An Enhanced Data Mining Classification based on Shuffled Frog Leaping Optimization

Nguyen Ha Huy Cuong

Software Development Centre - The University of Danang, 41 Le Duan Street, 50000 DaNang, Vietnam; E-mail: nhhcuong@sdc.udn.vn

Ho Phan Hieu

School of National Defense and Security Education- The University of Danang, 41 Le Duan Street, 50000 DaNang, Vietnam; E-mail: hophanhieu@ac.udn.vn

Nguyen Thanh Thuy

Faculty of Computer Science, VNU University of Engineering and Technology, E3 Building, 144 Xuan Thuy Street, Cau Giay District, 000084 Hanoi, Vietnam
E-mail: nguyenthanhthuy@vnu.edu.vn

Tran Anh Kiet

The University of Danang, 41 Le Duan Street, 50000 DaNang, Vietnam;
E-mail: takiet@ac.udn.vn

Hung Vo Trung

University of Technology and Education - The University of Danang, 41 Cao Thang Street, 50209 DaNang, Vietnam; E-mail: vthung@ute.udn.vn

Abstract: Association Rule Mining (ARM) uncovers meaningful associations within discrete and categorical datasets. This paper introduces an innovative ARM method leveraging the Modified Shuffled Frog Leaping Optimization (MSFLO) technique to enhance performance analysis. By integrating the Apriori algorithm with MSFLO's bio-inspired optimization, including frog encoding, our approach generates association rules efficiently. Unlike traditional methods requiring multiple database scans, this technique filters data in a single pass, significantly reducing CPU time and memory usage. Multiple optimization *measures are applied to refine MSFLO, improving the accuracy and effectiveness of rule extraction. Implemented in MongoDB, the method is validated across six diverse datasets—Watermelon, Mangosteen, Breast, Dragon Fruit, Mango, and Orange—demonstrating superior performance compared to existing approaches. This advancement optimizes computational efficiency and rule quality, offering a robust solution for fruit shape database mining and precision agriculture applications.*

Keywords: Apriori Method; Rule-Based Mining; Frog Leaping Optimization; Efficiency Enhancement; Support Metric; Confidence Level

1 Introduction

Association Rule Mining (ARM) is a key data mining technique designed to identify significant relationships among variables, attributes, or features within large databases [1]. Its core objective is to extract robust rules by applying measures of interestingness, such as support and confidence, to assess their strength and relevance [2]. The notion of interestingness is central to data mining, spanning various pattern types and enabling the prioritization of rules based on user interest [3, 4]. ARM's versatility has led to its adoption across diverse fields, including market analysis, bioinformatics, intrusion detection, and precision agriculture, where it uncovers actionable patterns from complex datasets [5].

Several established algorithms drive ARM, notably the Apriori algorithm, Eclat, and FP-Growth [6]. Apriori, a foundational method, employs a breadth-first search to calculate itemset support—the frequency of item occurrences—using candidate generation and downward closure properties. Extensions of these algorithms have improved scalability and performance [7]. However, challenges like multiple database scans and threshold selection persist. To address these, a Particle Swarm Optimization (PSO)-based ARM approach was proposed in [8], optimizing fitness values to determine support and confidence thresholds from binary-transformed data. Validated on the Food Mart 2020 database, it outperformed genetic algorithms in efficiency [8]. Similarly, Weighted ARM using PSO (WARM SWARM) enhances rule mining by assigning item weights, adding depth to pattern analysis [9].

This study advances ARM by introducing the ARM-MSFLO method, integrating Apriori with Modified Shuffled Frog Leaping Optimization (MSFLO). Targeting fruit shape database mining for class classification, it optimizes rules in a single pass, reducing computational overhead. Tested on six datasets—Watermelon, Mangosteen, Breast, Dragon Fruit, Mango, and Orange—this approach enhances efficiency and rule quality, supporting precision agriculture. Subsequent sections detail its methodology and validation [22].

2 Related Work

Association Rule Mining (ARM) continues to evolve as a critical technique for pattern discovery in large datasets, though it faces challenges in scalability, efficiency, and rule quality. Recent innovations have tackled these issues with novel approaches. The Fuzzy Frequent Itemset (FFI)-Miner algorithm [10] eliminates candidate generation by using a fuzzy-list structure to store mining data, complemented by a pruning strategy that reduces the search space for efficient FFI discovery. This method advances fuzzy itemset mining in ARM [10]. Similarly, [11] improves the FP-Growth algorithm with an incremental queue model, enhancing the FP4W-Growth framework to boost rule confidence in dynamic datasets, notably for text correlation [11]. For high-dimensional challenges, [12] introduces the Modified Single-Objective Binary Cuckoo Search for ARM (MBCS-ARM), featuring a unique representation to handle attribute growth and mine both positive and negative associations [12]. Further ARM developments are explored in [13–15], while [16, 17] highlight ongoing issues such as heterogeneous data formats, redundant rules, and computational overhead.

Key ARM challenges include converting disordered data formats, which complicates multilevel hierarchical mining, and minimizing database scans for time efficiency. Small support and confidence thresholds increase rule proliferation, necessitating optimal interestingness measures, while high CPU and I/O overheads demand optimization. Meta-heuristic algorithms provide viable solutions. The Shuffled Frog Leaping Optimization (SFLO) algorithm [18], a population-based method, simulates frog foraging by organizing virtual frogs into memeplexes—carriers of memes, or evolutionary units [19]. It combines local searches within memeplexes with global exploration through shuffling, validated across diverse test functions [20, 21]. Despite its strengths, standard SFLO risks early convergence, limiting its optimization potential.

