IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 8 July 2023, accepted 3 August 2023, date of publication 7 August 2023, date of current version 16 August 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3303360

RESEARCH ARTICLE

Grouped ABC for Feature Selection and Mean-Variance Optimization for Rule Mining: A Hybrid Framework

MRINALINI RANA[1], **OMDEV DAHIYA**[1], **PARMINDER SINGH**[1,2], **(Member, IEEE),**
WADII BOULILA[3,4], **(Senior Member, IEEE), AND ADEL AMMAR**[3]

[1]School of Computer Science and Engineering, Lovely Professional University, Phagwara, Punjab 144001, India
[2]School of Computer Science, Mohammed VI Polytechnic, Ben Guerir 43150, Morocco
[3]Robotics and Internet of Things Laboratory, Prince Sultan University, Riyadh 12435, Saudi Arabia
[4]RIADI Laboratory, National School of Computer Sciences, University of Manouba, Manouba 2010, Tunisia

Corresponding authors: Omdev Dahiya (Omdev.26990@lpu.co.in) and Wadii Boulila (wboulila@psu.edu.sa)

This work was supported by Prince Sultan University.

ABSTRACT Data mining has become a popular process in recent times. However, with the increase in data, traditional data mining methods are not sufficient to solve many problems. Therefore, advanced techniques are needed to provide better results without consuming more time during execution. Soft computing algorithms are used for mathematical optimization to achieve better results in less time. The primary purpose of this work is to propose a framework for rule mining that shall generalize the currently applied methods in rule mining. In this respect, this paper represents the R-miner using a soft computing algorithm. The Grouped -Artificial Bee Colony Optimization (G-ABC) was used to select the relevant attribute set and further verify the features. Mean-Variance optimization is used to find whether the selected rule is valid for further classification. Furthermore, a neural-based deep learning method is applied to validate the outcome. The investigation outcome indicates that the proposed algorithm provides more optimized results in terms of the number of rules generated, the time required for calculation, and obtaining supplementary information for rule mining.

INDEX TERMS Rule mining, feature selection, particle swarm optimization, artificial bee colony optimization.

I. INTRODUCTION

Mining involves finding relationships or patterns in a dataset to extract meaningful information. In this era, computers, data, and the internet have become key factors in many industries. Data mining is one of the functional steps in Knowledge Discovery Databases (KDD). An important point in KDD is how to manipulate the data or data elements. With the increase in data from various sources, some unstructured or irrelevant data may be collected, which can require more time for execution and memory space, with no or reduced fruitful outcomes. Hence, data mining is important before handling the data, preparing reports, or concluding the outcome for future trends and patterns [1]. Several algorithms can be used to mine the data, such as descriptive analysis, statistics, predictive analysis, Machine Learning (ML) algorithms, Evolutionary Algorithms, Soft Computing Algorithms, and many more [2]. There are various applications where researchers implement machine learning algorithms such as finding COVID patterns [3], IoT malware detection [4], [5], predicting students' performance [6], predicting the patterns for Vehicular sensor networks [7], and many more.

KDD is a process that involves mining data by looking for patterns and correlations in data [8]. This allows us to identify different features or characteristics of the dataset, which can then be used to generate reliable and actionable insights [9]. KDD is a collection of steps that analyze data

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

to generate usable information. The focus is to identify what customers want our company to know about us and provide them with a better product that is in line with their needs. This can be done through text mining or an algorithm such as a decision tree on historical data [10]. This KDD process is a three-step procedure to generate insights from data. In this process, data is first pre-processed or cleaned to remove any irrelevant information so that all the required columns are present in the dataset. Secondly, data is mined using a plethora of algorithms to identify patterns and relationships, which can help understand how the system works. Finally, the results are viewed in terms of reports, text, and visualizations.

Mining rules are generally made using Apriori, FP Tree, other statistical algorithms (such as linear regression, classification, and others), and soft computing algorithms [11]. Data mining is widely used in many industries, including finance, healthcare, retail, and telecommunications. It is used to analyze and identify patterns in large datasets to uncover useful information that can be used for decision-making or other purposes.

Soft computing, on the other hand, involves using computational techniques inspired by biological systems, such as artificial neural networks, fuzzy logic, and genetic algorithms. Soft computing constitutes a division of artificial intelligence that has been harnessed to design algorithms adept at dynamically accommodating the diverse strengths, weaknesses of inputs, and their interactions with the surrounding environment. Another term used in this field is fuzzy logic. A core principle underpinning soft computing is its capacity to handle data processing and problem-solving without the strict mandate for exact mathematical modeling or statistical analysis [12]. Key instances encompass recurring patterns, rules of learning, PSO, ABC, and the utilization of neural networks [13], [14].

Particle Swarm Optimization (PSO) is a complex optimization method based on particles moving around in a swarm, seeking to be near others with similar behaviors. PSO has applications in marketing research and management optimization, where it is used to solve combinatorial optimization problems [15]. PSO stands for Particle Swarm Optimization, a collective algorithm wherein individual particles endeavor to locate the most favorable solution within a given scenario. These particles originate from a precisely defined collection, and initially, all particles possess an equivalent likelihood of progressing towards their intended target. The chosen particles from this pool symbolize members of a group that strive to identify the optimal solution pertinent to their assigned task. A swarm comprises particles that delineate the fundamental "unit" guiding the swarm's journey towards solution discovery. The mean velocity, denoted as ρ, characterizes the rate at which each particle can traverse its current state. PSO is also used in conjunction with Genetic Algorithm for feature selection [16], and the SET-PSO method is used for mining positive and negative rules [17]. The gradient of ρ measures how far any individual particle can move toward the best state [18].

