IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 28 December 2023, accepted 1 February 2024, date of publication 6 February 2024, date of current version 13 February 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3362907

SURVEY

Preserving Privacy in Association Rule Mining Using Metaheuristic-Based Algorithms: A Systematic Literature Review

SHAHAD S. ALJEHANI AND YOUSEEF A. ALOTAIBI
Department of Computer Science, College of Computer and Information Systems, Umm Al-Qura University, Makkah 21955, Saudi Arabia

Corresponding author: Shahad S. Aljehani (s44380037@st.uqu.edu.sa)

This work was supported by the Deanship for Research & Innovation, Ministry of Education, Saudi Arabia, under Project Number IFP22UQU4281768DSR120.

ABSTRACT The current state of Association Rule Mining (ARM) technology is heading towards a critical yet profitable direction. The ARM process uncovers numerous association rules, determining correlations between itemsets, forming building blocks that have led to revolutionary scientific discoveries. However, a high level of privacy is vital for protecting sensitive rules, raising privacy concerns. Researchers have recently highlighted challenges in the Privacy-Preserving Association Rule Mining (PPARM) field. Many studies have proposed workarounds for the PPARM dilemma by using metaheuristics. This paper conducts a systematic literature review on metaheuristic-based algorithms addressing PPARM challenges. It explores existing studies, providing insights into diverse metaheuristic approaches tackling PPARM problems. A detailed taxonomy is presented, offering a structured classification of metaheuristic-based algorithms specific to PPARM. This classification facilitates a nuanced understanding of the field by categorizing these algorithms into metaphor-based and non-metaphor-based groups, with a discussion of the nature of the representation schemes for each category identified in the survey. The review extends its analysis to encompass the latest applied approaches, highlighting the diversification of existing metaheuristic algorithms in the PPARM context. Moreover, common datasets and evaluation metrics identified from selected studies are documented to provide a deeper understanding of the methodological choices made by researchers in this domain. Finally, a discussion of existing challenges and potential future directions is presented. This review serves as a helpful guide that outlines previous research and presents potential future opportunities for metaheuristic-based algorithms in the context of PPARM.

INDEX TERMS Association rule mining, metaheuristic, optimization, privacy preserving.

I. INTRODUCTION

Big data can be seen as a mine full of hidden insightful treasures, and the discovery of this insightful knowledge must be made to produce scientific and medical evolution. Numerous valuable findings can be discovered through the Data Mining (DM) process. In particular, a well-known type of data mining called Association Rule Mining (ARM) involves discovering associations between itemsets by finding patterns that occur frequently in the dataset [1]. Association Rules (AR) can be uncovered and used to determine the correlations between items. Using various optimization algorithms, frequent items can be discovered by scanning the database and collecting itemsets that meet the minimum support threshold and are checked against a minimum confidence level. These itemsets form association rules that go through the mining process to uncover hidden and useful insights [1], [2]. ARM has been utilized in various fields, such as healthcare informatics systems [2], [3], supply chain systems [4], malware detection systems [5], and coal mine monitoring systems [6].However, a high level of privacy is vital to protecting sensitive information when it is shared by

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

data owners for the association rule mining process. Privacy preservation is a data mining technique that addresses the security of the knowledge extracted through data mining methods [7]. Recently, researchers have shed light on the challenges of Privacy-Preserving Association Rule Mining (PPARM) due to the dilemma of choosing between disclosing sensitive association rules to generate highly accurate results and maintaining the privacy of these sensitive rules.

In recent decades, PPARM has been a focal point of interest for researchers, and many approaches have been proposed to address PPARM from different perspectives. PPARM is considered a dilemma because its objectives conflict with each other. In other words, it utilizes data while maintaining its privacy [8]. Moreover, PPARM is also referred to as an optimization problem because of the need to find the optimal result from a set of candidates [9]. To solve this optimization dilemma, researchers have utilized so-called approximation methods in PPARM problems to search for a good solution in the search space in a reasonable amount of time [10].

In general, approximation methods can be classified based on their dependency on given problems. Specifically, exact approaches depend on problem information to provide a solution, meaning they are heuristic. Heuristic-based algorithms look for a "good" solution by searching the solution space. However, heuristic methods consume large amounts of computational time and memory. On the other hand, metaheuristics are problem-independent approaches that provide a set of approximate solutions to given problems and evaluate them to find the optimal solution. This implies that metaheuristics are generic algorithm frameworks that can be utilized in almost all optimization problems.

The main aim of metaheuristic-based algorithms is to minimize side effects that arise during the mining process. These include Missing Cost (MC), Hiding Failure (HF), and Artificial Cost (AC) [11]. Additionally, the process of mining association rules is optimized while maintaining the privacy of the sensitive rules. Recently, several studies have proposed PPARM algorithms based on metaheuristics. Consequently, various reviews regarding PPARM have been published.

Existing literature surveys have conducted a Systematic Literature Review (SLR) on metaheuristic-based algorithms of association rule mining to investigate their performance. In 2020, Telikani et al. [12] reviewed evolutionary approaches to ARM and discussed their challenges. In 2021, Logeswaran et al. [13] surveyed recent studies on metaheuristic-based algorithms using ARM and other data mining fields. However, these studies focused on classifying metaheuristic-based algorithms in terms of their evolutionary approaches. Moreover, both addressed the metaheuristic-based algorithms used for the ARM process in general, but not specifically for PPARM. Thus, there is still a need to conduct an SLR in the PPARM field.

The motivation behind conducting a systematic literature review in this domain stems from the need to comprehensively understand and synthesize the existing body of knowledge. Critically examining the current state of research at the intersection of Privacy-Preserving Association Rules Mining (PPARM) and metaheuristic algorithms is undertaken to identify gaps, trends, and challenges in the existing literature. Therefore, this study bridges the gap in the existing literature by providing an advanced and comprehensive review of PPARM algorithms that depend on metaheuristic methods and techniques. The main contributions of this review are as follows:

- Perform SLR process on studies that address PPARM and metaheuristic-based algorithms.
- Provide a detailed topological structure of existing metaheuristic-based algorithms used for PPARM.
- Present a thorough review of these studies based on the applied metaheuristic types.
- Discuss the diversification of existing metaheuristic algorithms, including the newest approaches such as multi-objectivity, hybridization, discretization, and parallelism.
- Outline the characteristics of the most commonly used datasets.
- Analyze the metrics used for evaluating metaheuristic-based algorithms.
- Present existing challenges and discuss potential future directions.

This review provides new insightful findings related to metaheuristic-based algorithms in the PPARM context, which should be considered when proposing new versions. Thus, it bridges the gap between previous research studies and future opportunities and directions in the field of PPARM. Ultimately, this endeavor seeks to contribute to the development of advanced and privacy-aware association rules mining methodologies, fostering innovation in data mining practices while safeguarding individuals' sensitive information.

The remainder of this paper is organized as follows: Section II presents the methodology of this review. Section III reviews the existing metaheuristics used in PPARM algorithms. Section IV presents the types of diversification in these algorithms. Section V illustrates the characteristics of the datasets most frequently used in the literature. Section VI outlines the most common metrics for evaluating PPARM metaheuristic-based algorithms. Section VII outlines existing challenges and potential directions for future research. Finally, Section 8 concludes the study and describes future work.

II. METHODOLOGY

The methodology used to conduct the SLR process is based on the work of Kitchenham and Charters [14]. They proposed a review process consisting of three main stages: Planning, Conducting, and Reporting, each with a set of activities. Figure 1 illustrates a roadmap of the systematic reviews.

The key step in planning a systematic review is to specify the research questions that define the purpose of the review and formulate a review protocol. This approach was followed to ensure a clear and thorough review. A pre-defined protocol

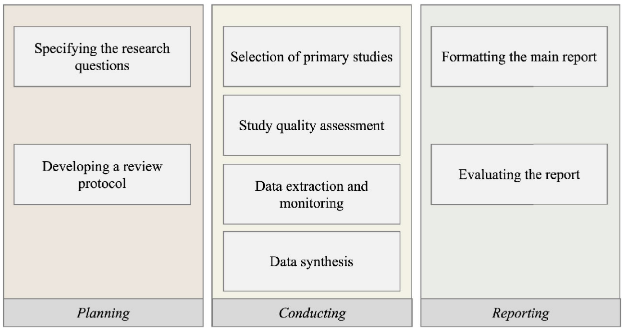
Specifying the research questions
Developing a review protocol
Planning
Selection of primary studies
Study quality assessment
Data extraction and monitoring
Data synthesis
Conducting
Formatting the main report
Evaluating the report
Reporting

FIGURE 1. **Stages of a systematic review.**

helps reduce the possibility of bias in the study selection process. The protocol includes six stages: (1) identifying the research questions, (2) designing the search approach, (3) defining the selection process, (4) designing the quality evaluation technique, (5) defining the strategies for extracting data, and (6) determining methodologies for synthesizing the extracted data [15]. Figure 2 outlines these stages. The following subsections present details of the review protocol.

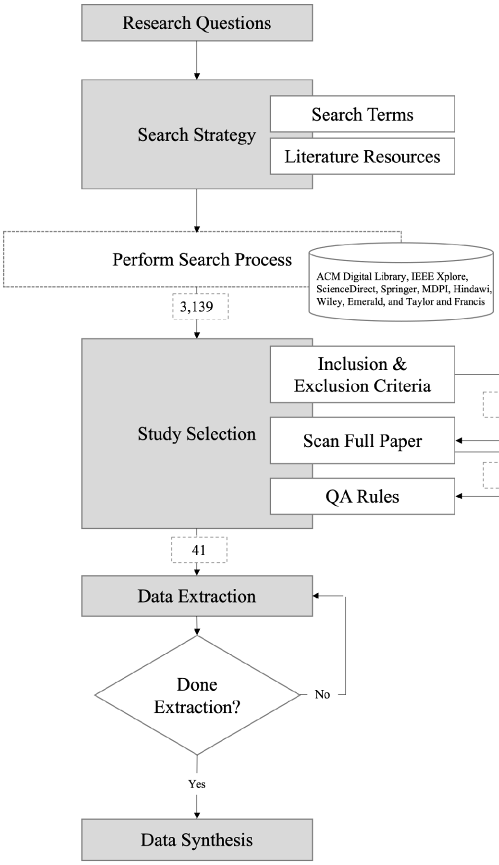

FIGURE 2. **Details stages of review protocol.**

A. RESEARCH QUESTIONS

The goal of this systematic review is to study and summarize the empirical evidence on metaheuristic-based algorithms used to preserve privacy in the association rules mining process. To fulfill this goal, the following questions were raised:

- RQ1: What types of metaheuristics are used in PPARM?
- RQ2: What are the most common features of datasets used in PPARM metaheuristic studies?
- RQ3: What are the most frequently used evaluation criteria in the PPARM metaheuristic studies?
- RQ4: What are the existing challenges and potential future directions?

B. SEARCH STRATEGY

A systematic review aims to search for and find numerous primary studies to answer research questions using a neutral and unbiased search strategy. The search strategy involves constructing search terms and identifying resources to extract related articles. Candidate articles were filtered based on the inclusion and exclusion criteria during the study selection phase. Further reduction occurred by scanning the full articles and applying quality assessment questions. Finally, related articles were extracted to generate answers to the research questions and summarize the findings through synthesis.

1) SEARCH TERMS

The following points were used to identify the most important research terms and keywords:

- Define the major terms from the research questions.
- Derive keywords from highly cited related research papers.
- Extract synonyms and alternative words for related terms.
- Establish links between search terms using Boolean AND.
- Combine synonyms and alternative words using Boolean OR.

Thus, the search terms were identified as follows: "preserve" AND "privacy" AND ("mining" OR "extracting" OR "hiding") AND ("sensitive rules" OR "association rules" OR "positive rules" OR "negative rules") AND ("metaheuristic" OR "evolutionary" OR "optimization" OR "genetic" OR "intelligent").

2) LITERATURE RESOURCES

Nine digital databases were used to retrieve related initial articles: ACM Digital Library, IEEE Xplore, ScienceDirect, Springer, MDPI, Hindawi, Wiley, Emerald, and Taylor and Francis. The search period was from January 2015, to October 2023. Using the search terms in these resources, the initial search returned 3156 articles covering titles, abstracts, and keywords. Table 1 summarizes the number of studies selected from each resource after the initial search.

3) STUDY SELECTION

The initial search yielded 3,156 articles. These articles may have duplicate versions, be irrelevant to the research questions, or be of insufficient quality. Therefore, additional selection criteria are necessary to filter out related articles.

TABLE 1. Number of studies selected per digital resource.

