WJCMS

Wasit Journal of Computer and Mathematics Science

Journal Homepage: https://wjcm.uowasit.edu.iq/index.php/WJCM
e-ISSN: 2788-5879   p-ISSN: 2788-5879

Optimization of Association Rule Using Ant Colony Optimization (ACO) Approach

Roni La'biran[1,*] and Muhammad Kristiawan[2]

[1]Indonesian Christian University, Toraja Jalan Nusantara No 12 Makale, Tana Toraja Indonesia
[2]University of Bengkulu Science of Education Jl. WR. Supratman, Kandang Limun Bengkulu, 38371A, Indonesia

*Corresponding Author: Roni La'biran

DOI: https://doi.org/10.31185/wjcms.190
Received: June 2023; Accepted: Septemper 2023; Available online: Septemper 2023

ABSTRACT: The Apriori algorithm creates all possible association rules between items in the database using the Association Rule Mining and Apriori Algorithm. Using Ant Colony Optimization, a new algorithm is proposed for improving association rule mining results. Using ant colony behaviour as a starting point, an optimization of ant colonies (ACO) is developed. The Apriori algorithm creates association rules. Determine the weakest rule set and reduce the association rules to find rules of higher quality than apriori based on the Ant Colony algorithm's threshold value. Through optimization and improvement of rules generated for ACO, the proposed research work aims to reduce the scanning of datasets.

Keywords: Data Mining, Association Rule Mining (ARM), Apriori Algorithm, Ant Colony Optimization (ACO), FP-Growth.

1. INTRODUCTION

This method analyses a database of association rules to identify those that satisfy predefined support and confidence criteria. Most association rules ARMs generate are not understandable or validated by end users. Therefore, data mining has limited use. The Apriori algorithm has made ARM very popular since it was introduced. In this study, frequent item sets are used to develop association rules. The research was based on an analysis of the market basket. The use of ARM is widespread, including business, medicine, cross-marketing, catalogue design, clustering, and classification, as well as risk analysis in businesses [1] (Chiclana et al., 2018).

A data mining technique allows large quantities of data to be analyzed to find patterns and trends [2] (Thong & Son, 2016). Many algorithmic techniques in this field include clustering, classification, association rule mining, pattern recognition, and sequence recognition [3](Vo et al., 2017). Graph-based ant colony optimization (ACO) is presented here as a method for mining association rules. ACO-ARM incorporates both Boolean and graph-based data representation schemes. Apriori and a scheme for representing data are enhanced in the first phase. Transaction data is represented through a Boolean matrix. The represented data is processed with an Apriori algorithm to generate n-frequent itemsets. An ACO-ARM graph of 2-frequent items is used in the second phase to estimate the final frequent itemset [4, 5] (Han & Kamber, 2006; Rani et al., 2022).

In data mining, revealed and previously unknown information is extracted from large databases to extract useful information. As an alternative to data mining, we can also analyze a large relational database from different angles to find correlations. Access to information about their customers' dealings is a powerful tool that can increase their sales and profits [4, 6](Han & Kamber, 2006; Hussain et al., 2022). Our lives depend on grouping in some ways [7–10](Berry & Linoff, 2004; Ceusters, 2001; Pesce, 2003; Tessmer, 1997). It has been used in a wide array of applications, including

*Corresponding author: ronilabran@ukitoraja.ac.id

swarm intelligence [11–16](Dorigo et al., 1996; Dorigo & Gambardella, 1997; Dorigo & Stützle, 2004; Eberhart et al., 2001; Engelbrecht, 2005, 2007).

Our data mining and archiving workflows use a high-performance cloud we developed [17] [18](Kumar et al., 2022; Yaginuma, 2000). Cloud computing involves using servers or infrastructure linked to the Internet to provide services and resources. Cloud computing services provide computing, while storage cloud services provide storage. This paper describes how the Sphere compute cloud relies on the Sector storage cloud for its storage needs. The Sphere compute cloud also supports a programming paradigm. Clustered computer clusters connected to high-speed wide-area networks (for example, 10Gbps) are used to analyze large data sets.

