International Journal of Engineering Trends and Technology
ISSN: 2231 – 5381 / https://doi.org/10.14445/22315381/IJETT-V71I4P202

Volume 71 Issue 4, 12-20, April 2023

Original Article

A Hybrid Grouped-Artificial Bee Colony Optimization (G-ABC) Technique for Feature Selection and Mean-Variance Optimization for Rule Mining

Mrinalini Rana[1], Omdev Dahiya[2]

[1,2]School of Computer Science and Engineering, Lovely Professional University, Kapurthala, Punjab, India

[2]Corresponding Author : Omdev.26990@lpu.co.in

Received: 14 January 2023 Revised: 24 March 2023 Accepted: 01 April 2023 Published: 25 April 2023

Abstract - *Data mining is a widely used method for analyzing and discovering knowledge in large data sets. It identifies the pattern hidden in the information by using mathematics and algorithms that can help discover hidden attributes. Data mining techniques are used to mine relevant patterns from large databases. It is broadly categorized into data preprocessing and mining results based on analysis outcomes. Soft computing procedures are used popularly these days for pattern predictions these days. This paper presents rule mining using Grouped - Artificial Bee Colony Optimization(G-ABC) technique for feature selection and mean-variance optimization for further rule mining. Classifiers are used to train and test the model for both feature selection and rule mining. For performing the experimental analysis of the work, Twitter and Baseball datasets were used. The proposed algorithm demonstrated the most optimized for the number of rules generated, the time required for calculation, and getting supplementary normalized information for rule mining. The best performer G-ABC with Neural Network (NN) classifier represents an average of 97.56% accuracy, a precision of 61.11, a recall of 96%, and an f-measure of 75% with G-ABC and mean-variance optimization technique with the Neural Network classifier.*

Keywords - *Machine learning, Artificial bee colony optimization, Feature selection, Particle swarm optimization, Mean-variance optimization, Rule mining.*

1. Introduction

Data mining has played a vital role in decision-making in the last few decades and is considered an essential tool for performing different operations. Data mining is essential to discover unknown patterns from a large database [1, 2]. Various functionalities, algorithms, models, and techniques are used to discover and extract the relevant patterns from the large database repository. Association rule mining is a well-known technique of data mining. This technique extracts the correlation between the data points, frequent patterns, and associations between the structures [3, 4]. Association rule mining will be used for the analysis of association among datasets for performance and correctness. The motive of this paper is to describe this new method by associating it with the original problem.

In machine learning and artificial intelligence, association rule mining is one technique and way to use big data for information extraction [5]. As the process is an unsupervised learner, it depends on the similarity or correlations of the dataset members used for training. Association rule mining, also known as soft clustering or dependency mapping, helps represent the data visualizing map generated from different types of representations such as diagrams, etc. [6, 7].

Several mining algorithms are used to mine rules. The Apriori algorithm uses a nested data structure to narrow the search space and achieve better results than other algorithms. FP tree is another efficient data structure for finding rules from large datasets [1]. In this approach, we need to build one single node for each customer and its relation to their budget limit. The file created in this way can be fed into the FP tree algorithm, which will compute the branch of interest and return the branch's value. This value should fall within the restricted range determined by the user's budget limits. Let us say we want to find customers who have below $100 available for subscription; then we would check only one or two branches at the motor to get higher accuracy; we need more nodes in the branch representing different conditions above $100 availability.

Soft computing is a branch of artificial intelligence. It has been used to create algorithms that are adaptive to different strengths, weaknesses, and drawbacks in their input and response to the environment. The concept was developed in the 1970s by researchers at the University of Toronto's School of Computer Science, Canada. Another name for soft computing is fuzzy logic. Soft computing deals with modeling and analysis, which use extra information considered beforehand in difficult scenarios or problems. One of the main concepts behind soft computing is that it does not require precise mathematical modeling or statistical analysis to process data for solving problems. Repetitive patterns, learning rules, PSO, ABC, and using neural networks are some main examples [9,10,27].

Swarm optimization is a type of metaheuristics that uses algorithms and Particle swarm optimization to solve optimization problems [11]. It is a kind of population-based algorithm in which individual particles search for the optimal solution to a given problem. These particles are known as "particles" and are created by a well-defined set. In the beginning, all particles have an equal probability of moving toward their destination. The selected particles from this pool represent group members searching for the optimal solution to their given task. Swarm optimization is a population-based method for designing optimal solutions to a given problem [12]. A swarm consists of particles that define the "minimal unit" through which the swarm should travel to find a solution. The average velocity, denoted by ρ, describes how fast each particle can go in its current state [13-15,28].

Artificial Bee Colony Optimization (ABCO) is a diagnostic and predictive algorithm for the optimal management of hives using artificial bees[8,16,24]. The algorithm was inspired by honey bee daily activities [10]. ABCO is divided into three bees, an employee bee and two scout bees. The first job of an employee bee is to search for food; otherwise, the task is passed to another scout bee. In case there are multiple sources of food, chances of approval increase if the final location has more nectar than other sources. In other words, it becomes more likely that the food will be accepted if the hive already contained some type of nectar earlier in its life cycle, giving this location a higher probability score than others with fewer available resources [18].