Digital Resource	No. of Articles
ACM Digital Library	457
IEEE Xplore	264
ScienceDirect	619
Springer	318
MDPI	483
Hindawi	69
Wiley	403
Emerald	233
Taylor and Francis	310
Total	**3,156**

As illustrated in figure 2, the selection of primary studies consisted of three filters: (1) removing any duplicated studies, (2) applying inclusion and exclusion criteria to select articles related to RQs and discarding irrelevant ones, and (3) performing quality assessment to select only high-quality studies. The study selection phase was accomplished by scanning not only titles, abstracts, and keywords, as in the initial search, but also the full text of candidate studies. The inclusion and exclusion criteria were initially defined when the review protocol was formed but were refined iteratively through the study selection stage. The inclusion and exclusion criteria are as follows:

Inclusion Criteria:

- Studies published in journals and commonly high-quality cited conference papers.
- The most recent versions of duplicated studies.
- Only publications in the English language.
- Studies that covered PPARM, Privacy Preserving in Data Mining (PPDM) or ARM optimization problems.

Exclusion Criteria:

- Surveys and review papers.
- Studies aimed at nonlinear problems or spatial association rules.
- Studies that applied ARM in graphs.

This survey aims to provide a comprehensive overview of the state of the field of privacy preservation in the mining process based on metaheuristic-based algorithms. For this purpose, the selection phase focused on including peer-reviewed journal articles and commonly high-quality cited conference papers. By applying these selection criteria, 281 studies were identified as candidates for answering the research questions.

4) FULL PAPER SCAN

Further reduction in the number of candidate studies was achieved by scanning the full text. The purpose of scanning an article was to quickly find answers to research questions. This step is mandatory when the exclusion criteria may not filter irrelevant studies from this review. For example, metaheuristic-based algorithms have been proposed for feature selection [16], [17] and rule classification [18], [19]. Thus, the results of this step included 59 candidate-related studies.

5) QUALITY ASSESSMENT CRITERIA

Quality assessment is used to evaluate the quality of the selected articles in terms of relevance, rigorousness and credibility [20]. The QA indicators were developed in the form of six questions or rules. QA questions were formulated based on the QA rules from [15] and modified to suit the objectives of this literature review, which are presented in Table 2. Each question has a weight of 1 with three options: full grade if the question was fully answered, half grade if it was partly answered, and no grade if it was not answered. The quality of this study was determined by the sum of the QA scores of the selected studies. The QA score ranges from 0 to 10, and the score of 5 is considered the minimum threshold for inclusion in this review. Based on the QA score, the outcome of this stage was 41 relevant studies with QA scores ≥ 5.

TABLE 2. Quality assessment questions.

No.	Question
QA1	Are the objectives of the study clearly defined?
QA2	Is the approach for PPARM, PPDM or ARM adequately described?
QA3	Is the algorithm depends on metaheuristics which is the scope of this review?
QA4	Is the type of association rules being mined was clearly stated?
QA5	Is the proposed PPARM model applied to a dataset with sufficient size and quality?
QA6	Is the performance of algorithm measured and reported?
QA7	Is the proposed metaheuristic model compared with other related models?
QA8	Are the results of the study comprehensibly reported and discussed?
QA9	Does the study discuss the applied approach in terms of privacy preservation?
QA10	Is there an added-value to the research field derived from the study

6) DATA EXTRACTION

Data extraction was performed for 41 eligible studies that addressed the research questions of this review. Extracting data from a vast number of studies can be challenging owing to the unstructured nature of the terminologies and synonyms of words. For example, some authors used different names for metaheuristic-based algorithms such as ‘‘genetic’’ or ‘‘revolutionary’’, while others used evaluation metrics that measured the performance of algorithms with different techniques. For instance, to measure the data utility level of an algorithm, some studies applied the Missing Cost (MC) metric, whereas others used the Artifact Pattern (AP) metric [5], [21]. To facilitate the gathering of necessary data and make it easier to trace, a binary checklist was used to fill the answers to the research questions for each study. Nevertheless, some articles did not address all the five research questions. The binary checklist provided a tracing form for each study in terms of the research questions that they answered.

7) DATA SYNTHESIS

The last stage specified in the review protocol was data synthesis, aiming to aggregate and summarize the findings of the selected primary studies that answered the research questions. In this SLR, both quantitative and qualitative data are extracted from studies, such as the evaluation performance and the characteristics of metaheuristic-based algorithms. Three approaches were used to synthesize the extracted data to answer the research questions in the next section.

The vote-counting method is used for RQ1, representing a comparison of different PPARM algorithms based on their approaches and adopted techniques. Narrative synthesis is used to describe the extracted data by formulating them into tables and visualization tools, such as charts, to highlight the similarities and differences between the findings. Hence, RQ2 and RQ3 were discussed using narrative synthesis to represent the characteristics of the most frequently used datasets and common evaluation metrics used in these studies. For data pertaining to RQ4 and RQ5, reciprocal translation was used to translate the diversification of existing metaheuristic-based algorithms extracted from multiple studies with similar meanings and create a uniform description for each category. For RQ5, the reciprocal translation method was used to project the existing challenges and potential future directions for similar research.

After presenting the review protocol used for this review, 41 studies were analyzed thoroughly to address and answer the research questions. Figure 3 outlines the number of eligible studies per digital resource that passed the review. The findings of SLR are presented and discussed in the following sections.

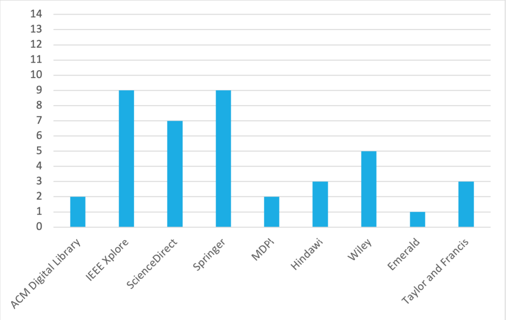

FIGURE 3. Number of eligible studies per digital resource.

III. REVIEW OF THE EXISTING METAHEURISTICS USED IN PPARM ALGORITHMS

To obtain a clear road map of this SLR and its findings, it is important to mention the existing taxonomies of the metaheuristics and adopt the most suitable one for classifying the algorithms found in the selected studies. The key functions of any metaheuristic are the exploration and exploitation processes, which play a vital role in determining the efficiency of the search process. Accordingly, various metaheuristic classifications have been presented based on how exploration and exploitation are used, and how the search procedures are symbolized. Osman [22] divided metaheuristics into three main categories: local search, construction-based, and population-based metaheuristics. Gendreau and Potvin [23] classified metaheuristics into trajectory-based and population-based. Abdel-Basset et al. [24] proposed a new classification of metaheuristics, encompassing a broad variety of algorithms. This classification covers the latest collection of algorithms developed based on metaheuristics. Therefore, this review adopts Abdel-Basset et al.'s [24] classification and applies it to studies addressing PPARM optimization problems. Figure 4 illustrates the taxonomy of metaheuristic-based algorithms identified from the 41 selected studies. Metaheuristics are divided into two groups: metaphor-based and non-metaphor-based metaheuristics. Metaheuristics based on metaphors are algorithms inspired by nature to determine their search strategy, such as chemistry, biology, or the simulation of the behavior of a swarm of living creatures. Two primary paradigms were identified in this survey: evolutionary and swarm systems. In contrast, non-metaphor-based metaheuristics are algorithms that do not mimic any behavior or phenomena as a search strategy. The nature of the representation schemes for each category and their implementation in the field of PPARM are discussed in the following sections.

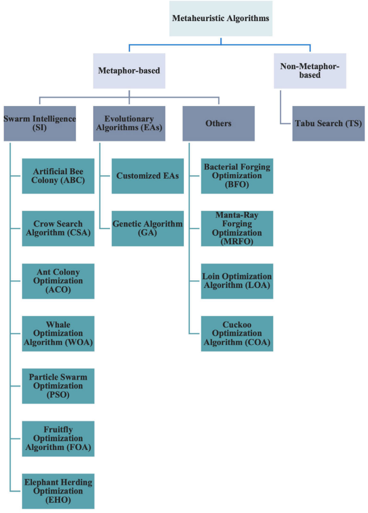
Metaheuristic Algorithms
Metaphor-based
Non-Metaphor-based
Swarm Intelligence (SI)
Evolutionary Algorithms (EAs)
Others
Tabu Search (TS)
Artificial Bee Colony (ABC)
Crow Search Algorithm (CSA)
Ant Colony Optimization (ACO)
Whale Optimization Algorithm (WOA)
Particle Swarm Optimization (PSO)
Fruitfly Optimization Algorithm (FOA)
Elephant Herding Optimization (EHO)
Customized EAs
Genetic Algorithm (GA)
Bacterial Forging Optimization (BFO)
Manta-Ray Forging Optimization (MRFO)
Loin Optimization Algorithm (LOA)
Cuckoo Optimization Algorithm (COA)

FIGURE 4. Taxonomy of the metaheuristic-based algorithms.

Furthermore, these metaheuristic algorithms were utilized in selected studies to address optimization challenges in the area of privacy preservation and data utility trade-off in the mining process. It is important to note that this review mainly concerns privacy preservation techniques used in the association rule mining process. However, other studies in similar fields, such as PPDM and ARM, were included in the review because of their significant contributions to the development of metaheuristic algorithms regarding the privacy and utility of mined data. As shown in figure 5, the majority of the selected studies focused on proposing metaheuristic algorithms related to PPARM, while ARM was the least discussed problem from 2015 to 2023. A review of these studies is presented based on the type of metaheuristics used.

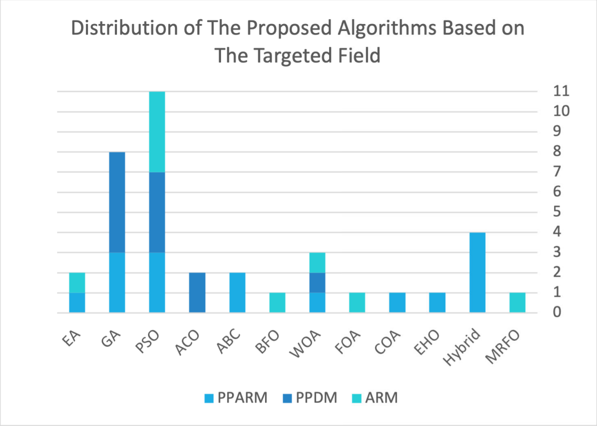

FIGURE 5. Distribution of the proposed algorithms based on the targeted field.

A. METAPHOR BASED METAHEURISTICS: BIO-INSPIRED

Nature has been a source of inspiration for researchers in various ways, enriching the scientific field. Most new algorithms are referred to as nature-inspired, and many existing metaheuristic algorithms are derived by simulating biological evolution principles. In particular, they imitate various metaphors such as the characteristics, structure, and components of biological systems [24], [25]. Therefore, many nature-inspired algorithms developed are bio-inspired.

In this review, the majority of bio-inspired algorithms identified from the selected studies were categorized into Evolutionary Algorithms (EA) and Swarm Intelligence (SI). The remaining algorithms are presented under the "Others" category.

1) EVOLUTIONARY ALGORITHMS (EAS)

The concept of evolutionary computation is based on Darwin's theory of biological evolution [26]. Evolutionary Algorithms (EA) are used to search for optimal solutions to optimization problems, considered NP-hard problems. Inspired by this, Holland developed Genetic Algorithms (GA), which are metaheuristics that adopt the process of natural selection of evolutionary algorithms. GA has three operations for repeatedly evaluating solutions during the evolutionary process: selection, crossover, and mutation [24].

To further improve the performance of the traditional GA and apply it to privacy preservation in data mining, several enhanced versions have been proposed. In the context of addressing PPARM using GA, Menaga and Saravanan [27] proposed a method by integrating GA with the FP-growth algorithm. FP-Growth extracts all frequent patterns from the original database and generates association rules using pre-defined minimum support and minimum confidence thresholds. These rules were evaluated using the proposed GA fitness function to produce the best solutions. Specifically, the fitness function was defined under two constraints: data privacy and utility.

Another study [28] proposed a constraint-based objective function for GA. The model, named Efficient Association Rules Hiding using a Genetic Algorithm (EARH-GA), incorporates a recursion process in its objective function to further improve the ARM process. The original database was mined to produce association rules, generating two subsets: a set of Sensitive Association Rules (SAR) and a set of Non-Sensitive Association Rules (NSAR). To preserve privacy, all SARs needed to be hidden. The EARH-GA model aimed to reduce computation time, hide SARs, and reduce NSARs while maximizing data utility and preventing the appearance of ghost rules. The GA encoded each transaction as a vector of indices (solution), and fitness criteria were used to examine the solution based on the lost NSARs. The EARH-GA model utilized GA in the ARM process by iteratively calling it to evaluate the candidate solution and produce new offspring through crossover and mutation.

In another study [29], a different formula for the GA fitness function was proposed to apply the hiding process to sensitive association rules using recursion. Three measures were defined: Availability, Sensitivity, and Conflict. The Availability and Sensitivity metrics were directly related to the fitness value, while the Conflict metric had a reverse relationship with an interdependent relationship with the total fitness values. These metrics served as inputs to the multi-objective fitness function of the proposed model. The Multi-Objective Strategy for Hiding Sensitive Association Rules (MOSAR) model iteratively executed the GA algorithm to hide only one sensitive association rule at a time.