Artificial bee colony optimization (ABCO) is an optimization algorithm based on the intelligent foraging behavior of a honey bee swarm. The algorithm was inspired by honey bee daily activities [17]. The ABCO algorithm is segmented into three types of bees: an active bee and two scout bees. The primary role of an active bee involves foraging for food, while an alternative scout bee takes over the task if needed. When multiple food sources are present, the likelihood of selection increases when the eventual site possesses a higher nectar content compared to other sources. This essentially means that the likelihood of food acceptance escalates if the hive had previously stored some amount of the specific nectar type in its earlier life stages. Consequently, this location is assigned a higher probability score compared to others with fewer accessible resources.

The foremost contributions of this study paper are as follows:

- In this study, an algorithm named Grouped Artificial Bee Colony Optimization (G-ABC) was introduced to enhance the preprocessing and optimization aspects of the feature selection procedure.
- In this study, a rule mining algorithm with a mean-variance approach was introduced, eliminating the necessity for minimum support and confidence thresholds.
- The validation of the proposed algorithm framework involves multiple metrics, including accuracy, precision, recall, and F-measure, utilizing both labeled and unlabeled data.

This study aims to explore the Artificial Intelligence (AI) application in the field of rule mining and proposed a novel rule mining algorithm called Grouped Artificial Bee Colony Optimization (G-ABC). The G-ABC algorithm is well suited for pre-processing the rules or selecting rule-mining features. The G-ABC algorithm was validated with the traditional rule mining algorithm, and the validation results are encouraging.

The remaining portions of the study are alienated into five segments. The state-of-the-art research on evolutionary or soft computing techniques for association rule mining is covered in Segment 2. The anticipated procedure is shown in Segment 3, the datasets used are discussed in Segment 4, and the results are discussed in Segment 5. Section VI recapitulates the study effort and then lists the sources used in the paper.

II. LITERATURE REVIEW

This section presents the work conducted by numerous researchers in laying down the efforts to develop many association rule mining algorithms. Moreover, it is analyzed that it is very important to have lesser meaningful rules rather than more mixed (relevant and irrelevant) ones. A number of research have been performed to optimize association rule mining or data mining output. This section also records the research findings concerning various association rule mining algorithms that employ optimization techniques to generate both positive and negative rules.

Classification is a fundamental problem in machine learning. It has applications in almost all domains, such as biomedical, industrial automation, biometric recognition, and advertising. The objective is to assign every data instance to one of the classes or categories with minimum error and/or maximum score error. Classification with a Genetic Algorithm (GA) has been successfully used for clustering problems rule mining. It can utilize the existing feature set and maximizes its predictive power [19], [21], [23]. The objective is to find the optimal number of clusters that minimize an appropriate distance measure. The proposed technique is based on a generalized GA as well as on an exhaustive search along a randomized tree algorithm which will explore all possible subsets of attributes available in a given data set or can be generated using some attribute ordering or cluster using rules such that one set can always be found at any point along the tree. Their outcome was to generate the rules needed to discretize the attributes without overlapping frequent item sets. Creighton and Hanash used clustering with parameters support and confidence. The objective was to generate more accurate rules [20]. They used a yeast data set. They have developed a database application to develop the rules. Badal et al. used NLP with text mining. The objective was to create a model to improve text clustering quality [22]. They used protein complexes data sets. They have achieved the quality of text clustering by using the proposed model.

In the study conducted by Huo et al. [24], the authors proposed an improved algorithm to analyze the problems related to the Apriori algorithm. They set up the frequent pattern tree structures that were used to maintain the fuzziness in the original datasets and transactions. The incremental strategy was used for implementation and frequent patterns were prioritized for both initial and new patterns. The proposed technique has the advantage of less execution time and memory cost was less when the support threshold was lower in comparison to existing algorithms. The limitation of the study was the weighting methods used that make the system complex [24].

In the conducted study by Pal & Kumar [25], the authors developed a MapReduce model using the distributed frequent itemset generation and using the association rule mining algorithm. The authors used distributed integrated technology to generate the association rules and frequent item sets. The mining of the rules in terms of frequent patterns was done in a distributed way and used the association rules with the weighted method. The proposed technique solved the problem of multifarious operation in the case of a large dataset [25].

An effective measurement method to improve the traditional rule mining methods has been used by Bao et al. [26]. The authors in this study considered several aspects and then find the defects of the underlying problem. The association rules were reviewed and application in different areas was discussed. The evaluation method in terms of Support and Confidence, Influence, Validity, and many other metrics was discussed. The numerical analysis was presented and different frameworks were compared and verification was done using the public datasets. The accuracy of the existing methods was improved but the limitations of the study are that the proposed technique was not valid for large datasets, and also in finding the robustness in different related fields [26].

Another parallel Frequent Pattern growth technique using Spark Streaming for association rules in real-time is proposed by Liu et al. [27] The authors determine the Support and Confidence, and Frequent Pattern growth algorithm was developed using the divide and conquer approach. The proposed algorithm works in two different steps such as database scanning to determine all the items in the database. The sorting was done in descending order as per the set threshold and the database was scanned. The second step was to construct the Frequent pattern tree and the root node was set as per FList. The performance metrics in terms of average time were computed for different public datasets. The limitation of the study was that there is a need to improve the proposed FP algorithm using a merging tree that speeds up the process [27].

In the study conducted by Shawkat et al. [28], avoids the performance gaps when processing frequent algorithms in case of huge databases. This paper provides a modified FP-growth technique to improve FP-The proposed approach aims to improve growth efficiency by eliminating the need for repeated conditional sub-tree generation, resulting in a reduction in the complexity of the entire frequent pattern tree. The proposed Mining Frequent Pattern (MFP)-growth algorithm in this study incorporates a header table configuration to improve operational efficiency. To evaluate the performance of this algorithm, it was compared with other state-of-the-art Machine Learning (ML) algorithms based on latency, memory requirements, and the effectiveness of generated rules. Further, four experimental series were carried out using various benchmark datasets. The experimental findings support the MFP-growth algorithm's superiority and emphasis its potential for use in a variety of situations. The limitation of the study was that accuracy for real-life datasets still needs to be computed using the association rule discovery [28].