The foremost contributions of this study paper are as follows:

- We proposed a novel Grouped Artificial Bee Colony Optimization (G-ABC) algorithm for preprocessing the rules or selecting rule-mining features.
- We proposed a rule mining algorithm where no minimum support and confidence threshold are required.
- We validate the proposed framework with parameters Accuracy, Precision, Recall, and F- measure on the proposed and traditional algorithm(s) using labeled or unlabelled data.

The paper is divided into five sections. Section 1 presented background information on evolutionary or soft computing techniques for association rule mining and the anticipated algorithm. Section 2 reviews related data structures and techniques. Section 3 discusses the proposed techniques and discusses its performance evaluation. Section 4 discusses results and suggestions for future work. The next section references list the sources that were used in this paper.

2. Related Work

This section documents the various survey related to different association rule mining algorithms using optimization techniques that generate positive and negative rules. Performance is analyzed by using a plethora of datasets. Four approaches are considered in this paper; Genetic algorithm, Bagging, a Combination of adaptive learning algorithms, and rule-based predictive filtering. Each algorithm is evaluated on different datasets and evaluated under various test scenarios.

The literature survey shows that a combination of support, confidence, lift, leverage, and conviction may be used to evaluate the interestingness. So, there is a scope for using such measures to generate appropriate rules. It has been observed that considering more metrics, such as amplitude, may obtain better rules. Therefore, these metrics can be used for better results. Performance may be improved by eliminating the need to determine the extent of the threshold for the criteria of support and confidence [19,29].

Many models have not used algorithms with categorical datasets [21]. So, this may add some scope to the research. It has been indicated that an increase in support value may give more appropriate rules [22]. Thus, there is some scope to enhance the efficiency of the rules.

3. Methodology

This section discusses the proposed methodology, which is divided into two parts. First, preprocessing is implemented, and then association rule mining is implemented. The proposed algorithm reduced the time of execution and the number of selected features. G-ABC used simple techniques to make it fast and easy to implement in an algorithm that works with the logarithmically sized set. In this research, we propose a new rule selection algorithm that consists of direct feature selection and association rule mining. The Fitness function used for finding the appropriate rule formula includes minimum and variance calculations. Two types of data sets are being used to verify the result. One is labeled data where direct feature selection and association rule mining can be implemented. The second data set is unlabeled data with the first K-means clustering algorithm implemented, followed by rest implementation.

Table 1. Summary of a few selected studies of existing work with techniques and datasets used

Author	Techniques	Dataset	Results
Zimbra *et al.*i 2016	ANN	Twitter dataset	The accuracy for the 3-class problem was 86%, while 85% was obtained for 5- the class problem.
Kale *et al.*, 2017	NAÏVE BAYES	Tweets	The accuracy using the Naïve Bayes classifier was 63.9%, while 27.8% was obtained using the Maximum Entropy.
Jianqiang *et al.*, 2018	CNN	Stanford Twitter Sentiment Dataset	The accuracy using the deep CNN model was 87.36%.
Alshariet et al., 2018	LEXICON BASED APPROACH	Movie Review Dataset	The accuracy for positive data was 85.4%, and 83.9% was obtained for negative data.
Bandana *et al.*, 2018	HYBRID TECHNIQUE BY INTEGRATING THE SENTIWORDNET, NAÏVE BAYES, AND SVM	Movie Review Dataset in which 250 samples were trained, and 100 samples were tested.	The accuracy using Naïve Bayes was 89%, and 76% using the SVM.
Ghosh and Sanyal, 2018	SEQUENTIAL MINIMAL OPTIMIZATION (SMO), MULTINOMIAL NAÏVE BAYES (MNB), RANDOM FOREST (RF) INTEGRATED WITH LOGISTIC REGRESSION (LR).	Movie Electronics Product Kitchenware	The F-measure using the SMO was 90.18. The accuracy for MNB was 88.18, 87.73 obtained using RF, and 87.32 obtained using the LR.
Sumit *et al.*, 2018	ANN	Facebook pages in the Bangladeshi language	The accuracy using the Skipgram technique was 83.79%, and 54.40% using Word to Index.
Zhang *et al.*, 2019	RULE MINING WITH GENETIC ALGORITHM	Various sizes of the road network	Concluded high accuracy with the proposed model
Zhou *et al.*, 2019	RULE MINING WITH CLUSTER ANALYSIS	MONKS problem ABC alphabet data set A–E alphabet data set SEA data set	Concluded that with large datasets proposed algorithm showed better results
Moslehi1 *et al.*, 2020	GA–PSO	Demographic and economic information Related to Iran	Concluded that the proposed algorithm showed better results for association rule mining
Pu *et al.*, 2021	PSO, ABC AND K-MEANS FOR CLUSTER ANALYSIS	Iris, Glass, CMC, Wine	Introduced K-means algorithm to generate initial clustering centers with high fitness, a formula based on the negative exponential function
Nagarajan *et al.*, 2021	GA-ABC	Heart Disease	For all original- features accuracy is 88.78% and for extracted feature, the accuracy is 92.34%.
Shreem *et al.*, 2022	ENHANCED BINARY GENETIC ALGORITHM	Student performance	For all classifiers used, performance increases between 1% to 11%.

Data was segregated into 3 labels: Positive, Negative, and Neutral. At the feature selection, stage proposed algorithm is compared with Basic PSO, ABC, and PSO-ABC algorithms. Four classifiers are used, KNN, NB, SVM, and NN, to validate the algorithm, with 70 to 30 percent of the training and testing division.

