal Saeed², Wad Ghaban³, Noraini Ibrahim¹, Muhammad Aamir⁴ and Wareesa Sharif⁵

¹ Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia, Parit Raja, Johor, Malaysia

² DAAI Research Group, Department of Computing and Data Science, School of Computing and Digital Technology, University of Birmingham, Birmingham, United Kingdom

³ Applied College, University of Tabuk, Tabuk, Saudi Arabia

⁴ School of Electronics, Computing and Mathematics,, University of Derby, Derby, United Kingdom

⁵ Faculty of Computing, The Islamia University of Bahawalpur, Bahawalpur, Pakistan

ABSTRACT

Feature selection (FS) is a critical step in many data science-based applications, especially in text classification, as it includes selecting relevant and important features from an original feature set. This process can improve learning accuracy, streamline learning duration, and simplify outcomes. In text classification, there are often many excessive and unrelated features that impact performance of the applied classifiers, and various techniques have been suggested to tackle this problem, categorized as traditional techniques and meta-heuristic (MH) techniques. In order to discover the optimal subset of features, FS processes require a search strategy, and MH techniques use various strategies to strike a balance between exploration and exploitation. The goal of this research article is to systematically analyze the MH techniques used for FS between 2015 and 2022, focusing on 108 primary studies from three different databases such as Scopus, Science Direct, and Google Scholar to identify the techniques used, as well as their strengths and weaknesses. The findings indicate that MH techniques are efficient and outperform traditional techniques, with the potential for further exploration of MH techniques such as Ringed Seal Search (RSS) to improve FS in several applications.

Submitted 17 August 2023

Accepted 3 May 2024

Published 12 June 2024

Corresponding authors

Norhalina Senan,

halina@uthm.edu.my

Faisal Saeed, faisal.saeed@bcu.ac.uk

Academic editor

Arkaitz Zubiaga

Additional Information and
Declarations can be found on
page 37

DOI [10.7717/peerj-cs.2084](https://doi.org/10.7717/peerj-cs.2084)

© Copyright

2024 Al-shalif et al.

Distributed under

Creative Commons CC-BY 4.0

OPEN ACCESS

Subjects Artificial Intelligence, Data Mining and Machine Learning, Text Mining

Keywords Feature selection, Meta-heuristic techniques, Text classification, Dimensionally reduction, Evolutionary algorithms, Ringed seal search, Ant colony optimization, Particle swarm optimization, Genetic algorithm, Learning accuracy

INTRODUCTION

The amount of high-dimensional data currently freely accessible online, such as text data, microarrays, and medical information, has drastically expanded in recent years. As a result, text classification has emerged as a crucial task in various domains, including natural language processing, information retrieval, sentiment analysis, and more. The effectiveness of text classification models heavily relies on the features used to represent the textual data. As the dimensionality of textual data increases, the need for feature selection (FS)

techniques becomes paramount to enhance model performance, reduce computational costs, and mitigate the curse of dimensionality (Larabi Marie-Sainte & Alalyani, 2020).

FS stands for the procedure of choosing the more important and relevant features from the datasets. Because it can increase learning accuracy, shorten learning times, and simplify learning outcomes, FS is significant and has shown significant growth. The FS approach aims to determine the most relevant and essential features to improve the classification accuracy and minimize these features without significantly affecting the performance of the classification. It increases classification effectiveness by reducing data dimensionality by removing unnecessary and redundant features (Larabi Marie-Sainte & Alalyani, 2020; Mojaveriyan, Ebrahimpour-komleh & Jalaleddin, 2016). FS technique is classified into three models: filter, wrapper, and embedded models. Filter models consider the statistical properties of the data to choose the optimal feature subset (Bertolazzi et al., 2016; García-Torres et al., 2016; Mohanty & Das, 2018; Chen, Zhou & Yuan, 2019; Ghosh et al., 2020). Filter models are also known as traditional FS techniques such as Information Gain, CHISquare, and ReliefF (Tubishat et al., 2019), they can be divided into two additional categories, namely feature ranking algorithms also known as univariate feature filters. In feature ranking algorithms, each feature is given a weight according to how relevant it is to the target concept. In contrast, subset search algorithms are known as multivariate filters (Ghimatgar et al., 2018). The filter model has no direct interaction with the classifier, while the wrapper model adopts optimization algorithms such as meta-heuristic (MH) that are capable of direct interaction with the features and classifier (Tubishat et al., 2019). In addition, embedded models interact with the classifier with a lower computational cost than the wrapper model (Ghimatgar et al., 2018). Due to FS's importance, many researchers have investigated its problems and proposed many techniques to improve it and remove irrelevant, redundant, and noisy features to choose a set of features that will provide the optimum accuracy and computational performance (Kashef & Nezamabadi-pour, 2015). Despite the growing interest in text classification and FS, there is a noticeable lack of comprehensive evaluations of meta-heuristic-based techniques specifically tailored for text classification. This review aims to bridge this gap by aggregating and critically analyzing the existing body of literature on this topic.

MH is a high-level, problem-independent algorithm framework that offers several methods for creating heuristic algorithms (Yong, Dun-wei & Wan-qiu, 2016). With an enormous number of features, it is computationally impossible to evaluate every state, necessitating MH search techniques. Recently, MH algorithms, such as genetic algorithms, particle swarm optimization, simulated annealing, and ant colony optimization, have shown promise in handling complex optimization challenges. These algorithms have the potential to uncover relevant features from high-dimensional text data, contributing to improved classification accuracy and model interpretability. In FS techniques, MH is used to improve the result of classification performance (Kashef & Nezamabadi-pour, 2015). Many researchers (Larabi Marie-Sainte & Alalyani, 2020; Tubishat et al., 2019; Jain et al., 2019; Ahmad, Bakar & Yaakub, 2019; Al-Rawashdeh, Mamat & Hafhizah Binti Abd Rahim, 2019; Chantar et al., 2020; Kumar & Jaiswal, 2019; Singh & Kaur, 2020;

Hassonah et al., 2020; Gokalp, Tasci & Ugur, 2020) attempt to utilize the advantages of the natural inspired MH search to discover the optimal subset feature to enhance classifier performance and decrease computational time and cost. These techniques make an effort to provide better solutions by using information from earlier iterations (*Kashef & Nezamabadi-pour, 2015*). The two stages of MH are exploration and exploitation, where various operators are placed to find the best solution. The potential solutions move with the search space during exploration. At the same time, the most popular regions in the search space are investigated in the exploitation. The interaction between exploration and exploitation should be balanced in a good MH, according to expectations (*Ibrahim et al., 2019*). In this review article, Meta-Heuristic Feature Selection (MH-FS) techniques have been analyzed in detail and RSS is investigated to be used as a feature subset selection technique for future direction. RSS is one of the MH techniques proposed by *Saadi et al. (2016)*. The natural behavior of the seal pup in selecting the ideal hiding lair to avoid predators served as the inspiration for RSS. As opposed to GA and PSO, when compared to its homologs, RSS is faster in locating the global optimum and maintaining the proper ratio of exploitation to exploration (*Saadi et al., 2016*). Previously, RSS is not used as FS but it can optimize the support vector machine (SVM) parameter leading to higher classification accuracy when compared to traditional SVM (*Sharif et al., 2019*).

Literature review articles might be classified into two groups: traditional literature review (TLR) and systematic literature review (SLR). TLR attends to show the research topic from a general point of view and looks at the research in general and from all directions. In comparison, SLR attempts to show the topic of the research from a broad perspective and tries to answer specific research questions through a systemic methodology (*Qasem et al., 2019*). Currently, no SLR focuses on MH-FS techniques for text classification, Therefore, this study seeks to identify the best practices, trends, and patterns in the application of MH-based feature selection techniques for text classification. By extracting insights from a diverse range of studies, researchers and practitioners can gain valuable guidance for selecting appropriate algorithms and parameters for their specific applications. As well as, this review not only evaluates the performance of various MH algorithms but also assesses their suitability for different text classification tasks, dataset characteristics, and evaluation metrics. By doing so, it contributes to the methodological advancement of text classification research. Specifically, this article focused on the SLR of FS using MH techniques published in the period from 2015 to 2022. This SLR aims to summarize and clarify available guidance related to (1) the MH techniques for FS techniques, (2) the MH-FS techniques that can be used for text classification, (3) the comparisons of the performance of the MH techniques over the traditional techniques, and (4) highlight the strengths and weaknesses of different MH techniques. While, the intended audience for this SLR encompasses a wide range of individuals with a shared interest in advancing the capabilities of text classification through the integration of MH-based feature selection techniques such as academics, data scientists, students, industry practitioners, decision-makers, and researchers in FS, text classification, and MH fields. Whether seeking theoretical insights, practical guidance, or interdisciplinary connections,

this review offers a valuable resource to foster informed decision-making, research, and innovation in the dynamic field of text analysis.

The rest of the article is organized as follows. “Survey Methodology” explains the methodology used in this review. “Bibliometric Analysis by Co-occurrence (Authors Keywords)” presents the bibliometric analysis maps. “Results and Discussion” presents and discusses the review results. The conclusion and future work are presented in “Conclusion”.

SURVEY METHODOLOGY

In this article, the planning, conducting, and reporting procedures are according to the procedure given by Kitchenham and Charters (*El-Gohary, Nasr & Wahaab, 2000*). The procedures consist of five steps which are: identifying research questions, search strategy and study selection criteria, quality assessment criteria, data extraction, and data synthesis processes (*Qasem et al., 2019*).

Firstly, the research questions were meticulously crafted to address the inherent challenges associated with FS-MH techniques. Subsequently, the search strategy has been elucidated, encompassing the identification of search terms, utilization of search resources, and the systematic execution of the search process aimed at identifying and selecting pertinent studies. The selection of relevant studies, in alignment with the research questions, was carried out based on well-defined inclusion and exclusion criteria. The next phase involved articulating quality assessment criteria, a pivotal component for rigorously analyzing and assessing the studies under consideration. Finally, the processes of data extraction and synthesis was detailed, which form the concluding steps in this SLR. The ensuing subsections provide an in-depth exposition of each of these procedural steps, and [Fig. 1](#) offers a visual representation of the sequential flow of the SLR methodology.

Research questions

This SLR is purposed to define the guidance gained from the previous studies using the MH techniques for FS in text classification. [Table 1](#) presents five research questions discussed in this SLR. From the previous studies, MH techniques that have been used for FS were identified in (RQ1). The analysis of these studies has been conducted to answer the questions. RQ2 determines which FS-MH techniques have been used for text classification. RQ3 compares the performance of MH-FS techniques with the traditional techniques. The purpose of this question is to discover if MH methods are superior to conventional methods. RQ4 identified the advantages and limitations of different MH techniques to guide the selection of appropriate MH techniques, while the final question (RQ5) investigated how RSS can be used as FS.

Search strategy and study selection

Detailed analysis of this step is described in four subsections which are search terms, literature sources, search process, and study selection.

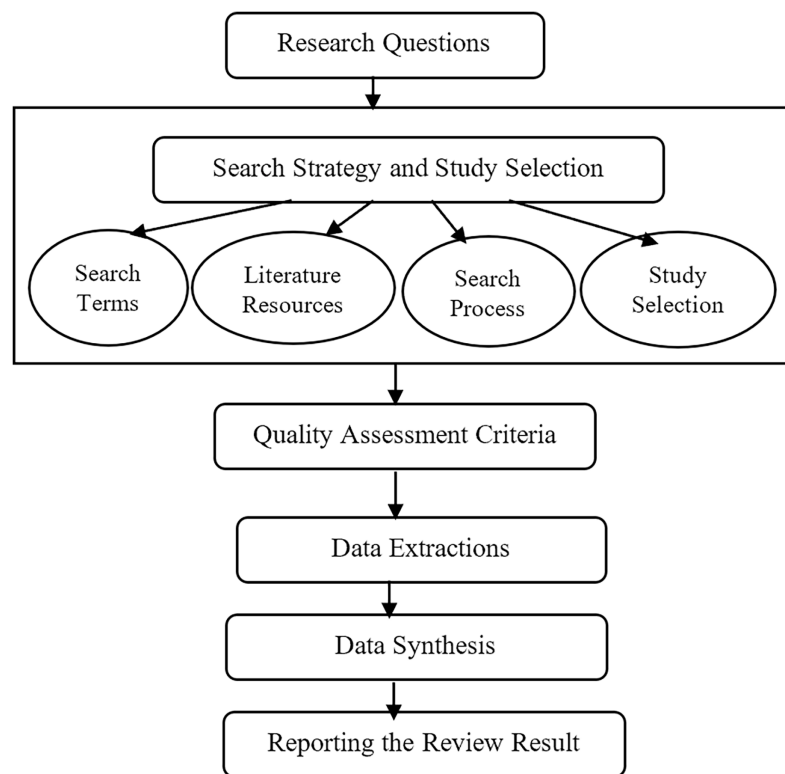


Figure 1 The steps of the SLR.

Full-size DOI: [10.7717/peerj-cs.2084/fig-1](https://doi.org/10.7717/peerj-cs.2084/fig-1)

Table 1 Research questions.

RQ#	Research questions	Reasons
RQ1	Which metaheuristic (MH) techniques have been utilized for feature selection (FS)	Identify which MH techniques are commonly used for FS.
RQ2	In the context of text classification, which specific metaheuristic techniques have been applied for feature selection?	Determine which MH techniques are commonly used for FS in text classification.
RQ2.1	What are the datasets employed in the application of metaheuristic feature selection (MH-FS) for text classification?	Identify common datasets used for text classification.
RQ2.2	Which classifiers have been used with MH-FS in text classification?	Identify common classifiers used for text classification.
RQ2.3	Which performance evaluation metrics are commonly utilized to assess the effectiveness of MH-FS in text classification?	Identify evaluation metrics reported to be appropriate for text classification.
RQ3	Is there empirical evidence indicating that MH-FS techniques outperform traditional FS methods in the domain of text classification?	Investigate the performance of the MH techniques over the traditional techniques
RQ4	What are the discernible strengths and weaknesses of MH techniques in the context of FS?	Highlighted the strengths and weaknesses of different MH techniques
RQ5	How can the RSS be effectively leveraged as FS technique?	Investigate the application of the RSS-FS technique algorithm.

Search terms

Five steps were conducted to extract the search terms which are as follows (*Malhotra, 2015*):

- 1) Extract the key terms from the research questions.
- 2) Determine the synonyms and alternative spellings for the main terms.
- 3) Explore the keywords and terminology from existing research articles.
- 4) Combine the synonyms and alternative spellings using the Boolean operator “OR”.
- 5) Connect the main terms using the Boolean operator “AND”.

All research terms were derived from the explored topic. These terms are feature selection, attribute selection, text classification, text categorization, meta-heuristics, and metaheuristics. The final search terms that connect with the Boolean operators were as follows: (“feature selection” OR “attribute selection”) AND (“text classification” OR “text categorization”) AND (“meta-heuristics” OR “metaheuristics”).

Literature sources

The relevant studies were investigated through the Scopus, Science Direct, and Google Scholar databases. The records identified in the period from 2008 to 2022 used “feature selection” and “attribute selection” as the main keywords and the rest of the keywords to specify and limit the selected studies.

Search process

The number of studies obtained from the Scopus, Science Direct, and Google Scholar databases using the main keywords were as follows: 45,236 for Scopus, 53,030 for Science Direct, and 19,300 for Google Scholar. These studies were then further filtered using specific keywords such as “text classification” or “text categorization,” resulting in 5,186 studies for Scopus, 1,922 studies for Science Direct, and 842 studies for Google Scholar. Another set of specific keywords, namely “meta-heuristic” or “metaheuristic,” yielded 226 studies for Scopus, 438 studies for Science Direct, and 34 studies for Google Scholar. After the initial search, the relevant studies were chosen according to predefined inclusion and exclusion criteria explained in the following subsection.

Study selection

The studies obtained were further narrowed down to include only articles published between 2015 and 2022. The resulting numbers were 194 studies for Scopus, 200 studies for Science Direct, and 25 studies for Google Scholar. The Inclusion-Exclusion criterion was then applied to limit the search scope. The purpose was to evaluate all selected studies that either facilitated or directly addressed at least one research question in the field. Additionally, the analysis focused on articles and review articles, and only English articles were included. Based on these criteria, the list of studies was reduced to 143 for Scopus, 38 for Science Direct, and 18 for Google Scholar. This initial list was further analyzed and filtered by examining the titles, keywords, and abstracts to remove irrelevant papers. The systematic literature review (SLR) was conducted on research published between 2015 and

Table 2 Inclusion and exclusion criteria.

Inclusion criteria	Exclusion criteria
English papers	Any other language papers
All paper discussing MH in FS.	Paper that did not have any link with the research question
Papers that can answer at least one research question	Papers with the same author and the same MH techniques because this information tended to be duplicated
Papers (≥ 3 pages)	Short papers (<3 pages)
Article, and review papers	Conference, books, and chapter in book

2022. The final list consisted of 91 studies for Scopus, 15 for Science Direct, and five for Google Scholar. Furthermore, any duplicate papers from Science Direct and Google Scholar were removed, resulting in a final list of 91, 12, and five studies for all databases. In addition, there are 15 selected studies were chosen out of limitation. [Table 2](#) presents the details of the inclusion and exclusion criteria.

Quality assessment criteria

The selected studies were assessed with great care and seriousness to maintain a high-quality standard. This assessment included evaluating the novelty of the proposed techniques and the ability of the studies to address at least one research question. Furthermore, special consideration was given to choosing high-quality studies from high-impact journals available in the Scopus, Science Direct, and Google Scholar digital libraries. These measures were taken to ensure a thorough quality check and maintain the overall standard of the selected studies.