ACO is a highly effective paradigm when identifying classification rules involving nominal attributes [19] (Grosan et al., 2006). In addition, ACO avoids converting the problem into a binary optimization problem because it deals directly with nominal attributes. With this algorithm, the user is freed from the problems associated with conversion compared to other combinatorial optimization algorithms (e.g., binary PSO). Using binary PSO requires converting nominal attributes with more than two values into binary ones.

FIGURE 1. The steps involved in knowledge discovery from data mining

Subterranean insect provinces energize ACO's search activity, making it ideal for discrete streamlining methods [19](Grosan et al., 2006). During ACO's existence, it has addressed numerous issues. Streamlining discrete issues, such as quadratic problems, is usually the best application for it [20] (Hardin & Usher, 2005), work booking [21](Lorpunmanee et al., 2007), subset issues [22](Chakraborty, 2008), system directing [23](Sim & Sun, 2002), vehicle steering [24](Tan et al., 2006), diagram shading issue [25] (Salari & Eshghi, 2005), bioinformatics [26–28](Lee et al., 2009; Lin et al., 2008; Parpinelli et al., 2001) and information mining [29](Duda et al., 2001) This proposition aims to address that issue.

2. LITERATURE REVIEW

Recent years have seen a significant increase in interest in PSO and ACO. In addition, swarm intelligence metaheuristics are quite popular in data mining. Our literature tables provide an overview of these nature-inspired computing methods and discuss popular data mining techniques that use these principles. Moreover, a framework that categorizes search and organizing procedures as two approaches based on swarm intelligence is developed [30](Weng & Liu, 2006).

This technique is based on observing ants' behaviour while searching for food. Food is the only thing that keeps ants from wandering randomly in nature. Their nests are returned to after they have succeeded. A chemical path is formed by the pheromones they deposit as they move. Other ants that locate and follow these trails also reinforce these trails. Because of this, shorter pathways to food have a higher concentration of pheromones and are more likely to be followed. The ants' search for food is imitated in ACO algorithms, which use probability to solve computational problems. Several problems have been solved successfully with such techniques. With the proliferation of large amounts of data available, Data Mining (DM), an area of study that consists of developing techniques for uncovering before unidentified, Analyzing large data sets for patterns and relationships, has become one of the world's most significant technologies. Advanced association rule mining is performed using Ant state streamlining, and apriori calculations are conducted [24](Tan et al., 2006). The Association Rule Mining (ARM) method is commonly used in data mining research. A hidden relationship between items can be discovered using ARM. A user-specified threshold, also known as minimum support, can be used to detect all frequent patterns. Frequent patterns can be identified if a minimum level of support is provided.

Traditional supercomputers and workstation clusters were the focus of parallel ACO proposals. ACO search results have been improved, and the computation time of ACO searches has been reduced due to the new emerging parallel computing architectures, including multicore processors, graphics processing units (GPUs), and grid environments. Compared to the traditional approach, Swarm Intelligence (SI) has shown to be more effective for almost all engineering domains. The social imitation of insects and animals is developed from imitation and mimicry, as with ACO, ABC, FF, and Honey Bot. "Ant Algorithms" are studies that use model learning based on observations of ant colonies. Novel optimization and distributed control algorithms are designed using ACO models. Foraging, division of labour, brood sorting, and cooperative transport are among the Ant Algorithms proposed recently [31](Chandra Mohan & Baskaran, 2012).

3. APRIORI ALGORITHM

Several other researchers later investigated association rules problems. The algorithms were optimized by adding random sampling, parallel thinking, reference points, reducing rules, and changing storing frameworks, among other things [32](Deneubourg et al., 1990). An item set in a database is accessed using Apriori. In numerous cases, Apriori calculations are based on disregarding the database. Broadness-first searches are done using an iterative process known as a broadness-first search. This involves investigating k itemsets as well as (L+1) itemsets. Based on the Apriori property, Apriori calculations operate reasonably if a successive item of nonempty subsets is not empty. The framework was also shown to fail all its supersets if it failed the base support test. So, if an occasional set has an occasional superset, then the supersets of the set are also occasional. A rare competitor component is pruned using this property.