Another model has been proposed by Rastogi and Bansal [29] for Diabetic prediction using classifiers techniques such as SVM, Random Forest and so. The author concluded that logistic regression represents better accuracy than other classifiers [29]. In the study conducted by Basha et. al. [30], the neutrosophic rule-based classification system and its hybridization with the genetic algorithm [30].

From the conducted literature survey, it has been concluded that Rule mining architecture is dependent upon the input data value, the processing rule sets, and the way the rules are formed in the system. The ruleset itself is of two types viz. rule base system and propagation-based system. The rule-based system is completely dependent on the set of rules that are formed for the processing but they consume a lot of time if the number of rules are more. In order to understand the concept of latency, consider a situation where "John" a normal human being, has to provide a tip to a waiter "Ali"

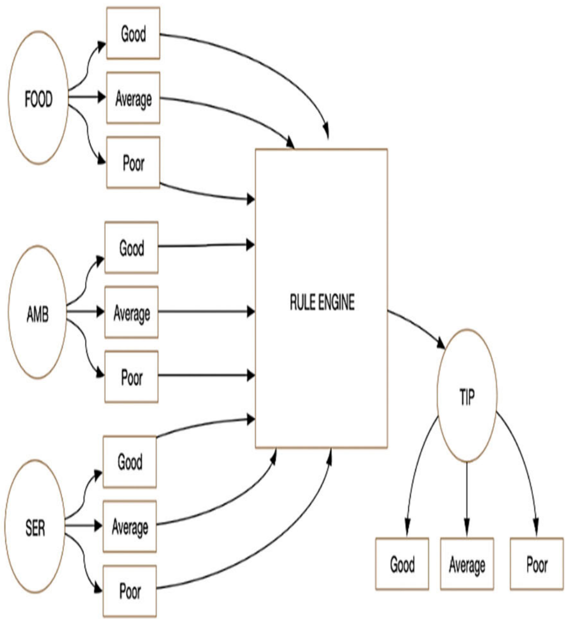


FIGURE 1. Fuzzy rule engine.

based on the type of service, food, and ambiance of the restaurant. Now, John has three input variables as shown in Figure 1 namely service, food, and ambiance.

Each input variable can have two or more two membership functions. As in the case of the illustrated example, each input variable has three membership functions. For each query, each membership function has to be analyzed and in the current case, at least 7 rules will be formed. In addition to this, there is no re-usability of the generated tip method. Every time, the engine will have to surf all the rules, and for sure that is going to consume a lot of time. Propagation-based learning methods are useful when it comes to latency reduction.

A propagation-based rule mining architecture contains four essential components as follows.

a. Data
b. Feature extraction method or extracted features
c. Mechanism of propagation engine
d. Validation of the outcome

As the rules in the case of propagation-based mining architecture are incorporated through propagation functions, the prediction time is quite low as compared to straight rule-based architecture as shown in Figure 2. The propagation engine converts the input variable's membership function into a property vector using a feature extraction mechanism or algorithm. These features are propagated through a propagation engine rather than getting propagated through a rule engine. The propagation engine uses a propagation function to circulate the data against its Ground Truth value. The propagation engine also has a stopping criterion that decides when the propagation engine has to stop the training. The user data is classified against the GT which in the case of the illustrated example is Tip-Good, Tip-Average, and Tip-Bad.

Prior to the illustration of the proposed framework, there are algorithm architectures that were completely based on static rule-based mechanisms and required to be illustrated as follows.

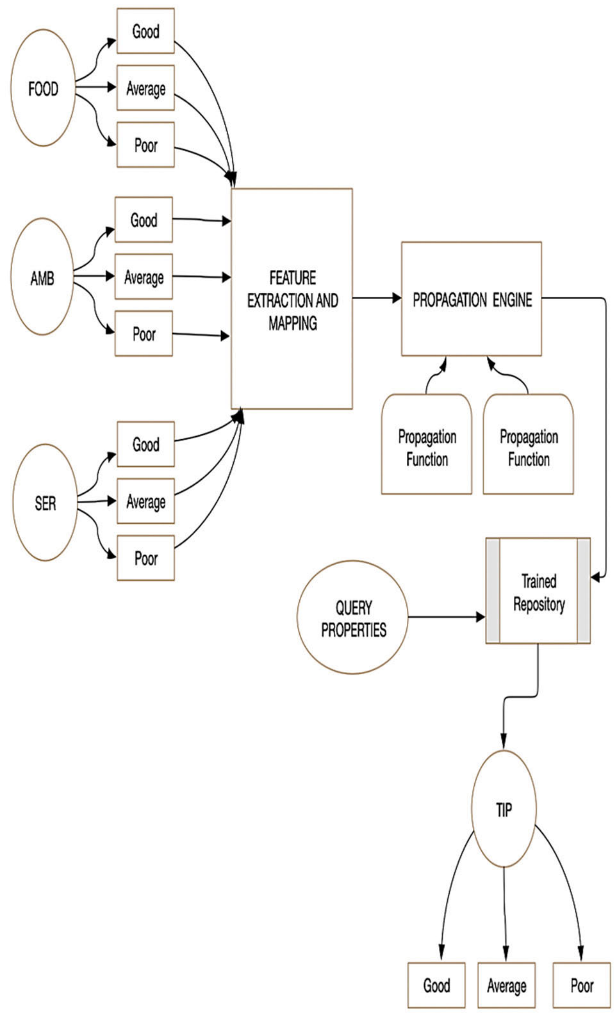


FIGURE 2. Propagation-based rule mining architecture.

Due to advancements in the propagation-based architecture and reduction in the computation complexity, the proposed architecture uses propagation-based rule mining architecture. Recent times had seen tremendous revolutionized work in the field of AI [32], [33], [34].

For sentimental analysis classification models has been used [35]. Moreover, recently for detecting COVID-19 many deep learning NLP models has been used and providing better outcomes as compare to the traditional models [36], [37].