Data extraction

During the data extraction phase, the selected studies were utilized to gather crucial information necessary to tackle the research questions. The extracted information encompassed various aspects such as author details, publication year, applied meta-heuristic (MH) techniques, utilized datasets, employed classifiers, performance measures used for evaluation, and the obtained results. This collected information was then organized and analyzed to facilitate further data synthesis.

Data synthesis

The data synthesis phase in the SLR involves summarizing and interpreting the collected information from the selected studies. This phase aims to address the research questions through analysis, discussion, and various forms of representation such as tables, graphs, charts, *etc.* The SLR processes are typically executed multiple times to ensure an effective review process that yields the most relevant and suitable studies. The process of this particular SLR began with the identification of research questions, followed by the application of relevant search terms on the Scopus, Science Direct, and Google Scholar digital libraries. This initial search yielded 143, 38, and 18 research papers, respectively. After analyzing and filtering the studies based on inclusion and exclusion criteria, as well as

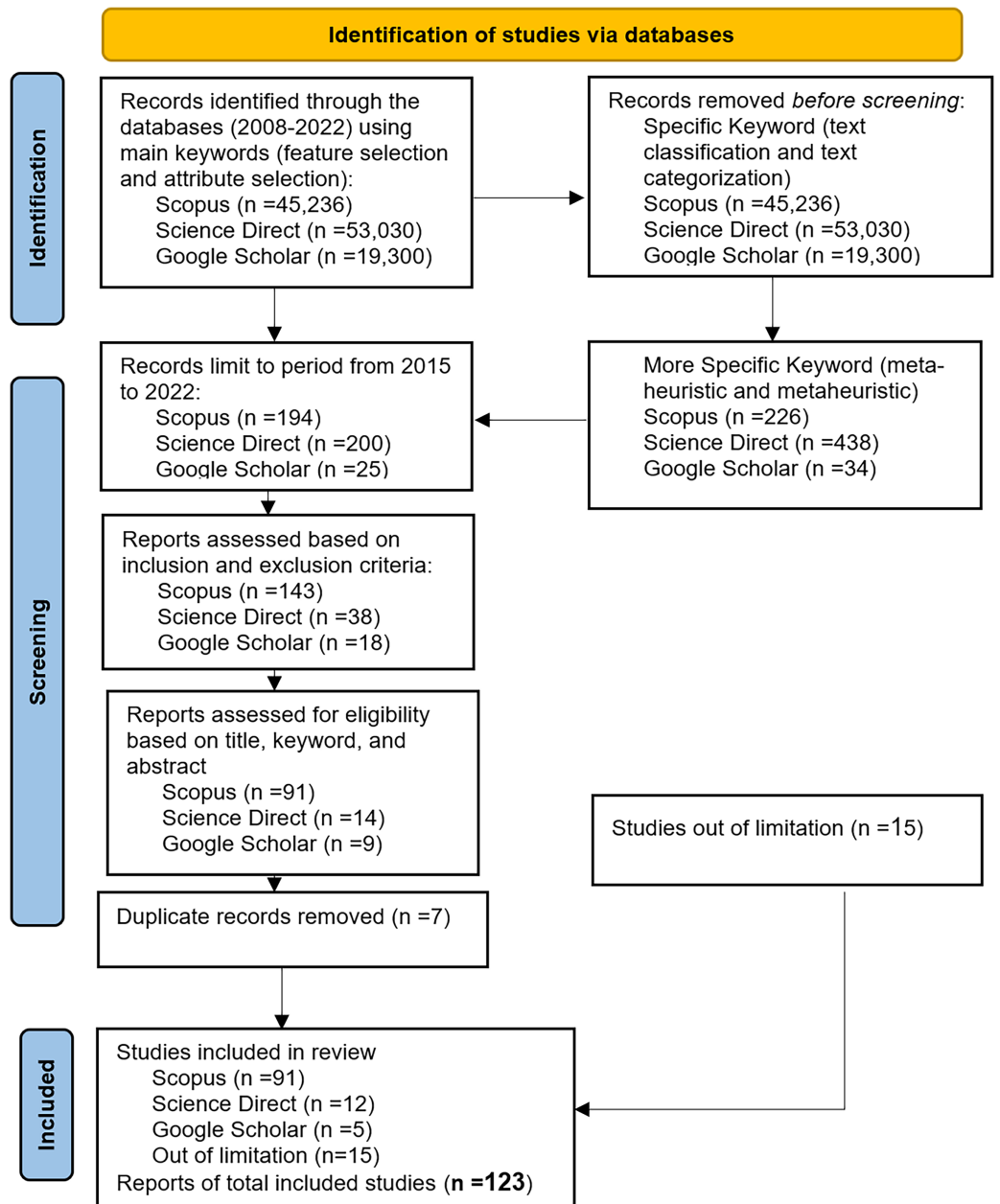


Figure 2 PRISMA flowchart of the review protocol.

Full-size DOI: 10.7717/peerj-cs.2084/fig-2

removing any duplicated papers, the final number of obtained studies was 91, 12, and five, a total of 108 articles. [Figure 2](#) illustrates the search steps protocol using the PRISMA flowchart, highlighting the progression from the initial search to the final selection of studies. After thoroughly searching through the selected studies, a total of 123 studies were found to be valuable for this SLR, as they exhibited a high level of relevance in addressing the chosen research domain.

BIBLIOMETRIC ANALYSIS BY CO-OCCURRENCE (AUTHORS KEYWORDS)

The keywords used by the authors of the article which occurred one time or more in the Scopus database were joined in the final analysis. The number of keywords used by the authors in the analysis was 417. The top three keywords that appeared were “feature selection”, “classification” and “ant colony optimization” with the following total number of occurrences 96, 28, and 14 including total strength 419, 152, and 54, respectively. Therefore, the network visualization demonstrates the three top keywords with higher weights of items, larger labels, and circles as presented in Fig. 3B. The keywords were presented in the map into 37 clusters with 788 of total length strength as shown in Fig. 3A. This finding can represent the wide applications of these keywords. On the other hand, the total occurrence of “text classification” and “metaheuristic” were three and two with total length strengths of 10 and 8, respectively.

Overly visualization map Fig. 3B determined the colors of the items, where blue (lowest score), green (middle score), and yellow (highest score) are the standard colors. The map shows that “feature selection” and “classification” were mostly used in the period between 2019 to 2021 however “ant colony optimization” was highly used between 2018 and 2019. In addition, “metaheuristic” was applied in 2021 more than any year before, and “text classification” was recently applied especially in 2020 as present in the map’s color and networks Fig. 3B.

RESULTS AND DISCUSSION

The primary contribution of this SLR is the systematic analysis of MH techniques in FS from 2015 to 2022, based on 108 primary studies from databases such as Scopus, Science Direct, and Google Scholar. The review highlights the efficiency of MH techniques compared to traditional ones and suggests the potential for further exploration of techniques like the Ringed Seal Search (RSS) to enhance FS in various applications. This section focuses on the results and findings of the review. Firstly, it provides a brief overview of the chosen studies. Subsequently, a separate subsection delves into a detailed discussion of the findings that address the research questions.

Overview of the selected study

After conducting a scan and filtering process on studies published between 2015 and 2022, a total of 108 relevant studies were initially obtained. However, 15 studies were out of limitations and were included. Therefore, the final number of studies included in this review amounted to 123. Figure 4 visually represents the distribution of these studies over the period from 2015 to 2022, indicating that the highest number of studies was observed in 2019, while the lowest number of studies occurred in 2017.

The quality of the selected studies was assessed by considering the quality and impact factors of the journals in which they were published. This ensured that high-quality studies were included. Table 3 presents a comprehensive list of the journals that published the selected studies, along with the corresponding number of studies in each journal. It also provides information on the Quartile ranking of these journals in the International

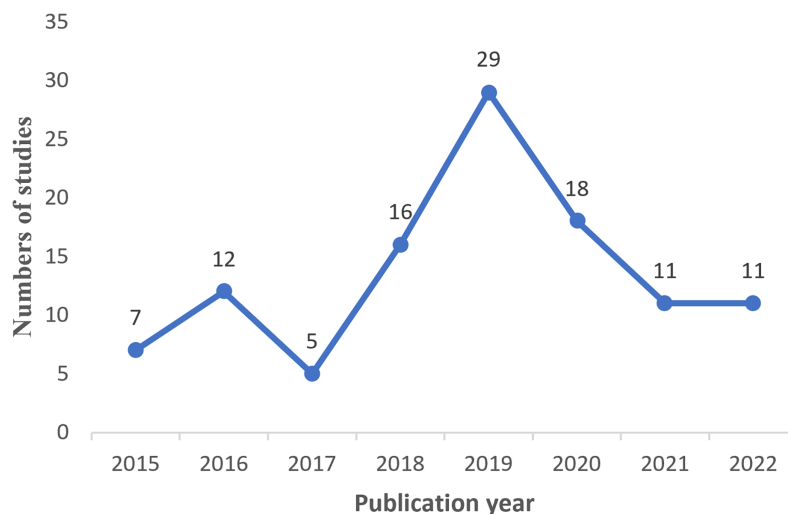


Figure 4 Studies distribution by year of publication. Full-size DOI: 10.7717/peerj-cs.2084/fig-4

Table 3 Summary of the journals of selected studies.

Publication journals	# of studies	Q# in ISI (2021)	Impact factor in ISI (2021)	Q# in Scopus indexed (2021)	SJR in Scopus (2021)
1 Artificial intelligence review	1	Q1	9.588	Q1	2.49
2 Expert systems with applications	9	Q1	8.665	Q1	1.87
3 Pattern recognition	3	Q1	8.518	Q1	2.09
4 Applied soft computing	8	Q1	8.263	Q1	1.88
5 Information sciences	2	Q1	8.233	Q1	2.29
6 Knowledge-based systems	10	Q1	8.139	Q1	2.07
7 Engineering applications of artificial intelligence	2	Q1	7.802	Q1	1.73
8 European journal of operational research	2	Q1	6.363	Q1	2.37
9 Computer networks	1	Q1	5.493	Q1	1.36
10 Neurocomputing	5	Q2	5.779	Q1	1.48
11 Neural computing and applications	6	Q2	5.102	Q1	1.17
12 Applied intelligence	5	Q2	5.019	Q2	1.15
13 International journal of approximate reasoning	1	Q2	4.452	Q1	0.98
14 International journal of machine learning and cybernetics	1	Q2	4.377	Q1	1
15 ACM transactions on knowledge discovery from data	1	Q2	4.157	Q1	1.27
16 PLoS ONE	2	Q2	3.752	Q1	0.89
17 Soft computing	4	Q2	3.732	Q2	0.82
18 Journal of ambient intelligence and humanized computing	2	Q2	3.662	Q1	0.91
19 IEEE access	8	Q2	3.476	Q1	0.93
20 Arabian journal for science and engineering	1	Q2	2.807	Q2	0.48
21 Connection science	1	Q2	0.641	Q2	0.85

(Continued)

Table 3 (continued)

Publication journals	# of studies	Q# in ISI (2021)	Impact factor in ISI (2021)	Q# in Scopus indexed (2021)	SJR in Scopus (2021)
22 Medical & biological engineering & computing	1	Q3	3.079	Q2	0.65
23 Applied artificial intelligence	2	Q3	2.777	Q3	0.49
24 Multimedia tools and applications	1	Q3	2.577	Q1	0.72
25 Knowledge and information systems	2	Q3	2.531	Q2	0.77
26 Cluster computing	2	Q3	2.303	Q2	0.62
27 Geotechnical and geological engineering	1	Q3	0.45	Q1	0.51
28 Pertanika journal of science and technology	1	Q3	0.13	Q3	0.19
29 Intelligent data analysis	2	Q4	1.321	Q3	0.38
30 International arab journal of information technology	1	Q4	0.967	Q3	0.31
31 IEICE TRANSACTIONS on information and systems	1	Q4	0.695	Q3	0.28
32 Journal of medical imaging and health informatics	1	Q4	0.659	Q4	0.19
33 Journal of king saud university—computer and information sciences	6	–	–	Q1	0.92
34 International journal of network security	1	–	–	Q2	0.336
35 Informatics in medicine unlocked	1	–	–	Q2	0.79
36 IISE transactions on healthcare systems engineering	1	–	–	Q2	0.4
37 International journal of intelligent systems and applications	1	–	–	Q3	0.241
38 Electronic notes in discrete mathematics	1	–	–	Q4	0.11
39 International journal of circuits, systems and signal processing	1	–	–	Q4	0.156
40 Journal of telecommunication, electronic and computer engineering	1	–	–	Q4	0.152
41 Future internet	1	–	–	Q2	0.77

Scientific Indexing (ISI) and their impact factors. Additionally, the Quartile ranking in Scopus is indexed with their Scientific Journal Ranking (SJR). It can be concluded that the ISI journals accounted for 85.19% of the total number of selected studies, further confirming the overall quality of the included literature.

The selected studies were assigned unique identifiers (IDs) for easy reference and consistency throughout the subsequent subsections. Table 4 presents a comprehensive list of all the studies included in the review, along with their corresponding IDs and references. Furthermore, Table 5 presents a summary of the selected studies' IDs along with the research questions they have addressed. The analysis reveals that all the selected studies have contributed to answering RQ1. For RQ2, 28 studies have provided relevant insights. Similarly, 26 studies have addressed RQ3, while 40 studies have tackled RQ4. It is worth mentioning that one paper, which was outside the predefined limitations, addressed RQ5 by utilizing RSS to enhance SVM in text classification. More detailed information on these findings can be found in the subsequent subsections.

Table 4 Selected primary studies along with their IDs and references.

ID	Paper author	Ref.	ID	Paper author	Ref.	ID	Paper author	Ref.
RP1	Tabakhi (2015a)	<i>Tabakhi & Moradi (2015)</i>	RP2	Wang (2015)	<i>Wang et al. (2015)</i>	RP3	Tabakhi (2015b)	<i>Tabakhi et al. (2015)</i>
RP4	Moshki (2015)	<i>Moshki, Kabiri & Mohebalhojeh (2015)</i>	RP5	Moradi (2015)	<i>Moradi & Rostami (2015)</i>	RP6	Inbarani (2015)	<i>Inbarani, Bagyamathi & Azar (2015)</i>
RP7	Kashef (2015)	<i>Kashef & Nezamabadi-pour (2015)</i>	RP8	Zorarpacı (2016)	<i>Zorarpacı & Özel (2016)</i>	RP9	Zawbaa (2016)	<i>Zawbaa, Emary & Grosan (2016)</i>
RP10	Zarshenas (2016)	<i>Zarshenas & Suzuki (2016)</i>	RP11	Bertolazzi (2016)	<i>Bertolazzi et al. (2016)</i>	RP12	Dadaneh (2016)	<i>Dadaneh, Markid & Zakerolhosseini (2016)</i>
RP13	Das (2016)	<i>Das, Mishra & Shaw (2016)</i>	RP14	Yong (2016)	<i>Yong, Dun-wei & Wan-qiu (2016)</i>	RP15	Saraswathi (2016)	<i>Saraswathi & Tamilarasi (2016)</i>
RP16	Salama (2016)	<i>Salama, Abdelbar & Anwar (2016)</i>	RP17	Mojaveriyan (2016)	<i>Mojaveriyan, Ebrahimpour-komleh & Jaleleddin (2016)</i>	RP18	Emary (2016)	<i>Emary & Zawbaa (2016)</i>
RP19	Garcia-Torres (2016)	<i>García-Torres et al. (2016)</i>	RP20	Ahmad (2017)	<i>Ahmad et al. (2017)</i>	RP21	Barani (2017)	<i>Barani, Mirhosseini & Nezamabadi-pour (2017)</i>
RP22	Deniz (2017)	<i>Deniz et al. (2017)</i>	RP23	Zhang (2017)	<i>Zhang, Song & Gong (2017)</i>	RP24	Kuo (2018)	<i>Kuo et al. (2018)</i>
RP25	Costa (2018)	<i>Costa et al. (2018)</i>	RP26	Ghimatgar (2018)	<i>Ghimatgar et al. (2018)</i>	RP27	Jadhav (2018)	<i>Jadhav, He & Jenkins (2018)</i>
RP28	Javidi (2018)	<i>Javidi & Zarisfi Kermani (2018)</i>	RP29	Kiziloz (2018)	<i>Kiziloz et al. (2018)</i>	RP30	Mafarja (2018a)	<i>Mafarja & Mirjalili (2018)</i>
RP31	Mafarja (2018b)	<i>Mafarja et al. (2018)</i>	RP32	Mohanty (2018)	<i>Mohanty & Das (2018)</i>	RP33	Oztekin (2018)	<i>Oztekin et al. (2018)</i>
RP34	Rais (2018)	<i>Rais & Mehmood (2018)</i>	RP35	Sayed (2018)	<i>Sayed, Khoriba & Haggag (2018)</i>	RP36	Singh (2018)	<i>Singh & Singh (2018)</i>
RP37	Yelmen (2018)	<i>Yelmen et al. (2018)</i>	RP38	Cheruku (2018)	<i>Cheruku et al. (2018)</i>	RP39	Abd El Aziz (2018)	<i>Aziz & Hassanien (2018)</i>
RP40	Ahmadi (2019)	<i>Ahmadi et al. (2019)</i>	RP41	Jain (2019)	<i>Jain et al. (2019)</i>	RP42	Ahmad (2019)	<i>Ahmad, Bakar & Yaakub (2019)</i>
RP43	Al-Rawashdeh (2019)	<i>Al-Rawashdeh, Mamat & Hafhizah Binti Abd Rahim (2019)</i>	RP44	Sayed (2019)	<i>Sayed, Hassanien & Azar (2019)</i>	RP45	Thiyagarajan (2019)	<i>Thiyagarajan & Shanthi (2019)</i>
RP46	Ghosh (2019)	<i>Ghosh et al. (2019)</i>	RP47	Arora (2019a)	<i>Arora et al. (2019)</i>	RP48	Arora (2019b)	<i>Arora & Anand (2019)</i>
RP49	Chantar (2019)	<i>Chantar et al. (2020)</i>	RP50	Chen (2019)	<i>Chen, Zhou & Yuan (2019)</i>	RP51	Dash (2019)	<i>Dash, Dash & Rautray (2019)</i>
RP52	Ghosh (2019)	<i>Ghosh et al. (2020)</i>	RP53	Han (2019)	<i>Han, Zhou & Zhou (2019)</i>	RP54	Mafarja (2019a)	<i>Mafarja & Mirjalili (2019)</i>
RP55	Tubishat (2019)	<i>Tubishat et al. (2019)</i>	RP56	Hichem (2019)	<i>Hichem et al. (2019)</i>	RP57	Ibrahim (2019)	<i>Ibrahim et al. (2019)</i>
RP58	Kumar (2019)	<i>Kumar & Jaiswal (2019)</i>	RP59	Liang (2019)	<i>Liang, Wang & Liu (2019)</i>	RP60	Mafarja (2019b)	<i>Mafarja et al. (2019)</i>