Proposed pseudo code:

1. Read the dataset
2. Labelling dataset 3 labels: Positive, Negative, and Neutral
3. Preprocessing
   a. Remove the stop words
   b. Stemming
   c. Tokenizer to generate numbers for each Term(s)
4. Grouped ABC
   a. Fitness function
      1. If(ebee >obee)
      2. Fitbee=1
      3. Else
      4. Fitbee=0
   b. Feature data= combine employee bee (randomly)
   c. Obee = mean (Ebee)
   d. Fitness (Ebee, Obee)
   e. Create final Data
5. Train and split the data
6. Apply KNN
7. Apply SVM
8. Apply NB
9. Apply NN
10. Apply Mean-Variance Optimization for Association Rule mining
11. Create 3 populations for each label or class
12. Calculate the mean and variance for each particle
13. If $\left(\frac{(mean*variance)-populationvalue}{(mean*variance)+populationvalue}\right)*100 < mean(complete\_population)$
    Value accepter
    Else
    Value rejected
14. Repeat steps 5 to 9 15.
15. Analyze the performance

3.1. Preprocessing

Words like "the" and "and" are commonly used in documents and make data analysis more easily understandable, but these words are not required for analysis[23]. These words can be ignored as they are called stop words. Other words, such as proper nouns, numbers, conjunctions, and prepositions, are used frequently in English. So, it may be a good idea to remove such common terms when converting data into text format. It is also possible to use stemmer, which removes any suffixes like –ed, –ing, etc., before processing them as a sentence or token based on their frequency of occurrence (English corpus). Some sample stop words are as follows

Tokenization is the process of breaking up each record into sentences through a machine called a tokenizer. Tokenization aims to take each text and break it down into an unchangeable sequence of words or tokens (procedures for which they exist). This has significant implications for data analysis, where text analysis must be performed independently on each token[25,30].

3.2. Feature Selection

After preprocessing the textual data, it was observed that the number of features to be used for the next phase remains the same. So to mine optimized features, further feature selection is implemented. For this proposed model, G-ABC is used. In the proposed model, five employees' bees work parallel instead of one at a time. With this implementation, more employees are working at one time. Therefore, the time of execution reduces.

The structure of the proposed algorithm is as follows:

a. Fitness function
b. If(ebee >obee)
   i. Fitbee=1
c. Else
   i. Fitbee=0
d. Feature data= combine employee bee (randomly)
e. Obee = mean (Ebee)
f. Fitness (Ebee, Obee)
g. Create final Data
h. Train and split the data
i. Apply KNN
j. Apply SVM
k. Apply NB
l. Apply NN

Fig. 1 Sample Stop Words

3.3. Rule Mining

Rule mining is a method where we search for patterns in the data. In the case of medical data, it could be used to predict something like age at the time of diagnosis and other personal attributes that can greatly impact human life. Rule mining is a commonly used technique for analyzing large datasets such as text, images, or audio files[26]. Applying an algorithm called the Mean-Variance Optimization Algorithm (MVA) to our dataset gives us an interesting statistical model that can help us with future predictions and classification. This can come in handy when we cannot immediately identify a pattern in exploratory analysis. However, we still want to know if there is any relationship between the two objects in our dataset. That can help us with future predictions and classification, such as patient mortality prediction etc. [21].

The structure of the proposed algorithm is as follows:

1. Apply Mean-Variance Optimization for Association Rule mining
2. Create 3 populations for each label or class
3. Calculate the mean and variance for each particle
4. If $\left(\frac{(mean*variance)-populationvalue}{(mean*variance)+populationvalue}\right)*100 < mean(complete\_population)$
   Value accepter
   Else
   Value rejected
5. Train and split the data
6. Apply KNN
7. Apply SVM
8. Apply NB
9. Apply NN
10. Analyze the performance

This framework is designed to perform rule mining tasks based on the theory that when there is a high level of confidence in rules and support requirement is low, no threshold needs to be applied. It also uses a feature selection technique to select relevant features from n number of irrelevant features.

This framework performs rule mining tasks using mean and variance optimization methods where data elements or population after feature selection is divided into two parts. First, it is divided into three classes using the mean * variance optimization method. Afterwards, mean and variance is calculated for each element or population, and if the condition is true, the value is accepted; otherwise, the value is rejected. After calculating the final rules again, four classifiers, KNN, SVM, NB, and NN, are used for a complete evaluation.

4. Results and Discussion

The paper presents rule mining, a technique to discover patterns in data automatically. It is based on data clustering and procedures to build generalizable rules from specific examples. Rule mining can be applied in many application areas where information is abundant but lacks clarity of meaning. For instance, in interpreting user intentions, medical tests, driving license or usage logs, and online advertising analytics. The application of rule mining shows its potential in exploring complex patterns of features that are hidden in large datasets without any prior assumptions or knowledge about the underlying structure.

This section discusses a machine learning-based approach for mining tweets, and the parameters for performance evaluations are high precision, recall, and accuracy. The performance of G-ABC and Mean-Variance Rule mining is evaluated based on 100 records. All the performance measures represent better outcomes with the proposed method than other existing algorithms. It is evident from the diagram that G-ABC outperforms all other methods in terms of all performance measures. In figure 4 accuracy value is represented where the proposed algorithms represent approximately 98 percent of accuracy with both datasets compared to existing algorithms.