The basic principle of association rule mining is dependent upon the maximum confidence generated within a given class interval. The highest confidence value will represent the closest co-relation among the data attributes and can be recommended on top of any other made recommendation. Keeping the architecture in mind, the proposed ABC algorithm is also based on central co-relation among the data attributes and selects the best attribute set that represents the best co-relation. The objective is to increase the overall

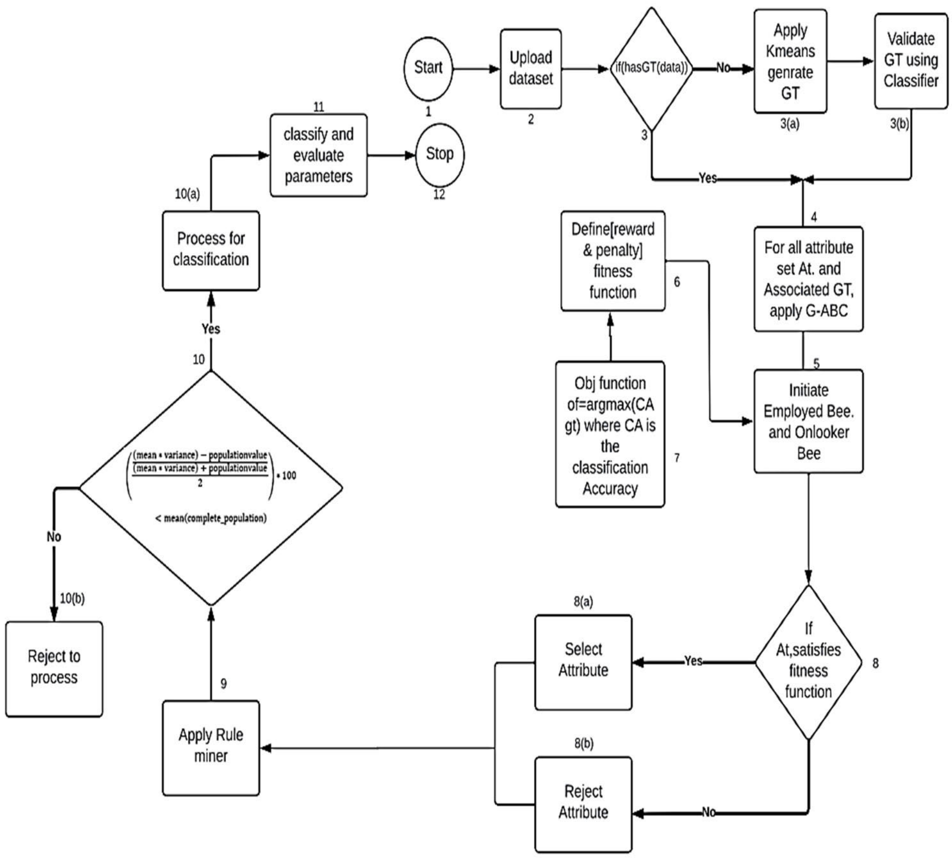


FIGURE 3. **The process of the proposed algorithm (R-Miner using soft computing).**

classification accuracy to maximize the association architecture among the data attributes and its ground truth value [38], [39], [40], [41].

III. METHODOLOGY

Intelligent software development that includes machine learning and big data has become a critical part of larger businesses. Companies are using soft computing algorithms to increase their efficiency, and the goal is to look for approaches and rules with the most optimization support. Threshold values for determining optimized solutions are entirely optional, but all options are carefully considered.

The proposed framework is represented in Algorithm 1, in which first the dataset is loaded, and then preprocessing is performed. After performing the preprocessing, features are selected by using the proposed Grouped ABC (G-ABC) method. Then data is divided into 70% to 30% ratio to validate the procedure by using four classifiers K-NN, NB, SVM, and NN. After getting selected features, association rule mining is performed using no minimum support and no minimum confidence, and again, results are validated by using four classifiers K-NN, NB, SVM, and NN.

This algorithm is divided into 2 parts. The first step indicates preprocessing. For this, in the proposed model, we used G-ABC (proposed model). This algorithm combines five randomly collected active bees to reduce the time of execution. Furthermore, this algorithm reduces the selected features for rule generation in step 2. Step 2 is used for rule mining where no threshold value is allocated. The fitness function includes the calculation of the mean and variance for each element in the population. For both steps, three classes of ground truth values are used: Positive, Negative, and Neutral. For each element, mean, variance, and mean*variance are calculated. Additionally, these values are advancing for calculating the fitness value for each element. If the condition is true, then the value is added to the final population. Otherwise, the element is rejected.

As shown in Figure 3, the proposed work aims to perform prediction for both labeled and unlabelled data. In the case of the labeled data, there is no need for Ground Truth (GT) generation whereas in the case of unlabelled data, the GT is generated via two subsequent processes. In the first Process, the data is divided into three groups considering the variance of the data and labeled as Good, Moderate, and Bad based on their co-relation present in the data. The proposed work applies an improved Artificial Bee Colony (G-ABC) with improved fitness function. Once the data attribute set is selected via the proposed Grouped ABC, an association rule mining is applied to check whether the data is suitable for classification or not. To do so, a mean and variance method