(Continued)

Table 4 (continued)

ID	Paper author	Ref.	ID	Paper author	Ref.	ID	Paper author	Ref.
RP61	Manbari (2019)	<i>Manbari, AkhlaghianTab & Salavati (2019)</i>	RP62	Krishnan (2019)	<i>Krishnan & Sowmya Kamath (2019)</i>	RP63	Selvarajan (2019)	<i>Selvarajan, Jabar & Ahmed (2019)</i>
RP64	Singh (2020)	<i>Singh & Kaur (2020)</i>	RP65	Xue (2019)	<i>Xue, Xue & Zhang (2019)</i>	RP66	Zakeri (2019)	<i>Zakeri & Hokmabadi (2019)</i>
RP67	Malar (2019)	<i>Malar, Nadarajan & Gowri Thangam (2019)</i>	RP68	Zhu (2019)	<i>Zhu et al. (2019)</i>	RP69	Hassonah (2020)	<i>Hassonah et al. (2020)</i>
RP70	Hu (2020)	<i>Hu, Pan & Chu (2020)</i>	RP71	Bhattacharyya (2020)	<i>Bhattacharyya et al. (2020)</i>	RP72	Oliva (2020)	<i>Oliva & Elaziz (2020)</i>
RP73	Too (2020)	<i>Too & Rahim Abdullah (2020)</i>	RP74	Arora (2020)	<i>Arora, Sharma & Anand (2020)</i>	RP75	Anand (2020)	<i>Anand & Arora (2020)</i>
RP76	Tawhid (2020)	<i>Tawhid & Ibrahim (2020)</i>	RP77	Anter (2020)	<i>Anter & Ali (2020)</i>	RP78	Marie-Sainte (2020)	<i>Larabi Marie-Sainte & Alalyani (2020)</i>
RP79	Gokalp (2020)	<i>Gokalp, Tasci & Ugur (2020)</i>	RP80	Ibrahim (2020)	<i>Ibrahim, Tawhid & Ward (2020)</i>	RP81	Tubishat (2020)	<i>Tubishat et al. (2020)</i>
RP82	Pan (2021)	<i>Pan et al. (2021)</i>	RP83	Mohan (2021)	<i>Mohan & Moorthi (2021)</i>	RP84	Sharaff (2021)	<i>Sharaff et al. (2021)</i>
RP85	Abualigah (2021)	<i>Abualigah & Dulaimi (2021)</i>	RP86	Ma (2021)	<i>Ma et al. (2021)</i>	RP87	Tubishat (2022)	<i>Tubishat et al. (2022)</i>
RP88	Osmani (2022)	<i>Osmani, Mohasefi & Gharehchopogh (2022)</i>	RP89	Das (2022)	<i>Das, Naik & Behera (2022)</i>	RP90	Feng (2022)	<i>Feng, Kuang & Zhang (2022)</i>
RP91	Zhao (2022)	<i>Zhao et al. (2022)</i>	RP92	Hosseinalipour (2022)	<i>Hosseinalipour & Ghanbarzadeh (2022)</i>	RP93	Pashaei (2017)	<i>Pashaei & Aydin (2017)</i>
RP94	Hammouri (2020)	<i>Hammouri et al. (2020)</i>	RP95	Souza (2020)	<i>Thom de Souza et al. (2020)</i>	RP96	Purushothaman (2020)	<i>Purushothaman, Rajagopalan & Dhandapani (2020)</i>
RP97	Agrawal (2020)	<i>Agrawal, Kaur & Sharma (2020)</i>	RP98	Sadeghian (2021)	<i>Sadeghian, Akbari & Nematzadeh (2021)</i>	RP99	Dash (2021)	<i>Dash (2021)</i>
RP100	Paul (2021)	<i>Paul et al. (2021)</i>	RP101	Wang (2022)	<i>Wang et al. (2022)</i>	RP102	Eluri (2022)	<i>Eluri & Devarakonda (2022)</i>
RP103	Allam (2022)	<i>Allam & Nandhini (2022)</i>	RP104	Liu (2022)	<i>Liu et al. (2022)</i>	RP105	Pandey (2020)	<i>Pandey, Rajpoot & Saraswat (2020)</i>
RP106	Ansari (2021)	<i>Ansari et al. (2021)</i>	RP107	Albashish (2021)	<i>Albashish et al. (2021)</i>	RP108	Al-Dyani (2022)	<i>Al-Dyani, Ahmad & Kamaruddin (2022)</i>

Table 5 RQs addressed in individual study.

Paper ID	RQ1	RQ2	RQ3	RQ4	Paper ID	RQ1	RQ2	RQ3	RQ4	Paper ID	RQ1	RQ2	RQ3	RQ4
RP1	√	-	√	√	RP38	√	-	-	-	RP75	√	-	-	√
RP2	√	√	√	√	RP39	√	-	-	√	RP76	√	-	-	-
RP3	√	-	√	√	RP40	√	-	-	√	RP77	√	-	-	√
RP4	√	-	-	√	RP41	√	√	√	-	RP78	√	√	√	-
RP5	√	-	√	√	RP42	√	√	√	√	RP79	√	√	√	-
RP6	√	-	√	√	RP43	√	√	-	√	RP80	√	-	-	-
RP7	√	-	-	√	RP44	√	-	-	√	RP81	√	-	-	√
RP8	√	-	√	√	RP45	√	√	-	√	RP82	√	-	-	-
RP9	√	-	-	√	RP46	√	-	√	-	RP83	√	√	-	-
RP10	√	-	√	-	RP47	√	-	-	√	RP84	√	√	-	-
RP11	√	-	-	-	RP48	√	-	-	√	RP85	√	-	-	-
RP12	√	-	√	√	RP49	√	√	-	√	RP86	√	-	-	-
RP13	√	-	√	√	RP50	√	-	√	√	RP87	√	√	-	-
RP14	√	-	-	-	RP51	√	-	-	√	RP88	√	√	-	-
RP15	√	√	-	-	RP52	√	-	-	-	RP89	√	-	-	-
RP16	√	-	-	-	RP53	√	-	-	√	RP90	√	-	-	-
RP17	√	√	√	-	RP54	√	-	-	√	RP91	√	-	-	-
RP18	√	-	-	-	RP55	√	√	-	√	RP92	√	√	-	-
RP19	√	√	-	-	RP56	√	-	-	-	RP93	√	√	-	√
RP20	√	√	-	-	RP57	√	-	-	√	RP94	√	-	-	-
RP21	√	-	-	-	RP58	√	√	√	-	RP95	√	-	-	-
RP22	√	-	-	-	RP59	√	-	-	-	RP96	√	√	-	-
RP23	√	-	√	√	RP60	√	-	√	-	RP97	√	√	-	-
RP24	√	-	-	-	RP61	√	-	√	√	RP98	√	-	-	-
RP25	√	-	-	-	RP62	√	-	√	-	RP99	√	-	-	-
RP26	√	-	√	√	RP63	√	-	-	-	RP100	√	√	-	-
RP27	√	-	-	-	RP64	√	√	-	-	RP101	√	-	-	-
RP28	√	-	-	√	RP65	√	-	-	-	RP102	√	-	-	-
RP29	√	-	-	√	RP66	√	-	-	√	RP103	√	-	-	-
RP30	√	-	√	-	RP67	√	-	-	-	RP104	√	-	-	-
RP31	√	-	√	-	RP68	√	-	-	√	RP105	√	-	-	-
RP32	√	-	-	-	RP69	√	√	√	-	RP106	√	-	-	-
RP33	√	-	-	-	RP70	√	-	-	√	RP107	√	-	-	-
RP34	√	-	√	-	RP71	√	-	-	-	RP108	√	√	-	-
RP35	√	-	-	-	RP72	√	-	-	√					
RP36	√	√	-	√	RP73	√	-	-	-					
RP37	√	√	-	-	RP74	√	-	-	-					

RQ1: which metaheuristic (MH) techniques have been utilized for feature selection (FS)?

In this section, the focus is on discussing and identifying the MH techniques that have been utilized for feature selection in various machine learning problems, including pattern recognition, email classification, microarray data classification, sentiment analysis, and text classification. The findings from the primary selected studies reveal that all the studies have provided insights into (RQ1). These studies have classified the MH techniques into three main groups according to their sources of inspiration. These groups are Evolutionary Algorithms (EA), Physics-Based (PB) Algorithms, and Swarm Intelligence (SI) Algorithms. This categorization provides a broad understanding of the different types of MH techniques employed for FS across various domains (*Kumar & Bawa, 2020*).

Evolutionary Algorithms (EAs) draw inspiration from the natural processes of evolution. One of the commonly used algorithms in this category is the Genetic Algorithm (GA) belongs to a class of optimization algorithms that draw inspiration from the principles of natural selection and genetics. They mimic the principles of evolution to solve complex problems by iteratively searching and refining a population of potential solutions, that has been utilized for feature selection in numerous studies. Specifically, in RP13, RP22, RP27, RP32, RP33, RP37, RP40, RP46, RP62, RP85, and RP106. Another algorithm in the EA category is Differential Evolution (DE), which is used to solve continuous optimization problems, and was utilized for feature selection in RP8 and RP55. Evolutionary Population Dynamics (EPD) combines concepts from evolutionary biology and population dynamics with computational methods, it was used in RP31, Imperialist Competitive Algorithm (ICA) draws inspiration from the socio-political behavior of imperialistic systems, it was applied in RP17 and RP88. RP95 utilized the Binary Coyote Optimization Algorithm (BCOA), which draws inspiration from the intelligent behavior exhibited by coyotes in their natural environment. The Golden Eagle Optimizer (GEO) inspired by the behavior and characteristics of golden eagles in nature, is used by RP102. Biogeography-based Optimization (BBO) is influenced by the principles of biogeography, which involves the study of the distribution of biological organisms across different geographic regions, it is used by RP107. These studies demonstrate the application of different evolutionary algorithms for feature selection in various machine-learning problems.

Secondly, is the Physics-Based (PBs) Algorithms, which aim to mimic physical rules in their search process. Several common algorithms in this category have been employed for feature selection in the selected studies. Simulated Annealing (SA) draws inspiration from the annealing process used in metallurgy, it was utilized in RP4, RP28, RP43, and RP87. The Harmony Search (HS) algorithm takes inspiration from the musical improvisation process, it was applied in RP2, RP6, RP13, RP71, and RP99. The Gravitational Search Algorithm (GSA) derives its inspiration from the fundamental law of gravity and the motion of celestial bodies, it was used in RP21, RP28, and RP68. The Teaching Learning Based Optimization (TLBO) algorithm draws inspiration from the teaching and learning processes that take place in a classroom, it was employed in RP29 and RP103. Water Cycle algorithms (WCA) were utilized in RP43. Atom Search Optimization (ASO) was applied in

RP73. The multi-Verse Optimizer (MVO) algorithm inspired by the concept of multiple universes and parallel universes in theoretical physics, was used in RP69. The Interior Search Algorithm (ISA) designed for solving constrained optimization problems, was employed in RP74. Lastly, RP93 utilizes the Black Hole Optimization (BHO) algorithm, a robust stochastic optimization technique that takes inspiration from the behavior of black holes in outer space. The key distinction between EAs and PBs lies in the mechanism of communication and movement of search agents within the search space. PB algorithms rely on physical rules to guide the search process, while EAs are inspired by evolutionary processes. This difference in approach enables PB algorithms to explore the search space using physics-inspired mechanisms.

The third group consists of Swarm intelligence (SI) algorithms, which draw inspiration from the collective behavior observed in swarms, herds, flocks, or schools of living organisms in nature. While these algorithms share similarities with evolutionary algorithms (EAs) and population-based (PB) algorithms in terms of their mechanism, SI algorithms leverage the simulated collective and social intelligence of these creatures to guide the interactions among search agents. The number of newly proposed SI algorithms exceeds those of EAs and PBs. One widely adopted SI algorithm is Ant Colony Optimization (ACO), which is utilized in several selected studies as a popular feature selection (FS) technique (RP1, RP3, RP5, RP7, RP12, RP15, RP16, RP20, RP26, RP34, RP42, RP52, RP59, RP61, RP63, and RP86). Additionally, Particle Swarm Optimization (PSO) is another notable SI algorithm employed in RP14, RP36, RP41, RP50, RP63, RP65, RP67, and RP100. It is worth emphasizing that the realm of Swarm Intelligence (SI) has witnessed a significant influx of novel algorithms in recent times, substantially broadening the array of choices available for optimization tasks. Among the more recent SI algorithms highlighted in the selected studies, notable examples include the Artificial Bee Colony (ABC) algorithm, applied in RP8, RP24, and RP88. Additionally, the Antlion Optimization (ALO) algorithm has been employed in RP9, RP18, and RP54, while the Crow Search Algorithm (CSA) has been identified in RP44, RP47, and RP77. Furthermore, the Whale Optimization Algorithm (WOA) has been utilized in RP30, RP55, RP76, and RP97. Other noteworthy SI algorithms encompass the Cuckoo Search (CS) algorithm, which has been employed in RP39, RP64, RP83, and RP105. Additionally, the Grasshopper Optimization Algorithm (GOA) has been found in RP31, RP56, RP60, RP66, and RP96, and the Grey Wolf Optimization (GWO) algorithm has been utilized in RP18, RP47, RP49, RP58, RP70, RP96, and RP101. and the Mayfly (MF) algorithm which is used in RP71 is introduced as a novel technique for addressing FS problems. This innovative method takes a hybrid approach, synergizing the strengths found in traditional optimization techniques like PSO, GA, and FA. These recent SI algorithms have substantially augmented the repertoire of options available for optimization tasks, providing researchers and practitioners with an expanded toolkit to effectively address intricate problems.

Furthermore, a diverse array of captivating swarm intelligence (SI) algorithms has emerged, significantly expanding the range of optimization techniques. These encompass the firefly algorithms (FFA) as scrutinized in RP23 and RP78, the Moth-Flame Optimizer (MFO) employed in RP18 and RP58, the Slap Swarm Optimization (SSO) explored in

RP35 and RP81, and the Brainstorm Optimization (BSO) utilized in RP59 and RP72. Additionally, the Bat Optimization Algorithm (BA) has been studied in RP38, RP90, and RP108, while the Water Wave Optimization (WWO) has been investigated in RP80. Noteworthy algorithms also include the Butterfly Optimization Algorithm (BOA) employed in RP48 and RP98, the Selfish Herd Optimizer (SHO) studied in RP75, the Social Spider Optimization (SSO) explored in RP57, and the Artificial Fish Swarm Algorithm (AFSA) examined in RP45. Moreover, the Shuffled Frog Leaping Algorithm (SFLA) has been discussed in RP51 and RP104, and the Symbiotic Organism Search (SOS) Algorithm has been utilized in RP53. Additionally, the Pigeon-inspired Optimization (PIO) algorithm has been studied in RP82, the Krill Herd Optimization (KHO) explored in RP84, and the Dandelion Algorithm (DA) utilized in RP91.

In addition, the Horse Herd Optimization Algorithm (HOA) was thoroughly investigated in RP92. RP94 employed the Dragonfly Algorithm (DA), which took inspiration from the natural behavior of dragonflies. These SI algorithms exemplify the ingenuity and diversity of drawing inspiration from various natural phenomena and collective behaviors. Each algorithm derives inspiration from distinct aspects of nature or collective behavior, with the shared goal of providing effective optimization solutions for diverse problem domains. By simulating the behavior of fireflies, moths, slaps, bats, water waves, butterflies, selfish herds, social spiders, artificial fish swarms, shuffled frogs, symbiotic organisms, pigeons, krill herds, dandelions, and horse herds, these algorithms strive to offer efficient optimization solutions for a wide spectrum of problem domains. The continuous advancement and exploration of such algorithms contribute to the ever-evolving field of optimization, offering promising avenues for addressing intricate optimization challenges across various domains.

Several metaheuristic (MH) techniques find inspiration from various mathematical theories and concepts. For instance, the Chaotic Optimization Algorithm (COA), applied in RP9, RP18, and RP44, draws influence from Chaos Theory. RP4 and RP11 utilize the Greedy Randomized Adaptive Search Procedure (GRASP), a metaheuristic algorithm tailored for solving combinatorial problems. GRASP involves the construction and local search phases in each iteration. The Binary Coordinate Ascent (BCA) algorithm, employed in RP10, takes inspiration from the well-known coordinate descent algorithm. Additionally, the Variable Neighborhood Search (VNS) technique (RP19, RP25) tackles global optimization and combinatorial optimization problems by modifying the neighborhood of the current solution during the search process in a systematic manner. RP79 employs the Iterated Greedy (IG) technique, which addresses challenging combinatorial optimization problems through two phases: destruction and construction. The Sine Cosine Algorithm (SCA), utilized in RP85 and RP87, emulates the behavior of sine and cosine functions to uncover optimal solutions for optimization problems. RP89 utilizes the Jaya Optimization Algorithm (JOA), which draws inspiration from the Sanskrit concept of “Jaya,” signifying success or victory. This algorithm iteratively improves a population of solutions to discover an optimal or nearly optimal solution.