As a first step, we arrange the successive 1 items. One thing in the arrangement meets the support limit: L. Therefore, each successive pass begins with vast itemsets as seeds. These seeds enable the creation of prospective considerable item sets, named candidate item sets, and the corresponding supporting data is numbered. So, for example, L is used to learn L, followed by consecutive 2-item sets, which are used to discover L, etc. Mining the incessant components begins with the following fundamental steps:

- **Generate and test:** Examine the database and remove any components from C that fail to meet the base bolster requirements. The next step is discovering the sequential components L of the 1 item.

- **Join venture:** $K_{L-1} * K_{L-1}$ is the Cartesian result of K_{L-1}, which combines the successive components independent of each other. In the past emphasis, joining K with itself produces new applicant L item sets. In the past, emphasis on joining K_{L-1} with itself produces new applicant k item sets. The hopeful k-itemset is shown by C_L, and the incessant k-item-set is shown by K_L.

- **Prune step:** With the help of Apriori property, the C_L, obtains L successive item sets. C_L Individuals as a superset from C_L It will be visiting, but all L "1 steady item sets are included in C_L. An item using the Apriori property of the applicant is killed in this progression. All competitors with tally numbers not exactly the base benchmark number would be included in K_L (i.e., all contestants with score numbers not precisely the dishonourable benchmark number would be included in K_L). Even so, C_L It can take a toll on an individual; therefore, grave calculations could be involved. Apriori property is used to reduce C_L The extent is as follows. (L-1) items can't be subsets of subsequent L itemsets if they haven't been visited. Accordingly, the competitor cannot be visited if some (L-1) subsection of confident k item is not in K_{L-1}. As a result, it is possible to exclude the competitor K. There is no need to create a new set of applicants after rehashing steps 2 and 3.

Numerous varieties have been occurring in Apriori calculations to minimize the impediments that arise due to the size of the database. Therefore, proposed calculations are compared with database analyses with the same level-by-level analysis used for Apriori calculations. There are various ways to deal with eager erasure and pruning, bolster numbering, and competitor representations.

3.1 ACO ALGORITHM

Social insect societies or colonies of ants are distributed systems. A person's behaviour may appear simple, but their social structure is highly structured. An ant's ability to handle difficult tasks would be impossible without the support of a swarm. Based on natural ant behaviour, swarm intelligence inspires population-based heuristic algorithms called ant colony optimization. Models of ants' food-foraging behaviour led to the development of the ACO algorithm. The ants use pheromones to communicate trail information. Other ants arriving on the journey will be attracted to a high concentration of pheromones. A prescient model is created by ordering inconspicuous tests based on the last found standard set [19](Grosan et al., 2006).

In the initial stages of looking for food, ants move randomly. Each ant leaves behind a chemical pheromone trail while it moves. Pheromones are detectable by all ants. The ants choose the paths characterized by a high concentration of pheromones since the more pheromone trails in ant colonies, the better the path is. Ants gather a food source, some of which is brought back to the nest when they find another food source. An ant's pheromone production may vary in response to food quality and quantity during the return trip. Food sources must be located via pheromone trails.

Whether the ant is in the simplest case (A) or the most complicated case (B), its path has no obstacles. It is purely coincidental that ants run around when placed in the path of an obstacle, as in (B). Running leaves a permanent trail of pheromones behind. With time, these disappear.