The results of the performance studies [27], [28], [29] showed that the privacy of sensitive association rules could be further improved by hybridizing optimization algorithms. In 2022, Navale and Mali [30] developed an integration approach called the Genetic Algorithm with Crow Search Algorithm (GA-CSA) to address not only PPARM data sanitization but also the restoration process. The novel fitness function was defined using the sum of six weighting objectives: the rate of hiding failure, false rules, information hiding, modification, compression, and tampering. The GA-CSA model preserved the privacy of sensitive data by generating an optimal key during the sanitization process. Sanitized data could be restored on the receiving side by inverting the generated key. Key generation was achieved by adopting the Khatri-Rao encoding process and then combining the GA and CSA algorithms to produce the optimal key. However, the proposed model did not consider the generation of association rules.

In the same year, Darwish et al. [2] investigated the issue of maintaining privacy while utilizing data by focusing on negative ARM. While most studies are concerned with positive ARM, negative association rules can provide more insightful knowledge for analyzing healthcare data. Negative association criteria serve as helpful diagnostic techniques for physicians and social organizations. For instance, a negative relationship between one symptom and another indicates the absence of symptoms of a certain disease. The proposed model integrated the Apriori algorithm with a combination of Genetic and Tabu (TG) algorithms to hide negative sensitive information by deleting it. The Tabu-Genetic optimization framework utilized GA and Tabu Search (TS) to find a solution that consists of a population of points and examines these points. The GA generates a set of candidate solutions from the preliminary solutions that form the population. TS enhances the solutions locally, preventing the process from entering local minima. Two novel fitness formulas were defined to reduce side effects by obtaining the optimum support and confidence values, thereby improving population survival fitness.

To broaden the investigation of the performance of GA optimization, this review covered studies that addressed PPDM problems. In particular, two studies [31], [32] utilized GA in PPDM by adopting the Non-Dominating Sorting Genetic Algorithm (NSGA-II) framework proposed by Deb et al. [33]. In [31], the Non-Dominating Sorting Genetic Algorithm to Data Mining (NSGA2DT) model employed two strategies for hiding sensitive information while minimizing side effects. Although it is common to factor the three side effects and use them as objectives to define the fitness function, the NSGA2DT model added database dissimilarity as the fourth objective to achieve flexibility in transaction selection for deletion. Moreover, no initial parameters must be set for the NSGA2DT's fitness function. Thus, the retrieved solutions were not affected by the user preferences. Furthermore, a pre-large concept was applied to improve the number of iterations of the search process. The Fast Sorting Strategy (FSR) is used to efficiently find optimized transactions for deletion. In addition, because the NSGA2DT is a multi-objective algorithm, Pareto solutions are discovered to avoid the problem of local optimization in single-objective approaches.

The second study [32] adopted a general process for item deletion to introduce an enhanced version that performs optimization at two levels. The Non-Dominating Sorting Genetic Algorithm II for Item Deletion (NSGAII4ID) model first generated a set of candidate transactions to be deleted at the transaction level. The model then applied a greedy Set Cover Problem (SCP) algorithm at the item level, in which each candidate transaction is searched within to find an optimal subset of items to be removed. Thus, the NSGAII4ID algorithm deletes only certain items from transactions instead of all transactions, and reduces the changes in the original database. A three-objective fitness function was used. NSGAII4ID is a multi-objective algorithm that produces several Pareto optimal solutions and evaluates these solutions to determine the optimal solution in which side effects are minimized.

In 2015, Lin et al. [34] addressed the PPDM problem using two GA-based algorithms, focusing on transaction deletions in a dynamic database. Accordingly, the two proposed GA-based algorithms are simple Genetic Algorithms for Deleting Transactions (sGA2DT) and a pre-large Genetic Algorithm for Deleting Transactions (pGA2DT). The pre-large concept was adopted in the chromosome evaluation phase to minimize rescans and efficiently handle transaction insertions and deletions. However, the requirement to pre-calculate the number of transactions to be deleted might not be applicable to real-world cases. Thus, the proposed algorithms need further investigation to dynamically determine the number of deleted transactions using a multi-threshold approach. Hence, in 2021, Wu et al. [35] proposed two multi-threshold GA-based algorithms, mGA2DR and pMGA2DR, which are extended versions of the sGA2DT and pGA2DT algorithms proposed in [34], respectively. Using the multi-threshold concept, various threshold values can be assigned to various lengths of sensitive patterns to provide more protection and avoid patient data identification in the health dataset. The MGA2DR algorithm uses a simple multi-threshold GA-based approach, necessitating re-scanning the database at each iteration and reevaluating all three side effects. To overcome the high computational cost of the MGA2DR algorithm, another approach is proposed that uses the pre-large concept to reduce rescans during transaction deletion. The designed pMGA2DR reduces the computational cost for evaluation by defining a new threshold for pre-large values.

Lin et al. [36] applied the GA algorithm to another field similar to DM, namely High-Utility Itemset Mining (HUIM). HUIM calculates both the unit profits and the purchased number of elements for each transaction for mining purposes. As with traditional ARM and DM, the same issue of privacy and utility trade-off is present in HUIM, leading to an important variation in PPDM: Privacy-Preserving Utility Mining (PPUM). Accordingly, this paper proposed a GA-based approach called PPUMGAT to address the issue of PPUM. The proposed model hides sensitive itemsets using transaction deletion and the pre-large concept.

Although GA is considered to be the most widely-used EA in the field of handling optimization problems, it has a few drawbacks, such as difficulty in debugging, high computation, the possibility of running into local optima, and the need for initial conditions [24]. Consequently, other studies have utilized the general workflow of EAs and proposed novel algorithms to address the issues of privacy preservation and rule mining.

Yang et al. [37] proposed and developed a customized Multi-Objective Evolutionary Algorithm (MOEA) to address the trade-off dilemma between preserving privacy and maintaining utility in shared datasets. The proposed scheme, Privacy-Preserved Minable Data Publication (PMDP), formulates a fitness formula based on two conflict parameters: privacy and utility. Accordingly, PMDP first uses a preprocessing mechanism to filter irrelevant data, narrowing down the search space and accelerating the convergence rate. The MOEA then randomly generates candidate solutions using an innovative mutation method. This approach guarantees the production of a variety of solutions, preventing them from falling into a local optimum. Furthermore, MOEA provides an elite learning strategy to generate solutions by learning from elite sets, known as Pareto-optimal solutions.

In addition to privacy preservation, some researchers have applied EAs to enhance the performance of the ARM process. For example, Wang et al. [38] applied an evolutionary optimization algorithm to ARM, introducing an alternative representation of ARs known as Functional ARs (FARs). Conventionally, ARM deals with discrete datasets and uses a discretization process when handling continuous datasets. Variable values are converted into intervals. However, the FARs presented in this study can handle continuity in a dataset without converting variables into intervals, eliminating the need for a discretization process. Furthermore, FARs are based on an Artificial Neural Network (ANN), serving as the foundation for identifying associative relations hidden in the dataset. In addition, ANN can predict the Right-Hand Side (RHS) variables, increasing the granularity of AR. In this study, FARs were mined using a Co-operative, Co-evolutionary Functional Association Rule Mining (CCFARM) algorithm. The CCFARM contains two sub-populations, one for the rules and the other for the ANN. A valid solution consists of a candidate interesting rule and a predictive ANN. Another study [39] utilized GA to enhance the Apriori algorithm for ARM, improving the efficiency and accuracy of the algorithm with better outcomings. Table 3 summarizes and compares existing EA approaches.

2) SWARM INTELLIGENCE(SI)

Swarm intelligence (SI) simulates the collective behavior of a natural or artificial system, relying on the principles of decentralization and self-organization. In SI systems, populations consist of particles interacting with each other and their surroundings. Nature has been the basis of inspiration for SI algorithms, especially biological metaphors such as the structure, components, and behaviors of natural systems. These include, bird swarms, ant colonies, animal herding, bacterial growth, and whale hunting [24], [40]. Furthermore, various SI-based metaheuristic algorithms have been developed to solve optimization problems. In this review, seven different SI-based metaheuristic algorithms were utilized to address privacy preservation and data utility issues in the mining process. Thus, the majority of the selected papers proposed solutions based on different variations in the SI algorithms. Seven SI-based algorithms were introduced and discussed in selected papers, namely Particle Swarm Optimization (PSO), Ant Colony System (ACS), Artificial Bee Colony (ABC), Crow Search Algorithm (CSA), Whale Optimization Algorithm (WOA), FruitFly Optimization Algorithm (FOA), and Elephant Herding Optimization (EHO). Details of these algorithms are presented in the following subsections.

TABLE 3. Metaheuristic-based EA algorithms.

Ref.	Year	Model	Approach	Advantages	Disadvantages
[27]	2022	GA-PPARM	GA, FP-Growth	New Fitness Function Based On Support And Confidence	Low Privacy
[28]	2018	EARH-GA	GA	Efficient Computation Time Using Recursion	Less Practicality; Hides One Rule At Each Run
[29]	2016	MOSAR	GA	High Accuracy Of Generating Sensitive Rules	Less Practicality; Hides One Rule At Each Run
[30]	2022	GA-CSA	GA,CSA	Optimal Key Generation For Sanitization And Restoration	Did Not Cover The Generation Of Sensitive Rules
[2]	2022	Tabu-Genetic	GA,TS	Enhanced Ability To Hide Negative Sensitive Rules	Not Suitable For Distributed Data Mining
[31]	2022	NSGA2DT	GA	Fast And Flexible Selection Of Optimal Transactions To Be Deleted	High Computation Cost
[32]	2022	NSGAII4ID	GA	Minimal Changes Are Performed On The Original Database	Poor Results In Minimizing The Missing Cost
[34]	2022	SGA2DT, PGA2DT	GA	Short Execution Time	The Number Of Transactions To Be Deleted Is Not Dynamic
[35]	2021	MGA2DR, PMGA2DR	GA	Highly Efficient In Hiding The Sensitive Patterns With Varying Lengths	High Computational Cost For A Large Dataset
[36]	2017	PPUMGAT	GA	Hides All Sensitive HUIs	Not Accurate Since Some False HUIs Can Appear In The Sanitized Dataset
[37]	2022	MOEA	EA	Provides A Diversity Of Candidate Solutions, Faster Convergence Speed	Low Performance In Term Of Optional Hiding
[38]	2017	CCFARM	EA	Suitable For Handling Continuity In A Dataset	High Computational Cost

a: PARTICLE SWARM OPTIMIZATION (PSO)

The most widely used metaheuristic-based algorithm to search for optimal solutions is Particle Swarm Optimization (PSO), which was proposed by Kennedy and Eberhart [40].PSO simulates the flocking activity of birds, where each particle in the swarm represents a candidate solution. Each particle has attributes–velocity and position–that define its movement. [41]. This movement is influenced by two parameters: pbest, the particle's best-known position locally, and gbest, the current known position across the search space. These are updated to moving particles in the search-space during each iteration to guide the swarm towards the best solutions [5]. As shown in figure 6, the majority of the studies in this review utilized the PSO algorithm to propose their models, indicating the scalability and adaptability of the PSO algorithm.

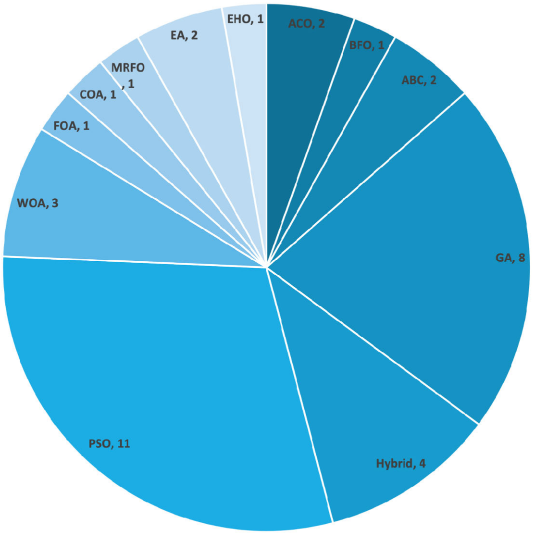

FIGURE 6. The existing pparm metaheuristic-based algorithms.

Lin et al. [42] suggested a general PSO-based solution for PPDM. The Particle Swarm Optimization to Data Mining (PSO2DT) model aims to sanitize sensitive itemsets through transaction deletion while minimizing side effects. The fitness function of the PSO2DT model is built using the weighted sum of three objectives: Hiding Failure, Missing Cost, and Artificial Cost. Usually, the threshold that specifies the item sets to be hidden is either a fixed pre-defined threshold or a multi-value threshold. However, the PSO2DT model uses a user-specific minimum support threshold, which is a new method for setting a threshold that can be manually adjusted by users or experts. For example, the weight of Hiding Failure is set higher if the user wants to hide more sensitive information. The missing cost weight is assigned a higher value when preserving non-sensitive information that is important to the user. The Artificial Cost weight is set higher when the user wants to minimize the appearance of the artificial information. Another strength of the model is that it adopted the concept of discrete PSO, which can reduce the three side effects by finding transactions that are deleted to protect sensitive information. In other words, the particle size in the PSO algorithm is automatically determined by the maximum number of transactions to be deleted, which was generated by the model. The concept of pre-large is applied to buffer infrequent item sets that have a large support value, which may become more common during transaction deletion. This application of the pre-large concept helps skip many database scans during side-effect evaluation, accelerating the process. Overall, the PSO2DT model demonstrates good performance in preserving privacy while generating more association rules through transaction deletion in the PSO algorithm.