[Table 6](#) presents a concise overview of the metaheuristic (MH) techniques utilized in the selected studies to address Research Question 1, focusing on Feature Selection (FS). These

Table 6 Distribution of studies across MH techniques.

#	MH techniques	# of study	Studies ID	#	MH techniques	# of study	Studies ID
1	ACO	15	RP1, RP3, RP5, RP7, RP12, RP15, RP16, RP20, RP26, RP34, RP42, RP52, RP59, RP61, RP63	26	EPD	1	RP31
2	GA	9	RP13, RP22, RP27, RP32, RP33, RP37, RP40, RP46, RP62, RP106	27	WCFS	1	RP43
3	PSO	8	RP14, RP36, RP41, RP50, RP63, RP65, RP67, RP100	28	AFSA	1	RP45
4	GWO	5	RP18, RP47, RP49, RP58, RP70, RP96	29	BOA	1	RP48
5	GOA	4	RP31, RP56, RP60, RP66, RP96	30	SFLA	1	RP51, RP104
6	COA	3	RP9, RP18, RP44	31	SOS	1	RP53
7	SA	3	RP4, RP28, RP43	32	SSO	1	RP57
8	ALO	3	RP9, RP18, RP54	33	BCA	1	RP10
9	CSA	3	RP44, RP47, RP77	34	ASO	1	RP73
10	WOA	3	RP30, RP55, RP76, RP97	35	ISA	1	RP74
11	HS	3	RP2, RP6, RP13, RP71, RP99	36	SHO	1	RP75
12	GSA	3	RP21, RP28, RP68	37	IG	1	RP79
13	CS	2	RP39, RP64, RP83, RP105	38	WVO	1	RP80
14	FFA	2	RP23, RP78	39	MVO	1	RP69
15	ABC	2	RP8, RP24, RP88	40	MF	1	RP71
16	VNS	2	RP19, RP25	41	PIO	1	RP82
17	SSO	2	RP35, RP81	42	KHO	1	RP84
18	MFO	2	RP18, RP58	43	JOA	1	RP89
19	GRASP	2	RP4, RP11	44	DA	1	RP91
20	DE	2	RP8, RP55	45	HOA	1	RP92
21	SCA	2	RP85, RP87	46	BHO	1	RP93
22	BA	1	RP38, RP90, RP108	47	DA	1	RP94
23	BSO	2	RP59, RP72	48	BCOA	1	RP95
24	ICA	1	RP17, RP88	49	GEO	1	RP102
25	TLBO	1	RP29, RP103	50	BBO	1	RP107

techniques have been utilized in the selected studies to tackle FS challenges and provide solutions in the context of Research Question 1.

Statistical analysis

The statistical analysis plan has been developed to assess the significance of differences between MH models in Table 6, which involves calculating summary statistics such as odds ratios (OR) for meta-analysis. OR for the data provided in Table 6 were calculated by constructing a 2×2 contingency table for each MH technique compared to the reference group (ACO). The ACO MH technique was designated as the reference group because it was the most frequently used technique. Then, the odds of studies using each technique were compared to the odds of studies using the reference technique (ACO). The odds ratio

Table 7 Odds ratios (OR) meta-analysis for MH-FS in text classification.

MH technique	Odds ratio (vs. ACO)	Interpretation
GA	9/15 = 0.6	The odds of studies using GA compared to ACO are 0.6 times as likely.
PSO	8/15 = 0.533	The odds of studies using PSO compared to ACO are 0.533 times as likely.
GWO	5/15 = 0.333	The odds of studies using GWO compared to ACO are 0.333 times as likely.
GOA	4/15 = 0.267	The odds of studies using GOA compared to ACO are 0.267 times as likely.
COA	3/15 = 0.2	The odds of studies using COA compared to ACO are 0.2 times as likely.
SA	3/15 = 0.2	The odds of studies using SA compared to ACO are 0.2 times as likely.
ALO	3/15 = 0.2	The odds of studies using ALO compared to ACO are 0.2 times as likely.
CSA	3/15 = 0.2	The odds of studies using CSA compared to ACO are 0.2 times as likely.
WOA	3/15 = 0.2	The odds of studies using WOA compared to ACO are 0.2 times as likely.
HS	3/15 = 0.2	The odds of studies using HS compared to ACO are 0.2 times as likely.
GSA	3/15 = 0.2	The odds of studies using GSA compared to ACO are 0.2 times as likely.
CS	2/15 = 0.133	The odds of studies using CS compared to ACO are 0.133 times as likely.
FFA	2/15 = 0.133	The odds of studies using FFA compared to ACO are 0.133 times as likely.
ABC	2/15 = 0.133	The odds of studies using ABC compared to ACO are 0.133 times as likely.
VNS	2/15 = 0.133	The odds of studies using VNS compared to ACO are 0.133 times as likely.
SSO	2/15 = 0.133	The odds of studies using SSO compared to ACO are 0.133 times as likely.
MFO	2/15 = 0.133	The odds of studies using MFO compared to ACO are 0.133 times as likely.
GRASP	2/15 = 0.133	The odds of studies using GRASP compared to ACO are 0.133 times as likely.
DE	2/15 = 0.133	The odds of studies using DE compared to ACO are 0.133 times as likely.
SCA	2/15 = 0.133	The odds of studies using SCA compared to ACO are 0.133 times as likely.
BA	1/15 = 0.067	The odds of studies using BA compared to ACO are 0.067 times as likely.
BSO	2/15 = 0.133	The odds of studies using BSO compared to ACO are 0.133 times as likely.
WCFS	1/15 = 0.067	The odds of studies using WCFS compared to ACO are 0.067 times as likely.
AFSA	1/15 = 0.067	The odds of studies using AFSA compared to ACO are 0.067 times as likely.
BOA	1/15 = 0.067	The odds of studies using BOA compared to ACO are 0.067 times as likely.
SFLA	1/15 = 0.067	The odds of studies using SFLA compared to ACO are 0.067 times as likely.
SOS	1/15 = 0.067	The odds of studies using SOS compared to ACO are 0.067 times as likely.
GE	1/15 = 0.067	The odds of studies using GE compared to ACO are 0.067 times as likely.
PIO	1/15 = 0.067	The odds of studies using PIO compared to ACO are 0.067 times as likely.
KHO	1/15 = 0.067	The odds of studies using KHO compared to ACO are 0.067 times as likely.
JOA	1/15 = 0.067	The odds of studies using JOA compared to ACO are 0.067 times as likely.
DA	1/15 = 0.067	The odds of studies using DA compared to ACO are 0.067 times as likely.
HOA	1/15 = 0.067	The odds of studies using HOA compared to ACO are 0.067 times as likely.

was calculated by dividing the number of studies using each technique by the number of studies using ACO as shown in Eq. (1). Then, the results were organized into Table 7.

$$\text{Odds ratio (each technique vs. ACO)} = \frac{\text{Odds of studies with each technique}}{\text{Odds of studies with ACO}} \quad (1)$$

Table 8 Datasets used for MH-FS in text classification.

Text classification types	Datasets	Referred studies
Sentiment analysis text classification	IMDb movie and Initial tour medical blogs	RP15
	Nikon, Nokia, Apex, Canon, and Creative from Amazon	RP20, RP42
	OCA, Twitter, Political, and Software which is Arabic sentiment analysis datasets	RP55
	Two Twitter benchmark corpora (SemEval 2016 and SemEval 2017)	RP58
	Nine datasets that belong to four different contexts from Twitter social network	RP69
	The Product Opinion Dataset from Amazon	RP71
	Nine public sentiment analysis datasets (doctor, lawyer, drug, laptop, camera, radio, music, camp, and TV). four Amazon review datasets are DVD, electronics, books, and kitchen	RP79
	The SemEval-2014	RP83
	Two datasets were gathered from Amazon1 reviews (Electronic and Movie), Sixteen UCI datasets (Iris, Heart, Hepatitis, Lung Cancer, Yelp, Lymph, Pima, Cancer, Diabetes, Heart-Stalog, Dermatology, Thyroid, Sonar, Gene, IMDB, and Amazon,), and three Twitter datasets (SOMD, STS-Test, and Sanders).	RP88
	Spam text classification	PU2, PU3, Lingspam, CSDMC2010, Trec2007, and Enron-spam.
WEBSpAM UK-2006		RP36
Spam-Base dataset and Enron spam email <i>corpus</i>		RP43, RP92
Public shared <i>corpus</i>		RP84
Text classification	Retures-21578 from UCI Repository	RP17, RP96
	Alt, Structure, Disease, Function, Subcell, Acq, Money-fx, Corn, Earn, Ship, Grain, and Crude	RP19
	Three open-source web applications (qaManager, bitWeaver, and WebCalendar) Two play store web application (Dineout: Table, Reserve, and Wynk Music)	RP41
	OHSUMED	RP45
	Twitter, ASKfm, and Formspring	RP64
	Three public Arabic datasets, namely Akhbar-Alkhaleej, Alwatan, and Al-jazeera-News	RP49
	OSAC which is collected from BBC and CNN Arabic websites	RP78
	Turkish tweets obtained from three various GSM operators	RP37
	Three Hadiths datasets	RP86
	Chess and Email word subject	RP93
	20Newsgroups from UCI Repository	RP96, RP108
	TR11WC and TR23WC	RP97
	Society, Science, Reference, Recreation, Health, Entertainment, Enron, Education, Computer, Business, and Arts.	RP100
	News Aggregator, News articles, RSS news feed, and Facebook news posts.	RP108

RQ2: in the context of text classification, which specific MH techniques have been applied for FS?

In this section, the aim is to identify the metaheuristic-feature selection (MH-FS) techniques utilized in the selected studies for text classification. Additionally, we provide an overview of the commonly used datasets, classifiers, and performance evaluation metrics in the context of MH-FS for text classification machine learning techniques. As depicted in Table 5, a total of 28 studies (RP2, RP15, RP17, RP19, RP20, RP36, RP37, RP41,

RP42, RP43, RP45, RP49, RP55, RP58, RP64, RP69, RP78, RP79, RP83, RP84, RP87, RP88, RP92, RP93, RP96, RP97, RP100, and RP108) have employed MH techniques for FS in text classification. These studies offer insights into the application of MH-FS techniques, highlighting the datasets commonly used, the classifiers employed, and the frequently utilized performance evaluation metrics in the domain of text classification machine learning.

RQ2.1: what are the datasets employed in the application of MH-FS for text classification?

Different datasets have been employed for MH-FS in text classification, depending on the specific classification task, such as sentiment analysis, spam classification, or general text classification. Additionally, the choice of datasets has been influenced by the language used, including English, Arabic, and Turkish. In [Table 8](#), ten selected studies (RP15, RP20, RP42, RP55, RP58, RP69, RP79, RP83, and RP88) utilized sentiment analysis for text classification. All of these studies used the English language, except for RP55, which employed Arabic. Similarly, five studies (RP2, RP36, RP43, RP84, and RP92) focused on spam text classification. Among the studies that performed English text classification, RP17, RP19, RP41, RP45, RP64, RP49, RP78, RP37, RP93, RP96, RP97, RP100, and RP108 were identified, while RP49 and RP78 also incorporated Arabic text classification, and RP37 included Turkish text classification.

RQ2.2: which classifiers have been used with MH-FS in text classification?

The text classification process encompasses three primary stages: text preprocessing, feature selection (FS), and constructing a text classification model with a machine learning classifier to evaluate the performance of different FS techniques. The selected studies have utilized different classifiers for FS-MH in text classification. [Table 9](#) presents eight classifiers along with their definitions and the studies where they were applied for text classification. These classifiers include support vector machine (SVM), naïve Bayesian (NB), k-nearest neighbor (KNN), decision tree (DT), multilayer perceptron (MLP), artificial neural networks (ANN), centroid based algorithm (CBA), and AdaBoost. As shown in [Table 9](#), SVM, NB, and KNN are the most commonly used classifiers in text classification, followed by DT and MLP, with AdaBoost being used to a lesser extent. The least utilized classifiers are ANN and CBA. [Figure 5](#) provides a visual representation of the number of studies employing different classifiers in text classification.

RQ2.3: which performance evaluation metrics are commonly utilized to assess the effectiveness of MH-FS in text classification?

Among the studies that have been reviewed in the field of MH-FS for text classification, a range of evaluation metrics have been utilized. These metrics serve to evaluate and compare the performance of diverse models developed through a variety of machine learning and statistical methods. The metrics in question include Precision, Recall, F-measure, Accuracy, AUC (area under the curve), the number of selected features, and

Table 9 Classifiers used for MH-FS in text classification.

Classifiers	Definitions	Referred studies
SVM	SVM is the most successful supervised machine learning algorithm used for either classification or regression problems to determine the decision boundary between two classes to the maximum extent away from a point in the training dataset. By applying the kernel approach to transform the data, SVM may carry out either linear classification or non-linear classification and according to these transformations, it can determine an optimal boundary between the possible outputs (<i>Saraswathi & Tamlaras, 2016; Alshalif et al., 2023</i>).	RP2, RP15, RP19 RP36, RP37, RP41, RP43, RP45, RP49, RP55, RP58, RP64, RP69, RP78, RP84, RP87, RP92, RP97
NB	NB is among the popular practical supervised machine learning algorithms that is universally utilized especially for text classification and medical diagnosis because it is capable of scaling features of the large dimensions of space. NB is a simple probabilistic model according to the Bayes theorem. it depends on the hypothesis that attribute values are conditionally independent by looking at the targeted labels (<i>Chantar et al., 2020; Kumar & Jaiswal, 2019; Thiyagarajan & Shanthi, 2019</i>).	RP2, RP15, RP19 RP36, RP41, RP43, RP45, RP49, RP55, RP58, RP64, RP79, RP84, RP92.
KNN	KNN is the most basic and easiest supervised machine learning algorithm and is widely used in the text classification model. In the training data, it collects new data according to the shortest distance between k neighbors. The Euclidean distance formula is used to determine this distance. The basic concept of this classifier is that an object is classified based on the votes of the majority of its neighbors (<i>Ahmad, Bakar & Yaakub, 2019; Chantar et al., 2020; Kumar & Jaiswal, 2019</i>).	RP17, RP41, RP42, RP43, RP49, RP55, RP58, RP87, RP92, RP97
DT	DT is a supervised machine-learning technique that resembles a tree and builds the classification tree using a set of training instances and it includes branches, root, and leaf nodes. Generally, the most widely used decision tree algorithm is C4.5, which is an improvement over the decision tree technique from the previous version (<i>Chantar et al., 2020; Kumar & Jaiswal, 2019</i>).	RP36, RP41, RP49, RP58, RP93, RP97
MLP	MLP is a supervised machine learning algorithm it is one type of neural network. it includes three main layers which are input, hidden, and output layers. MLP is a self-adaptive and data-driven technique that can arrange them according to the data without explicitly defining a distributional or appropriate format for the underlying model (<i>Kumar & Jaiswal, 2019</i>).	RP36, RP41, RP58, RP92.
AdaBoost	AdaBoost is an appropriate algorithm for building a strong classifier from a combination of weak classifiers. it is considered to be suitable for real-time applications. Another advantage of AdaBoost it uses fewer features and less memory (<i>Thiyagarajan & Shanthi, 2019</i>).	RP36, RP45, RP83.
ANN	ANN is the organization and function of this model inspired by the biological neural networks of the human brain. it is a collection of interconnected processing units named neurons or nodes. It consists of five main elements which are inputs, weight, bias, activation function, and output. Each input is multiplied by weight to create the weighted inputs. the bias along with all weighted inputs are then added. Then, on the output neuron, an activation function is applying to the summary of prior weighted inputs and bias (<i>Dwivedi, 2018; Alshalif, Ibrahim & Waheeb, 2017; Alshalif, Ibrahim & Herawan, 2017</i>).	RP37, RP87.
CBA	CBA has been used to solve text classification issues. In this approach, the vector-space model is used to represent the documents. In the term space, each document is therefore viewed as a vector (<i>Ferrandin et al., 2015</i>).	RP37

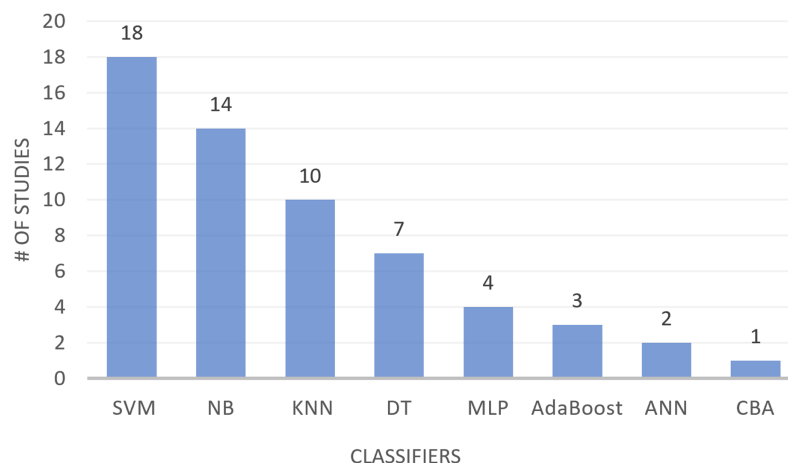


Figure 5 Studies using different classifiers in text classification.

Full-size DOI: 10.7717/peerj-cs.2084/fig-5

Table 10 Evaluation metrics used for MH-FS in text classification.