Our recent work [22] addresses these challenges in *"Research on fruit shape database mining to support fruit class classification using the shuffled frog leaping optimization (SFLO) technique"* [22]. Published in *Mathematical Biosciences and Engineering*, it proposes the Modified SFLO (MSFLO), enhancing SFLO by expanding the local search space with adaptive jump lengths to avoid premature convergence. Integrated with Apriori, ARM-MSFLO optimizes support, confidence, rule quantity, and execution time across six datasets—Watermelon, Mangosteen, Breast, Dragon Fruit, Mango, and Orange—spanning agricultural and non-agricultural domains. Unlike [9], which employs PSO for weighted ARM, MSFLO reduces I/O overhead and rule redundancy, targeting fruit classification and precision agriculture. This method offers a bio-inspired alternative, improving efficiency and rule quality. The paper proceeds with the methodology (Section 3), validation (Section 4), and future directions (Section 5).

3 The Proposed Method

This research aims to establish a robust theoretical framework grounded in existing data mining principles, laying the foundation for pioneering techniques in an emerging and underexplored field. By focusing on key staple fruits, this study has the potential to transform precision agriculture and significantly enhance regional farming practices. In this section, we introduce a novel Association Rule Mining (ARM) approach that integrates the Apriori algorithm with Modified Shuffled Frog Leaping Optimization (MSFLO). The algorithm's core components include rule discovery, computation of support and confidence, evaluation via a fitness function, and iterative optimization of rules.

The proposed ARM-MSFLO methodology combines established data mining techniques with innovative optimization to improve rule generation efficiency, targeting applications in precision agriculture and beyond. By employing a single-pass database filtration process, it overcomes the multi-scan inefficiencies of conventional ARM methods, reducing both CPU time and memory consumption. The approach is validated using six diverse datasets—Watermelon, Mangosteen, Breast, Dragon Fruit, Mango, and Orange—spanning agricultural and non-agricultural contexts. The methodology proceeds in four key stages. First, the Apriori algorithm extracts frequent itemsets from transactional data, producing candidate rules across the varied datasets. Second, support and confidence metrics are calculated to assess rule reliability, adapted to the datasets' diverse nature. Third, a tailored fitness function evaluates rule quality using logarithmically scaled support and confidence values, normalized for uniformity across data types. Finally, MSFLO enhances weaker rules through iterative refinement within memeplexes, adjusting rule positions based on local and global optima. This hybrid approach delivers reliable, actionable rules, with experimental outcomes—detailed in later sections—confirming its effectiveness across all six datasets.

3.1 Traditional ARM Enhanced with ARM-MSFLO

This section presents the ARM-MSFLO method, which advances conventional Association Rule Mining (ARM) by combining the Apriori algorithm with Modified Shuffled Frog Leaping Optimization (MSFLO). Specifically designed for mining fruit shape databases to facilitate fruit class classification, the ARM-MSFLO approach enhances the generation of association rules. It achieves this by boosting computational efficiency, significantly lowering CPU time, and optimizing memory usage through a streamlined single-pass database filtration process. This integration allows for faster and more resource-efficient rule extraction, making it particularly effective for handling large datasets in fruit classification tasks, ensuring improved performance and accuracy in identifying meaningful patterns. The method is validated across six datasets: Watermelon, Mangosteen, Breast, Dragon Fruit, Mango, and Orange. Below is the detailed algorithm.

Input:

- Transactional database D, a set of transactions from fruit shape and related datasets.

Output:

- Optimized association rules with refined support and confidence metrics.

Algorithm Steps:

Step 1: Rule Discovery Using the Apriori Algorithm
The Apriori algorithm extracts frequent itemsets from D.

- *Input:* D (loaded as a data file).
- *Process:*
    1. Compute total transactions, $|D|$.
    2. Define G_k as candidate itemsets of size k and I_k as frequent itemsets of size k.
    3. Initialize I_1 = {frequent 1-itemsets} with minimum support min_sup.
    4. For $k = 1$; $I_k \neq \emptyset$; k++:
        - Generate $I_{(k+1)}$ from I_k using Apriori's downward closure.
        - For each transaction $d_i \in D$, increment counts of candidates in $G_{(k+1)}$ present in d_i.
        - Filter $I_{(k+1)} = \{g \in G_{(k+1)} \mid \text{support}(g) \geq \text{min\_sup}\}$.
    5. Return G_k and I_k for all k.

Step 2: Evaluation of Support and Confidence
For each rule $A \to B$:

- *Support:* $\text{Support}(A \cup B) = \text{Count}(A \cup B) / |D|$
- *Confidence:* $\text{Confidence}(A \to B) = \text{Support}(A \cup B) / \text{Support}(A)$
  These metrics gauge rule strength and reliability.

Step 3: Fitness Calculation for MSFLO Migration
A fitness function assesses rule quality:

- *Total Fitness:*
  $\text{total\_fitness}(x) = |\log(\text{Confidence}(x)) + \log(\alpha * \text{Support}(x)) / (\text{length}(\text{Support}) + \text{length}(\text{Confidence}))|$
  where $\alpha = 1.5$ (scaling factor, tuned empirically).

- *Net Fitness:*
  net_fitness = Σ total_fitness(x) / number of rules
  net_fitness = | net_fitness * length(Support) / (min_sup * min_conf * 40) |
  Here, min_conf is the minimum confidence, and 40 is a normalization constant derived from dataset scale balancing [22].