In Figure 6, the precision value represents where it is analyzed that the proposed algorithm represents approx. 68 percent precision with both datasets for the Twitter dataset and approx. 76 percent with the Baseball dataset, respectively, compared to existing algorithms. In the proposed method, the recall value is approx. 98% and 96% for the Twitter dataset, while the baseball dataset represents approx—94%.

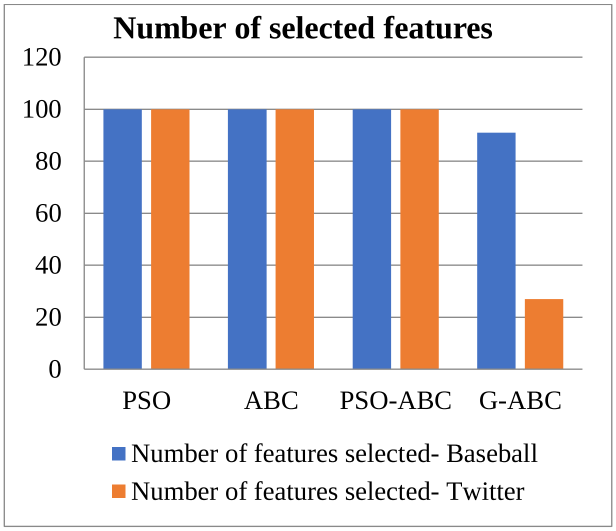

Fig. 2 Analysis of the number of selected features

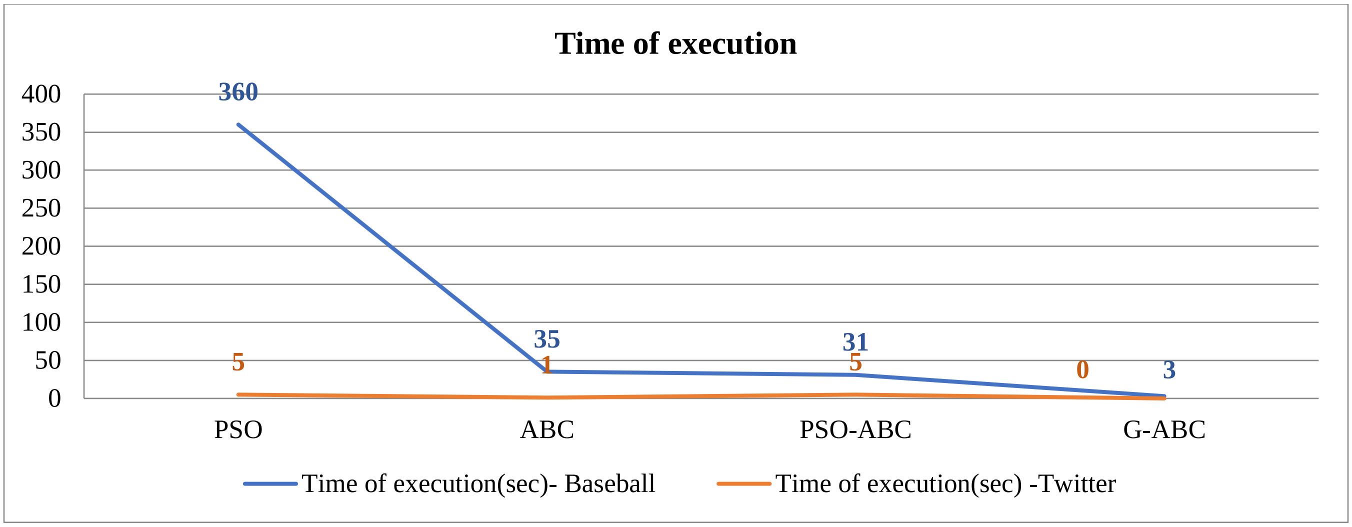

Fig. 3 Analysis of execution time taken by proposed and existing techniques

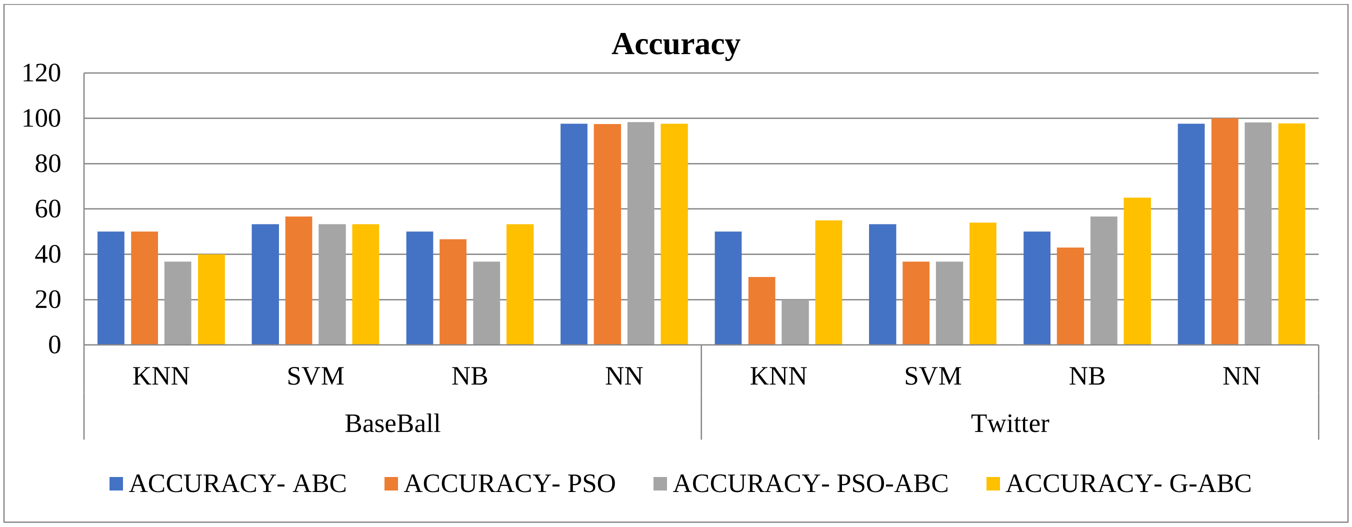

Fig. 4 Analysis of accuracy by proposed and existing techniques

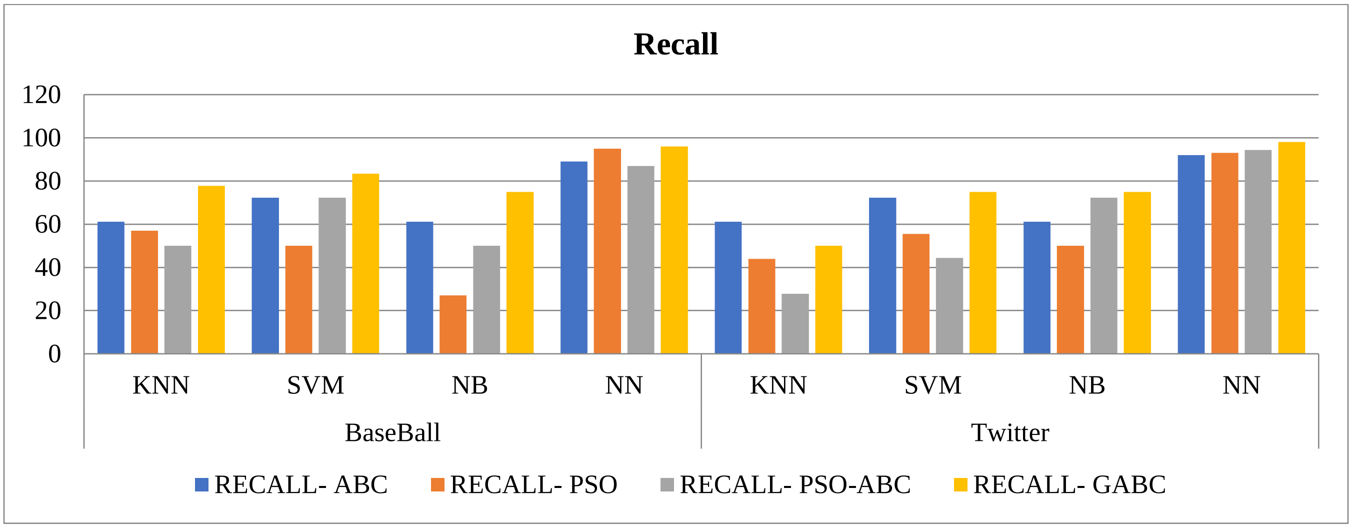

Fig. 5 Analysis of recall by proposed and existing techniques

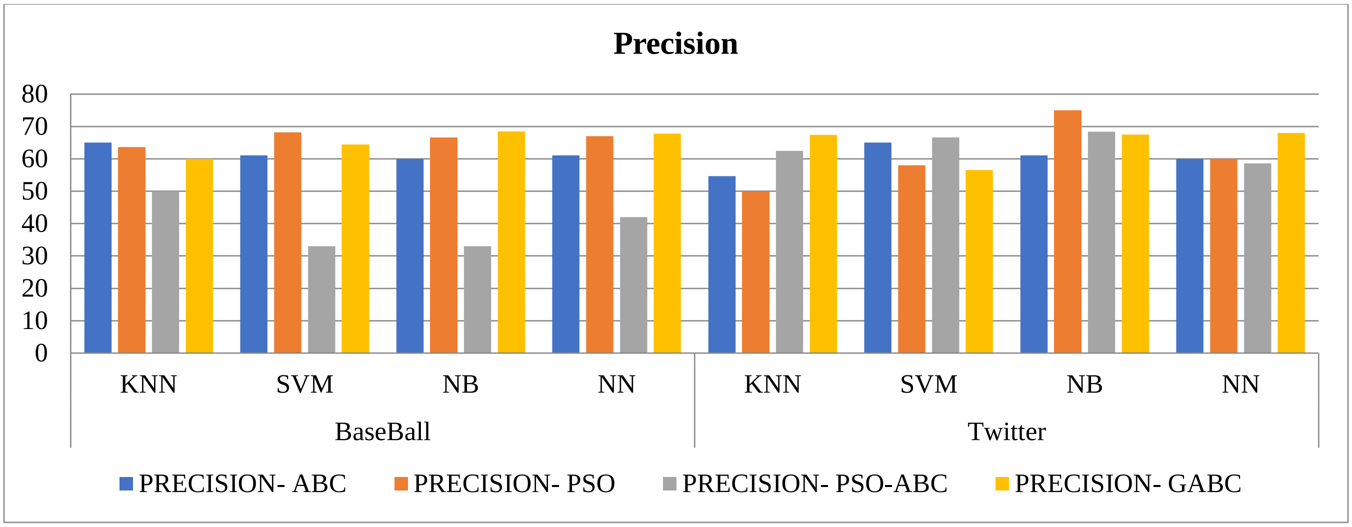

Fig. 6 Analysis of precision by proposed and existing techniques

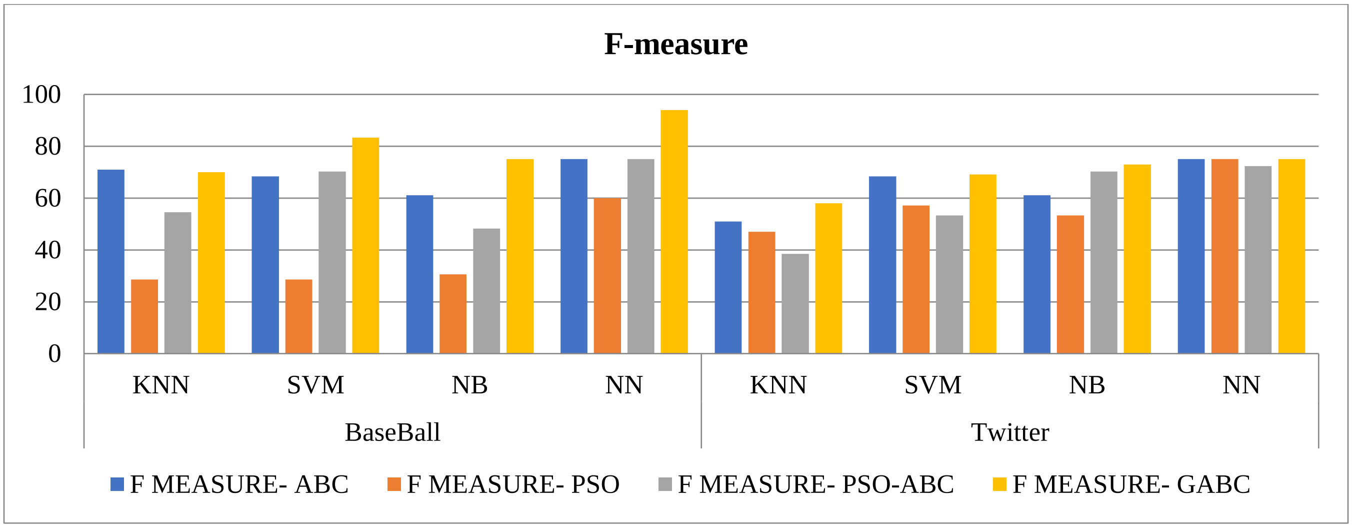

Fig. 7 Analysis of f-measure by proposed and existing techniques