Because of the variety of lengths of the patterns, having one fixed pre-defined minimum support threshold is not practical for determining how sensitive information of different lengths is hidden. Moreover, sensitive information with long patterns can be easily identified. Thus, different thresholds are required for real-world applications. Wu et al. [43] extended the PSO2DT [42] model by introducing a multi-threshold technique and proposed a new model called Multi-threshold Particle Swarm Optimization to Data Mining (MPSO2DT). A multi-threshold technique allows the model to protect more critical items and minimize side effects as much as possible. The model adjusted the minimum support threshold value to accommodate various patterns. That is, a loose threshold (higher threshold) is set for short patterns and a tight threshold (lower threshold) is set for long patterns. The minimum support threshold was automatically obtained by applying a normal distribution function during the experiments. The MPSO2DT's fitness function is the same as pPSO2DT model [42]. However, MPSO2DT uses a new technique to compare any two rounds that have an equal fitness value and then choose the best one. The model first finds large and pre-large itemsets by performing the Apriori process. The projected database is then generated by applying the dynamic minimum support threshold function. A set of swarm particles representing each possible solution is generated and evaluated using the weighted fitness function. The size of each particle was set to a suitable number of deleted transactions to hide sensitive information. In general, MPSO2DT showed a better performance than the pPSO2DT model. The MPSO2DT model can be improved by adopting a multi-objective approach to cover non-dominated constraints. However, both MPSO2DT and pPSO2DT models can be improved to handle the problem of PPDM as a multi-objective optimization problem, which is more applicable in real-world applications.

Multi-objective optimization is designed to achieve a trade-off between multiple conflicting objectives; hence, it identifies many non-dominated Pareto optimal solutions, all considered equally good. The goal is to choose solutions that satisfy the trade-offs for the different objectives. Accordingly, Wu et al. [44] presented a solution for the PPDM to find a trade-off between four objectives: hiding failure, missing cost, artificial cost, and database dissimilarity. Their new approach, called Grid-based Multi-objective Particle Swarm Optimization (GMPSO), extends the traditional PSO algorithm by adapting the grid method. The primary benefit of the grid approach is to maintain an external archive that keeps the particles at the limit of the search space. From the perspective of PSO, the data are sanitized by modifying the global best and personal best attributes to obtain a better diversity of non-dominated Pareto solutions. The strategy for determining the personal best (pbest) of each particle is obtained using the non-domination relation. The global updating strategy for gbest is obtained using the grid-based method, which assigns probabilities to each particle and then selects one as gbest based on a random function. The model utilizes the pre-large concept to handle dynamic data mining, particularly in the case of transaction insertion, thus speeding up the computations.

Another study addressing the PPDM using the PSO algorithm was presented by Jangra and Toshniwal [45]. In this study, the proposed Victim Item Deletion PSO (VIDPSO) model utilized the PSO algorithm to hide sensitive information from a dense dataset with minimal data loss and miss cost. A dense dataset contains cells populated with non-zero values and is measured using the density factor. VIDPSO examined the impact of the support count value on the efficiency of the item deletion process.

In the context of PPARM, Yang and Liao [46] presented a new optimized Sanitization Approach for Minable Data Publication (SA-MDP) to address the PPARM problem. Because the evolution method of any metaheuristic-based algorithm significantly affects the search space, the original PSO algorithm was modified by adopting two mechanisms to update the location of the particles. The directional learning and random splitting updating methods are the strengths of the SA-MDP algorithm because they can improve the exploitation and exploration abilities of the algorithm, respectively. The directional learning and random splitting updating methods are strengths of the SA-MDP algorithm because they can improve the exploitation and exploration abilities of the algorithm, respectively. The directional learning method enables the particles to update their velocity and position by learning from the best solution (gbest). On the other hand, the SA-MDP model adopted the concept of particle splitting to ensure that the particles are not trapped in a local optimum. Specifically, the particles produce several child particles through splitting. Furthermore, the convergence rate can be guaranteed to find the optimum solution because of the learning and preprocessing methods adopted in the SA-MDP model. A pre-defined fixed minimum support threshold was used for mining the association rules. Furthermore, the fitness function of the SA-MDP model was defined by three typical side effects as well as a new metric, namely, the optimal sanitation distance.

Another study on PPARM proposed a hybrid approach that combined the conventional Apriori algorithm with the PSO algorithm [4]. Herein, association rules are mined by Apriori and identified as sensitive or non-sensitive using a pre-set minimum support threshold (MST). The selection of key transactions and victim items was performed before the rule-hiding process. After defining the population of particles, the model constructs a discrete binary space to minimize the search space, thereby enhancing the performance of the process. The objectives of the fitness function were obtained to mine the database during the iterative process. Although the proposed model succeeded in improving the efficiency and effectiveness of hiding sensitive rules, it could not be applied to a database with a dynamic update.

The remaining PSO-based studies focused on improving the efficiency of mining association rules without mentioning privacy concerns. However, it is good to cover and discuss them, as they provide valuable contributions to the field of ARM and can be used for further investigations in PPARM studies. Moreover, studies have applied PSO algorithms to mine different types of data, such as manufacturing systems [47], underground coal mine monitoring systems [6], malware detection systems [5], and big data environments [48].

Kou and Xi in [47] proposed a Binary Particle Swarm Optimization for Association Rule Mining (BPSO-ARM) approach to explore the association rules between machine capacities and product characteristics. The novelty of BPSO-ARM lies in the absence of the need for pre-defined thresholds, making it more practical for real-world industrial applications. Additionally, a new overlapping measurement method is introduced to eliminate low-quality rules by evaluating the similarity between association rules with a maximum similarity threshold. The fitness function, based on Support and Confidence, assesses the interestingness of a rule. The binary PSO algorithm proves convenient for solving discrete problems, particularly in mining association rules, as it accelerates database scans and the calculation of particle vector values.

Muduli et al. [6] proposed another binary PSO-based model for effectively monitoring underground coal mines. This optimized fuzzy system, an extension of their previous work [49], employs the BPSO algorithm to enhance the detection accuracy and efficiency of a fuzzy-logic-based fire-monitoring system. The fitness function considers two main objectives: reducing redundant rules and improving the accuracy of the fuzzy system. To address conflicting objectives, the weight sum approach is employed, resulting in a single-objective optimization function. Simulation results demonstrate that the proposed model generates more efficient optimal fuzzy rules compared to the previous model, at a lower fitness cost.

Researchers have explored integrating the PSO algorithm with other optimization techniques. Su et al. [48] integrated FP-growth with PSO in a big-data environment to replace traditional methods of setting support and confidence thresholds. The proposed PSOFP-growth algorithm determines an optimal support value to evaluate obtained rules, incorporating information entropy in the fitness function to assess rule effectiveness and filter out invalid rules.

Adebayo and AbdulAziz [5] proposed a PSO algorithm for malware detection on the Android platform. They developed an Apriori Association Rule and adaptive Particle Swarm Optimization (AAR-aPSO) approach to enhance detection rates. Large itemsets, extracted from features with a pre-defined minimum support threshold, are used to generate association rules for features surpassing the minimum confidence threshold. The fitness function, modified based on three objectives (feature length, population, and classification quality), extracts the best features from the search space, generates signature rules, and classifies applications as benign or malware.

b: ANT COLONY SYSTEM (ACS)

The Ant Colony System (ACS) is a popular swarm intelligence algorithm, known as one of the well-established ant colony optimization methods. Initially proposed by Dorigo and Gambardella [50], ACS is an extension of the original Ant System, drawing inspiration from ant colonies and their behaviors. Ants exhibit intelligent foraging behavior by taking the shortest route from a food source to the nest, serving as the basis for ACS–a search approach that relies on probability and simulates ant foraging behavior [51]. In the literature, two studies utilized ACS to address the PPDM for different applications, such as the medical environment and IoT network environment.

Wu et al. [52] addressed the PPDM issue by mining sensitive rules in an identifiable health dataset containing patterns of varying lengths. Traditionally, PPDM models use a pre-defined single support threshold for discovering frequent itemsets, which may be unsuitable for health information due to the easy identification of long-length sensitive rules using specific attributes. The proposed Ant Colony System to Data Mining (ACS2DTM) model utilized a multi-threshold function, adjusting the minimum support threshold value and defining different thresholds for different pattern lengths. This model, combining the multi-threshold function with the ACS-based algorithm, aimed to hide sensitive rules while mitigating side effects that could affect sanitization process quality. Unlike other ACS-based algorithms with a designated destination to terminate the search process, the ACS2DTM model utilized a special ant routing graph, reaching termination when an ant concludes a tour after selecting the minimum number of records that allows hiding all sensitive patterns. The pre-large concept was employed to optimize new transitions and avoid multiple database scans.

Lin et al. [53] utilized ACS algorithms to enhance the security of 6G IoT networks and to protect confidential and sensitive information from exposure. Specifically, the integration of the IoT industry with 5G\6G networks can increase the efficiency of data exchange and sharing between the connected devices. Simultaneously, there is an urgent need for privacy techniques to protect sensitive information from illegal exposure and leakage. Thus, this study employed the PPDM approach in the context of 6G IoT networks to sanitize confidential data before allowing them to be shared and exchanged. The study proposed a model based on the Pareto Ant Colony Optimization (PACO) approach, which differs from the conventional approach in terms of the number of pheromone vectors and random choices of weights for the fitness objectives. The PACO2DT model minimizes the three side effects of the PPDM by using transaction deletion with a multi-objective fitness function that considers these side effects equally. Each node in the ant routing graph in the PACO2DT model represents a transaction in the dataset. This model deletes transactions that contain sensitive information. The tour continues until each ant reaches the maximum number of deleted transactions MaxDT, that is, the termination condition. Moreover, PACO2DT reduced the computational cost by utilizing a pre-large concept. Above all, the model efficiently obtained a set of well-known Pareto solutions by saving global optimal solutions in an external archive model.

c: ARTIFICIAL BEE COLONY (ABC)

The Artificial Bee Colony (ABC) algorithm was proposed by Karaboga and Basturk [54], which mimics the foraging behavior of honey bees to find food sources. It is a successful swarm intelligence algorithm for solving optimization problems. The ABC algorithm divides bees into three groups: employed bees responsible for specific food sources, onlooker bees as decision-makers choosing food sources, and scout bees as employed bees that abandon their food sources [55]. ABC has been successfully applied to binary problems and many variant versions have been proposed, such as XOR-based algorithms [56].

The selected studies discussed the PPARM problem and solved it using variant modifications of ABC algorithms. For instance, a study [57] utilized a modified version of ABC called Discrete ABC (DisABC) and proposed an advanced approach called Improved Binary ABC (IBABC). The DisABC algorithm was developed on the basis of the Jaccard coefficient similarity measure. Therefore, it can be applied to the ARM process to select suitable transactions for sanitization. In this study [57], the IBABC algorithm is presented, in which the exploration and exploitation phases are balanced using two modifications. Moreover, integration with the ABC4ARH algorithm was performed to delete victim items from sensitive transactions that were selected by the IBABC algorithm. Although ABC4ARH protected the sensitive association rules with a small number of side effects, the results showed that ABC4ARH is less scalable than other related algorithms. The reason for this drawback is that the phases of the employee and onlooker bees require high execution time. Another study attempted to address this by applying parallel processing to ABC4ARH.

Evolutionary PPARM approaches often require high computation time due to the large number of iterations in the evolution phase. Accordingly, the study [58] proposed a new way to improve the performance of PPARM algorithms and overcome the limitations of PPARM approaches through parallelization using the GPU platform. The GPU-based Evolutionary Privacy-Preserving (GEPP) model is an extension of the ABC4ARH approach [57]. Herein, the proposed model exploits the benefit of GPU hosts in two specific phases of ABC4ARH that require a large amount of computation time: index list generation and fitness computation. First, the generation of index lists is parallelized using the Parallel Indexing Machine (PIM) mechanism, in which dataset scans are distributed between GPU blocks, and the related lists are generated. The second strategy concerns the parallelization of the transaction selection phase, where the search process takes place on the CPU hosts. The evaluation of the solutions using the fitness function was performed on the GPU hosts. The results showed that the proposed algorithm outperformed the ABC4ARH algorithm.