Evaluation metric	Definition	Referred studies
Precision	Precision is the positive predictive value and it is the ratio of documents to documents that are correctly categorized. By dividing the total number of true positives by the sum of true positives and false positives, it is calculated (<i>Mojaveriyan, Ebrahimpour-komleh & Jalaleddin, 2016; Gokalp, Tasci & Ugur, 2020</i>).	RP2, RP15, RP17, RP20, RP36, RP42, RP43, RP45, RP49, RP64, RP78, RP79, RP83, RP84, RP88, RP92, RP93, RP96
Recall	Recall, often referred to as sensitivity, measures how well a model can identify all relevant instances within a dataset. It is measured as the ratio of true positives to both true positives and false negatives (<i>Mojaveriyan, Ebrahimpour-komleh & Jalaleddin, 2016; Gokalp, Tasci & Ugur, 2020</i>).	RP2, RP15, RP17, RP20, RP36, RP42, RP43, RP45, RP49, RP64, RP78, RP79, RP83, RP84, RP88, RP92, RP93, RP96
F-measure	The weighted harmonic means of recall and precision, often known as the F-measure or F-score, is a metric for assessing the correctness of a test (<i>Singh & Kaur, 2020; Gokalp, Tasci & Ugur, 2020; Singh & Singh, 2018</i>).	RP2, RP15, RP17, RP20, RP36, RP42, RP43, RP45, RP49, RP64, RP78, RP79, RP83, RP84, RP88, RP96
Accuracy	The accuracy rate (ACC) is the most general evaluation measure used in practice; it is used to evaluate classifier effectiveness according to the percentage of its correct predictions. Generally, it is determined by dividing the total number of true positives and true negatives by the total number of true positives, true negatives, false negatives, and false positives. A data point that the algorithm correctly identified as true or untrue is referred to as a true positive or true negative. Moreover, a data point that the algorithm misclassified as a false positive or false negative (<i>Tubishat et al., 2019; Al-Rawashdeh, Mamat & Hafhizah Binti Abd Rahim, 2019</i>).	RP15, RP19, RP37, RP41, RP43, RP55, RP58, RP69, RP79, RP84, RP87, RP88, RP92, RP93, RP96, RP97
Number of selected features.	It is a parameter examined to measure the performance of the proposed FS technique. Low number of selected features indicated a better FS technique (<i>Senan et al., 2012</i>).	RP19, RP45, RP58, RP84.
AUC	AUC stands for Area under the ROC Curve, however, ROC stands for Receiver Operator Characteristic which is a probability curve that plots the True Positive Rate against False Positive Rate at different threshold values, while AUC determines degree or measure of separability it shows the capability of the model to distinguish between the positive and negative classes. Higher AUC shows a better performance of the model (<i>Narkhede, 2019</i>).	RP36, RP64, RP79, RP93, RP97
Stability	The stability of a classification algorithm is the degree to which the same procedure may produce repeatable results when different batches of data are specified (<i>Turney, 1995</i>).	RP19.

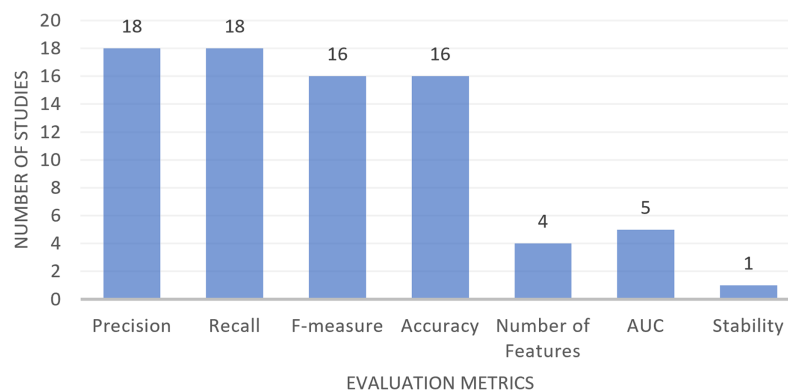


Figure 6 Studies using different evaluation metrics in text classification.

Full-size DOI: [10.7717/peerj-cs.2084/fig-6](https://doi.org/10.7717/peerj-cs.2084/fig-6)

Stability. Table 10, detailed the descriptions and definitions of these evaluation metrics, along with information on which studies made use of them. Figure 6 has been prepared to offer a visual overview of the usage frequency of these metrics. As shown in Fig. 6, Precision and Recall are the most commonly employed metrics, with F-measure and Accuracy following closely behind. AUC and the count of selected features are also frequently employed for evaluation purposes. In contrast, Stability is a less commonly utilized metric in the analyzed studies.

RQ3: is there empirical evidence indicating that MH-FS techniques outperform traditional FS methods in the domain of text classification?

According to the selected studies, the comparison was done based on two types of studies which are: (1) studies compared MH-FS and traditional FS techniques and (2) studies compared MH-FS techniques with existing MH-FS techniques. In this section, the focus is on comparing MH with the traditional one. The MH-FS techniques perform better than the traditional FS. In this review, 26 articles extracted from the chosen studies (RP1, RP2, RP3, RP5, RP6, RP8, RP10, RP12, RP13, RP17, RP23, RP26, RP30, RP31, RP34, RP41, RP42, RP46, RP50, RP58, RP60, RP61, RP62, RP69, RP78, and RP79) undertake comparisons of their methodologies with various traditional FS techniques such as information gain, ReliefF, Laplacian score (L-score), Fisher score (F-score), relevance-redundancy feature selection (RRFS), random subspace method (RSM), Chi-square, mutual Information, symmetrical-uncertainty, minimal-redundancy-maximal-relevance (mRMR), sequential feature selection (SFS), correlation-based feature selection (CFS), and so many other techniques.

To provide a scientific explanation without the need to delve into detailed descriptions of all the techniques, two examples representing the mentioned technologies have been presented such as RP1 and RP2. These two studies were chosen to ensure a clear and comprehensive understanding for the reader without being overwhelmed with additional details. In RP1, the RRFS-ACO MH technique consistently outperformed traditional techniques, demonstrating superior results when compared to information gain, gain ratio, symmetrical-uncertainty, Gini index, F-score, term variance (TV), L-score, mRMR,

Table 11 Traditional FS techniques.

Conventional methods	Definitions	Referred studies
Information gain	By assessing the gain in relation to the class, information gain determines the value of an attribute. Entropy, which is a measure of how chaotic or unpredictable a system is, is what it depends on to function. Information gain indicates the amount of knowledge left over after removing ambiguity (<i>Gokalp, Tasci & Ugur, 2020</i>).	RP2, RP8, RP17, RP30, RP31, RP34, RP42, RP50, RP60, RP78, RP79
Relieff	Relieff is a single-variate, multiple-class, supervised, filter-based feature-weighting technique that can handle noisy and imperfect data. By periodically sampling an instance and seeing the value of the provided property for k of its closest examples belonging to the same and distinct classes, it determines the value of the attributes. Additionally, it uses a feature weighting scheme to investigate features with the greatest ability to distinguish between classes. with little computing complexity and unaffected by feature interactions (<i>Ghimatgar et al., 2018; Gokalp, Tasci & Ugur, 2020</i>).	RP5, RP12, RP26, RP46, RP50, RP60, RP69, RP79
L-Score	The L-Score is a univariate method that evaluates each characteristic independently while ignoring interdependencies and relying on locality to preserve power. The local geometric structure of the data space is more significant in this approach than the overall structure. The nearest neighbor graph is used to model the local geometric structure (<i>Ghimatgar et al., 2018</i>).	RP1, RP3, RP5, RP12, RP13, RP26, RP61
F-score	The F-Score is a univariate algorithm that operates solely on relevance analysis. To maintain power, the F-Score investigates variables with the greatest discrimination potential and the greatest number of locales. Similar to the L-Score, the F-Score evaluates features independently without taking into account how they are related (<i>Ghimatgar et al., 2018</i>).	RP5, RP12, RP26, RP30, RP31, RP50, RP60
RRFS	A multivariate approach called RRFS evaluates features based on their maximal relevance to classes and their minimal redundancy with respect to one another. Mean Absolute Difference (MAD) or Mutual Information (MI) are two supervised or unsupervised relevance criteria that can be employed in this algorithm to assess the significance of each feature (<i>Ghimatgar et al., 2018</i>).	RP1, RP3, RP5, RP13, RP26, RP61
Correlation-based Feature Selection (CFS)	CFS is an algorithm that ranks a subset of features based on a heuristic evaluation function that relies on correlations. This algorithm starts to build a correlation matrix between the features in the dataset. Then, a search metaheuristic is utilized to build subsets of features to be ranked. The rank assigned to the subsets that result is the correlation between the features and the class divided by the intercorrelation of the features between themselves (<i>Salama, Abdelbar & Anwar, 2016</i>).	RP8, RP30, RP31, RP50, RP60
Chi-square	Chi-square also known as χ^2 Statistic is determined by computing the value of the chi-squared statistic in relation to the class, it determines the worth of an attribute. It is compared to the χ^2 distribution with one degree of freedom in order to assess the lack of independence between terms and classes (<i>Gokalp, Tasci & Ugur, 2020</i>).	RP2, RP8, RP46, RP79
RSM	To better manage the noise in high-dimensional datasets, the Random Subspace Method (RSM) applied a multivariate search methodology to a randomly chosen subset of features (<i>Tabakhi & Moradi, 2015</i>).	RP1, RP3, RP13, RP61
TF-IDF	It is a numerical statistical method for determining the significance of a term for a set of documents (<i>Saraswathi & Tamilarasi, 2016</i>).	RP2, RP41, RP58, RP78
mRMR	A multivariate method called mRMR uses assessments of redundancy and relevance. It evaluates a feature subset with the least amount of overlap between features and the most amount of class relevance. Average F-statistic values are calculated over various time steps to execute the relevance analysis. The dynamic time-warping method employs the redundancy analysis (<i>Ghimatgar et al., 2018</i>).	RP5, RP12, RP26

Table 11 (continued)

Conventional methods	Definitions	Referred studies
Sequential Feature Selection (SFS)	SFS is a wrapper algorithm that begins with an empty set and iteratively tries to add features that maximize the current predictive accuracy of the learning algorithms (<i>Salama, Abdelbar & Anwar, 2016</i>).	RP10, RP23, RP62
Mutual information	It is an information theory basic concept. It is a measurement of general interdependence among two random variables (<i>Tourassi et al., 2001</i>).	RP17, RP62
symmetrical-uncertainty	Symmetrical-uncertainty coefficient is an improved version of information gain that minimizes the bias across the multivalued features. It assesses the value of an attribute by measuring the symmetrical uncertainty with regard to the class (<i>Gokalp, Tasci & Ugur, 2020</i>).	RP46, RP79

mutual correlation (MC), RSM, and RRFs. In RP2, the global best harmony-oriented harmony search (GBHS) achieved superior results when compared to conventional methods, including Chi-square, feature selection based on comprehensive measures, t-test-based feature selection, information gain using term frequency, and an improved term frequency-inverse document frequency approach. Table 11 lists the most frequently used traditional FS techniques along with their definitions and their studies.

RQ4: what are the discernible strengths and weaknesses of MH techniques in the context of FS?

This section identifies and summarizes the strengths and weaknesses of metaheuristic (MH) techniques as reported by researchers. It is focusing on the strengths and weaknesses that have been supported by multiple studies. MH techniques have demonstrated strong performance in feature selection (FS) problems. They are particularly praised for their ability to effectively handle redundant features and high-dimensional data. Table 12 summarizes the strengths of MH techniques based on the selected studies, along with the studies that support each strength. Concurrently, Table 13 summarizes the weaknesses of MH techniques, accompanied by the studies that provide evidence for each weakness. In summary, it is important to note that different MH techniques have varying advantages, and there is no universal solution for MH-FS techniques that fits all scenarios.

According to the previous discussion it is expected that some different solutions that can efficiently improve the MH-FS which are as follows: use MH with the binary system, integrate with traditional FS techniques, integrate with some mathematics theories such as Chaos theory and Rough Set theory, and combine two MH techniques to take benefit from their advantages to complement each other. In addition, there are several reasons which inspire FS techniques to adopt MH techniques as highlighted by the research articles in RP3, RP4, RP6, RP42, RP53, RP57, and RP66. The reasons that motivate researchers to adopt MH in FS are summarized in Fig. 7. Briefly, there are various aspects of MH for FS following section highlights their strengths and discusses each aspect.

Table 12 Strength of MH techniques.

Technique	Strengths	Studies
ACO	<ul style="list-style-type: none"> • Positive feedback that leads to find rapid and good solutions. • Easy and nature implementation in a parallel way. • low execution time in optimization. • Can increase the local and global search capabilities due to the greedy and stochastic natures of the algorithm. • Used of distributed long-term memory. • Used the same structure of reinforcement learning schema • A population-based algorithm that used the colony of ants which leads to raising the robustness of this algorithm. • Efficient and competent in the convergence process. 	RP1, RP3, RP5, RP12, RP26, RP42 RP61
Advanced binary ACO	<ul style="list-style-type: none"> • ABACO allows ants to search for all features. • In the ABACO algorithm ants are authorize to select or deselect visiting features. • ABACO is not constrained to preselect or deselect specific features. • ABACO incorporates heuristic desirability to enhance the exploration of the search process and guide ants towards more prominent features. 	RP7
HS	<ul style="list-style-type: none"> • Free from divergence. • No initial value settings of the decision variables are required. • The algorithm has the capability to identify and select the most suitable individuals, ensuring that their optimal harmonies are preserved and carried forward to subsequent iterations. • Parallel exploration of the search space for the given data. 	RP2, RP6 RP13
PSO	<ul style="list-style-type: none"> • Easy to implement. • Fast convergence speed. • Global communication between the particles. • Able to produce quick solutions for nonlinear optimization problems. 	RP36, RP54, RP81
ABC	<ul style="list-style-type: none"> • Easy implementation. • Demonstrates significant robustness. • Exhibits high flexibility. • Requires fewer control parameters. • Excels in exploitation during the onlooker bee processing phase. 	RP8
DE	<ul style="list-style-type: none"> • Requires a smaller number of parameters. • Operates at a high speed. • Exhibits robustness. • Suitable for tackling high-dimensional and complex optimization problems. 	RP8
GA	<ul style="list-style-type: none"> • Able to search solution space in combinatorial optimization problems. • Able to solve the nonlinear and complex problems. 	RP40, RP54,
CS	<ul style="list-style-type: none"> • Less parameters to be tuned. • Can adapt to a wider class of optimization problems. • Fast convergence. • Global optima achievement. 	RP39

Table 12 (continued)

Technique	Strengths	Studies
ALO	<ul style="list-style-type: none"> ALO has the ability to deliver very competitive and promising performance. 	RP54
Chaotic (CALO)	<ul style="list-style-type: none"> CALO demonstrates the capability to converge towards the same optimal solution across a wide range of applications. 	RP9
CSA	<ul style="list-style-type: none"> When addressing complex, high-dimensional, and multimodal problems, it is straightforward to avoid local optima. 	RP47
SSO	<ul style="list-style-type: none"> Utilizes basic mathematical operators to discover the optimal solution. Offers a cost-effective approach in terms of time complexity and space complexity. 	RP57
BSO	<ul style="list-style-type: none"> Use of clustering in the iterative process to create an optimization algorithm. 	RP72
SA	<ul style="list-style-type: none"> Can escape the local minimums. In each iteration, it requires a single evaluation of the wrapper. Enables control over the trade-off between solution length and result precision. 	RP4
TLBO	<ul style="list-style-type: none"> Needs only the general control parameters to be tuned. 	RP29
WCA	<ul style="list-style-type: none"> Characterized by a minimal number of control parameters (only 3). Effectively tackles the issue of rapid convergence towards local optima entrapment through the implementation of an evaporation technique. 	RP43
GWO	<ul style="list-style-type: none"> Simple and easy to use. Fast convergence. Adaptable and capable of scaling to different contexts or sizes. Few parameters to tune. Demonstrates a certain degree of capability to prevent stagnation in local optima. Achieves a favorable balance between exploration and exploitation through a straightforward approach. Inspired by the intelligent leadership and hunting behaviors observed in grey wolves in nature. 	RP47, RP49, RP70
SFLA	<ul style="list-style-type: none"> Characterized by a straightforward structure. Involves a reduced number of controlling parameters. Features a simple implementation of the algorithm. 	RP51
SSA	<ul style="list-style-type: none"> Requires a reduced number of parameters. Straightforward to implement. Ability to solve large-and-small-scale problems. Flexible and strong stochastic nature. 	RP81
AFSA	<ul style="list-style-type: none"> Demonstrates resilience and robustness. Straightforward and easy to comprehend. Prone to being influenced by initial values due to its dependent on heuristic global optimization. Great influence by the fish behavior in the convergence speed and global search. 	RP45

(Continued)

Table 12 (continued)

Technique	Strengths	Studies
BOA	<ul style="list-style-type: none"> • Provides a comprehensive explanation of competitive results by considering factors such as exploration, exploitation, convergence, and avoidance of local optima. • Demonstrates strong performance across a wide range of unimodal and multimodal benchmark functions. • High convergence rate. • Assists in achieving high exploration by utilizing fragrance attenuation, which facilitates an efficient search across the solution space. 	RP48
SOS	<ul style="list-style-type: none"> • Simple structure. • Straightforward implementation. • No parameter requirements. • Exhibits remarkable stability. • Rapid convergence speed. • Produces high-precision solutions. • Avoids getting trapped in locally optimal solutions. • The mutualism and commensalism phases facilitate the population's quick focus on the vicinity of the optimal solution. 	RP53
WOA	<ul style="list-style-type: none"> • Demonstrates the ability to achieve a balance between exploration and exploitation. 	RP55

Table 13 Weaknesses of MH techniques.