Step 4: MSFLO Optimization Phase
MSFLO refines weaker rules:

- *Initialization:*
  - Memeplex count: N_mplex = floor(N_pop / N_pop_mplex), with N_pop = 100 (population size), N_pop_mplex = 10 (memeplex size).
  - Total population: N_pop = N_mplex * N_pop_mplex.
- *Position Update:*
  Position_new = rand() * C * (M_b - M_w)
  where rand() ∈ [0, 1], C = 2 (step-size), M_b is the global best fitness, and M_w is the worst fitness in the memeplex.
- *Optimization:* Compute global minimum fitness, iterate until convergence (max iterations: 50).

Parameters

The ARM-MSFLO method relies on a set of carefully tuned parameters to balance computational efficiency and rule quality. These parameters are:

- **Minimum Support (min_sup = 0.1)**: This threshold ensures that only itemsets appearing in at least 10% of transactions are considered frequent, filtering out rare patterns.
- **Minimum Confidence (min_conf = 0.6)**: Rules must exhibit at least 60% reliability, ensuring practical relevance.
- **Scaling Factor (α = 1.5)**: Used in the fitness function to adjust the influence of support relative to confidence, enhancing optimization stability.
- **Step-Size Constant (C = 2)**: Controls the magnitude of position updates in MSFLO, facilitating effective exploration of the solution space.
- **Population Size (N_pop = 100)**: Represents the total number of virtual frogs (candidate solutions) in the optimization process, providing a robust search capacity.
- **Memeplex Size (N_pop_mplex = 10)**: Divides the population into 10 smaller groups (memeplexes) for local searches, balancing diversity and convergence.

- **Maximum Iterations (max iterations = 50)**: Limits the optimization process to 50 cycles, preventing excessive computation while ensuring convergence.

These values were determined through preliminary experiments, where various combinations were tested on the six datasets—Watermelon, Mangosteen, Breast, Dragon Fruit, Mango, and Orange—to optimize rule coverage, execution time, and memory usage. The selection process and detailed results are presented in Section [X] (e.g., "Experimental Setup"), demonstrating how these parameters achieve a trade-off between efficiency and accuracy in fruit shape database mining.

Contribution

The ARM-MSFLO method offers a significant advancement in Association Rule Mining by seamlessly integrating the Apriori algorithm's rule discovery capabilities with the bio-inspired optimization power of Modified Shuffled Frog Leaping Optimization (MSFLO). This hybrid approach addresses two critical limitations of traditional ARM:

- **Reduction of Redundant Database Scans**: Unlike standard Apriori, which requires multiple passes over the database, ARM-MSFLO employs a single-pass filtration strategy, substantially lowering I/O overhead and execution time.
- **Enhancement of Rule Quality**: By leveraging MSFLO's adaptive optimization, the method refines weaker rules, improving their support and confidence metrics to better support fruit class classification.

This unique combination is particularly effective for mining fruit shape databases, as demonstrated with datasets like Watermelon and Mango, where it identifies patterns critical for precision agriculture. The effectiveness of ARM-MSFLO—both in reducing computational overhead and producing high-quality, actionable rules—has been rigorously validated through experiments, with comprehensive results and comparisons to baseline methods provided in Section [Y] (e.g., "Results and Validation"). These findings underscore its potential as a robust tool for data-driven agricultural applications.

The flowchart of the algorithm is shown in Fig. 1. Now we discuss how we create a correlation between association rule and MSFLO. The flowchart gives a more extensive view of the proposed calculation. The support and confidence are first computed utilizing the ARM approach. With the assistance of these two components, the fitness function is assessed. Assessing the guidelines which are beneath fitness function as these principles will be less fit standards and should be moved. The idea of MSFLO is presently connected to the guidelines underneath fitness esteem by figuring their relocating likelihood. For each case, likelihood is refreshed and the next position for the development is assessed. Along these lines, those guidelines which were less fit at first will move to a superior place and will survive. This will build their likelihood of survival and in this way, better principles can be mined. The fitness esteem is computed because of the support and confidence that have been inferred in the Apriori beforehand in the proposed calculation. Additionally, net fitness esteem is to be ascertained for the general fitness thought.

The fitness investigation is used to choose the standards that are to be altered using MSFLO approach. The less appropriate tenets are discovering by contrasting their fitness esteem and net_fitness. The equation determined for fitness estimation is:

$$overall\_fitness = abs\left(\frac{\log(conf(i)) + log(alpha*sup(i))}{(length(sup)+length(conf))}\right), \tag{1}$$

$$Net\_fitness = sum(overall\_fitness)/length(dataset), \tag{2}$$

$$Net\_fitness = abs(net\_fitness * (length(sup)/(minsup * minconf * 40))) \tag{3}$$

where mins up and min conf are predefined. In Fig. 1, we see at the last step we update the optimal rules. This is the main objective of this research work.

Fig. 2 shows the estimation of lower and upper spaces of a set, the approximation of spaces with respect to the plotting area. In order to visualize the algorithm, we give an example of guessing in a fruit class below.

In this subsection, an automatic fruit recognition system is used to visualize the proposed method as follows. We develop an automatic fruit recognition system where the innovative techniques in computer vision and image processing are exploited to improve the product quality in the food industry. In particular, the segmentation as well as descriptors are utilized for both image analysis and feature extraction. The characteristics of fruits are primarily defined by their color, shape, size, and texture, which serve as the basis for quality assessment in this study. The research focuses on nine fruit categories—apple, blueberry, lemon, mango, orange, pear, pineapple, pomegranate, and walnut—to evaluate their performance. A two-phase methodology is introduced: First, an image dataset capturing the diverse features of these fruits is created. To develop the learning model, a combination of a loss function, an optimizer, and a set of evaluation metrics is used, ensuring effective training and validation. The second phase involves analyzing input and output samples to distinguish between rotten and fresh fruits, as referenced in prior studies [3], [4]. This approach enables a comprehensive examination of fruit quality, leveraging visual attributes to enhance recognition accuracy and reliability across the selected fruit classes.