Regarding the matter of parallelization using GPU platform, Telikani et al. [59] leveraged the GPU technology to address the PPDM. They introduced an advanced data sanitization algorithm called HEDS4IoT for managing extensive streaming IoT data on edge computing platforms. HEDS4IoT offered index retrieval lists for identifying sensitive transactions and the necessary components for fitness calculations. Additionally, the model facilitated parallelized fitness computation, processing index lists efficiently on GPU devices. When compared to ABC4ARH algorithm [57] and similar algorithms, HEDS4IoT demonstrated significant advantages in reducing side effects, achieving high speed improvements, and better scalability, especially with dense datasets. Consequently, the GPU platform exhibits promising results in real-time privacy protection for PPDM through evolutionary algorithms.

d: CROW SEARCH ALGORITHM (CSA)

The Crow Search Algorithm (CSA) is another swarm intelligence optimization algorithm inspired by the behavior of a flock of crows in hiding and retrieving food. Introduced by Askarzadeh in 2016 [60], CSA stands out for its advantages over other nature-inspired algorithms, offering a simpler structure, fewer control parameters, and easier implementation [61]. Although CSA has interesting strong features, it has some drawbacks, including slow convergence and a tendency to fall into the local optima. To address these limitations, researchers often opt to combine CSA with other optimization algorithms [30].

e: WHALE OPTIMIZATION ALGORITHM (WOA)

Whale Optimization Algorithm (WOA) is another optimization algorithm inspired by nature which was introduced by Mirjalili and Lewis [25]. The WOA simulates humpback whales and their hunting strategies. WOA simulates the foraging behavior of humpback whales by searching for optimal solutions from a set of candidate solutions. WOA has been applied to various types of optimization problems to further investigate its efficiency and scalability.

Sharmila and Vijayarani [62] applied WOA to fuzzy association rule mining in their Fuzzy Rules Using Whale Optimization Algorithm (FRUWOA) model. This model leverages the three phases of WOA to generate frequent items and rules. In the encircling the prey phase, recurrent items are discovered from a dimensionality-reduced database. The exploitation phase, resembling a bubble-net attack, identifies item occurrences and their membership values, which are updated within specified ranges using a spiral updating position mechanism. Finally, fuzzy rules are generated in the exploration phase.

Another application of the WOA to a different optimization problem was introduced by Karlekar and Gomathi [63]. They utilized the WOA to classify privately preserved medical data in a cloud environment. The OW-SVM model integrates an ontology technique with a whale optimization algorithm based on Support Vector Machine SVM to produce an effective classification of medical data. The proposed model first generates a privately preserved medical database using the Kronecker product-based bat algorithm. It builds an ontology structure of different heart diseases by producing rules that correspond to the features of the records in the database and combining them with SVM for classification. Here comes the adaptation of WOA algorithm, the novel multi-objective fitness function of WOA plays a vital role in finding the optimal kernel space for the classification and optimally selecting the feasible kernel parameters. These four parameters formed the basis of the kernel function of the SVM for classification.

On the other hand, Shailaja and Rao [64] introduced a PPDM hybrid model for generating the optimal key to sanitize and restore the datasets. This model combines WOA and PSO algorithms to overcome limitations of traditional versions, addressing issues like run-time and local optimal solutions. TU-WPA employs a novel method for updating the evaluation phase using a Trial Value (TR). Herein, the Trial-based Update of the Whale and Particle Swarm Algorithm (TU-WPA) model sets an initial value of zero. The current value was then compared with the previous value in each iteration. This evaluation method alters the trial value by one whenever the current fitness value is not improved compared with the previous one. Otherwise, the trial value remains the same. All iterations end when the evaluation of the fitness value and update trial value are completed. This is different and more efficient than the updating condition used in traditional WOA and PSO, in which the update occurs based on an arbitrary number and locations of the particles, respectively.

f: FRUITFLY OPTIMIZATION ALGORITHM (FOA)

The FruitFly Optimization Algorithm (FOA) is a nature-inspired swarm intelligence optimization algorithm that was first proposed by Pan [65] to address the optimization issues. FOA imitates the food search activities of a swarm of fruit flies, incorporating their sensing and perception (osphresis and vision). Despite having desirable features, such as a simple structure and good performance, FOA exhibits drawbacks in high-dimensional and large-scale optimization problems, including route instability and lower convergence speed. Thus, researchers have focused on proposing new modifications to FOA, aiming to enhance its search ability and introduce new parameters [66].

FOA algorithms were utilized in diverse data-mining problems. For instance, Reddy et al. [67] applied FOA to fuzzy association rule mining, optimizing the identification of frequent items and fuzzy rules. The model demonstrated excellent performance in reducing redundant rules. Dhinakaran and Prathap [68] utilized FOA to address PPARM by introducing the FruitFly Whale Optimization Algorithm (FWOA) model. In traditional FOA, random generation of directions for particles may lead to an immature convergence problem and potential entrapment in local optima. To overcome these limitations, the proposed FWOA model integrates the whale hunting technique from the Whale Optimization Algorithm (WOA) to guide fruitfly agents toward the food source. This approach replaces the random selection of distance and direction with a more directed strategy, enhancing the effectiveness of the optimization process.

g: ELEPHANT HERDING OPTIMIZATION (EHO)

Another bio-inspired optimization algorithm is used in the context of the PPARM, Elephant Herding Optimization (EHO), proposed by Wang et al. [69], to simulate the behavior of elephants. The control of the optimization parameters mimics the way elephants are herded, that is, by updating and separating the operators. Separation of these operators can enhance the population diversity of the search space. Accordingly, Gopagoni and S K [70] proposed a modified version of EHO by integrating it with a distributed concept. The distributed Adaptive Elephant Herding Optimization (AEHO) generate the privacy rules for grid-based ARM. Although the ARM process is based on confidence and support thresholds, the AEHO utilizes other factors for ARM: probability-based confidence and holo-entropy. Another study [71] utilized a swarm-based algorithm called Bat algorithm. The Bat algorithm for ARM (BatMiner), was developed to mine association rules from sports training sessions datasets. Table 4 and 5 summarizes and compares the existing SI approaches.

3) OTHER BIO-INSPIRED ALGORITHMS

There are four other additional bio-based algorithms were introduced and discussed in selected papers: Lion Optimization Algorithm (LOA), Cuckoo Search Algorithm (CSA), Bacterial Foraging Optimization (BFO), and Manta-Ray Foraging Optimization (MRFO). Details of these algorithms are presented in the following subsections.

a: LION OPTIMIZATION ALGORITHM (LOA)

The Lion Optimization Algorithm (LOA) is another metaheuristic algorithm that simulates natural phenomena. LOA was presented by Yazdani and Jolai [72] to mimic the unique lifestyle of lions, including their cooperative characteristics such as prey-catching, reproduction, and territory-marking.

In a study by Menaga and Revathi [73], the WOA was integrated with LOA to mine association rules while preserving privacy. The proposed model comprises two phases. Firstly, the WOA is applied for ARM, utilizing a fitness function with support and confidence thresholds as objectives to validate the mined rules. Subsequently, a modified version of the LOA algorithm, incorporating the Least Mean Square (LMS), is introduced as the Least Lion Optimization Algorithm (LLOA). LLOA plays a crucial role in privacy preservation during the mining process by generating a secret key. These keys ensure that sensitive information is hidden in a sanitized database. The study's novelty lies in the separation of the two objectives of Privacy-Preserving Association Rule Mining

TABLE 4. Metaheuristic-based SI algorithms.

Ref.	Year	Model	Approach	Advantages	Disadvantages
[42]	2016	PSO2DT	PSO	High flexibility, fast computation	Poor results for sparse datasets
[43]	2021	MPSO2DT	PSO	Better performance for ARM	No guarantee to protect all necessary information, high execution time
[44]	2019	GMPSO	PSO	Enhanced the diversity of the candidate solutions, short execution time	inefficient for database dissimilarity
[45]	2020	VIDPSO	PSO	Highly efficient for dense datasets, optimal masking	High execution time
[46]	2022	SA-MDP	PSO	Improved the effectiveness of exploitation and exploration processes	Relatively high execution time
[4]	2021	BPSO-Apriori	PSO	High efficiency of hiding ability	Manually adding threshold values
[47]	2018	BPSO-ARM	PSO	High applicability, dynamic thresholds	Missing some rules
[6]	2019	BPSO-fuzzy	PSO	Enhanced detection accuracy, minimized the redundant rules	Limited to a low number of rules
[48]	2019	PSOFP-growth	PSO,FP-growth	Improved the effectiveness of the fitness function	Algorithm is affected by the value of support
[5]	2019	AAR-aPSO	PSO, Apriori	Good results in accuracy and detection rates	Less applicable, high execution time
[52]	2021	ACS2DTM	ACS	Efficient in generating better fitness values	Not suitable for large data sets
[53]	2021	PACO	ACS	Provides a diversity of candidate solutions	Poor results for dense datasets
[57]	2020	IBABC	ABC	Efficient in hiding rules, dynamic thresholds	High execution time
[58]	2021	GEPP	ABC	Better performance using GPU parallelization	Suitable only for dense data sets
[59]	2023	HEDS4IoT	EA	lower side effects	not efficient for Spark datasets
[30]	2022	GA-CSA	CSA,GA	Optimal Optimal key generation	Did not cover the generation of sensitive rules
[63]	2018	OW-SVM	WOA	Improved data confidentiality, optimal selection of fitness parameters	Missing some values

TABLE 5. Metaheuristic-based SI algorithms, cont.

Ref.	Year	Model	Approach	Advantages	Disadvantages
[42]	2016	PSO2DT	PSO	High flexibility, fast computation	Poor results for sparse datasets
[64]	2022	TU-WPA	WOA, PSO	Efficient at hiding rules with minimum side effects, improved the generation of the optimal key	High execution time
[67]	2022	IFRS-FFA	FOA	Smaller search space using dimensionality reduction	Poor results for accuracy of ARM
[68]	2022	FWOA	FOA	Higher accurate results, increased data privacy	Impractical for some cases
[70]	2020	AEHO	EHO	High accuracy and privacy, less computation time	Not scalable

TABLE 6. Metaheuristic-based bio-inspired algorithms.

Ref.	Year	Model	Approach	Advantages	Disadvantages
[73]	2018	LLOA	LOA, WOA	Effectively preserve privacy by generating an optimal secret key	Not suitable for large datasets
[76]	2016	COA4-ARH	CSA	Enhanced preservation while minimized the side effects	Low performance in data utility
[77]	2020	Hybrid	CSA	Better hiding capacity using perturbation	Complex and tedious
[21]	2018	BaCARO-II	BFO	Enhanced the search space exploration	Not suitable for all types of stream data models
[81]	2022	FMRF	MRFO	High accuracy rate and low time and memory cost	limited to a low number of rules

(PPARM): the WOA-based rule mining algorithm for ARM and the LLOA-based sanitization for privacy preservation. This separation provides both privacy and improved search performance.

b: CUCKOO SEARCH ALGORITHM (CSA)

The Cuckoo Search Algorithm (CSA), initially proposed by Yang and Deb [74] for solving optimization problems, was later enhanced by Rajabioun [75] to tackle continuous nonlinear optimization problems. The algorithm models the reproductive mechanisms of cuckoo birds, with the Cuckoo Optimization Algorithm (COA) commencing with a random population of grown-up cuckoos and eggs. The ability to survive defines the next generation, continuing until only one cuckoo society remains, representing the optimal solution in the search process.

Afshari et al. [76] introduced COA for Sensitive Association Rules Hiding (COA4ARH) within the context of PPARM. This model aims to minimize three side effects while maintaining the privacy of sensitive association rules. A novel method is employed to prevent being trapped into local optima by adjusting solutions to be more similar and closer to the optimal solution.

In another study [77], COA was combined with Particle Swarm Optimization (PSO) and Chemical Reaction Optimization (CRO) to enhance the performance of association rule hiding. The model achieves privacy preservation by concealing association rules through perturbation instead of data sanitization. Perturbation involves adding ‘noise’ to a database, protecting its records and generating a perturbed dataset. This is distinct from data sanitization, which preserves privacy by removing sensitive information from a database, producing a sanitized dataset. The model integrates four modules: (a) Association Rule Mining (ARM) for rule generation, (b) a perturbed dataset processing the original and perturbed datasets, (c) a modified PSO algorithm determining gbest values for Sensitive Association Rules (SAR) and pbest values for Non-Sensitive Association Rules (NSAR), and (d) a modified C4ARH calculating the fitness function to evaluate sensitive association rules.

c: BACTERIAL FORAGING OPTIMIZATION (BFO)

Bacterial Foraging Optimization (BFO) is a bio-inspired optimization algorithm that emulates the foraging strategy of Escherichia coli (E. coli). It replicates mechanisms such as chemotaxis, reproduction, elimination, and dispersion. First proposed by Passino [78], BFO has seen various modifications from researchers to tackle diverse discrete and continuous optimization problems [79]. For instance, in selected papers, Cunha et al. [21] introduced an algorithm called Bacterial Colony Association Rule Optimization-II (BaCARO-II) that adopted the BFO algorithm. This model leverages intra-cellular and extra-cellular communication principles inspired by bacterial foraging mechanisms to solve association rule-mining tasks.

d: MANTA-RAY FORAGING OPTIMIZATION (MRFO)

Manta-Ray Foraging Optimization (MRFO), proposed by Zhao et al. [80], is a bio-inspired optimization algorithm that replicates the food exploration behavior of manta rays. MRFO introduces three distinct foraging mechanisms: chain foraging, cyclone foraging, and somersault foraging, each contributing to different aspects of the search process. In a study by Lakshmi and Krishnamurthy [81], MRFO was adopted to create a Fuzzy Manta Ray Foraging (FMRF) optimization algorithm for generating association rules from large itemsets. In the FMRF approach, the chain foraging phase retrieves recurrent itemsets from transactional databases. In the cyclone foraging phase, items are updated by calculating their occurrences with respect to membership values. In the last somersault phase, fuzzy rules are generated. Table 6 summarizes and compares the existing bio-inspired approaches.