Technique	Weaknesses	Studies
ACO	<ul style="list-style-type: none"> • Limited efficiency when dealing with datasets containing a large number of features. • Slow convergence speed. • Time-consuming execution. • High space complexity leading to premature convergence. • Computationally expensive in terms of memory requirement and speed. • Computational complexity • The efficiency of this algorithm is strongly influenced by the size of the selected feature subsets. 	RP1, RP5 RP26, RP39, RP42
HS	<ul style="list-style-type: none"> • The primary limitation of HS is the excessive number of iterations required to find an optimal solution. 	RP2
PSO	<ul style="list-style-type: none"> • Slow convergence speed. • Time-intensive. • Elevated space complexity and premature convergence. • Inefficient trade-off between local and global search performance. • Exhibits a weakness in fine-tuning near locally optimal positions. • Not as effective when applied to large-scale problems. 	RP39, RP50, RP81, RP93
ABC	<ul style="list-style-type: none"> • The algorithm is time-consuming for convergence, and it fails to showcase its true performance adequately. 	RP8.
DE	<ul style="list-style-type: none"> • Exhibits unstable convergence. • Proneness to getting stuck in local optima. 	RP8

Table 13 (continued)

Technique	Weaknesses	Studies
GA	<ul style="list-style-type: none"> The GA algorithm has a limitation related to the crossover operator, which can cause sudden and drastic changes to the solutions during the search process. The GA can result in slow convergence due to its lack of guidance, hindering effective exploration of the search space. GA requires the tuning of multiple parameters. 	RP81, RP93.
CS	<ul style="list-style-type: none"> Requires much time to compute the fitness function. slow convergence rate. 	RP39
ALO	<ul style="list-style-type: none"> Select Sub-optimal selection due to unbounded random walk in the search space. Stagnation because the exploration capability is very limited. Local optima and premature convergence problems. 	RP9, RP81
CALO	<ul style="list-style-type: none"> The optimization results may not be exactly repeatable. Careful consideration is required when transitioning to a different classifier, particularly in real-time applications. Increasing the running time when switching to another classifier. 	RP9
CSA	<ul style="list-style-type: none"> Inefficient local search strategy. Low convergence rate due to trap in local optima. The stochastic nature of CSA introduces ambiguity in distinguishing between exploitation and exploration, leading to an unclear boundary between the two. 	RP44, RP47, RP77
SSO	<ul style="list-style-type: none"> Explores the search space predominantly in one direction. Lacks information about other regions of the search space. Fall into sub-optimal solutions that affect algorithm performance. 	RP57
BSO	<ul style="list-style-type: none"> The shortage of exploration relies directly on the algorithm internal configuration. Configuring the control parameters of BSO is a challenging task. The process of exploitation is influenced by the method used to create clusters. 	RP72
WCA	<ul style="list-style-type: none"> The effectiveness of the algorithm as a spam classifier remains uncertain or ambiguous. 	RP43
GWO	<ul style="list-style-type: none"> Cannot always perform exploration well. The algorithm is not always capable of successfully addressing the problem and may fail to find the global optimal solution. 	RP47
SFLA	<ul style="list-style-type: none"> Negative affected on convergence speed and solving precision. 	RP51
SSA	<ul style="list-style-type: none"> The algorithm is susceptible to issues related to population diversity and can become trapped in locally optimal solutions. 	RP81
FFA	<ul style="list-style-type: none"> Waste of computation resources. Exhibits low efficiency in searching for optimal regions. Requires the tuning of numerous control parameters. 	RP23

(Continued)

Table 13 (continued)

Technique	Weaknesses	Studies
GSA	<ul style="list-style-type: none"> Experiences premature convergence as a result of rapid deduction diversity. Demonstrates fast initial convergence during the early stages of the search process, which gradually slows down as the global solution approaches the local solution. Challenges arise in achieving a proper balance between exploration and exploitation. 	RP28, RP68
SHO	<ul style="list-style-type: none"> Prone to getting trapped in local optimal solutions. Demonstrates low precision. Exhibits slow convergence speeds. 	RP75

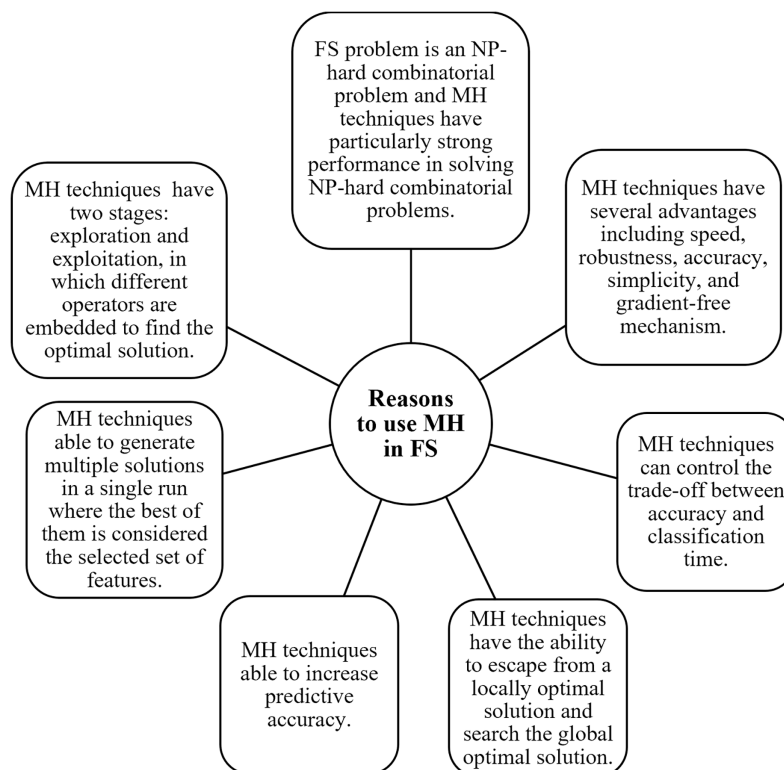


Figure 7 Reasons to use MH in FS.

Full-size DOI: 10.7717/peerj-cs.2084/fig-7

Aspects of MH based on FS

The various strengths of MH approaches for feature selection, including their robust search strategies, global optimization capabilities, parallelizability, adaptability, innovation potential, and cross-domain applicability, make them a powerful choice for addressing feature selection challenges across a wide range of applications. Researchers and practitioners often turn to MH when seeking efficient and effective solutions for feature subset selection.

- 1) Search strategy: MH approaches are known for their powerful search strategies. They excel at navigating complex, high-dimensional search spaces to find the optimal subset of features. These strategies balance exploration and exploitation to efficiently explore the space while exploiting promising solutions. This dynamic search process is a significant strength of MH, enabling them to find feature subsets that traditional methods might overlook due to their deterministic nature.
- 2) Global optimization: One of the key strengths of MH approaches is their ability to perform global optimization. In FS, it is crucial to find the best combination of features that leads to optimal model performance. MH can efficiently search for solutions across the entire feature space, which traditional methods may struggle to do. This global exploration capability is particularly valuable when dealing with large and complex feature sets.
- 3) Parallelization and scalability: MH approaches often lend themselves well to parallelization. This means they can be distributed across multiple processors or machines to expedite the search process, making them scalable for large datasets and high-dimensional feature spaces. This scalability is crucial for handling big data and complex applications, further showcasing the practicality of MH feature selection.
- 4) Adaptability: Another advantage of MH is its adaptability to various optimization objectives and constraints. Researchers can easily customize the objective function to reflect the specific goals of FS problem, such as maximizing classification accuracy, minimizing model complexity, or considering trade-offs.
- 5) Exploration of new techniques: MH approaches encourage the development and exploration of new techniques. Researchers continually innovate and propose novel metaheuristic algorithms specifically designed for FS. For example, the RSS may represent such an innovation. This innovation-driven aspect of MH contributes to the field's dynamism and evolution.
- 6) Cross-domain applicability: MH approaches are not limited to a single domain. They can be applied to FS in various fields, not just text classification. This cross-domain applicability demonstrates the versatility and effectiveness of these techniques in solving FS problems in diverse application areas.

RQ5: how can the RSS be effectively leveraged as an FS technique?

RSS, which stands for Ringed Seal Search, is a metaheuristic (MH) technique introduced in reference ([Saadi et al., 2016](#)). It draws inspiration from the natural behavior of seal pups when selecting a secure lair to evade predators. In comparison to other MH techniques like genetic algorithms (GA) and particle swarm optimization (PSO), RSS demonstrates faster convergence towards the global optimum and maintains a better trade-off between exploration and exploitation ([Saadi et al., 2016](#)). Although RSS has not been widely employed as a feature selection (FS) technique according to the existing literature, it possesses the capability to optimize SVM parameters. Consequently, this optimization leads to enhanced classification accuracy when compared to traditional SVM approaches ([Sharif et al., 2019](#)). The RSS algorithm is primarily inspired by the search behavior of seal

pups seeking optimal lairs to evade predators. It adopts a similar approach where the algorithm continually searches for better solutions and moves towards them. In the context of RSS, these “lair” correspond to problem-specific representations, and the algorithm aims to optimize the quality or fitness of these representations. By iteratively improving the representations, RSS strives to identify the best possible solution. The scenario begins once the female seal gives birth to a pup in a birthing lair that has been created for this purpose. The seal pup technique entails of searching for and choosing the ideal lair by conducting a randomized walk to discover a new lair. The seal pup’s random walk alternates between the normal and urgent search modes because seals are sensitive to outer noise produced by predators. The pup’s normal mode is an intensive search among closely spaced lairs, which is described by Brownian walk. In urgent state, the pup leaves the proximity area and implements extensive search to discover new lair from scatter targets; this movement is described as Levy walk. The change among these two modes is realized by the random noise released *via* predators. The algorithm holds changing between normal and urgent modes until the global optimum is reached.

RSS is especially based on seal pup search for optimal lairs to escape predators. Each time a new lair that has perfect quality is found; the pup moves into it. In the end, the lair (habitat) with the optimal fitness (quality) it is the term that RSS is going to optimize. The RSS concepts is represented on the following depictions.

- i) Each female seal gives birth to a single pup at a time, selecting a random habitat for the pup.
- ii) The seal pup randomly explores its ecosystem to locate a suitable lair for protection against predators.
- iii) The movement of the seal pup can be categorized into two states: Normal, where the search is focused and follows a Brownian walk, and Urgent, where the search is expansive and follows a Levy walk.
- iv) If the best-seen lair $L^{best,t}$ among the current set of lairs K is superior in terms of fitness value compared to the best lair $L^{best,t-1}$ from the previous iteration, L^{best} is updated to $L^{best,t-1}$. Otherwise, L^{best} remains unchanged.

Over time, inferior lairs will be discarded, and the seals will continue to explore and move towards better lairs or chambers, leading to convergence towards good solutions. The RSS algorithm will be adapted for feature selection by using the following steps:

1. Input: Provide the initial number of lairs for the search.
2. Output: The algorithm aims to find the best lairs based on some evaluation criteria.
3. Initialization: Generate birthing lairs’ initial number: Initialize the lairs as $L_1 = (f = 1, 2, 3, \dots, n)$, where n is the initial number of lairs.
4. While (Stopping criterion): Repeat the following steps until a specific stopping criterion is met. This criterion could be a maximum number of iterations, reaching a satisfactory solution, or other conditions specific to the problem being solved.

Algorithm: Ringed Seal Search (RSS) Feature Selection Algorithm**Begin**

Input: Initial Ranked Terms
 Output: Best Features
 Purpose= To select the best subset of features

Generate the birthing lairs' initial number,

$$L_1 = (f = 1, 2, 3, \dots, n)$$

While (Stopping criterion)**If** *noise* = *false*

Search in the proximity for a new lair using Brownian walk;

Else

Expand the search for a way for a new layer using levy walk;

End if

Evaluate the fitness of every new lair using AUC and compare with previous;

If $L^{best,T} > L^{best,T-1}$

Choose the new lair

$$L^{best} = L^{best,T}$$

Go to 4

End if

Select the lair;

Return the best lair

The global best lair is fed to classifiers for training.

End**Figure 8** RSS feature selection algorithm.Full-size  DOI: 10.7717/peerj-cs.2084/fig-8

5. If *noise* = *false*: Check if the noise parameter is set to false. If so, perform a Brownian walk in the proximity to search for a new lair. A Brownian walk is a random process where the next step is determined by a random direction.
6. Else: If the noise parameter is set to true, expand the search for a new lair using a Levy walk. A Levy walk is a random process that allows for long-range exploration and can provide more global search capabilities.
7. Evaluation: Evaluate the fitness of every new lair and compare with previous: Assess the fitness of each newly generated lair using an appropriate evaluation function metrics such as accuracy, recall, precision, *etc.* Compare the fitness of these lairs with the previously evaluated lairs.
8. If $L^{best,T} > L^{best,T-1}$: Check if the fitness of the current best lair ($L^{best,T}$ is greater than the fitness of the previous best lair ($L^{best,T-1}$)).
9. Choose the new lair: If the fitness of the current best lair is greater, select it as the new best lair ($L^{best} = L^{best,T}$).
10. Else: If the fitness of the current best lair is not greater, go to step 4 and continue the search.
11. Rank the lairs: Once the stopping criterion is met, rank the lairs based on their fitness evaluations.
12. End the loop.
13. Termination: Return the best feature subset found as the result of the feature selection.

RSS can be employed for feature selection by following the previous steps. Firstly, the problem should be defined, and the features must be represented appropriately. Secondly, a population of potential feature subsets is initialized, and their fitness is evaluated using a suitable metric. Thirdly, the RSS algorithm is applied iteratively, with the search space being explored, the fitness function being assessed, and the best solutions being exploited. A termination criterion is defined to determine when the algorithm should stop. Finally, the selected features can be extracted from the best solution after the algorithm finishes. [Figure 8](#) summarizes the RSS feature selection algorithm.

Future research directions

In this SLR, future research direction could focus on the further refinement and development of existing MH techniques and investigate how this technique can be optimized to enhance their performance in feature selection. It was noted that there are still some MH techniques, such as RSS, which have not been extensively explored for FS despite their effectiveness. RSS employs two search states, normal and urgent, and dynamically switches between them until the optimal solution is reached. This balance between exploitation and exploration enables RSS to find global optima faster than other techniques. Moreover, RSS has shown high accuracy in text classification problems, making it a promising choice for FS. Therefore, as a future direction, the study suggests the utilization of RSS as an FS technique in this research.

Further considerations addressing the challenges in feature selection research based on MH techniques require a multifaceted approach. Researchers can develop novel algorithms suitable for handling large-scale datasets efficiently while maintaining robustness and scalability. Additionally, strategies for handling dynamic environments and evolving datasets must be devised to ensure the adaptability of feature selection methods over time. Improving the interpretability and explainability of MH-based feature selection models is essential for gaining insights into the decision-making process. Hybrid approaches combining MH with other optimization techniques, machine learning algorithms, binary systems, or mathematical theories such as chaos theory can leverage the strengths of each method, leading to enhanced performance and flexibility. Additionally, the limitations of MH techniques require targeted strategies tailored to each algorithm's weaknesses. For instance, to overcome the challenges associated with ACO, efforts can focus on developing enhanced variants that improve efficiency and convergence speed, possibly through parameter tuning or hybridization with other optimization methods. Similarly, for HS, PSO, and ABC, optimizations could target convergence speed and space complexity by refining the search strategies or introducing adaptive mechanisms. DE's instability and GA's issues with crossover could be mitigated by incorporating diversity maintenance mechanisms or alternative operators. In addition, for algorithms like CSA and SSO, refining the balance between exploration and exploitation is crucial, possibly through algorithmic modifications or parameter adjustments. Finally, interdisciplinary research can uncover applications beyond text classification, while comparing MH techniques with deep learning methods can offer insights into their relative strengths and limitations.

CONCLUSIONS

This study conducted a comprehensive SLR to assess the efficiency and effectiveness of MH techniques in the context of FS. The review rigorously followed a structured approach, encompassing the identification and quality assessment of 108 primary studies conducted between 2015 and 2022. These studies' characteristics were summarized based on the defined research questions, revealing that MH techniques have seen widespread adoption for FS across diverse domains, notably in text classification. Comparative analysis of MH techniques against traditional methods demonstrated their substantial enhancements in the performance of machine learning techniques, specifically within the field of classification. The strengths and weaknesses of MH techniques were meticulously scrutinized, with insights drawn exclusively from the selected studies. Moreover, this research unveils a promising avenue for future investigations, particularly emphasizing the potential for further exploration of MH techniques, as exemplified by the RSS, to refine feature selection across various application domains. This research significantly contributes to our comprehension of the central role MH techniques play in the realm of FS and their broader implications for the fields of data science and text classification. In summary, the findings underscore the compelling case for the adoption of MH techniques in feature selection, emphasizing their superior performance in text classification and serving as a catalyst for ongoing innovation and advancement in this crucial domain.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

Funding was provided by the Research Management Center at Universiti Teknologi Malaysia (Vot No: Q.J130000.21A6.00P48) and the Data Analytics and Artificial Intelligence (DAAI) Research Group in Birmingham City University, UK. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
Research Management Center at Universiti Teknologi Malaysia: Q.J130000.21A6.00P48.
Data Analytics and Artificial Intelligence (DAAI) Research Group in Birmingham City University, UK.

Competing Interests

Faisal Saeed is an Academic Editor for PeerJ.

Author Contributions

- Sarah Abdulkarem Al-shalif conceived and designed the experiments, performed the experiments, analyzed the data, performed the computation work, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.

- Norhalina Senan conceived and designed the experiments, prepared figures and/or tables, and approved the final draft.
- Faisal Saeed performed the experiments, authored or reviewed drafts of the article, and approved the final draft.
- Wad Ghaban analyzed the data, prepared figures and/or tables, and approved the final draft.
- Noraini Ibrahim analyzed the data, authored or reviewed drafts of the article, and approved the final draft.
- Muhammad Aamir performed the computation work, authored or reviewed drafts of the article, and approved the final draft.
- Wareesa Sharif analyzed the data, authored or reviewed drafts of the article, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

This is a literature review.