Algorithm 1 Proposed Algorithm: R-Miner Using Soft Computing

Require: ds as data set gt as Ground Truth
Ads = ds.sort (gt) Arrange data as per gt
IF *isempty* (gt)
k = 2 : 5 initate Cluster size from 2 and ending upto 5, represents the possibility of the clusters
$[kindex, kcent] = kmeans(A-ds, k)$ *Divide the data into k number of clusters*
Initiate $Nb = Naive\ Bayes\ (Ads, kindex)$
Initate Naive Bayes Classifier for the validation of the $k-clusters$
Classification A
$= Classify.NaiveBayes(Nb, Ads, randomsample())$
$Ma = Find\ Max\ (ClassificationA)$;
Find maximum classification acuracy from all clusters
$gt = kindex.Ma$;
end IF
Require: *AST as Attribute Set*, *K as total number of GT Classes*
$k = 1$;
initialize a state variable where
$k \| = \{1, \ldots, k$
While $k \leq K$
$Eb = Find(Ast.index-of(k))$ *Extracting the feature vectors of the specified class*
lf initialize levy flight variable where $lf = \{1, \ldots, l\}$ $ANDL = 10$
while $lf \leq L$
$Bp = Generate\ Population\ (AST.Random())$
$Obl = means\ (Bp)$;
if Bp satifies Obl
$reward + +$
else
$plenty + +$
endif
if $reward \geq plenty$
accept attribute
else
reject attribute
EndWhile
Initiate Rule Mining Engine
$f = Choose\ Mining\ Engine$
for $m = 1 : gt$
$m = 1 : gt$
$rn = Generate\ Random - pop$;
generate a random population
$Popmean = Calculate\ Mean\ (Rn)$;
$PopVar = Calculate\ Variance\ (Rn)$;
$MeanA = Calculate\ Mean\ (gt.record)$

Algorithm 1 *(Continued.)* Proposed Algorithm: R-Miner Using Soft Computing

Variance=Calculate Variance (gt.record);
$f = Choose\ Mining\ Engine$
for $m = 1 : gt$
$rn = Generate\ random -pop$; *generate a random population*
$PopMean = Calculate\ Mean\{Rn\}$;
$PopVar = Calculate\ Variance\ (Rn)$;
$MeanA = Calculate\ Mean\ (gt.record)$
Variance=Calculate Variance (gt.record);
if $\left(\frac{(m.Mean*m.variance)-(PopMean*PopVar)}{(PopMean*PopVar)+(Mean*Variance)}\right) * 100 < mean(complete\_population)$
Accept for processing,
Else
Reject for processing
END FOR
a.NearestNeighbor
b. SVM
c. Naive Bayes
d. Neural Propagation
$Engine\ [xtrain, xtext.ytrain, ytest] = SplitTrainTest\ (Ads, 70 : 30)$
xtrain : *Training Dataset*
xtest : *Test Dataset*
ytrain : *Test Label*
if $f == a$
$Nc = 3$; *Neighbor Count*
$tR = InitiateTraining\ (xtrain, ytrain, Nc)$;
$Classifiedresult = Simulate(tR, xtest)$;
$Classificationscore\ (ClassificationResult, ytest)$
Elseif $f == b$
$KernalFunction =' linear', ' polynomial', ' rbf'$
$tR = InitiateTraining(xtrain, ytrain, KernalFunction)$;
$ClassifiedResult = Simulate\ (tR, xtest)$;
$Classificationscore\ (ClassificationResult, ytest)$
Elseif $f == c$
$BCI =' Gaussian'$;
$tR = InitiateTraining\ (xtrain, ytrain, BCI)$;
$ClassifiedResult = Simulate\ (tR, xtest)$;
$Classifiicationscore\ (ClassificationResult, ytest)$
Else
$Neural\ Layer = 5 : 20$
$InitiateTraining\ (xtrain, ytrain, BCI)$;
$ClassifiedResult = Simulate\ (tR, xtest)$;
$Classificationscore\ (ClassificationResult, ytest)$

is applied. If the variance in the selected data is greater than the complete population means, then it can be processed for classification. Furthermore, a neural-based deep learning method is applied.

IV. DATASET

A. DATASET 1

The first dataset,'' NAME: Pay for Play: Are Baseball Salaries Based on Performance? TYPE: Census, SIZE: 337 observations, 18 variables. The features are divided into two datasets, one for training and the other for testing, with a 70 to 30 ratio once the feature optimization technique has been used. 100 records total, divided into 3 labels, make up the dataset. neutral, negative, or positive [31].

B. DATASET 2

The same analysis is also performed on another dataset. The second data set is the ''Twitter'' dataset. The data set includes the tweets during the first GOP debate as well as the sentiment of each tweet, among other attributes. ''SIZE: 13872 observations, 21 variables. These are the ones that we used: candidate, candidate: confidence, sentiment (of tweet), subject_matter, subject_matter: confidence, retweet_count, tweet_coord, tweet_created, tweet_location, user_timezone, and text (Tweet). URL used: First GOP Debate Twitter Sentiment | Kaggle''. After using the feature optimization technique, features are divided into 2 datasets: 1 training, and 1 testing set with a 70 to 30% ratio. The dataset includes 100 records which are divided into 20 Positive sentiments, 20 Neutral sentiments, and 60 negative sentiments [31].

For both datasets, there were 3 classes of Ground Truth values. There are 60 values for Negative and 20 for positive, neutral labels, as represented in Figure 3.

V. PERFORMANCE EVALUATION AND DISCUSSION

Information Retrieval (IR) is a domain in Artificial Intelligence (AI). AI has taken a part of information retrieval, i.e., classification, to the next level by developing algorithms that are smarter than humans. Computational Intelligence researchers use AI techniques for IR problems, and it is an important by-product of artificial intelligence. Although several approaches can be used for IR problems, text processing techniques remain among the most popular choices because they are easy to implement, communicate results easily and quickly with humans, etc. In this paper, we propose an algorithm that uses group-wise entity extraction with phrase and sentence detection applied to textual data.

Practical analysis implementation has involved the usage of two datasets. The test computer system is set up with all of the Python 3.9 algorithms that are used in the experiment written in Anaconda Spyder Version 5. We used an 11th Gen Intel(R) Core (TM) i5-1135G7 @ 2.40GHz processor. Comparisons are made between the proposed algorithm and PSO, ABC, hybrid PSO, and group ABC (proposed algorithm). On two datasets, the proposed algorithm is assessed. The Twitter dataset was one dataset that contained textual data with terms. The second dataset was for clustered data and was a dataset. The k-Means algorithm is initially used with three labels: Positive, Neutral, and Negative. A smaller number of features that were chosen for the proposed technique suggests that more precise or effective features were extracted rather than choosing all the features for future implementation. Additionally, the proposed approach runs faster than other algorithms. In the proposed model, two datasets are used. This study compared the proposed algorithm with PSO, ABC, hybrid PSO, and group ABC to evaluate the accuracy [31]. The proposed algorithm had satisfactory results on both of the datasets except for when performance was assessed on the whole dataset.