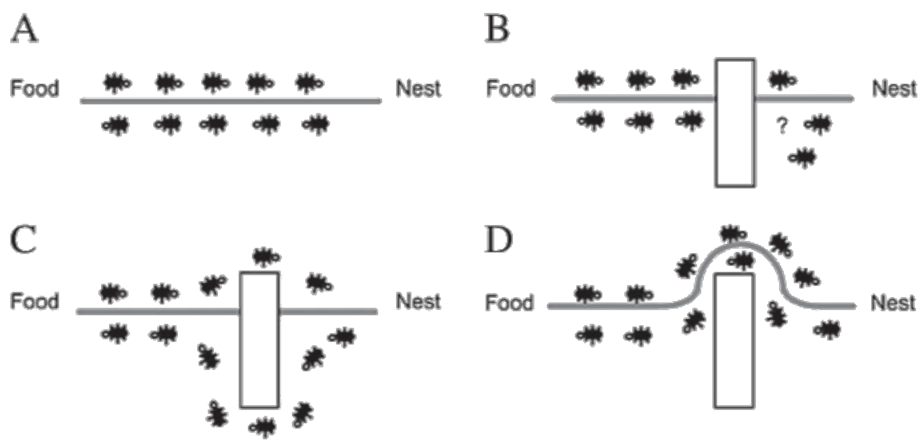

FIGURE 2. The working procedure of the ACO algorithm.

3.2 ARTIFICIAL BEE COLONY ALGORITHM WITH CROSSOVER

Adding the crossover operative of genetic algorithms to the ABC algorithm adds one more component. Another step is added to the overall process by the ABC algorithm after the employed bee phase. The crossover operator comes after the employed bee phase of the standard ABC algorithm. A crossover, a watcher, and a scout phase follow employed bee phases.

Adapting Deb's constrained handling method to ABC, we solved constrained optimization problems [19] (Grosan et al., 2006) rather than greedy selection [33](Kumar et al., 2023). Deb's method compares two solutions simultaneously, with the following criteria always being met:

- It is always preferable to find a feasible solution over an infeasible one,
- It is preferable to choose the feasible solution with the higher objective function value between two feasible solutions,
- A solution with fewer constraint violations is preferred between two infeasible solutions.

This algorithm consists of the following stages:

Initialization stage.

REPEAT

- Food sources in the Recollection are occupied by working bees;
- After applying the crossover operator to older offspring, produce new offspring.
- Memory bees act as onlookers to food bases;
- Scout bees should search the search area for new food sources.

4. RESULT SIMULATION

Using the execution time, we can also compare the number of comparisons made against the probabilistic factor with the time taken for mining the frequent itemsets. In Figure 2 and Figure 3, experimental results are illustrated graphically for the number of rules comparison time comparison. The black line indicates Apriori's performance with the ACO-based algorithm, and Apriori's performance with Artificial Bee Colonies is indicated by Figures 4 to 5.

.

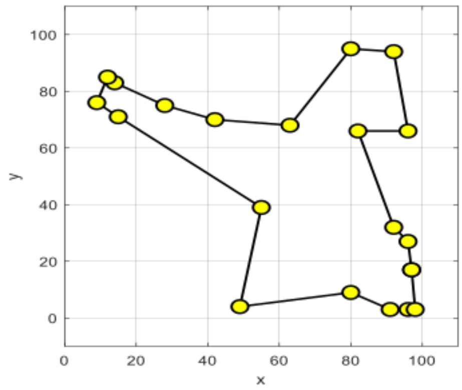

FIGURE 3. **Algorithms for optimizing ACO over Apriori.**

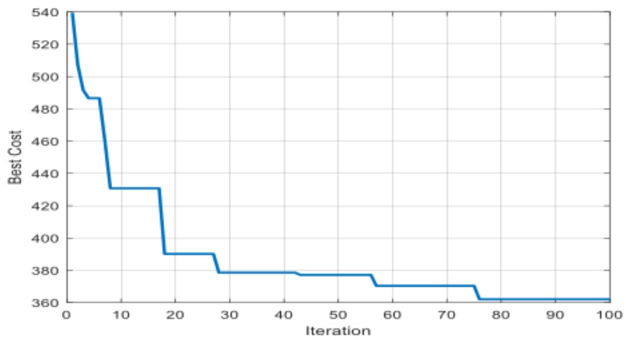

FIGURE 4. **Optimizing with ACO is more efficient than optimizing with Apriori.**