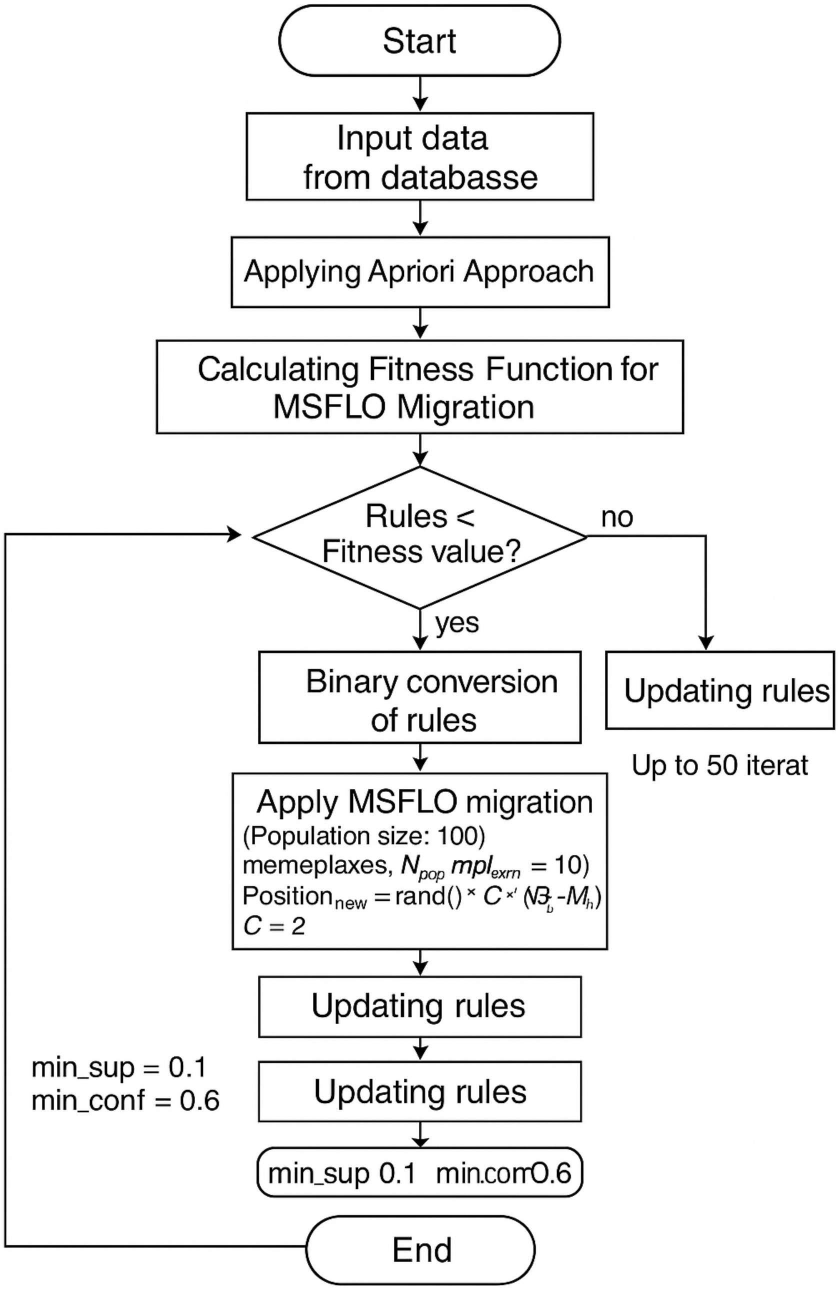

Figure 1
Flow chart of the proposed algorithm

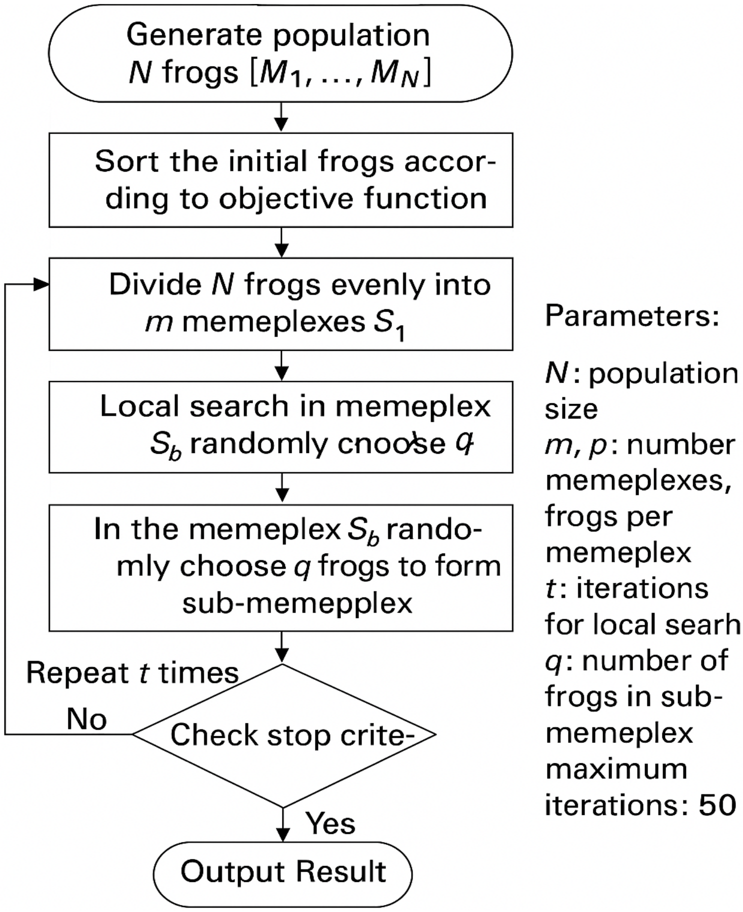

Figure 2
Rough set conversion

a_1-**Apple:** va_1 = {**128**; **128**; **16**}

a_2-**Blueberry:**va_2={**64**; **64**; **32**}

a_3-**Lemon:**va_3={**32**; **32**; **32**}

a_4-**Mango:**va_4={**128**; **128**; **16**}

a_5-**Orange:**va_5={**64**; **64**; **32**}

a_6-**Pear:**va_6={**32**;**32**;**32**}

a_7-**Pineapple:**va_7={**128**;**128**;**16**}

a_8-**Pomegranate:**va_8={**64**;**64**;**32**}

a_9-**Walnut:**va_9={**32**;**32**;**32**}

a_{10}-**Mean in other subjects:**va_{10} **= {16; 16; 128}**

a_{11}-**Fresh fruit recognition:** va_{11} **= {128; 128; 32}**

a_{12}-**Motivation:**va_{12}**={16;16;128}**

a_{13} **- Opinion from previous rotten fruit recognition:** va_{13} **= {128; 128; 32**

d **- Decision of system fresh fruit recognition:** v_d $\{A; R\}$:= **Now, the set of state attributes is** $A = \{a_1; a_2; a_3; a_4; a_5; a_6; a_7; a_8; a_9; a_{10}; a_{11}; a_{12}; a13\}$**.**

Table 1

Decision table

Condition / Applicants	A_1	A_2	A_3	A_4	A_5	A_6	A_7	Decision d
Y1	3	4	3	3	1	2	2	R
Y2	5	3	5	4	2	1	2	R
Y3	4	4	4	4	2	2	1	A
Y4	3	3	4	3	2	1	1	R
Y5	4	4	5	4	2	2	2	A
Y6	4	4	4	4	2	2	2	A
Y7	4	4	5	4	2	1	2	A
Y8	3	4	3	3	2	1	3	R
Y9	4	3	3	3	3	2	2	A
Y10	4	5	5	4	2	1	1	A
Y11	3	3	4	3	2	3	3	R
Y12	4	4	4	4	1	1	2	R
Y13	5	4	4	4	1	1	2	A
Y14	5	3	5	4	2	1	2	A

The dataset for training and validation comprises nine distinct fruit classes, namely apple, blueberry, lemon, mango, orange, pear, pineapple, pomegranate, and walnut, forming the example set. The decision attributes are defined as D = {A, R}, where A represents admission and R denotes rejection. These attributes are used to categorize the fruits based on specific criteria, and the detailed information is systematically presented in Table 1, providing a clear overview of the classification process and the structure of the dataset used in the study.

Table 1 shows ZB = {Y3 Y5 Y6 Y7 Y9 Y10 Y13 Y14} (accepted candidates) and ZR = {Y1 Y2 Y4 Y8 Y11 Y12} (rejected candidates). This example illustrates the final stage of our algorithm, demonstrating rule updates and the removal of high-migration rules.

3.2 Innovative Algorithm based on Optimization