Parameters used for comparison-

1. **Accuracy:** Accuracy represents the ratio of correct prediction over all the predictions made by the model. The accuracy formula is represented AS follows

$$\text{Accuracy} = ((\text{TP} + \text{TN}))/((\text{TP} + \text{TN} + \text{FP} + \text{FN})) \tag{1}$$

2. **Precision:** Precision represents the percentage of how many elements are positively predicted by the system. The precision formula is represented as follows

$$\text{Precision} = \text{TP}/((\text{TP} + \text{FP})) \tag{2}$$

3. **Recall or Sensitivity:** Recall represents the percentage of how many elements are positive predict by the system. The recall formula is represented AS follows

$$\text{Sensitivity(recall)} = \text{TP}/((\text{TP} + \text{FN})) \tag{3}$$

4. **F1- Score:** The F1-score represents the mean of precision and recall or sensitivity. F1- score formula is represented AS follows

$$\text{FMeasure} = ((2 * \text{Precision} * \text{Recall})) /((\text{Precision} + \text{Recall})) \tag{4}$$

A. FEATURE SELECTION

We propose G-ABC as an alternate algorithm to other prevalent feature selection algorithms. The proposed G-ABC algorithm can produce diverse sets of highly-accurate rules in a short time. It has also been found that although both selection and correction inflate the size of the final ruleset, selecting by using lower-quality rules results in more accurate predictions than selecting by using higher-quality rules. This is because selecting from multiple low-complexity rules will outperform higher-complexity rules as long as they are trained at a low training sample rate, which can be achieved in low-cost hardware. The experimental analysis represents that the proposed algorithm represents better results in terms of the number of rules selected, time of execution, and accuracy. Figure 4 represents the total time of execution for selecting the features. For the analysis total of 100 records are used, and the following graph represents the proposed G-ABC required minimum time of execution. Figure 5 represents the total number of selected features. The graph 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 6 represents the comparison between the algorithms in terms of better accuracy, and again visualization

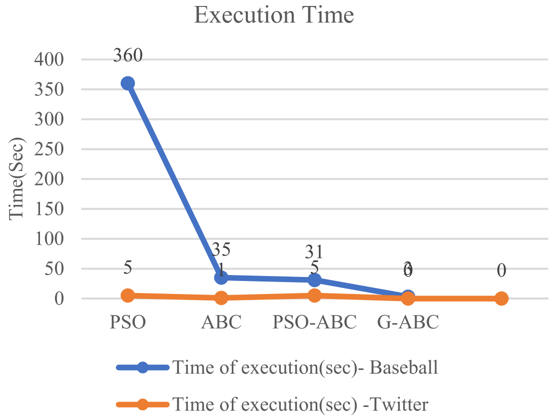


FIGURE 4. Represents the total time of execution.

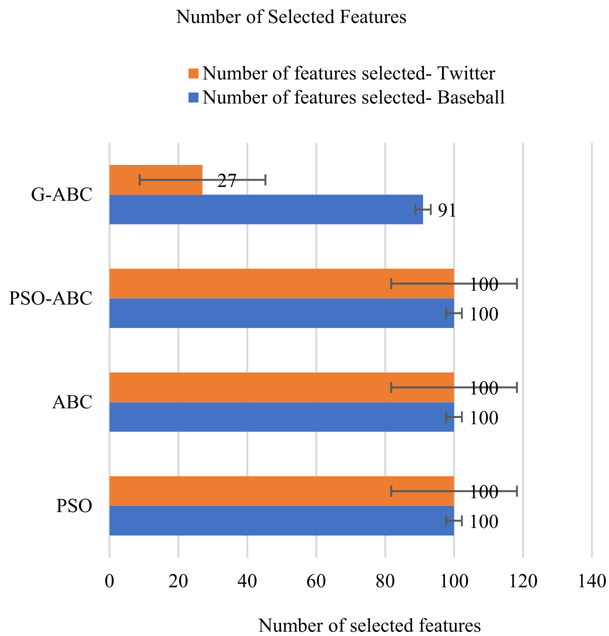


FIGURE 5. A total number of selected features.

represents that the proposed algorithm shows better results. In Figure 6 first four values are from the first dataset, which is the baseball dataset, and the next four are based on the second Twitter dataset

The proposed algorithm, G-ABC, stands out among all other algorithms. The first parameter is accuracy, and it shows that G-ABC shows better results rather than all other algorithms. The second parameter is execution time and it shows that G-ABC takes less time to do its operations. The third parameter is the variance of values, and it also proves that G-ABC is the best-fit algorithm. Moreover, this paper also covers the classification algorithm used in our process. These are K-Nearest Neighbour (KNN), Support Vector Machine (SVM), Naïve Bayes (NB), and Neural Network (NN).

B. RULE MINING

This paper outlines the idea of rule mining, an automated method for uncovering patterns within data. Grounded in data clustering and methodologies for constructing generalizable rules from specific instances, rule mining's application demonstrates its capacity to unearth intricate feature patterns concealed within extensive datasets, all without relying on prior assumptions or preexisting knowledge of the underlying structure. Important characteristics of rule mining algorithms are generated by a combination of two different factors: the exploration process (maintain diversity and avoid overfitting) and dataset selection criterion (sample size, dimensionality, available data). In our proposed algorithm second architecture is to generate a rule using MEAN VARIANCE optimization where minimum support and minimum confidence are not required.

TABLE 1. Represents rule set.