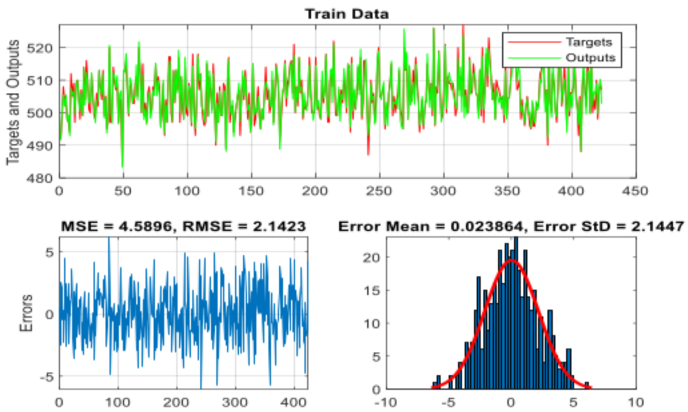

FIGURE 5. **An example of train data with error and performance value representations.**

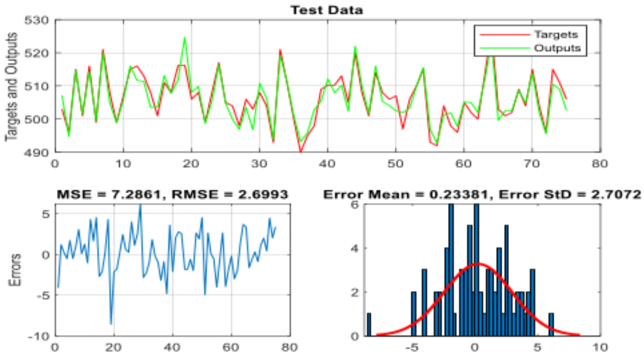

FIGURE 6. **The performance and error values of the test are displayed with their representations.**

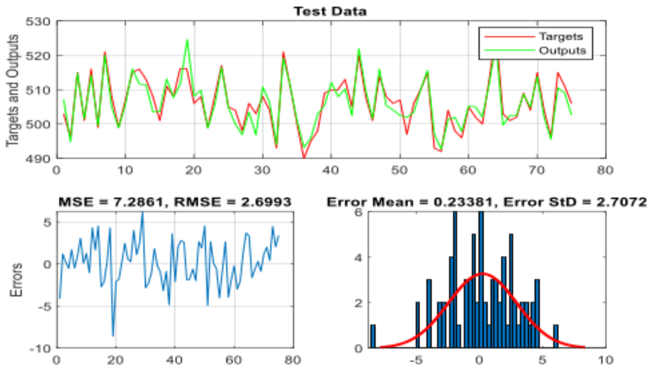

FIGURE 7. **A combination of data, performance information, and error values, represented by representations, is used in Train & Test.**

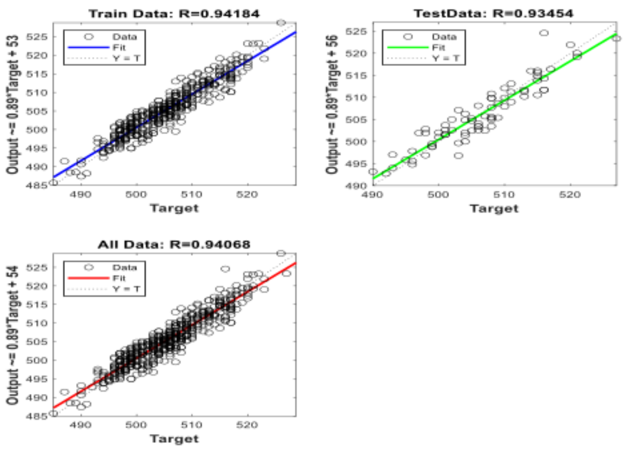

FIGURE 8. **All three data are represented by regression plots.**

5. CONCLUSION

Data mining relies heavily on association rule mining. Association rules can be mined using the Apriori algorithm. The apriori algorithm generates all significant associations between database items. A better Ant Colony Optimization algorithm will be proposed based on association rule mining and Apriori algorithms. An ACO-based associative classification algorithm is proposed in this paper by combining two important paradigms of classification and association rules mining. Compared to artificial bee colonies and apriori, artificial bee colonies take less time to process algorithms on datasets than apriori or artificial bee colonies. In comparison with existing techniques, the proposed method was very useful.