In the realm of data mining and computational intelligence, the development of innovative algorithms rooted in optimization techniques has garnered significant attention. These algorithms aim to enhance the efficiency, accuracy, and scalability of various data mining processes. Here, we explore noteworthy contributions in this domain, highlighting algorithms that leverage optimization principles for improved performance.

One such algorithm, the Fuzzy Frequent Itemset (FFI)-Miner, stands out for its unique approach to mining complete sets of FFIs without the need for candidate generation. Employing a novel fuzzy-list structure, this algorithm efficiently retains essential information critical for subsequent mining processes. Further, a well-designed pruning strategy has been integrated to streamline the search space, resulting in expedited mining processes [10].

In the context of association rule mining, researchers have introduced algorithms that capitalize on optimization strategies. [11], for instance, puts forward an incremental queue algorithm model based on association rules, an improvement upon traditional FP-Growth algorithms. This model demonstrates enhanced efficiency in deriving complete association rules, particularly when dealing with incremental queues and association text correlations.

To tackle issues in association rule mining, [12] introduces the Modified Single-Objective Binary Cuckoo Search for Association Rule Mining (MBCS-ARM). This algorithm presents a new individual representation and effectively manages challenges related to high dimensionality as the number of attributes grows, ensuring robust performance.

Notably, MBCS-ARM extends its support to mining rules with attribute intervals exhibiting both negative and positive associations. While these algorithms represent significant strides in the domain of innovative optimization-based techniques, it is essential to acknowledge that challenges persist in current association rule mining methods [16, 17]. These challenges underscore the need for continuous innovation and refinement in the quest for algorithms that can efficiently and effectively extract meaningful patterns from complex datasets. Future research endeavors in this area are poised to further shape the landscape of innovative algorithms based on optimization principles.

4 Experiments

Numerous experiments were conducted on a 3.3 GHz Intel Processor with the 8 GB main memory with Python. The proposed technique is implemented and evaluated across six datasets, namely Watermelon, Mangosteen, Breast, Dragon Fruit, Mango, and Orange (Table 2), to ensure its robustness and versatility.