The F-measure value presented in Figure 7 shows that the proposed method has a better F-measure result than other algorithms, which indicates higher performance in mining big data due to its efficient processing of mean-variance rule mining.

The proposed technique is used to select the best features from a dataset. This can be used to identify links between two variables based on a previous conclusion of the data. In the baseball dataset, G-ABC selected the most relevant features, with 91 features selected from 100 records, as evident from figure 2. This proposed GABC's takes a minimum time of execution for selecting features. For the analysis, 100 records are used, and figure 3 represents the required minimum time for the proposed algorithm. The graph clearly represents that for both data sets proposed algorithm selects a smaller number of features that are directly proportionate to reduce the time of execution.

Figure 4 shows the accuracy comparison for G-ABC analysis using the NB, KNN, SVM, and NN for the baseball and Twitter datasets. The analysis result shows that the least F-measure is shown by KNN, followed by the NB and SVM classifiers for both datasets. NN shows a better performance for both datasets. Thus, the NN classifier provides better results with G-ABC for rule mining.

Figure 5 shows the recall comparison for G-ABC analysis using the NB, KNN, SVM, and NN for the baseball and Twitter datasets. The analysis result shows that the least F-measure is shown by KNN, followed by the NB and SVM classifiers for both datasets. NN shows a better performance for both datasets. Thus, the NN classifier provides better results with G-ABC for rule mining.

Figure 6 shows the precision comparison for G-ABC analysis using the NB, KNN, SVM, and NN for the baseball and Twitter datasets. The analysis result shows that the least F-measure is shown by KNN, followed by the NB and SVM classifiers for both datasets. NN shows a better performance for both datasets. Thus, the NN classifier provides better results with G-ABC for rule mining.

Figure 7 shows the f-measure comparison for G-ABC analysis using the NB, KNN, SVM, and NN for the baseball and Twitter datasets. The analysis result shows that the least F-measure is shown by KNN, followed by the NB and SVM classifiers for both datasets. NN shows a better performance for both datasets. Thus, the NN classifier provides better results with G-ABC for rule mining.