B. NON-METAPHOR-BASED METAHEURISTICS

Many other metaheuristic-based algorithms do not use any simulation to optimize their search strategies, such as Variable Neighborhood Search (VNS) and Partial Optimization Metaheuristic Under Special Intensification Conditions (POPMUSIC) [24]. However, in this study, only one non-metaphor based metaheuristic algorithm was identified which is Tabu Search (TS).

1) TABU SEARCH (TS)

Tabu Search (TS) is a metaheuristic search method first proposed by Glover and McMillan [82]. The algorithm is rooted in the local search method, applied to mathematical optimization. TS penalizes movements that revisit previously explored search spaces (referred to as "tabu"). This technique ensures exploration of the solution space while avoiding entrapment in local optima. In the literature, TS has often been integrated with other metaheuristic algorithms, such as the Tabu-Genetic (TG) optimization framework proposed by Darwish et al. [2], which has been discussed in the GA section.

C. DIVERSIFICATION OF THE EXISTING METAHEURISTIC ALGORITHMS

Original metaheuristic algorithms have undergone improvements by researchers through various modifications and techniques. This section discusses the variety of metaheuristics identified in the studies selected in this review.

D. MULTI-OBJECTIVITY IN METAHEURISTICS

Traditionally, most metaheuristic-based algorithms have been employed to optimize problems involving a single-objective fitness function. However, real-world applications often require decisions based on multiple conflicting objectives [83]. Applying traditional single-objective metaheuristics to such problems is inadequate. Therefore, the concept of multi-objective functions has been incorporated into metaheuristics. A multi-objective algorithm generates several optimal solutions, known as Pareto-optimal solutions, which are then evaluated to determine the best solution [32].

Moreover, the development of a multi-objective algorithm requires the redesign of the fitness function of the metaheuristic to accommodate the multiple conflicting objectives inherent in optimization problems, such as PPARM. In PPARM and PPDM problems, the objectives of the fitness function generally represent the side effects stemming from the trade-offs between preserving privacy and utilizing mined data. These side effects include hiding failure, missing cost, and artificial cost. Table 7 presents the various parameters used to define the fitness functions of the proposed algorithms.

TABLE 7. Multi-objective metaheuristic-based algorithms.

Field	Ref.	Meta-heuristic	No. Obj.	Fitness Function Objectives
PPARM	[29]	GA	3	Availability, Sensitivity, and Conflict
PPARM	[30]	GA, CSA	6	The rate of HF, false rules, information hiding, modification, compression and tampering
PPARM	[4]	PSO	3	hiding failure (HF), lost rules (LR), and average hide deviation (AHD).
PPARM	[46]	PSO	3	Hiding-Failure Rate (HFR), Lost-Rule Rate (LRR), Fake-Rule Rate (FRR)
PPARM	[64]	PSO, WOA	4	HF, DM, IP, and FR
PPARM	[73]	WOA, LOA	2	WOA: support, and confidence, LLOA: privacy and utility.
PPARM	[37]	EA	2	privacy and utility
PPDM	[52]	ACO	3	Weighted Sum of HF, MC and AC
PPDM	[53]	ACO	3	Minimum of HF, MC and AC
PPDM	[31]	GA	4	Weighted Sum of HF, MC, AC and Dis
PPDM	[34]	GA	3	Weighted Sum of HF, MC and AC
PPDM	[44]	PSO	4	Weighted Sum of HF, MC, AC and Dis
PPDM	[63]	WOA	3	Accuracy, Sensitivity, Specificity
PPDM	[36], [32], [35]	GA	3	Weighted Sum of HF, MC and AC

E. HYBRIDIZATION OF METAHEURISTICS

It is evident from the above discussion that no single metaheuristic can independently address all types of optimization problems. Moreover, the integration of two metaheuristics into a unified approach has shown significant improvements in the efficiency of the search process. Consequently, future studies may explore the hybridization of metaheuristics as a promising avenue. Hybridization aims to combine the strengths of multiple algorithms, creating a hybrid model while mitigating inherent limitations. Typically, the hybridization of metaheuristics enhances either the computational speed or the accuracy of the search process. In this review, seven studies proposed hybrid models based on metaheuristics, as presented and explained in the previous sections. These are GA-CSA [30], PSO-WOA [64], WOA-LOA [73], GA-TS [2], HFPSO [84], PSO-CRO-CSA [77], and FOA-WOA [68].

F. DISCRETIZATION OF METAHEURISTICS

In general, metaheuristic-based algorithms typically address continuous optimization problems. However, they are insufficient for binary and integer-valued combinatorial problems. For this reason, many versions of these algorithms have been developed using the concept of discretization to accommodate combinatorial problems, such as the knapsack problem. Many discretization methods are utilized to develop metaheuristic-based algorithms that convert a continuous search space into a binary one, such as Random Key (RK), Smallest Position Value (SPV), and Sigmoid Function (SF) methods [24]. Table 8 summarizes the related studies that proposed metaheuristic-based algorithms to handle discrete optimization problems.

TABLE 8. Discrete metaheuristic-based algorithms.

Field	Ref.	Metaheuristic
PPARM	[57]	ABC
PPDM	[44]	PSO
PPDM	[42]	PSO
PPARM	[4]	PSO
PPDM	[45]	PSO
ARM	[81]	MRFO

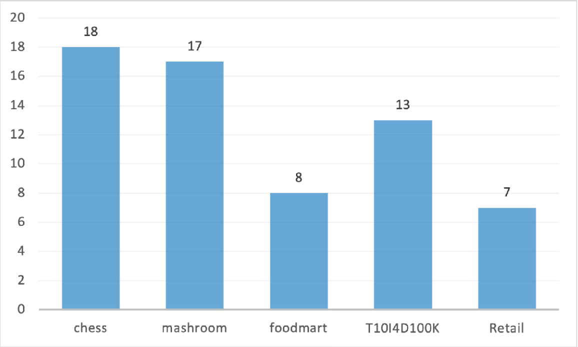

FIGURE 7. Most commonly used datasets in the selected studies.

G. PARALLELISM IN METAHEURISTICS

Although metaheuristics offer solid solutions for various types of complex problems, they face many limitations when solving real-world problems, such as searching for an optimal solution in a relatively short computational time. Moreover, the design of metaheuristic-based algorithms is constrained by the problem types, such as dynamic or high-dimensional problems [85]. Recent trends in metaheuristic design are directed towards adopting parallel schemes to solve such problems. Parallel metaheuristic algorithms can guarantee the delivery of high-quality solutions while maintaining the search time at a low rate. Two main models employ parallelism in metaheuristic algorithms: population-based and trajectory-based models [24]. In this review, only one study adopted a parallel scheme to propose a metaheuristic-based algorithm to address PPARM using a GPU platform [58].

IV. CHARACTERISTICS OF THE MOST USED DATA SETS

In the selected studies, the datasets used to evaluate the proposed algorithms and models were identified and analyzed. A total of 49 different datasets were used in PPARM, PPDM, and ARM in 41 studies that addressed this research question. These datasets were classified into two categories: dense and sparse datasets. In a dense dataset, cells are populated with non-zero values and are measured by the density factor. A dataset is referred to as sparse when the stored data contain more 0s than 1s [45]. As shown in figure 7, the most used datasets in the experiments are Chess, Mushroom, Foodmart, T10I4D100K, and Retail datasets, which were collected from the SPMF library [86]. Tables 9 and 10 list the parameters of the most commonly used datasets and their corresponding characteristics, respectively.

TABLE 9. Parameters of used datasets.

Parameter	Description
#|D|	Total number of dataset records
#|I|	Number of unique items
AvgLen	Average length of record
MaxLen	Maximal length of record
Type	Dataset type

TABLE 10. Characteristics of the most used datasets.

Dataset	Type	#|D|	Item count (I)	AvgLen	MaxLen
Chess	dense	3,196	76	37	37
Mushroom	dense	8,124	120	23	23
Foodmart	sparse	21,557	1,559	4	11
T10I4D100K	sparse	100,000	870	10.1	29
Retail	sparse	88,162	16,470	10.3	176

V. MOST COMMON METRICS FOR EVALUATING PPARM METAHEURISTIC-BASED ALGORITHMS

Identifying suitable evaluation metrics is an important step in designing new metaheuristic-based algorithms for PPARM. In this review, many different metrics were applied to evaluate the performance of the algorithms. However, a set of metrics is commonly used in selected studies. To summarize the findings, only the most common evaluation metrics are discussed in this review. For this purpose, the existing metrics were grouped based on the aspect of the PPARM algorithms being measured.

A. DATA PRIVACY METRICS

In terms of data privacy, the hiding failure (HF) rate measures the balance between privacy and data utility. HF refers to sensitive information that has been labeled as 'hidden' after the data sanitization process but can still be discovered during the mining process. The formula for HF is defined as the percentage of sensitive rules discovered from the sanitized database D' to the number of sensitive rules generated from the original database [37].

B. DATA UTILITY METRICS

HF is considered one of the three side effects of PPARM. The other two side effects correspond to the rate of data utility, namely, the lost rule rate and the ghost rule rate. The lost rule, also known as the Missing Cost (MC), refers to information that is already discovered but might be missed or gone after the data sanitization process. The ghost rule rate, known as the Artificial Cost (AC), is any unnecessary information that has not been discovered and that might be mined after the data sanitization process. Furthermore, another data utility metric used in some studies is Data Dissimilarity (Dis). It measures the difference between the original and sanitized databases [44].

C. EFFICIENCY METRICS

The efficiency of the algorithms was measured by the amount of time and space required for the given algorithm. The time is measured in terms of the the computational time, CPU time, or communication time. The space is measured based on the amount of memory required to execute a given algorithm [8].

VI. EXISTING CHALLENGES AND POTENTIAL FUTURE DIRECTIONS

The research above indicates that metaheuristic algorithms still have great potential for growth and practical application in PPARM. However, as the volume of data expands, the challenge of preserving both data privacy and utility becomes increasingly formidable. Consequently, this domain continues to confront significant hurdles. The current challenges, as well as future research directions in PPARM, can be summarized as follows:

- It is commonly known that no PPARM metaheuristic-based algorithm surpasses all other algorithms in terms of data utility, data privacy, and efficiency. This is because the PPARM is an optimization problem with conflicting objectives that require a trade-off between them. This is the main challenge faced by researchers when developing metaheuristics for PPARM. To find a workaround for this issue, researchers are investigating the application of the multi-threshold concept along with a multi-objective approach. In this review, two studies [35], [52] have applied this solution to address the PPDM problem. Hence, the development of metaheuristic-based algorithms for PPARM remains open to future studies.
- The majority of studies that addressed PPARM or ARM are concerned about positive association rules. Traditionally, ARM algorithms mine positive association rules, that are positively related to each other. On the other hand, negative association rules are as sensitive and useful as positive rules. Two items are said to be negatively correlated if there is an independent relationship between them, that is, they have never occurred together [2]. Accordingly, the negative association rules may provide useful information. However, studies focusing on mining negative rules concerning the privacy and utility of data are rare.
- Given the significant effects of both types of rules, however, when developing models for PPARM, most studies focus on one type, either positive or negative. It is important to develop a metaheuristic-based model that utilizes both types of sensitive rules, while applying the highest standards of privacy [2].
- Association rules mining methods based on uncertain theories, such as soft set, rough set, and fuzzy set, can be further explored in the context of preserving privacy using metaheuristic-based algorithms. Metaheuristic algorithms can enhance the robustness and scalability of association rules mining methods under uncertainty. This would involve investigating how these uncertain theories contribute to more effective and reliable pattern discovery in the PPARM domain. It presents a promising avenue for future exploration in the realm of data mining and knowledge discovery.
- In the majority of current mining schemes, the interestingness of association rules and patterns has been traditionally assessed using support and confidence. Depending on the characteristics of particular applications, various metrics can be employed to gauge the interestingness of association rules [8], [87].
- Ensuring the accuracy and performance of the PPARM metaheuristic-based algorithm can be challenging when dealing with missing or erroneous datasets, unless artificial preprocessing is applied, specifically to introduce a degree of fault tolerance.
- Many existing evaluation metrics are specific to some cases, which can be leads to a difficult comparison among the advantages and disadvantages of the existing PPARM schemes. Therefore, more universal applicable metrics are required for an effective comparison of different PPARM schemes.