REFERENCES

- Abualigah L, Dulaimi AJ. 2021.** A novel feature selection method for data mining tasks using hybrid Sine cosine algorithm and genetic algorithm. *Cluster Computing* **24**(3):2161–2176 DOI [10.1007/s10586-021-03254-y](https://doi.org/10.1007/s10586-021-03254-y).
- Agrawal RK, Kaur B, Sharma S. 2020.** Quantum based whale optimization algorithm for wrapper feature selection. *Applied Soft Computing Journal* **89**(1):106092 DOI [10.1016/j.asoc.2020.106092](https://doi.org/10.1016/j.asoc.2020.106092).
- Ahmad SR, Bakar AA, Yaakub MR. 2019.** Ant colony optimization for text feature selection in sentiment analysis. *Intelligent Data Analysis* **23**(1):133–158 DOI [10.3233/IDA-173740](https://doi.org/10.3233/IDA-173740).
- Ahmad SR, Bakar AA, Yaakub MR, Yusop NMM. 2017.** Statistical validation of aco-knn algorithm for sentiment analysis. *Journal of Telecommunication, Electronic and Computer Engineering* **9**:165–170 DOI [10.1063/1.5005351](https://doi.org/10.1063/1.5005351).
- Ahmadi E, Garcia-Arce A, Masel DT, Reich E, Puckey J, Maff R. 2019.** A metaheuristic-based stacking model for predicting the risk of patient no-show and late cancellation for neurology appointments. *IJSE Transactions on Healthcare Systems Engineering* **9**(3):272–291 DOI [10.1080/24725579.2019.1649764](https://doi.org/10.1080/24725579.2019.1649764).
- Al-Dyani WZ, Ahmad FK, Kamaruddin SS. 2022.** Adaptive binary bat and markov clustering algorithms for optimal text feature selection in news events detection model. *IEEE Access* **10**(2):85655–85676 DOI [10.1109/ACCESS.2022.3198654](https://doi.org/10.1109/ACCESS.2022.3198654).
- Al-Rawashdeh G, Mamat R, Hafhizah Binti Abd Rahim N. 2019.** Hybrid water cycle optimization algorithm with simulated annealing for spam E-mail detection. *IEEE Access* **7**:143721–143734 DOI [10.1109/ACCESS.2019.2944089](https://doi.org/10.1109/ACCESS.2019.2944089).
- Albashish D, Hammouri AI, Braik M, Atwan J, Sahran S. 2021.** Binary biogeography-based optimization based SVM-RFE for feature selection. *Applied Soft Computing* **101**(2):107026 DOI [10.1016/j.asoc.2020.107026](https://doi.org/10.1016/j.asoc.2020.107026).
- Allam M, Nandhini M. 2022.** Optimal feature selection using binary teaching learning based optimization algorithm. *Journal of King Saud University—Computer and Information Sciences* **34**(2):329–341 DOI [10.1016/j.jksuci.2018.12.001](https://doi.org/10.1016/j.jksuci.2018.12.001).

- Alshalif SA, Ibrahim N, Herawan T. 2017.** Artificial neural network with hyperbolic tangent activation function to improve the accuracy of COCOMO II model. *Advances in Intelligent Systems and Computing* **549**:81–90 DOI [10.1007/978-3-319-51281-5_9](https://doi.org/10.1007/978-3-319-51281-5_9).
- Alshalif SA, Ibrahim N, Waheeb W. 2017.** Improving the accuracy of COCOMO II effort estimation based on neural network with hyperbolic tangent activation function. *Journal of Telecommunication, Electronic and Computer Engineering* **9**:77–82.
- Alshalif SA, Senan N, Saeed F, Ghaban W, Ibrahim N, Aamir M, Sharif W. 2023.** Alternative relative discrimination criterion feature ranking technique for text classification. *IEEE Access* **11**:71739–71755 DOI [10.1109/ACCESS.2023.3294563](https://doi.org/10.1109/ACCESS.2023.3294563).
- Anand P, Arora S. 2020.** A novel chaotic selfish herd optimizer for global optimization and feature selection. *Artificial Intelligence Review* **53**(2):1441–1486 DOI [10.1007/s10462-019-09707-6](https://doi.org/10.1007/s10462-019-09707-6).
- Ansari GJ, Shah JH, Farias MCQ, Sharif M, Qadeer N, Khan HU. 2021.** An optimized feature selection technique in diversified natural scene text for classification using genetic algorithm. *IEEE Access* **9**:54923–54937 DOI [10.1109/ACCESS.2021.3071169](https://doi.org/10.1109/ACCESS.2021.3071169).
- Anter AM, Ali M. 2020.** Feature selection strategy based on hybrid crow search optimization algorithm integrated with chaos theory and fuzzy c-means algorithm for medical diagnosis problems. *Soft Computing* **24**(3):1565–1584 DOI [10.1007/s00500-019-03988-3](https://doi.org/10.1007/s00500-019-03988-3).
- Arora S, Anand P. 2019.** Binary butterfly optimization approaches for feature selection. *Expert Systems with Applications* **116**(3):147–160 DOI [10.1016/j.eswa.2018.08.051](https://doi.org/10.1016/j.eswa.2018.08.051).
- Arora S, Sharma M, Anand P. 2020.** A novel chaotic interior search algorithm for global optimization and feature selection. *Applied Artificial Intelligence* **34**(4):292–328 DOI [10.1080/08839514.2020.1712788](https://doi.org/10.1080/08839514.2020.1712788).
- Arora S, Singh H, Sharma M, Sharma S, Anand P. 2019.** A new hybrid algorithm based on grey wolf optimization and crow search algorithm for unconstrained function optimization and feature selection. *IEEE Access* **7**:26343–26361 DOI [10.1109/ACCESS.2019.2897325](https://doi.org/10.1109/ACCESS.2019.2897325).
- Aziz MAE, Hassanien AE. 2018.** Modified cuckoo search algorithm with rough sets for feature selection. *Neural Computing and Applications* **29**(4):925–934 DOI [10.1007/s00521-016-2473-7](https://doi.org/10.1007/s00521-016-2473-7).
- Barani F, Mirhosseini M, Nezamabadi-pour H. 2017.** Application of binary quantum-inspired gravitational search algorithm in feature subset selection. *Applied Intelligence* **47**(2):304–318 DOI [10.1007/s10489-017-0894-3](https://doi.org/10.1007/s10489-017-0894-3).
- Bertolazzi P, Felici G, Festa P, Fiscon G, Weitschek E. 2016.** Integer programming models for feature selection: new extensions and a randomized solution algorithm. *European Journal of Operational Research* **250**(2):389–399 DOI [10.1016/j.ejor.2015.09.051](https://doi.org/10.1016/j.ejor.2015.09.051).
- Bhattacharyya T, Chatterjee B, Singh PK, Yoon JH, Geem ZW, Sarkar R. 2020.** Mayfly in harmony: a new hybrid meta-heuristic feature selection algorithm. *IEEE Access* **8**:195929–195945 DOI [10.1109/ACCESS.2020.3031718](https://doi.org/10.1109/ACCESS.2020.3031718).
- Chantar H, Mafarja M, Alsawalqah H, Heidari AA, Aljarah I, Faris H. 2020.** Feature selection using binary grey wolf optimizer with elite-based crossover for Arabic text classification. *Neural Computing and Applications* **32**(16):12201–12220 DOI [10.1007/s00521-019-04368-6](https://doi.org/10.1007/s00521-019-04368-6).
- Chen K, Zhou FY, Yuan XF. 2019.** Hybrid particle swarm optimization with spiral-shaped mechanism for feature selection. *Expert Systems with Applications* **128**(1–2):140–156 DOI [10.1016/j.eswa.2019.03.039](https://doi.org/10.1016/j.eswa.2019.03.039).
- Cheruku R, Edla DR, Kuppili V, Dharavath R. 2018.** RST-BatMiner: a fuzzy rule miner integrating rough set feature selection and Bat optimization for detection of diabetes disease. *Applied Soft Computing* **67**(4):764–780 DOI [10.1016/j.asoc.2017.06.032](https://doi.org/10.1016/j.asoc.2017.06.032).

- Costa H, Galvão LR, Merschmann LHC, Souza MJF. 2018.** A VNS algorithm for feature selection in hierarchical classification context. *Electronic Notes in Discrete Mathematics* **66**:79–86 DOI [10.1016/j.endm.2018.03.011](https://doi.org/10.1016/j.endm.2018.03.011).
- Dadaneh BZ, Markid HY, Zakerolhosseini A. 2016.** Unsupervised probabilistic feature selection using ant colony optimization. *Expert Systems with Applications* **53**(3):27–42 DOI [10.1016/j.eswa.2016.01.021](https://doi.org/10.1016/j.eswa.2016.01.021).
- Das K, Mishra D, Shaw K. 2016.** A metaheuristic optimization framework for informative gene selection. *Informatics in Medicine Unlocked* **4**(10):10–20 DOI [10.1016/j.imu.2016.09.003](https://doi.org/10.1016/j.imu.2016.09.003).
- Das H, Naik B, Behera HS. 2022.** A Jaya algorithm based wrapper method for optimal feature selection in supervised classification. *Journal of King Saud University—Computer and Information Sciences* **34**(6):3851–3863 DOI [10.1016/j.jksuci.2020.05.002](https://doi.org/10.1016/j.jksuci.2020.05.002).
- Dash R. 2021.** An adaptive harmony search approach for gene selection and classification of high dimensional medical data. *Journal of King Saud University—Computer and Information Sciences* **33**(2):195–207 DOI [10.1016/j.jksuci.2018.02.013](https://doi.org/10.1016/j.jksuci.2018.02.013).
- Dash R, Dash R, Rautray R. 2019.** An evolutionary framework based microarray gene selection and classification approach using binary shuffled frog leaping algorithm. *Journal of King Saud University—Computer and Information Sciences* **34**(3):880–891 DOI [10.1016/j.jksuci.2019.04.002](https://doi.org/10.1016/j.jksuci.2019.04.002).
- Deniz A, Kiziloz HE, Dokeroglu T, Cosar A. 2017.** Robust multiobjective evolutionary feature subset selection algorithm for binary classification using machine learning techniques. *Neurocomputing* **241**(4):128–146 DOI [10.1016/j.neucom.2017.02.033](https://doi.org/10.1016/j.neucom.2017.02.033).
- Dwivedi AK. 2018.** Artificial neural network model for effective cancer classification using microarray gene expression data. *Neural Computing and Applications* **29**(12):1545–1554 DOI [10.1007/s00521-016-2701-1](https://doi.org/10.1007/s00521-016-2701-1).
- El-Gohary FA, Nasr FA, Wahaab RA. 2000.** Integrated low-cost wastewater treatment for reuse in irrigation. *Biomedical and Environmental Sciences: BES* **13**:37–43 DOI [10.1145/1134285.1134500](https://doi.org/10.1145/1134285.1134500).
- Eluri RK, Devarakonda N. 2022.** Binary golden eagle optimizer with time-varying flight length for feature selection. *Knowledge-Based Systems* **247**(2):108771 DOI [10.1016/j.knosys.2022.108771](https://doi.org/10.1016/j.knosys.2022.108771).
- Emary E, Zawbaa HM. 2016.** Impact of chaos functions on modern swarm optimizers. *PLOS ONE* **11**(7):e0158738 DOI [10.1371/journal.pone.0158738](https://doi.org/10.1371/journal.pone.0158738).
- Feng J, Kuang H, Zhang L. 2022.** EBBA: an enhanced binary bat algorithm integrated with chaos theory and lévy flight for feature selection. *Future Internet* **14**(6):178 DOI [10.3390/fi14060178](https://doi.org/10.3390/fi14060178).
- Ferrandin M, Enembreck F, Nievola JC, Scalabrin EE, Ávila BC. 2015.** A centroid-based approach for hierarchical classification. In: *ICEIS 2015—17th International Conference on Enterprise Information Systems, Proceedings*. Washington, D.C.: ICEIS, 25–33.
- García-Torres M, Gómez-Vela F, Melián-Batista B, Moreno-Vega JM. 2016.** High-dimensional feature selection via feature grouping: a variable neighborhood search approach. *Information Sciences* **326**(12):102–118 DOI [10.1016/j.ins.2015.07.041](https://doi.org/10.1016/j.ins.2015.07.041).
- Ghimatgar H, Kazemi K, Helfroush MS, Aarabi A. 2018.** An improved feature selection algorithm based on graph clustering and ant colony optimization. *Knowledge-Based Systems* **159**(12):270–285 DOI [10.1016/j.knosys.2018.06.025](https://doi.org/10.1016/j.knosys.2018.06.025).
- Ghosh M, Adhikary S, Ghosh KK, Sardar A, Begum S, Sarkar R. 2019.** Genetic algorithm based cancerous gene identification from microarray data using ensemble of filter methods. *Medical and Biological Engineering and Computing* **57**(1):159–176 DOI [10.1007/s11517-018-1874-4](https://doi.org/10.1007/s11517-018-1874-4).

- Ghosh M, Guha R, Sarkar R, Abraham A. 2020. A wrapper-filter feature selection technique based on ant colony optimization. *Neural Computing and Applications* 32(12):7839–7857 DOI 10.1007/s00521-019-04171-3.
- Gokalp O, Tasci E, Ugur A. 2020. A novel wrapper feature selection algorithm based on iterated greedy metaheuristic for sentiment classification. *Expert Systems with Applications* 146(11):113176 DOI 10.1016/j.eswa.2020.113176.
- Hammouri AI, Mafarja M, Al-Betar MA, Awadallah MA, Abu-Doush I. 2020. An improved dragonfly algorithm for feature selection. *Knowledge-Based Systems* 203(3):106131 DOI 10.1016/j.knosys.2020.106131.
- Han C, Zhou G, Zhou Y. 2019. Binary symbiotic organism search algorithm for feature selection and analysis. *IEEE Access* 7:166833–166859 DOI 10.1109/ACCESS.2019.2953800.
- Hassonah MA, Al-Sayyed R, Rodan A, Al-Zoubi AM, Aljarah I, Faris H. 2020. An efficient hybrid filter and evolutionary wrapper approach for sentiment analysis of various topics on twitter. *Knowledge-Based Systems* 192(1):105353 DOI 10.1016/j.knosys.2019.105353.
- Hichem H, Elkamel M, Rafik M, Mesaaoud MT, Ouahiba C. 2019. A new binary grasshopper optimization algorithm for feature selection problem. *Journal of King Saud University—Computer and Information Sciences* 34(2):316–328 DOI 10.1016/j.jksuci.2019.11.007.
- Hosseinalipour A, Ghanbarzadeh R. 2022. A novel approach for spam detection using horse herd optimization algorithm. *Neural Computing and Applications* 34(15):13091–13105 DOI 10.1007/s00521-022-07148-x.
- Hu P, Pan J-S, Chu S-C. 2020. Improved binary grey wolf optimizer and its application for feature selection. *Knowledge-Based Systems* 195(3):105746 DOI 10.1016/j.knosys.2020.105746.
- Ibrahim RA, Elaziz MA, Oliva D, Cuevas E, Lu S. 2019. An opposition-based social spider optimization for feature selection. *Soft Computing* 23(24):13547–13567 DOI 10.1007/s00500-019-03891-x.
- Ibrahim AM, Tawhid MA, Ward RK. 2020. A binary water wave optimization for feature selection. *International Journal of Approximate Reasoning* 120(10):74–91 DOI 10.1016/j.ijar.2020.01.012.
- Inbarani HH, Bagyamathi M, Azar AT. 2015. A novel hybrid feature selection method based on rough set and improved harmony search. *Neural Computing and Applications* 26(8):1859–1880 DOI 10.1007/s00521-015-1840-0.
- Jadhav S, He H, Jenkins K. 2018. Information gain directed genetic algorithm wrapper feature selection for credit rating. *Applied Soft Computing* 69:541–553 DOI 10.1016/j.asoc.2018.04.033.
- Jain DK, Kumar A, Sangwan SR, Nguyen GN, Tiwari P. 2019. A particle swarm optimized learning model of fault classification in web-apps. *IEEE Access* 7:18480–18489 DOI 10.1109/ACCESS.2019.2894871.
- Javidi MM, Zarisfi Kermani F. 2018. Utilizing the advantages of both global and local search strategies for finding a small subset of features in a two-stage method. *Applied Intelligence* 48(10):3502–3522 DOI 10.1007/s10489-018-1159-5.
- Kashef S, Nezamabadi-pour H. 2015. An advanced ACO algorithm for feature subset selection. *Neurocomputing* 147(5):271–279 DOI 10.1016/j.neucom.2014.06.067.
- Kiziloz HE, Deniz A, Dokeroglu T, Cosar A. 2018. Novel multiobjective TLBO algorithms for the feature subset selection problem. *Neurocomputing* 306(4):94–107 DOI 10.1016/j.neucom.2018.04.020.
- Krishnan GS, Sowmya Kamath S. 2019. A novel GA-ELM model for patient-specific mortality prediction over large-scale lab event data. *Applied Soft Computing Journal* 80(1):525–533 DOI 10.1016/j.asoc.2019.04.019.