5. Conclusion and Future Work

This study explored how artificial intelligence can be applied in the field of rule mining and presented a new algorithm called Grouped Artificial Bee Colony Optimization (G-ABC) which has been adapted from natural bee colony optimization. G-ABC is well suited for preprocessing or selection features when it comes to mining rules. It also offered promising results during its evaluation against traditional algorithms. This research was conducted to reduce the time of execution and reduce the itemset generated without any support and confidence threshold value. The core mechanism is performing the feature selection first using G-ABC algorithms. The proposed algorithm reduced the time of execution and the number of selected features. G-ABC used simple techniques to make it fast and easy to implement in an algorithm that works with the logarithmically sized set. In this research, we propose a new rule selection algorithm that consists of direct feature selection and association rule mining. The Fitness function used for finding the appropriate rule formula includes minimum and variance calculations. Two types of data sets are being used to verify the result. One is labeled data where direct feature selection and association rule mining can be implemented. The second data set is unlabeled data with the first K-means clustering algorithm implemented, followed by rest implementation. Data was segregated into 3 labels: Positive, Negative, and Neutral. At the feature selection, stage proposed algorithm is compared with Basic PSO, ABC, and PSO-ABC algorithms. Four classifiers are used, KNN, NB, SVM, and NN, to validate the algorithm, with 70 to 30 percent of the training and testing division.

The G-ABC algorithm based on the neural network has obtained the best possible accuracy, precision, and recall performance. The proposed algorithm combines the greedy meta-heuristic algorithm with minimum length and feature selection techniques to derive intuitive rules for association rule mining. This paper aims to evaluate the G-ABC-based rule extraction method against other methods used in Computer Science literature. The results have been evaluated with two metrics - accuracy and precision, as well as f measure, which measures generalization over fitness function built from previous rules to predict new data instances. The proposed technique combines different computational methods and performs better than the simple addition of operations. Experimental results show an accuracy of 97.56%, a precision of 61.11, a recall of 96%, and an f-measure of 75% with G-ABC and mean-variance optimization technique with the Neural Network classifier. We use different datasets, each containing different information, in order to improve our algorithm and therefore produce a better solution that reaches an approximate solution faster than in an environment where only one dataset is used. More datasets and meta-heuristic algorithms may be used for further study to get a better solution.

References

[1] Rakesh Aggarwal, Tomasz Imielinksi, and Arun Swami, "Mining Association Rules between Sets of Items in Large Databases," *ACM SIGMOD International Conference on Management of Data*, vol. 22, no. 2, pp. 207-216, 1993. [CrossRef] [Google Scholar] [Publisher Link]

[2] Krzysztof J. Cios, Witold Pedrycz, and Roman W. Swiniarski, "*Data Mining and Knowledge Discovery,*" Data Mining Methods for Knowledge Discovery, Springer, vol. 458, pp. 1-26, 1998. [CrossRef] [Google Scholar] [Publisher Link]

[3] M. V. Fidelis, H. S. Lopes, and A. A. Freitas, "Discovering Comprehensible Classification Rules with a Genetic Algorithm," *Proceedings of the 2000 Congress on Evolutionary Computation*, vol. 1, pp. 805-810, 2000. [CrossRef] [Google Scholar] [Publisher Link]

[4] Ujjwal Maulik, and Sanghamitra Bandyopadhyay, "Genetic Algorithm-Based Clustering Technique," *Pattern Recognition,* vol. 33, no. 9, pp.1455-1465, 2000. [CrossRef] [Google Scholar] [Publisher Link]

[5] Celia C. Bojarczuk, Heitor S. Lopes, and Alex A. Freitas, "Genetic Programming for Knowledge Discovery in Chest Pain Diagnosis," *IEEE Engineering in Medicine and Biology Magazine*, vol. 19, no. 4, pp. 38-44, 2000. [CrossRef] [Google Scholar] [Publisher Link]

[6] Jacinto Mata, Jose Luis Alvarez, and Jose- Cristobal Riquelme, *Discovering Numeric Association Rules via Evolutionary Algorithm*, Advances in Knowledge Discovery and Data Mining, PAKDD, Lecture Notes in Computer Science, Springer, vol. 2336. pp. 40–51, 2002. [CrossRef] [Google Scholar] [Publisher Link]

[7] Vahid Beiranvand, Mohamad Mobasher-Kashani, and Azuraliza Abu Bakar, "Multi-Objective PSO Algorithm for Mining Numerical Association Rules without a Priori Discretization," *Expert Systems with Applications,* vol. 41, no. 9, pp. 4259–4273, 2014. [CrossRef] [Google Scholar] [Publisher Link]

[8] N.Deexit and A.Surya Prakasa Rao, "Solving Economic Load Dispatch Problem using Particle Swarm Optimization and Artificial Bee Colony Optimization Algorithms," *SSRG International Journal of Electrical and Electronics Engineering*, vol. 5, no. 7, pp. 5-8, 2018. [CrossRef] [Publisher Link]