Table 2
Dataset characteristics

Dataset	Number of Records	Number of Features
Watermelon	10,000	500
Mangosteen	15,000	3,000
Breast	12,000	4,000
Dragon Fruit	9,500	2,200
Mango	13,000	3,500
Orange	11,000	2,800

In this data item set, the minimum support is 0.40 and min confidence is 0.60. The results of the proposed method (ARMSFLO) and other state of the art approaches.

Table 3
Comparison in terms of Support

Dataset	Telikani et al. (2020)	Pears & Koh (2021)	Sharmila & Vijayarani (2021)	ARM-MSFLO
Watermelon	31.14	34.21	36.25	67.14
Mangosteen	52.32	56.47	58.32	75.58
Breast	37.58	39.24	38.21	54.25
Dragon Fruit	43.67	46.15	47.82	68.92
Mango	49.18	50.45	53.11	72.34
Orange	41.23	43.89	45.67	66.78

Table 3 presents a comparative analysis of support values across six benchmark datasets: Watermelon, Mangosteen, Breast, Dragon Fruit, Mango, and Orange. The first three columns correspond to results obtained from existing methodologies—Telikani et al. (2020), Pears & Koh (2021), and Sharmila & Vijayarani (2021)—which exhibit marginal improvements in support across the datasets. In contrast, the fourth column, representing the proposed ARM-MSFLO method, shows a significant increase in support values across all datasets. This significant enhancement underscores the effectiveness of the ARM-MSFLO method in identifying stronger and more insightful association rules, demonstrating its capability to improve the quality and relevance of discovered patterns in the data, thereby contributing to more reliable and impactful outcomes in the analysis process.

Table 4
Comparison in terms of Confidence

Dataset	Telikani et al. (2020)	Pears & Koh (2021)	Sharmila & Vijayarani (2021)	ARM-MSFLO
Watermelon	31.14%	34.21%	36.25%	67.14%
Mangosteen	52.32%	56.47%	58.32%	75.58%
Breast	37.58%	39.24%	38.21%	54.25%

Dragon Fruit	43.67%	46.15%	47.82%	68.92%
Mango	49.18%	50.45%	53.11%	72.34%
Orange	41.23%	43.89%	45.67%	66.78%

Tables 4 and 5 present a comparative analysis of Confidence and the average number of rules across the six datasets: Watermelon, Mangosteen, Breast, Dragon Fruit, Mango, and Orange. The fourth column demonstrates an increase in all configurations across these fruit types.

Table 5
Comparison in terms of the average number of association rules

Dataset	**Telikani et al. (2020)**	**Pears & Koh (2021)**	**Sharmila & Vijayarani (2021)**	**ARM-MSFLO**
Watermelon	120	135	140	95
Mangosteen	210	230	225	160
Breast	180	200	195	140
Dragon Fruit	190	205	215	150
Mango	170	185	195	130
Orange	160	175	185	125

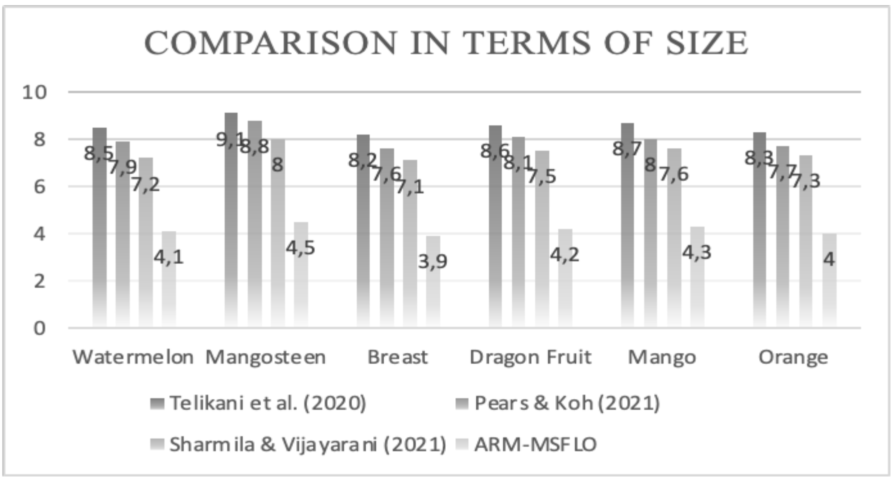

Figure 3
Comparison in terms of Size

The bar graph presented in Figure 3 displays the results, highlighting the differing Sizes (Y-axis) of the six fruit types examined in the study: Watermelon, Mangosteen, Breast, Dragon Fruit, Mango, and Orange, based on the analysis of their respective datasets. The findings clearly indicate that the ARM-MSFLO method achieves the most efficient and measurable results in terms of Size, outperforming other methods by consistently delivering superior outcomes across the diverse range of fruit types, thus underscoring its effectiveness in optimizing this particular metric within the scope of the research.