Rule	Antecedent	Consequent
Rule 1	high_positive_valence, high_arousal	Ground_Truth(2)
Rule 2	low_positive_valence, low_arousal	Ground_Truth (1)
Rule 3	moderate_positive_valence, moderate_arousal	Ground_Truth (3)
Rule 4	high_positive_valence, low_arousal	Ground_Truth(2)
Rule 5	low_positive_valence, high_arousal	Ground_Truth (3)
Rule 6	high_positive_valence, moderate_arousal	Ground_Truth(2)
Rule 7	low_positive_valence, moderate_arousal	Ground_Truth (3)
Rule 8	moderate_positive_valence, low_arousal	Ground_Truth (1)
Rule 9	moderate_positive_valence, high_arousal	Ground_Truth (3)
Rule 10	moderate_positive_valence, low_arousal	Ground_Truth(2)

We introduced a soft computing-driven strategy for extracting tweets with elevated precision, recall, and accuracy. The effectiveness of G-ABC and Mean-Variance Rule mining was assessed using a dataset of 100 records. Across all performance metrics—precision, recall, accuracy, and f-measure—the proposed approach consistently outperformed other established algorithms. It is evident from the diagram that G-ABC outperforms all other methods in terms of all performance measures. The neural engine uses the weighted method to predict the emotion of the provided input. The rules for the detection can be illustrated using the following table. The rules have been formed in such a way that Ground_Truth(2) is encoded with emotion value 2, Ground_Truth is encoded as 1 and Ground_Truth is encoded as 3.

In the above table, the antecedent represents the combination of input features related to sentiment scores (e.g., positive sentiment score, negative sentiment score) and arousal scores.

The consequent represents the predicted sentiment class as Ground_Truth(1), Ground_Truth(2), Ground_Truth(3).

Please note that in sentiment analysis with Neural Networks, weights are learned during the training process, and associations between features and sentiment classes are determined based on the network's learned parameters. The rulesets can be encoded as follows.

1. Rule 1 1, 0 2
2. Rule 2 0, 1 1
3. Rule 3 2, 2 3
4. Rule 4 1, 2 2
5. Rule 5 2, 1 1
6. Rule 6 1, 3 2
7. Rule 7 0, 4 1
8. Rule 8 2, 3 3
9. Rule 9 0, 3 2
10. Rule 10 1, 4 2

Rule 1:
Antecedent:
high_positive_sentiment_score,
low_negative_sentiment_score
Consequent: Ground_Truth(2)

Explanation: Rule 1 states that if an input has a high positive sentiment score and a low negative sentiment score, it is classified as Ground_Truth(2). This rule assumes that a high positive sentiment score indicates a positive sentiment and a low negative sentiment score implies a lack of negative sentiment. Therefore, based on these criteria, the sentiment is classified as Ground_Truth(2).

Rule 2:
Antecedent:
low_positive_sentiment_score,
high_negative_sentiment_score
Consequent: Ground_Truth (1)

Explanation: Rule 2 suggests that if an input has a low positive sentiment score and a high negative sentiment score, it is classified as Ground_Truth (1). This rule assumes that a low positive sentiment score indicates a lack of positive sentiment, and a high negative sentiment score implies a strong presence of negative sentiment. Therefore, based on these criteria, the sentiment is classified as Ground_Truth.

Rule 3:
Antecedent:
moderate_positive_sentiment_score,
moderate_negative_sentiment_score
Consequent: Ground_Truth (3)

Explanation: Rule 3 states that if an input has both a moderate positive sentiment score and a moderate negative sentiment score, it is classified as Ground_Truth (3). This rule assumes that moderate values for both positive and negative sentiment scores imply a balance between positive and negative sentiments, resulting in a Ground_Truth sentiment.

Rule 4:
Antecedent:
high_positive_sentiment_score,
moderate_negative_sentiment_score
Consequent: Ground_Truth(2)

Explanation: Rule 4 suggests that if an input has a high positive sentiment score and a moderate negative sentiment score, it is classified as Ground_Truth(2). This rule assumes that a high positive sentiment score indicates a positive sentiment and a moderate negative sentiment score implies a relatively low presence of negative sentiment. Therefore, based on these criteria, the sentiment is classified as Ground_Truth(2).

Rule 5:
Antecedent:
moderate_positive_sentiment_score,
high_negative_sentiment_score
Consequent: Ground_Truth (1)

Explanation: Rule 5 states that if an input has a moderate positive sentiment score and a high negative sentiment score, it is classified as Ground_Truth (1). This rule assumes that a moderate positive sentiment score indicates some positive sentiment and a high negative sentiment score implies a strong presence of negative sentiment. Therefore, based on these criteria, the sentiment is classified as Ground_Truth.

Rule 6:
Antecedent:
high_positive_sentiment_score, high_arousal_score
Consequent: Ground_Truth(2)

Explanation: Rule 6 suggests that if an input has both a high positive sentiment score and a high arousal score, it is classified as Ground_Truth(2). This rule assumes that both high positive sentiment and high arousal indicate a positive and energetic sentiment, aligning with the classification of happiness.

Rule 7:
Antecedent:
high_negative_sentiment_score, low_arousal_score
Consequent: Ground_Truth (1)

Explanation: Rule 7 states that if an input has a high negative sentiment score and a low arousal score, it is classified as Ground_Truth (1). This rule assumes that a high negative sentiment score indicates a strong presence of negative sentiment, and a low arousal score suggests a lack of energy or excitement. Therefore, based on these criteria, the sentiment is classified as Ground_Truth.

Rule 8:
Antecedent:
moderate_positive_sentiment_score,
moderate_arousal_score
Consequent: Ground_Truth (3)

Explanation: Rule 8 suggests that if an input has both a moderate positive sentiment score and moderate arousal score, it is classified as Ground_Truth (3). This rule assumes that moderate values for both positive sentiment and arousal imply a balanced and Ground_Truth sentiment.

TABLE 2. Accuracy comparison between algorithm.