FUNDING

None

ACKNOWLEDGEMENT

None

CONFLICTS OF INTEREST

The author declares no conflict of interest.

REFERENCES

[1] F. Chiclana, R. Kumar, M. Mittal, M. Khari, J. M. Chatterjee, and S. W. Baik, “ARM-AMO: An efficient association rule mining algorithm based on animal migration optimization,” *Knowledge-Based Systems*, vol. 154, pp. 68–80, 2018.

[2] P. H. Thong, “A novel automatic picture fuzzy clustering method based on particle swarm optimization and picture composite cardinality,” *Knowledge-Based Systems*, vol. 109, pp. 48–60, 2016.

[3] B. Vo, S. Pham, T. Le, and Z. H. Deng, “A novel approach for mining maximal frequent patterns,” *Expert Systems with Applications*, vol. 73, pp. 178–186, 2017.

[4] J. Han, M. Kamber, and D. Mining, “Concepts and techniques,” *Morgan kaufmann*, vol. 340, pp. 94104–3205, 2006.

[5] P. Rani, S. Verma, S. P. Yadav, B. K. Rai, M. S. Naruka, and D. Kumar, “Simulation of the Lightweight Blockchain Technique Based on Privacy and Security for Healthcare Data for the Cloud System,” *International Journal of E-Health and Medical Communications (IJEHMC)*, vol. 13, no. 4, pp. 1–15, 2022.

[6] N. Hussain, P. Rani, N. Kumar, and M. G. Chaudhary, “A deep comprehensive research architecture, characteristics, challenges, issues, and benefits of routing protocol for vehicular ad-hoc networks,” *International Journal of Distributed Systems and Technologies (IJDST)*, vol. 13, no. 8, pp. 1–23, 2022.

[7] G. S. Linoff and M. J. Berry *Data mining techniques: for marketing, sales, and customer relationship management*, 2011.

[8] W. Ceusters, “Medical natural language understanding as a supporting technology for data mining in healthcare,” *Studies In Fuzziness And Soft Computing*, vol. 60, pp. 41–71, 2001.

[9] J. Pesce, “Stanching hospitals' financial hemorrhage with information technology,” *Health Management Technology*, vol. 24, no. 8, pp. 12–12, 2003.

[10] A. C. Tessmer, “What to learn from near misses: an inductive learning approach to credit risk assessment,” *Decision Sciences*, vol. 28, no. 1, pp. 105–120, 1997.

[11] M. Dorigo, V. Maniezzo, and A. Colorni, “Ant system: optimization by a colony of cooperating agents,” *IEEE transactions on systems, man, and cybernetics, part b (cybernetics)*, vol. 26, pp. 29–41, 1996.

[12] M. Dorigo and L. M. Gambardella, “Ant colony system: a cooperative learning approach to the traveling salesman problem,” *IEEE Transactions on evolutionary computation*, vol. 1, no. 1, pp. 53–66, 1997.

[13] M. Dorigo, M. Birattari, and T. Stutzle, “Ant colony optimization,” *IEEE computational intelligence magazine*, vol. 1, no. 4, pp. 28–39, 2006.

[14] R. C. Eberhart, Y. Shi, and J. Kennedy *Swarm Intelligence (Morgan Kaufmann series in evolutionary computation)*, 2001.

[15] A. Engelbrecht, *Fundamentals of Computational Swarm Intelligence*. Chichester, UK: John Wiley & Sons, 2005.

[16] A. P. Engelbrecht *Computational intelligence: an introduction*, 2007.