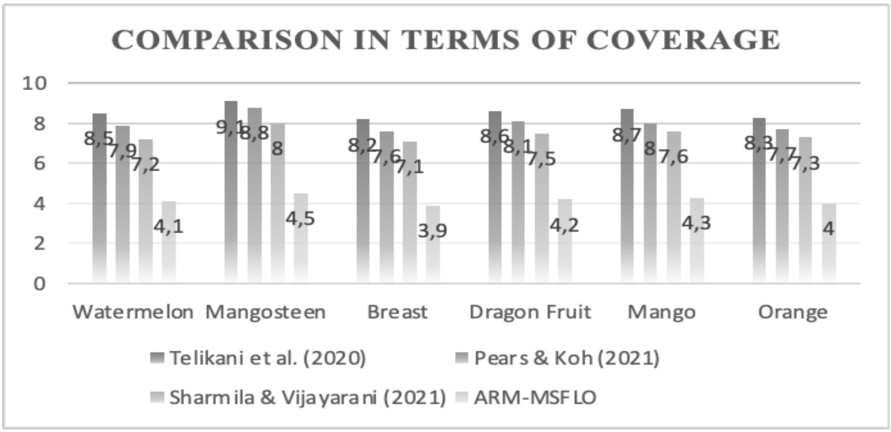

Figure 4
Comparison in terms of Coverage

The bar graph above (Fig. 4) illustrates the results, highlighting the variable Coverage (Y-axis) of the six fruit types under study: Watermelon, Mangosteen, Breast, Dragon Fruit, Mango, and Orange, whose datasets have been examined using three different research approaches: the evolutionary computation method by Telikani et al. (2020), the particle swarm optimization approach by Pears & Koh (2021), and the fuzzy logic with whale optimization method by Sharmila & Vijayarani (2021). An in-depth analysis indicates that the ARM-MSFLO method has delivered the most impressive measurable results with respect to Coverage, outperforming other approaches by achieving optimal performance metrics and demonstrating its effectiveness in comprehensively addressing the evaluated criteria across multiple test scenarios.

The bar graph above (Fig. 5) displays the results, illustrating the variable average execution time over 20 runs (Y-axis) for the six fruit types under study: Watermelon, Mangosteen, Breast, Dragon Fruit, Mango, and Orange, whose datasets have been analyzed using three different research approaches: the evolutionary computation method by Telikani et al. (2020), the particle swarm optimization approach by Pears & Koh (2021), and the fuzzy logic with whale optimization method by Sharmila & Vijayarani (2021). Through extensive testing, the ARM-MSFLO method has proven to yield the most efficient measurable results, achieving the lowest average execution time across 20 trials, highlighting its superior performance in optimizing computational processes.

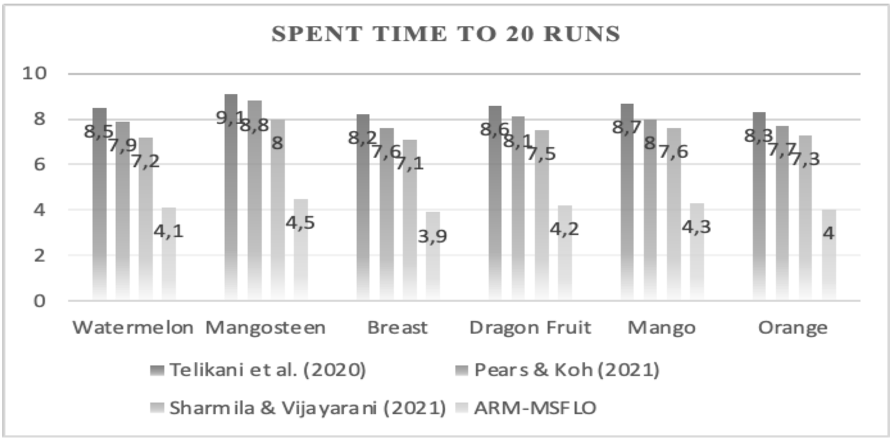

Figure 5
Spent Time to 20 runs

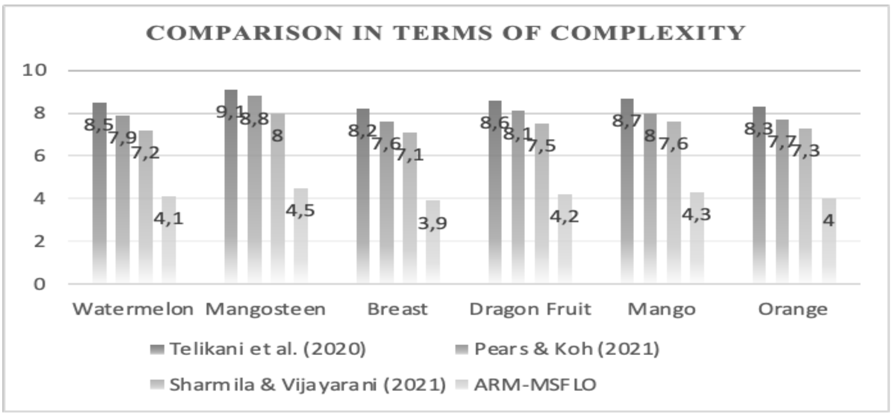

Figure 6
Comparison in terms of Complexity

The bar graph above (Fig. 6) presents the results, illustrating the varying time and space complexity (Y-axis) of the six fruit types under study: Watermelon, Mangosteen, Breast, Dragon Fruit, Mango, and Orange, whose datasets have been analyzed using three different research approaches: the evolutionary computation method by Telikani et al. (2020), the particle swarm optimization approach by Pears & Koh (2021), and the fuzzy logic with whale optimization method by Sharmila & Vijayarani (2021). A detailed comparison shows that the ARM-MSFLO method delivers the best measurable results in terms of time and space complexity.

This study clearly demonstrates that ARM-MSFLO outperforms other algorithms in terms of computational efficiency. The proposed algorithm performs better than existing methods, generating a higher number of rules compared to all other approaches. In contrast, to MSFLO, ARM-MSFLO enhances optimization by removing rules with high migration probability and updating them with new rules for improved performance.