VII. CONCLUSION

Despite the extensive research on protecting sensitive rules while maintaining their availability for mining processes, PPARM continues to captivate the attention of researchers. As the digital landscape evolves, the perpetual challenge of striking a delicate balance between data utility and privacy in the intricate process of mining association rules remains at the forefront of scholarly endeavors. This systematic literature review represents a crucial contribution to the ongoing discourse on PPARM, specifically focusing on the role of metaheuristic algorithms in this domain. By delving into empirical studies reported in esteemed journals from 2015 to 2023, our review meticulously examined the landscape to identify 41 pertinent papers from an initial pool of 3156 articles across nine digital databases. The stringent application of the review protocol ensured the selection of studies that provided valuable insights into solutions based on metaheuristics. The SLR findings, harnessed to address five pivotal research questions, uncovered a diverse array of metaheuristic-based algorithms employed in PPARM. This review took a step further by classifying these algorithms into two distinct categories: metaphor-based and non-metaphor-based. Subsequently, the proposed solutions from the selected studies were systematically presented and discussed, offering a comprehensive understanding of their metaheuristic approaches. In addition to the classification and analysis of metaheuristics, our review scrutinized various modifications introduced to metaheuristic-based algorithms in the context of PPARM. By shedding light on these adaptations, we aimed to contribute not only to the theoretical advancement of metaheuristics but also to their practical applicability in the field. Furthermore, the review outlined and illustrated the datasets and evaluation metrics commonly employed in these studies, providing valuable insights into the methodological choices made by researchers. This contextualization is essential for understanding the generalizability and applicability of the proposed solutions. Finally, by addressing the existing challenges and forecasting future trends in metaheuristic-based algorithms for PPARM, this review serves as a guiding compass for researchers. It not only aids in better comprehension of the development of metaheuristics in the PPARM field but also plays a pivotal role in bridging the gap between theoretical solutions and their real-world applications. As we navigate the complex terrain of privacy-aware association rule mining, the synthesis of findings in this review lays the foundation for future research endeavors, offering a road map for innovative solutions and applications in the pursuit of privacy-preserving data mining.

REFERENCES

[1] N. Domadiya and U. P. Rao, "Privacy-preserving association rule mining for horizontally partitioned healthcare data: A case study on the heart diseases," *Sadhana*, vol. 43, no. 8, p. 127, Aug. 2018.

[2] S. M. Darwish, R. M. Essa, M. A. Osman, and A. A. Ismail, "Privacy preserving data mining framework for negative association rules: An application to healthcare informatics," *IEEE Access*, vol. 10, pp. 76268–76280, 2022.

[3] C. Masciocchi, B. Gottardelli, M. Savino, L. Boldrini, A. Martino, C. Mazzarella, M. Massaccesi, V. Valentini, and A. Damiani, "Federated Cox Proportional Hazards model with multicentric privacy-preserving LASSO feature selection for survival analysis from the perspective of personalized medicine," in *Proc. IEEE 35th Int. Symp. Comput.-Based Med. Syst. (CBMS)*, Jul. 2022, pp. 25–31.

[4] H. Cheng, W. Zhang, Z. Wang, F. Zuo, and Z. Zhang, "An integrated method for hiding sensitive association rules of the supply chains," *IET Collaborative Intell. Manuf.*, vol. 3, no. 4, pp. 324–333, Dec. 2021.

[5] O. S. Adebayo and N. Abdul Aziz, "Improved malware detection model with apriori association rule and particle swarm optimization," *Secur. Commun. Netw.*, vol. 2019, pp. 1–13, Aug. 2019.

[6] L. Muduli, D. P. Mishra, and P. K. Jana, "Optimized fuzzy logic-based fire monitoring in underground coal mines: Binary particle swarm optimization approach," *IEEE Syst. J.*, vol. 14, no. 2, pp. 3039–3046, Jun. 2020.

[7] M. Chaudhari and J. Varmora, "Advance privacy preserving in association rule mining," in *Proc. Int. Conf. Electr., Electron., Optim. Techn. (ICEEOT)*, 2016, pp. 2527–2530.

[8] L. Zhang, W. Wang, and Y. Zhang, "Privacy preserving association rule mining: Taxonomy, techniques, and metrics," *IEEE Access*, vol. 7, pp. 45032–45047, 2019.

[9] N. Abd-Alsabour, "Nature as a source for inspiring new optimization algorithms," in *Proc. 9th Int. Conf. Signal Process. Syst.* New York, NY, USA: Association for Computing Machinery, Nov. 2017, pp. 51–56.

[10] M. Díaz-Madroñero, J. Mula, and D. Peidro, "A review of discrete-time optimization models for tactical production planning," *Int. J. Prod. Res.*, vol. 52, no. 17, pp. 5171–5205, Sep. 2014, doi: 10.1080/00207543.2014.899721.

[11] C.-W. Lin, T.-Y. Wu, P. Fournier Viger, G. Lin, J. Zhan, and M. Vozňák, "Fast algorithms for hiding sensitive high-utility itemsets in privacy-preserving utility mining," *Eng. Appl. Artif. Intell.*, vol. 55, pp. 269–284, Oct. 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0952197616301282

[12] A. Telikani, A. H. Gandomi, and A. Shahbahrami, "A survey of evolutionary computation for association rule mining," *Inf. Sci.*, vol. 524, pp. 318–352, Jul. 2020.

[13] K. Logeswaran, R. Andal, S. Ezhilmathi, A. Khan, P. Suresh, and K. P. Kumar, "A survey on metaheuristic nature inspired computations used for mining of association rule, frequent itemset and high utility itemset," *IOP Conf. Ser., Mater. Sci. Eng.*, vol. 1055, Feb. 2021, Art. no. 012103.

[14] B. Kitchenham and S. Charters, "Guidelines for performing systematic literature reviews in software engineering," *J. Softw. Eng. Appl.*, vol. 2, pp. 16–49, Jan. 2007.

[15] F. Salo, M. Injadat, A. B. Nassif, A. Shami, and A. Essex, "Data mining techniques in intrusion detection systems: A systematic literature review," *IEEE Access*, vol. 6, pp. 56046–56058, 2018.

[16] S. S. Kareem, R. R. Mostafa, F. A. Hashim, and H. M. El-Bakry, "An effective feature selection model using hybrid metaheuristic algorithms for IoT intrusion detection," *Sensors*, vol. 22, no. 4, p. 1396, Feb. 2022. [Online]. Available: https://www.mdpi.com/1424-8220/22/4/1396

[17] A. K. Dubey, "Optimized hybrid learning for multi disease prediction enabled by lion with butterfly optimization algorithm," *Sadhana*, vol. 46, no. 2, p. 63, Jun. 2021.

[18] E. Zorarpaci and S. A. Özel, "Differentially private 1R classification algorithm using artificial bee colony and differential evolution," *Eng. Appl. Artif. Intell.*, vol. 94, Sep. 2020, Art. no. 103813. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0952197620301925

[19] E. Zorarpaci and S. Ayse Özel, "Privacy preserving rule-based classifier using modified artificial bee colony algorithm," *Exp. Syst. Appl.*, vol. 183, Nov. 2021, Art. no. 115437. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417421008526

[20] E. Merkohitaj, "Identification and analysis of practices for organizing development teams," Ph.D. dissertation, Dept. Comput. Sci., Gottfried Wilhelm Leibniz Univ., Hannover, Germany, 2021.

[21] D. S. D. Cunha, R. S. Xavier, D. G. Ferrari, F. G. Vilasboas, and L. N. de Castro, "Bacterial colony algorithms for association rule mining in static and stream data," *Math. Problems Eng.*, vol. 2018, pp. 1–14, Nov. 2018. [Online]. Available: https://www.hindawi.com/journals/mpe/2018/4676258/

[22] I. H. Osman, "Focused issue on applied meta-heuristics," *Comput. Ind. Eng.*, vol. 44, no. 2, pp. 205–207, Feb. 2003.

[23] M. Gendreau and J.-Y. Potvin, "Metaheuristics in combinatorial optimization," *Ann. Oper. Res.*, vol. 140, no. 1, pp. 189–213, 2005.

[24] M. Abdel-Basset, L. Abdel-Fatah, and A. K. Sangaiah, "Chapter 10—Metaheuristic algorithms: A comprehensive review," in *Computational Intelligence for Multimedia Big Data on the Cloud With Engineering Applications* (Intelligent Data-Centric Systems), A. K. Sangaiah, M. Sheng, and Z. Zhang, Eds. Cambridge, MA, USA: Academic Press, Jan. 2018, pp. 185–231. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780128133149000104

[25] S. Mirjalili and A. Lewis, "The whale optimization algorithm," *Adv. Eng. Softw.*, vol. 95, pp. 51–67, May 2016, doi: 10.1016/j.advengsoft.2016.01.008.

[26] J. I. Serrano and M. D. del Castillo, "On the origin of the evolutionary computation species influences of Darwin's theories on computer science," *Artif. Intell. Rev.*, vol. 38, no. 1, pp. 41–54, Jun. 2012.

[27] D. Menaga and S. Saravanan, "GA-PPARM: Constraint-based objective function and genetic algorithm for privacy preserved association rule mining," *Evol. Intell.*, vol. 15, no. 2, pp. 1487–1498, Jun. 2022, doi: 10.1007/s12065-021-00576-z.

[28] N. Khuda Bux, M. Lu, J. Wang, S. Hussain, and Y. Aljeroudi, "Efficient association rules hiding using genetic algorithms," *Symmetry*, vol. 10, no. 11, p. 576, Nov. 2018. [Online]. Available: https://www.mdpi.com/2073-8994/10/11/576

[29] F. N. Motlagh and H. Sajedi, "MOSAR: A multi-objective strategy for hiding sensitive association rules using genetic algorithm," *Appl. Artif. Intell.*, vol. 30, no. 9, pp. 823–843, Oct. 2016, doi: 10.1080/08839514.2016.1268038.

[30] G. S. Navale and S. N. Mali, "A multi-analysis on privacy preservation of association rules using hybridized approach," *Evol. Intell.*, vol. 15, no. 2, pp. 1051–1065, Jun. 2022, doi: 10.1007/s12065-019-00277-8.

[31] J. C.-W. Lin, Y. Zhang, B. Zhang, P. Fournier-Viger, and Y. Djenouri, "Hiding sensitive itemsets with multiple objective optimization," *Soft Comput.*, vol. 23, no. 23, pp. 12779–12797, Dec. 2019, doi: 10.1007/s00500-019-03829-3.

[32] M. Lefkir, F. Nouioua, and P. Fournier-Viger, "Hiding sensitive frequent itemsets by item removal via two-level multi-objective optimization," *Appl. Intell.*, vol. 53, no. 9, pp. 10027–10052, Aug. 2022, doi: 10.1007/s10489-022-03808-6.

[33] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Trans. Evol. Comput.*, vol. 6, no. 2, pp. 182–197, Apr. 2002.

[34] C.-W. Lin, T.-P. Hong, K.-T. Yang, and S.-L. Wang, "The GA-based algorithms for optimizing hiding sensitive itemsets through transaction deletion," *Int. J. Speech Technol.*, vol. 42, no. 2, pp. 210–230, Mar. 2015, doi: 10.1007/s10489-014-0590-5.

[35] J. M.-T. Wu, G. Srivastava, A. Jolfaei, P. Fournier-Viger, and J. C.-W. Lin, "Hiding sensitive information in eHealth datasets," *Future Gener. Comput. Syst.*, vol. 117, pp. 169–180, Apr. 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167739X20330594

[36] J. C.-W. Lin, T.-P. Hong, P. Fournier-Viger, Q. Liu, J.-W. Wong, and J. Zhan, "Efficient hiding of confidential high-utility itemsets with minimal side effects," *J. Experim. Theor. Artif. Intell.*, vol. 29, no. 6, pp. 1225–1245, Nov. 2017, doi: 10.1080/0952813x.2017.1328462.

[37] F. Yang, X. Lei, J. Le, N. Mu, and X. Liao, "Minable data publication based on sensitive association rule hiding," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 6, no. 5, pp. 1247–1257, Oct. 2022.

[38] B. Wang, K. E. Merrick, and H. A. Abbass, "Co-operative coevolutionary neural networks for mining functional association rules," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 6, pp. 1331–1344, Jun. 2017.

[39] H. Wu, "Data association rules mining method based on improved apriori algorithm," in *Proc. 4th Int. Conf. Big Data Res. (ICBDR)*. New York, NY, USA: Association for Computing Machinery, Nov. 2020, pp. 12–17, doi: 10.1145/3445945.3445948.

[40] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proc. IEEE ICNN*, vol. 4, Dec. 1995, pp. 1942–1948.

[41] M. Aviles, J. Rodríguez-Reséndiz, and D. Ibrahimi, "Optimizing EMG classification through metaheuristic algorithms," *Technologies*, vol. 11, no. 4, p. 87, Jul. 2023. [Online]. Available: https://www.mdpi.com/2227-7080/11/4/87

[42] J. C.-W. Lin, Q. Liu, P. Fournier-Viger, T.-P. Hong, M. Voznak, and J. Zhan, "A sanitization approach for hiding sensitive itemsets based on particle swarm optimization," *Eng. Appl. Artif. Intell.*, vol. 53, pp. 1–18, Aug. 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0952197616300653

[43] J. M. Wu, G. Srivastava, U. Yun, S. Tayeb, and J. C. Lin, "An evolutionary computation-based privacy-preserving data mining model under a multithreshold constraint," *Trans. Emerg. Telecommun. Technol.*, vol. 32, no. 3, p. e4209, Mar. 2021.