- Kumar A, Bawa S. 2020.** A comparative review of meta-heuristic approaches to optimize the SLA violation costs for dynamic execution of cloud services. *Soft Computing* **24(6)**:3909–3922 DOI [10.1007/s00500-019-04155-4](https://doi.org/10.1007/s00500-019-04155-4).
- Kumar A, Jaiswal A. 2019.** Swarm intelligence based optimal feature selection for enhanced predictive sentiment accuracy on twitter. *Multimedia Tools and Applications* **78(20)**:29529–29553 DOI [10.1007/s11042-019-7278-0](https://doi.org/10.1007/s11042-019-7278-0).
- Kuo RJ, Huang SBL, Zulvia FE, Liao TW. 2018.** Artificial bee colony-based support vector machines with feature selection and parameter optimization for rule extraction. *Knowledge and Information Systems* **55(1)**:253–274 DOI [10.1007/s10115-017-1083-8](https://doi.org/10.1007/s10115-017-1083-8).
- Larabi Marie-Sainte S, Alalyani N. 2020.** Firefly algorithm based feature selection for arabic text classification. *Journal of King Saud University—Computer and Information Sciences* **32(3)**:320–328 DOI [10.1016/j.jksuci.2018.06.004](https://doi.org/10.1016/j.jksuci.2018.06.004).
- Liang H, Wang Z, Liu Y. 2019.** A new hybrid ant colony optimization based on brain storm optimization for feature selection. *IEICE Transactions on Information and Systems* **E102D(7)**:1396–1399 DOI [10.1587/transinf.2019EDL8001](https://doi.org/10.1587/transinf.2019EDL8001).
- Liu Y, Heidari AA, Cai Z, Liang G, Chen H, Pan Z, Alsufyani A, Bourouis S. 2022.** Simulated annealing-based dynamic step shuffled frog leaping algorithm: optimal performance design and feature selection. *Neurocomputing* **503(6)**:325–362 DOI [10.1016/j.neucom.2022.06.075](https://doi.org/10.1016/j.neucom.2022.06.075).
- Ma W, Zhou X, Zhu H, Li L, Jiao L. 2021.** A two-stage hybrid ant colony optimization for high-dimensional feature selection. *Pattern Recognition* **116(1)**:107933 DOI [10.1016/j.patcog.2021.107933](https://doi.org/10.1016/j.patcog.2021.107933).
- Mafarja M, Aljarah I, Faris H, Hammouri AI, Al-Zoubi AM, Mirjalili S. 2019.** Binary grasshopper optimisation algorithm approaches for feature selection problems. *Expert Systems with Applications* **117(3)**:267–286 DOI [10.1016/j.eswa.2018.09.015](https://doi.org/10.1016/j.eswa.2018.09.015).
- Mafarja M, Aljarah I, Heidari AA, Hammouri AI, Faris H, Al-Zoubi AM, Mirjalili S. 2018.** Evolutionary population dynamics and grasshopper optimization approaches for feature selection problems. *Knowledge-Based Systems* **145(2)**:25–45 DOI [10.1016/j.knosys.2017.12.037](https://doi.org/10.1016/j.knosys.2017.12.037).
- Mafarja M, Mirjalili S. 2018.** Whale optimization approaches for wrapper feature selection. *Applied Soft Computing* **62(1)**:441–453 DOI [10.1016/j.asoc.2017.11.006](https://doi.org/10.1016/j.asoc.2017.11.006).
- Mafarja MM, Mirjalili S. 2019.** Hybrid binary ant lion optimizer with rough set and approximate entropy reducts for feature selection. *Soft Computing* **23(15)**:6249–6265 DOI [10.1007/s00500-018-3282-y](https://doi.org/10.1007/s00500-018-3282-y).
- Malar B, Nadarajan R, Gowri Thangam J. 2019.** A hybrid isotonic separation training algorithm with correlation-based isotonic feature selection for binary classification. *Knowledge and Information Systems* **59(3)**:651–683 DOI [10.1007/s10115-018-1226-6](https://doi.org/10.1007/s10115-018-1226-6).
- Malhotra R. 2015.** A systematic review of machine learning techniques for software fault prediction. *Applied Soft Computing Journal* **27**:504–518 DOI [10.1016/j.asoc.2014.11.023](https://doi.org/10.1016/j.asoc.2014.11.023).
- Manbari Z, AkhlaghianTab F, Salavati C. 2019.** Hybrid fast unsupervised feature selection for high-dimensional data. *Expert Systems with Applications* **124(4)**:97–118 DOI [10.1016/j.eswa.2019.01.016](https://doi.org/10.1016/j.eswa.2019.01.016).
- Mohan I, Moorthi M. 2021.** Topic flexible aspect based sentiment analysis using minimum spanning tree with cuckoo search. *Journal of Ambient Intelligence and Humanized Computing* **12(7)**:7399–7406 DOI [10.1007/s12652-020-02416-x](https://doi.org/10.1007/s12652-020-02416-x).
- Mohanty R, Das SK. 2018.** Settlement of shallow foundations on cohesionless soils based on SPT value using multi-objective feature selection. *Geotechnical and Geological Engineering* **36(6)**:3499–3509 DOI [10.1007/s10706-018-0549-0](https://doi.org/10.1007/s10706-018-0549-0).

- Mojaveriyan M, Ebrahimpour-komleh H, Jalaleddin MS. 2016.** IGICA: a hybrid feature selection approach in text categorization. *International Journal of Intelligent Systems and Applications* 8(3):42–47 DOI 10.5815/ijisa.2016.03.05.
- Moradi P, Rostami M. 2015.** Integration of graph clustering with ant colony optimization for feature selection. *Knowledge-Based Systems* 84:144–161 DOI 10.1016/j.knosys.2015.04.007.
- Moshki M, Kabiri P, Mohebalhojeh A. 2015.** Scalable feature selection in high-dimensional data based on GRASP. *Applied Artificial Intelligence* 29(3):283–296 DOI 10.1080/08839514.2015.1004616.
- Narkhede S. 2019.** Understanding AUC—ROC curve. Towards Data Science, 6–11. Available at <https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5>.
- Oliva D, Elaziz MA. 2020.** An improved brainstorm optimization using chaotic opposite-based learning with disruption operator for global optimization and feature selection. *Soft Computing* 24(18):14051–14072 DOI 10.1007/s00500-020-04781-3.
- Osmani A, Mohasefi JB, Gharehchopogh FS. 2022.** Sentiment classification using two effective optimization methods derived from the artificial bee colony optimization and imperialist competitive algorithm. *Computer Journal* 65(1):18–66 DOI 10.1093/comjnl/bxz163.
- Oztekin A, Al-Ebbini L, Sevkli Z, Delen D. 2018.** A decision analytic approach to predicting quality of life for lung transplant recipients: a hybrid genetic algorithms-based methodology. *European Journal of Operational Research* 266(2):639–651 DOI 10.1016/j.ejor.2017.09.034.
- Pan JS, Tian AQ, Chu SC, Li JB. 2021.** Improved binary pigeon-inspired optimization and its application for feature selection. *Applied Intelligence* 51(12):8661–8679 DOI 10.1007/s10489-021-02302-9.
- Pandey AC, Rajpoot DS, Saraswat M. 2020.** Feature selection method based on hybrid data transformation and binary binomial cuckoo search. *Journal of Ambient Intelligence and Humanized Computing* 11(2):719–738 DOI 10.1007/s12652-019-01330-1.
- Pashaei E, Aydin N. 2017.** Binary black hole algorithm for feature selection and classification on biological data. *Applied Soft Computing* 56(1):94–106 DOI 10.1016/j.asoc.2017.03.002.
- Paul D, Jain A, Saha S, Mathew J. 2021.** Multi-objective PSO based online feature selection for multi-label classification. *Knowledge-Based Systems* 222(1):106966 DOI 10.1016/j.knosys.2021.106966.
- Purushothaman R, Rajagopalan SP, Dhandapani G. 2020.** Hybridizing gray wolf optimization (GWO) with grasshopper optimization algorithm (GOA) for text feature selection and clustering. *Applied Soft Computing Journal* 96(4):106651 DOI 10.1016/j.asoc.2020.106651.
- Qasem YAM, Abdullah R, Jusoh YY, Atan R, Asadi S. 2019.** Cloud computing adoption in higher education institutions: a systematic review. *IEEE Access* 7:63722–63744 DOI 10.1109/ACCESS.2019.2916234.
- Rais HM, Mehmood T. 2018.** Dynamic ant colony system with three level update feature selection for intrusion detection. *International Journal of Network Security* 20:184–192 DOI 10.6633/IJNS.201801.20(1).20.
- Saadi Y, Yanto ITR, Herawan T, Balakrishnan V, Chiroma H, Risnumawan A. 2016.** Ringed seal search for global optimization via a sensitive search model. *PLOS ONE* 11(1):e0144371 DOI 10.1371/journal.pone.0144371.
- Sadeghian Z, Akbari E, Nematzadeh H. 2021.** A hybrid feature selection method based on information theory and binary butterfly optimization algorithm. *Engineering Applications of Artificial Intelligence* 97(5):104079 DOI 10.1016/j.engappai.2020.104079.
- Salama KM, Abdelbar AM, Anwar IM. 2016.** Data reduction for classification with ant colony algorithms. *Intelligent Data Analysis* 20(5):1021–1059 DOI 10.3233/IDA-160855.

- Saraswathi K, Tamilarasi A. 2016.** Ant colony optimization based feature selection for opinion mining classification. *Journal of Medical Imaging and Health Informatics* **6(7)**:1594–1599 DOI [10.1166/jmihi.2016.1856](https://doi.org/10.1166/jmihi.2016.1856).
- Sayed GI, Hassanien AE, Azar AT. 2019.** Feature selection via a novel chaotic crow search algorithm. *Neural Computing and Applications* **31(1)**:171–188 DOI [10.1007/s00521-017-2988-6](https://doi.org/10.1007/s00521-017-2988-6).
- Sayed GI, Khoriba G, Haggag MH. 2018.** A novel chaotic salp swarm algorithm for global optimization and feature selection. *Applied Intelligence* **48(10)**:3462–3481 DOI [10.1007/s10489-018-1158-6](https://doi.org/10.1007/s10489-018-1158-6).
- Selvarajan D, Jabar ASA, Ahmed I. 2019.** Comparative analysis of PSO and ACO based feature selection techniques for medical data preservation. *International Arab Journal of Information Technology* **16**:731–736.
- Senan N, Ibrahim R, Nawi NM, Yanto ITR, Herawan T. 2012.** Rough and soft set approaches for attributes selection of traditional malay musical instrument sounds classification. *International Journal of Software Science and Computational Intelligence* **4**:14–40 DOI [10.4018/IJSSCI](https://doi.org/10.4018/IJSSCI).
- Sharaff A, Kamal C, Porwal S, Bhatia S, Kaur K, Hassan MM. 2021.** Spam message detection using danger theory and krill herd optimization. *Computer Networks* **199(4)**:108453 DOI [10.1016/j.comnet.2021.108453](https://doi.org/10.1016/j.comnet.2021.108453).
- Sharif W, Yanto ITY, Samsudin NA, Deris MM, Khan A, Mushtaq MF, Ashraf M. 2019.** An optimised support vector machine with ringed seal search algorithm for efficient text classification. *Journal of Engineering Science and Technology* **14**:1601–1613.
- Singh A, Kaur M. 2020.** Detection framework for content-based cybercrime in online social networks using metaheuristic approach. *Arabian Journal for Science and Engineering* **45(4)**:2705–2719 DOI [10.1007/s13369-019-04125-w](https://doi.org/10.1007/s13369-019-04125-w).
- Singh S, Singh AK. 2018.** Detection of spam using particle swarm optimisation in feature selection. *Pertanika Journal of Science and Technology* **26**:1355–1372.
- Tabakhi S, Moradi P. 2015.** Relevance-redundancy feature selection based on ant colony optimization. *Pattern Recognition* **48(9)**:2798–2811 DOI [10.1016/j.patcog.2015.03.020](https://doi.org/10.1016/j.patcog.2015.03.020).
- Tabakhi S, Najafi A, Ranjbar R, Moradi P. 2015.** Gene selection for microarray data classification using a novel ant colony optimization. *Neurocomputing* **168**:1024–1036 DOI [10.1016/j.neucom.2015.05.022](https://doi.org/10.1016/j.neucom.2015.05.022).
- Tawhid MA, Ibrahim AM. 2020.** Feature selection based on rough set approach, wrapper approach, and binary whale optimization algorithm. *International Journal of Machine Learning and Cybernetics* **11(3)**:573–602 DOI [10.1007/s13042-019-00996-5](https://doi.org/10.1007/s13042-019-00996-5).
- Thiyagarajan D, Shanthi N. 2019.** A modified multi objective heuristic for effective feature selection in text classification. *Cluster Computing* **22(S5)**:10625–10635 DOI [10.1007/s10586-017-1150-7](https://doi.org/10.1007/s10586-017-1150-7).
- Thom de Souza RC, de Macedo CA, dos Santos Coelho L, Pierezan J, Mariani VC. 2020.** Binary coyote optimization algorithm for feature selection. *Pattern Recognition* **107(5)**:107470 DOI [10.1016/j.patcog.2020.107470](https://doi.org/10.1016/j.patcog.2020.107470).
- Too J, Rahim Abdullah A. 2020.** Binary atom search optimisation approaches for feature selection. *Connection Science* **32(4)**:1–25 DOI [10.1080/09540091.2020.1741515](https://doi.org/10.1080/09540091.2020.1741515).
- Tourassi GD, Frederick ED, Markey MK, Floyd CE. 2001.** Application of the mutual information criterion for feature selection in computer-aided diagnosis. *Medical Physics* **28(12)**:2394–2402 DOI [10.1118/1.1418724](https://doi.org/10.1118/1.1418724).
- Tubishat M, Abushariah MAM, Idris N, Aljarah I. 2019.** Improved whale optimization algorithm for feature selection in Arabic sentiment analysis. *Applied Intelligence* **49(5)**:1688–1707 DOI [10.1007/s10489-018-1334-8](https://doi.org/10.1007/s10489-018-1334-8).

- Tubishat M, Idris N, Shuib L, Abushariah MAM, Mirjalili S. 2020.** Improved salp swarm algorithm based on opposition based learning and novel local search algorithm for feature selection. *Expert Systems with Applications* **145**(13):113122 DOI [10.1016/j.eswa.2019.113122](https://doi.org/10.1016/j.eswa.2019.113122).
- Tubishat M, Ja'afar S, Idris N, Al-Betar MA, Alswaitti M, Jarrah H, Ismail MA, Omar MS. 2022.** Improved sine cosine algorithm with simulated annealing and singer chaotic map for Hadith classification. *Neural Computing and Applications* **34**(2):1385–1406 DOI [10.1007/s00521-021-06448-y](https://doi.org/10.1007/s00521-021-06448-y).
- Turney P. 1995.** Technical note: bias and the quantification of stability. *Machine Learning* **20**(1–2):23–33 DOI [10.1007/BF00993473](https://doi.org/10.1007/BF00993473).
- Wang J, Lin D, Zhang Y, Huang S. 2022.** An adaptively balanced grey wolf optimization algorithm for feature selection on high-dimensional classification. *Engineering Applications of Artificial Intelligence* **114**:105088 DOI [10.1016/j.engappai.2022.105088](https://doi.org/10.1016/j.engappai.2022.105088).
- Wang Y, Liu Y, Feng L, Zhu X. 2015.** Novel feature selection method based on harmony search for email classification. *Knowledge-Based Systems* **73**(8):311–323 DOI [10.1016/j.knosys.2014.10.013](https://doi.org/10.1016/j.knosys.2014.10.013).
- Xue YU, Xue B, Zhang M. 2019.** Self-adaptive particle swarm optimization for large-scale. *ACM Transactions on Knowledge Discovery from Data (TKDD)* **13**:1–27 DOI [10.1145/3340848](https://doi.org/10.1145/3340848).
- Yelmen I, Zontul M, Kaynar O, Sonmez F. 2018.** A novel hybrid approach for sentiment classification of TURKISH tweets for GSM operators. *International Journal of Circuits, Systems and Signal Processing* **12**:637–645.
- Yong Z, Dun-wei G, Wan-qiu Z. 2016.** Feature selection of unreliable data using an improved multi-objective PSO algorithm. *Neurocomputing* **171**(6):1281–1290 DOI [10.1016/j.neucom.2015.07.057](https://doi.org/10.1016/j.neucom.2015.07.057).
- Zakeri A, Hokmabadi A. 2019.** Efficient feature selection method using real-valued grasshopper optimization algorithm. *Expert Systems with Applications* **119**(6):61–72 DOI [10.1016/j.eswa.2018.10.021](https://doi.org/10.1016/j.eswa.2018.10.021).
- Zarshenas A, Suzuki K. 2016.** Binary coordinate ascent: an efficient optimization technique for feature subset selection for machine learning. *Knowledge-Based Systems* **110**:191–201 DOI [10.1016/j.knosys.2016.07.026](https://doi.org/10.1016/j.knosys.2016.07.026).
- Zawbaa HM, Emary E, Grosan C. 2016.** Feature selection via chaotic antlion optimization. *PLOS ONE* **11**(3):e0150652 DOI [10.1371/journal.pone.0150652](https://doi.org/10.1371/journal.pone.0150652).
- Zhang Y, Song X, Gong D. 2017.** A return-cost-based binary firefly algorithm for feature selection. *Information Sciences* **418-419**(3):561–574 DOI [10.1016/j.ins.2017.08.047](https://doi.org/10.1016/j.ins.2017.08.047).
- Zhao Y, Dong J, Li X, Chen H, Li S. 2022.** A binary dandelion algorithm using seeding and chaos population strategies for feature selection. *Applied Soft Computing* **125**(5):109166 DOI [10.1016/j.asoc.2022.109166](https://doi.org/10.1016/j.asoc.2022.109166).
- Zhu L, He S, Wang L, Zeng W, Yang J. 2019.** Feature selection using an improved gravitational search algorithm. *IEEE Access* **7**:114440–114448 DOI [10.1109/ACCESS.2019.2935833](https://doi.org/10.1109/ACCESS.2019.2935833).
- Zorarpaçlı E, Özel SA. 2016.** A hybrid approach of differential evolution and artificial bee colony for feature selection. *Expert Systems with Applications* **62**(1):91–103 DOI [10.1016/j.eswa.2016.06.004](https://doi.org/10.1016/j.eswa.2016.06.004).