[9] Fateme Moslehi, and Abdorrahman Haeri, "A Novel Hybrid Wrapper–Filter Approach Based on Genetic Algorithm, Particle Swarm Optimization for Feature Subset Selection," *Journal of Ambient Intelligence and Humanized Computing*, vol. 11, no. 3, pp.1105-1127, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[10] Qiumei Pu et al., "An Efficient Hybrid Approach Based on PSO, ABC, and k-means for Cluster Analysis," *Multimedia Tools and Applications*, vol. 81, pp.19321-19339, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[11] Jitendra Agrawal et al., "SET-PSO-based Approach for Mining Positive and Negative Association Rules," *Knowledge and Information Systems, Springer*, vol. 45, pp. 453–47, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[12] M. Karunyalakshmi, and N. Tajunisha, "Classification of Cancer Datasets Using Artificial Bee Colony and Deep Feed-Forward Neural Networks," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 6, no. 2, pp. 33–41, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[13] Beatriz de la Lglesia et al., "Data Mining Rules Using Multi-Objective Evolutionary Algorithms," *Congress on Evolutionary Computation*, pp. 1552-1559, 2003. [CrossRef] [Google Scholar] [Publisher Link]

[14] Chad Creighton, and Samir Hanash, "Mining Gene Expression Databases for Association Rules," *Bioinformatics*, vol. 19, no. 1, pp.79-86, 2003. [CrossRef] [Google Scholar] [Publisher Link]

[15] Ashish Ghosh, and Bhabesh Nath, "Multi-Objective Rule Mining using Genetic Algorithms," *Information Sciences*, vol.163, no. 1-3, pp.123–133, 2004. [CrossRef] [Google Scholar] [Publisher Link]

[16] Vanam Satyanarayana, and Vairavasamy Jayasankar, "Advanced Modeling and Optimization of Hybrid Renewable Energy Management Strategy Based on Artificial Bee Colony Algorithm in Micro Grid," *International Journal of Engineering Trends and Technology*, vol. 70, no. 10, pp. 329-338, 2022. [CrossRef] [Publisher Link]

[17] Bilal Alatas, and Erhan Akin, "An Efficient Genetic Algorithm for Automated Mining of both Positive and Negative Quantitative Association Rules," *Soft Computing, Springer*, vol. 10, pp. 230–237, 2006. [CrossRef] [Google Scholar] [Publisher Link]

[18] Daniela Zaharie, Diana Lungeanu, and Flavia Zamfirache, "Influence of Missing Values Handling on classification Rules Evolved from Medical Data," *8th Industrial Conference on Data Mining ICDM*, pp. 86-95, 2008. [Google Scholar] [Publisher Link]

[19] S. Meera, and C. Sundar, "A Hybrid Metaheuristic Approach for Efficient Feature Selection Methods in Big Data," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, pp. 3743-3751, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[20] Satish S. Banait et al., "Literature Review: An Efficient Clustering Approach to Big Data," *International Journal of Computer Trends and Technology*, vol. 71, no. 2, pp. 25-31, 2023. [CrossRef] [Publisher Link]

[21] Mrinalini Rana, and Jimmy Singla, "A Framework for Selecting Features using Various Soft Computing Algorithms," *6th International Conference on Trends in Electronics and Informatics (ICOEI)*, pp. 1743-1748, 2022, [CrossRef] [Google Scholar] [Publisher Link]

[22] Salam Salameh Shreem et al., "Enhanced Binary Genetic Algorithm as a Feature Selection to Predict Student Performance," *Soft Computing*, vol. 26, no. 4, pp.1811-1823, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[23] Omdev Dahiya, and Kamna Solanki, "An Efficient Requirement-Based Test Case Prioritization Technique Using Optimized TFC-SVM Approach," *International Journal of Engineering Trends and Technology*, vol. 69, no. 1, pp. 5-16, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[24] C.Ambika, M.Karnan, and R.Sivakumar, "Resolving Dynamic Shortest Path Routing Problems in Mobile Adhoc Networks Using ABCAnd ACO," *International Journal of Computer & organization Trends*, vol. 3, no. 1, pp. 58-62, 2013. [Google Scholar] [Publisher Link]

[25] Mrinalini Rana, and Jimmy Singla, "A Comprehensive Review of Data Mining Rules Generation Algorithms," *Turkish Online Journal of Qualitative Inquiry*, vol. 12, no. 6, pp. 5269-5276, 2021. [Google Scholar] [Publisher Link]

[26] Mrinalini Rana, and Jimmy Singla, "A Systematic Review on Data Mining Rules Generation Optimizing via Genetic Algorithm," *International Conference on Innovative Computing & Communications (ICICC)*, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[27] R. J. Kuo, Monalisa Gosumolo, and Ferani E. Zulvia, "Multi-Objective Particle Swarm Optimization Algorithm Using Adaptive Archive Grid for Numerical Association Rule Mining," *Neural Computing and Applications*, vol. 31, no. 8, pp. 3559-3572, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[28] Shady Shehata, Fakhri Karray, and Mohamed Kamel, "Enhancing Text Clustering using Concept-based Mining Model," *Sixth International Conference on Data Mining (ICDM'06)*, pp. 1043-1048, 2006. [CrossRef] [Google Scholar] [Publisher Link]

[29] Senthil Murugan Nagarajan et al., "Feature Selection Model for Healthcare Analysis and Classification Using Classifier Ensemble Technique," *International Journal of System Assurance Engineering and Management*, pp. 1-12, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[30] Omdev Dahiya, and Kamna Solanki, "An Efficient APHT Technique for Requirement-Based Test Case Prioritization," *International Journal of Engineering Trends and Technology*, vol. 69, no. 4, pp. 215-227, 2021. [CrossRef] [Google Scholar] [Publisher Link]