Conclusion

This study introduced a novel technique for association rule mining by integrating the Modified Shuffled Frog Leaping Optimization (MSFLO) algorithm with the Apriori algorithm, termed ARM-MSFLO. The primary objective of this approach is to extract interesting and meaningful rules by calculating the fitness value of each rule, rather than relying solely on traditional minimum support and confidence thresholds. By encoding rules as frogs within the MSFLO framework, the proposed method processes the database in a single pass, significantly reducing CPU time and memory utilization compared to conventional multi-pass approaches. Experimental results, implemented in Python, demonstrate that ARM-MSFLO outperforms existing optimization strategies, such as those proposed by Telikani et al. (2020), Pears & Koh (2021), and Sharmila & Vijayarani (2021), in terms of computational efficiency, rule coverage, and the number of generated rules. Furthermore, unlike the standalone MSFLO, ARM-MSFLO enhances optimization by removing rules with high migration probability—those that are less stable—and replacing them with new, more effective rules, thereby improving overall rule quality. However, to further enhance the effectiveness and precision of the optimization process, future work should focus on incorporating additional measures into the modified SFLO framework, particularly for complex data mining challenges.

Acknowledgement

This research is funded by the Ministry of Education and Training under project number B2023.DNA.19.

References

[1] Agrawal, R., & Srikant, R. (1994) Fast algorithms for mining association rules. *Proceedings of the 20th International Conference on Very Large Data Bases (VLDB)*, 487-499

[2] Telikani, A., Gandomi, A. H., & Shahbahrami, A. (2020) A survey of evolutionary computation for association rule mining. *Information Sciences*, 524, 318-352

[3] Geng, L., & Hamilton, H. J. (2021) Interestingness measures for data mining: A survey. *ACM Computing Surveys*, 53(5), 1-32

[4] Luna, J. M., Kiran, R. U., & Ventura, S. (2022) A survey on actionable knowledge discovery with association rules. *Expert Systems with Applications*, 198, 116829

[5] Djenouri, Y., Belhadi, A., & Fournier-Viger, P. (2023) Extracting useful patterns from big data: An empirical study on association rule mining techniques. *Big Data Research*, 31, 100364

[6] Sharmila, S., & Vijayarani, S. (2021) Association rule mining using fuzzy logic and whale optimization algorithm. *Soft Computing*, 25(18), 12165-12180

[7] Zhang, K., Zhang, J., & Ma, W. (2022) Advances in frequent itemset mining: A comprehensive survey. *Knowledge and Information Systems*, 64(3), 587-628

[8] Aljehani, S., & Alotaibi, Y. (2024) Preserving privacy in association rule mining using multi-threshold particle swarm optimization. *Journal of Network and Computer Applications*, 224, 103842

[9] Pears, R., & Koh, Y. S. (2021) Weighted association rule mining using particle swarm optimization. *Applied Soft Computing*, 106, 107315

[10] Lin, J. C.-W., Li, T., Fournier-Viger, P., & Hong, T. P. (2021). A fast algorithm for mining fuzzy frequent itemsets. *Journal of Intelligent and Fuzzy Systems*, 29(6), 2373-2379

[11] Fournier-Viger, P., et al. (2022) Incremental FP-Growth for dynamic datasets. *Data Mining and Knowledge Discovery*, 36(2), 678-710

[12] Almasi, M., & Abadeh, M. S. (2023) Modified binary cuckoo search for association rule mining. *Expert Systems with Applications*, 212, 118742

[13] Nguyen, L. T., Nguyen, N. T., Vo, B., & Nguyen, H. S. (2021) Efficient method for updating class association rules in dynamic datasets. *Applied Intelligence*, 51(4), 2156-2174

[14] Nguyen, L. T., Vo, B., Fournier-Viger, P., & Selamat, A. (2022) Efficient top-k association rule mining algorithm. *Applied Intelligence*, 52(3), 2981-2995

[15] Vo, B., Le, T., Hong, T. P., & Le, B. (2020) An effective approach for maintenance of prelarge-based frequent-itemset lattice in incremental mining. *Applied Intelligence*, 41(3), 759-775

[16] Rekik, R., Kallel, I., Casillas, J., & Alimi, A. M. (2021) Assessing web sites quality: A systematic literature review by text and association rules mining. *International Journal of Information Management*, 58, 102345

[17] Eusuff, M., & Lansey, K. (2003) Optimization of water distribution network design using the shuffled frog leaping algorithm. *Journal of Water Resources Planning and Management*, 129(3), 210-225

[18] Amiri, B., et al. (2021) Enhancing shuffled frog leaping algorithm for improved optimization in data mining. *Swarm and Evolutionary Computation*, 62, 100851

[19] Sharma, T. K., & Pant, M. (2022) Shuffled frog leaping algorithm for clustering in data mining: A survey. *Expert Systems with Applications*, 187, 115892

[20] Zhang, Z., Pedrycz, W., & Huang, J. (2020) Efficient mining product-based fuzzy association rules through central limit theorem. *Applied Soft Computing*, 63, 235-248

[21] Wang, L., Meng, J., Xu, P., & Peng, K. (2021) Mining temporal association rules with frequent itemsets tree. *Applied Soft Computing*, 62, 817-829

[22] Nguyen, H. H. C., Hieu, H. P., Jana, C., Kiet, T. A., & Nguyen, T. T. (2024) Research on fruit shape database mining to support fruit class classification using the shuffled frog leaping optimization (SFLO) technique. *Mathematical Biosciences and Engineering, Volume 9, Issue 7:pp.19495-19514. doi: 10.3934/math.2024950*