[44] T.-Y. Wu, J. C.-W. Lin, Y. Zhang, and C.-H. Chen, "A grid-based swarm intelligence algorithm for privacy-preserving data mining," *Appl. Sci.*, vol. 9, no. 4, p. 774, 2019.

[45] S. Jangra and D. Toshniwal, "VIDPSO: Victim item deletion based PSO inspired sensitive pattern hiding algorithm for dense datasets," *Inf. Process. Manag.*, vol. 57, no. 5, Sep. 2020, Art. no. 102255.

[46] F. Yang and X. Liao, "An optimized sanitization approach for minable data publication," *Big Data Mining Anal.*, vol. 5, no. 3, pp. 257–269, Sep. 2022.

[47] Z. Kou and L. Xi, "Binary particle swarm optimization-based association rule mining for discovering relationships between machine capabilities and product features," *Math. Problems Eng.*, vol. 2018, pp. 1–16, Oct. 2018.

[48] T. Su, H. Xu, and X. Zhou, "Particle swarm optimization-based association rule mining in big data environment," *IEEE Access*, vol. 7, pp. 161008–161016, 2019.

[49] L. Muduli, D. P. Mishra, and P. Jana, *Wireless Sensor Network Based Underground Coal Mine Environmental Monitoring Using Machine Learning Approach*. Singapore: Springer, Jan. 2019, pp. 776–786.

[50] M. Dorigo and L. M. Gambardella, "Ant colony system: A cooperative learning approach to the traveling salesman problem," *IEEE Trans. Evol. Comput.*, vol. 1, no. 1, pp. 53–66, Apr. 1997.

[51] G. Dhiman, "ESA: A hybrid bio-inspired metaheuristic optimization approach for engineering problems," *Eng. Comput.*, vol. 37, no. 1, pp. 323–353, Jan. 2021.

[52] J. M.-T. Wu, G. Srivastava, J. C.-W. Lin, and Q. Teng, "A multi-threshold ant colony system-based sanitization model in shared medical environments," *ACM Trans. Internet Technol.*, vol. 21, no. 2, pp. 1–26, Jun. 2021, doi: 10.1145/3408296.

[53] J. C.-W. Lin, G. Srivastava, Y. Zhang, Y. Djenouri, and M. Aloqaily, "Privacy-preserving multiobjective sanitization model in 6G IoT environments," *IEEE Internet Things J.*, vol. 8, no. 7, pp. 5340–5349, Apr. 2021.

[54] D. Karaboga and B. Basturk, "Artificial bee colony (ABC) optimization algorithm for solving constrained optimization problems," in *Foundations of Fuzzy Logic and Soft Computing*, vol. 4529. Berlin, Germany: Springer, Jan. 2007, pp. 789–798.

[55] B. Yang, J. Wang, X. Zhang, T. Yu, W. Yao, H. Shu, F. Zeng, and L. Sun, "Comprehensive overview of meta-heuristic algorithm applications on PV cell parameter identification," *Energy Convers. Manage.*, vol. 208, Mar. 2020, Art. no. 112595.

[56] M. S. Kiran and M. Gündüz, "XOR-based artificial bee colony algorithm for binary optimization," *Turkish J. Electr. Eng. Comput. Sci.*, vol. 21, pp. 2307–2328, 2013. [Online]. Available: https://journals.tubitak.gov.tr/elektrik/vol21/iss8/15

[57] A. Telikani, A. H. Gandomi, A. Shahbahrami, and M. Naderi Dehkordi, "Privacy-preserving in association rule mining using an improved discrete binary artificial bee colony," *Exp. Syst. Appl.*, vol. 144, Apr. 2020, Art. no. 113097. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417419308140

[58] A. Telikani, A. Shahbahrami, and A. H. Gandomi, "High-performance implementation of evolutionary privacy-preserving algorithm for big data using GPU platform," *Inf. Sci.*, vol. 579, pp. 251–265, Nov. 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0020025521007982

[59] A. Telikani, A. Shahbahrami, J. Shen, G. Gaydadjiev, and J. C.-W. Lin, "An edge-aided parallel evolutionary privacy-preserving algorithm for Internet of Things," *Internet Things*, vol. 23, Oct. 2023, Art. no. 100831. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2542660523001543

[60] A. Askarzadeh, "A novel metaheuristic method for solving constrained engineering optimization problems: Crow search algorithm," *Comput. Struct.*, vol. 169, pp. 1–12, Jun. 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0045794916300475

[61] Q. Cheng, H. Huang, and M. Chen, "A novel crow search algorithm based on improved flower pollination," *Math. Problems Eng.*, vol. 2021, pp. 1–26, Oct. 2021. [Online]. Available: https://www.hindawi.com/journals/mpe/2021/1048879/

[62] S. Sharmila and S. Vijayarani, "Association rule mining using fuzzy logic and whale optimization algorithm," *Soft Comput.*, vol. 25, no. 2, pp. 1431–1446, Jan. 2021, doi: 10.1007/s00500-020-05229-4.

[63] N. P. Karlekar and N. Gomathi, "OW-SVM: Ontology and whale optimization-based support vector machine for privacy-preserved medical data classification in cloud," *Int. J. Commun. Syst.*, vol. 31, no. 12, p. e3700, Aug. 2018. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1002/dac.3700

[64] G. K. Shailaja and C. V. G. Rao, "Robust and lossless data privacy preservation: Optimal key based data sanitization," *Evol. Intell.*, vol. 15, no. 2, pp. 1123–1134, Jun. 2022, doi: 10.1007/s12065-019-00309-3.

[65] W.-T. Pan, "A new fruit fly optimization algorithm: Taking the financial distress model as an example," *Knowl.-Based Syst.*, vol. 26, pp. 69–74, Feb. 2012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0950705111001365

[66] L. Wang, R. Liu, and S. Liu, "An effective and efficient fruit fly optimization algorithm with level probability policy and its applications," *Knowl.-Based Syst.*, vol. 97, pp. 158–174, Apr. 2016.

[67] T. S. Reddy, R. Sathya, and M. Nuka, "Intuitionistic fuzzy rough sets and fruit fly algorithm for association rule mining," *Int. J. Syst. Assurance Eng. Manag.*, vol. 13, no. 4, pp. 2029–2039, Aug. 2022, doi: 10.1007/s13198-021-01616-8.

[68] D. Dhinakaran and P. M. J. Prathap, "Protection of data privacy from vulnerability using two-fish technique with apriori algorithm in data mining," *J. Supercomput.*, vol. 78, no. 16, pp. 17559–17593, Nov. 2022, doi: 10.1007/s11227-022-04517-0.

[69] G.-G. Wang, S. Deb, and L. D. S. Coelho, "Elephant herding optimization," in *Proc. 3rd Int. Symp. Comput. Bus. Intell. (ISCBI)*, Dec. 2015, pp. 1–5.

[70] P. K. Gopagoni and S. K. M. Rao, "Distributed elephant herding optimization for grid-based privacy association rule mining," *Data Technol. Appl.*, vol. 54, no. 3, pp. 365–382, May 2020, doi: 10.1108/dta-07-2019-0104.

[71] I. Fister, A. Galvez, E. Osaba, J. D. Ser, A. Iglesias, and I. Fister, "Discovering dependencies among mined association rules with population-based metaheuristics," in *Proc. Genetic Evol. Comput. Conf. Companion*. New York, NY, USA: Association for Computing Machinery, 2019, pp. 1668–1674, doi: 10.1145/3319619.3326833.

[72] M. Yazdani and F. Jolai, "Lion optimization algorithm (LOA): A nature-inspired metaheuristic algorithm," *J. Comput. Des. Eng.*, vol. 3, no. 1, pp. 24–36, 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2288430015000524

[73] D. Menaga and S. Revathi, "Least lion optimisation algorithm (LLOA) based secret key generation for privacy preserving association rule hiding," *IET Inf. Secur.*, vol. 12, no. 4, pp. 332–340, Jul. 2018. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1049/iet-ifs.2017.0634

[74] X.-S. Yang and S. Deb, "Cuckoo search via Lévy flights," in *Proc. World Congr. Nature Biologically Inspired Comput. (NaBIC)*, Dec. 2009, pp. 210–214.

[75] R. Rajabioun, "Cuckoo optimization algorithm," *Appl. Soft Comput.*, vol. 11, no. 8, pp. 5508–5518, Dec. 2011. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1568494611001670

[76] M. H. Afshari, M. N. Dehkordi, and M. Akbari, "Association rule hiding using cuckoo optimization algorithm," *Exp. Syst. Appl.*, vol. 64, pp. 340–351, Dec. 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417416303980

[77] T. S. Murthy, M. S. Roy, and M. K. Varma, "Improving the performance of association rules hiding using hybrid optimization algorithm," *J. Appl. Secur. Res.*, vol. 15, no. 3, pp. 423–437, Jul. 2020, doi: 10.1080/19361610.2020.1756155.

[78] K. M. Passino, "Bacterial foraging optimization," in *Innovations and Developments of Swarm Intelligence Applications*. Pennsylvania, PA, USA: IGI Global, 2012, pp. 219–234. [Online]. Available: https://www.igi-global.com/chapter/bacterial-foraging-optimization/www.igi-global.com/chapter/bacterial-foraging-optimization/65815

[79] B. Niu and H. Wang, "Bacterial colony optimization," *Discrete Dyn. Nature Soc.*, vol. 2012, 2012, Art. no. e698057. [Online]. Available: https://www.hindawi.com/journals/ddns/2012/698057/

[80] W. Zhao, Z. Zhang, and L. Wang, "Manta ray foraging optimization: An effective bio-inspired optimizer for engineering applications," *Eng. Appl. Artif. Intell.*, vol. 87, 2020, Art. no. 103300. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0952197619302593

[81] N. Lakshmi and M. Krishnamurthy, "Association rule mining based fuzzy manta ray foraging optimization algorithm for frequent itemset generation from social media," *Concurrency Comput., Pract. Exper.*, vol. 34, no. 10, p. e6790, May 2022. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1002/cpe.6790

[82] F. Glover and C. McMillan, "The general employee scheduling problem. An integration of MS and AI," *Comput. Oper. Res.*, vol. 13, no. 5, pp. 563–573, 1986. [Online]. Available: https://www.sciencedirect.com/science/article/pii/030505488690050X

[83] J. Pierezan, L. D. S. Coelho, V. C. Mariani, S. K. Goudos, A. D. Boursianis, N. V. Kantartzis, C. S. Antonopoulos, and S. Nikolaidis, "Multiobjective ant lion approaches applied to electromagnetic device optimization," *Technologies*, vol. 9, no. 2, p. 35, May 2021. [Online]. Available: https://www.mdpi.com/2227-7080/9/2/35

[84] I. B. Aydilek, "A hybrid firefly and particle swarm optimization algorithm for computationally expensive numerical problems," *Appl. Soft Comput. J.*, vol. 66, pp. 232–249, May 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S156849461830084X

[85] T. G. Crainic and M. Toulouse, *Parallel Meta-heuristics*. Boston, MA, USA: Springer, 2010, pp. 497–541.

[86] *SPMF: An Open-Source Data Mining Library*. Accessed: Aug. 10, 2023. [Online]. Available: https://www.philippe-fournier-viger.com/spmf/index.php?link=datasets.php

[87] P. Fournier-Viger, W. Gan, Y. Wu, M. Nouioua, W. Song, T. Truong, and H. Duong, "Pattern mining: Current challenges and opportunities," in *Database Systems for Advanced Applications*. Berlin, Germany: Springer, 2022, pp. 34–49.

SHAHAD S. ALJEHANI received the B.S. degree in computer science from Taibah University, Saudi Arabia, in 2018. She is currently pursuing the master's degree with the Department of Computer Science, College of Computer and Information Systems, Umm Al-Qura University, Saudi Arabia. She has two years of research experience and has published three research articles in peer-reviewed journals and conference proceedings. Her research interests include data mining, the IoT-based systems, blockchain, human–computer interaction, and machine learning.

YOUSEEF A. ALOTAIBI received the master's degree in information technology (computer network) from La Trobe University, Melbourne, Australia, in 2009, and the Ph.D. degree from the Department of Computer Science and Computer Engineering, La Trobe University, in 2014. He is currently a Professor with the Department of Computer Science, College of Computer and Information Systems, Umm Al-Qura University, Saudi Arabia. He is the author of more than 90 SCIE journal publications and more than 15 conference publications. His research interests include software engineering, sustainable development, system analysis and design, business process and information technology, process model and process reengineering, global software development, sustainability and smart cities, the IoT, wireless communication, and project management. He is a member of IEEE Computer Science Society, Australian Computer Society, Association of Information Systems, IAITCD, and IACSIT. He is also a reviewer of more than 20 journals.