[17] N. Kumar, P. Rani, V. Kumar, S. V. Athawale, and D. Koundal, “THWSN: Enhanced energy-efficient clustering approach for three-tier heterogeneous wireless sensor networks,” *IEEE Sensors Journal*, vol. 22, no. 20, 2022.

[18] Y. Yaginuma, “High-performance data mining system,” *Fujitsu Scientific and Technical Journal*, vol. 36, no. 2, pp. 201–210, 2000.

[19] C. Grosan, A. Abraham, and M. Chis, “Swarm intelligence in data mining,” *Swarm Intelligence in Data Mining*, pp. 1–20, 2006.

[20] C. T. Hardin and J. S. Usher, “Facility layout using swarm intelligence,” *Proceedings 2005 IEEE Swarm Intelligence Symposium, 2005. SIS 2005*, pp. 424–427, 2005.

[21] S. Lorpunmanee, M. N. Sap, A. H. Abdullah, and C. Chompoo-Inwai, “An Ant Colony Optimization for Dynamic JobScheduling in Grid Environment,” *International Journal of Computer and Information Engineering*, vol. 1, no. 5, pp. 1343–1350, 2007.

[22] B. Chakraborty, “Feature subset selection by particle swarm optimization with fuzzy fitness function,” *2008 3rd international conference on intelligent system and knowledge engineering*, vol. 1, pp. 1038–1042, 2008.

[23] K. M. Sim and W. H. Sun, “Multiple ant-colony optimization for network routing,” *First International Symposium on Cyber Worlds, 2002. Proceedings*, pp. 277–281, 2002.

[24] X. Tan, X. Zhuo, and J. Zhang, “Ant colony system for optimizing vehicle routing problem with time windows (VRPTW),” *International Conference on Intelligent Computing*, pp. 33–38, 2006.

[25] E. Salari and K. Eshghi, "An ACO algorithm for graph coloring problem," *2005 ICSC Congress on computational intelligence methods and applications*, pp. 5–5, 2005.

[26] M. E. Lee, S. H. Kim, W. H. Cho, S. Y. Park, and J. S. Lim, "Segmentation of brain MR images using an ant colony optimization algorithm," *2009 Ninth IEEE international conference on bioinformatics and bioengineering*, pp. 366–369, 2009.

[27] C. J. Lin, C. Chen, and C. Lee, "Classification and medical diagnosis using wavelet-based fuzzy neural networks," *International Journal of Innovative Computing, Information and Control*, vol. 4, no. 3, pp. 735–748, 2008.

[28] R. S. Parpinelli, H. S. Lopes, and A. A. Freitas, "An ant colony based system for data mining: applications to medical data," *Proceedings of the 3rd annual conference on genetic and evolutionary computation*, pp. 791–797, 2001.

[29] R. Duda, P. Hart, and D. Stork, *Pattern classification, edition wiley interscience*. 2001.

[30] S. S. Weng and Y. H. Liu, "Mining time series data for segmentation by using Ant Colony Optimization," *European Journal of Operational Research*, vol. 173, no. 3, pp. 921–937, 2006.

[31] B. C. Mohan and R. Baskaran, "A survey: Ant Colony Optimization based recent research and implementation on several engineering domain," *Expert Systems with Applications*, vol. 39, no. 4, pp. 4618–4627, 2012.

[32] J. L. Deneubourg, S. Goss, N. Franks, A. Sendova-Franks, C. Detrain, and L. Chrétien, "The dynamics of collective sorting robot-like ants and ant-like robots," *From animals to animats: proceedings of the first international conference on simulation of adaptive behavior*, pp. 356–365, 1991.

[33] N. Kumar, P. Rani, V. Kumar, P. K. Verma, and D. Koundal, "TEEECH: Three-Tier Extended Energy Efficient Clustering Hierarchy Protocol for Heterogeneous Wireless Sensor Network," *Expert Systems with Applications*, vol. 216, pp. 119448–119448, 2023.