BASEBALL				TWITTER			
KNN	SVM	NB	NN	KNN	SVM	NB	NN
50	53.3	50	97.56	50	53.3	50	97.56
50	56.67	46.67	97.47	30	36.67	43	100
36.67	53.33	36.67	98.3	20	36.67	56.67	98.19
40	53.33	53.33	97.56	55	54	65	97.77

TABLE 3. Precision Comparison between algorithms.

BASEBALL				TWITTER			
KNN	SVM	NB	NN	KNN	SVM	NB	NN
65	61.11	60	61.11	54.62	65	61.11	60
63.64	68.18	66.67	67	50	58	75	60
50	33	33	42	62.5	66.67	68.42	58.62
60	64.4	68.45	67.78	67.4	56.5	67.55	68

TABLE 4. Recall comparison between algorithm.

BASEBALL				TWITTER			
KNN	SVM	NB	NN	KNN	SVM	NB	NN
61.11	72.22	61.11	89	61.11	72.22	61.11	92
57	50	27	95	44	55.56	50	93
50	72.22	50	87	27.78	44.44	72.22	94.44
77.78	83.33	75	96	50	75	75	98.1

TABLE 5. F-measure comparison between algorithm.

BASEBALL				TWITTER			
KNN	SVM	NB	NN	KNN	SVM	NB	NN
70.97	68.42	61.11	75	50.97	68.42	61.11	75
28.57	28.57	30.57	60	47.06	57.14	53.33	75
54.55	70.27	48.28	75	38.46	53.33	70.27	72.34
70	83.33	75	94	58	69.1	73	75

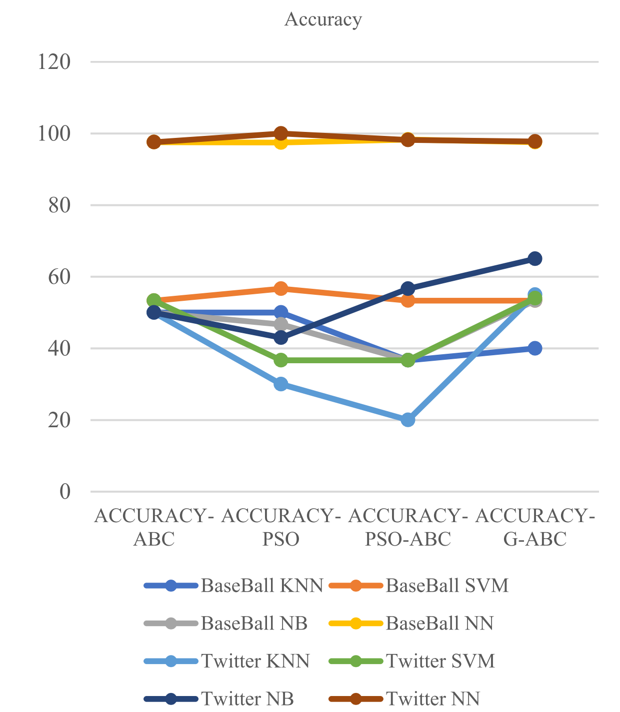


FIGURE 6. Accuracy comparison between algorithms.

Rule 9:
Antecedent:
low_positive_sentiment_score, high_arousal_score
Consequent: Ground_Truth(2)

Explanation: Rule 9 states that if an input has a low positive sentiment score and a high arousal score, it is classified as Ground_Truth(2). This rule assumes that a low positive sentiment score indicates a lack of positive sentiment, but a high arousal score suggests a high level of energy or excitement. Therefore, based on these criteria, the sentiment is classified as Ground_Truth(2).

Rule 10:
Antecedent:
high_positive_sentiment_score, low_arousal_score
Consequent: Ground_Truth(2)

Explanation: Rule 10 suggests that if an input has a high positive sentiment score and a low arousal score, it is classified as Ground_Truth(2). This rule assumes that a high positive sentiment score indicates a positive sentiment and a low arousal score suggests a lack of energy or excitement. Therefore, based on these criteria, the sentiment is classified as Ground_Truth(2).

These rules are designed to capture relationships between different sentiment features and sentiment classes. Each rule specifies the conditions (antecedent) under which a particular sentiment class (consequent) is assigned. However,

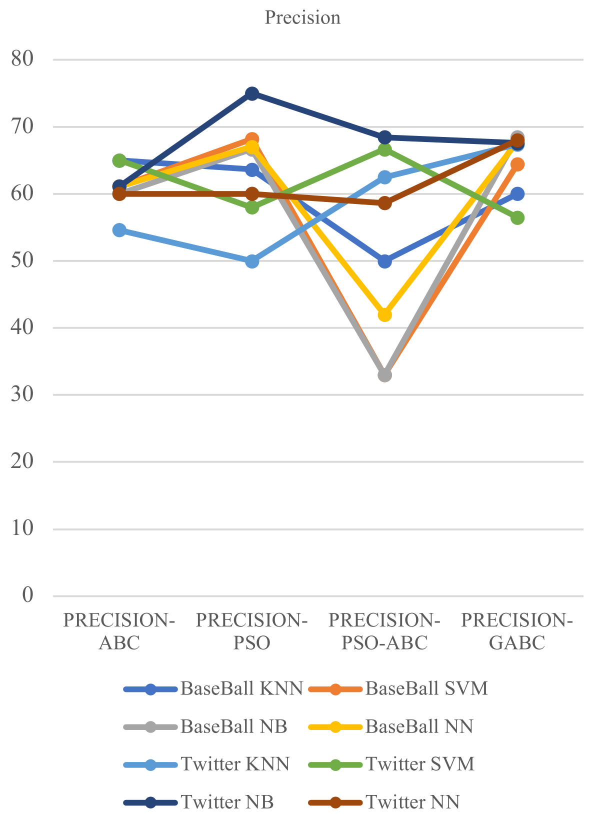


FIGURE 7. Precision comparison between algorithms.

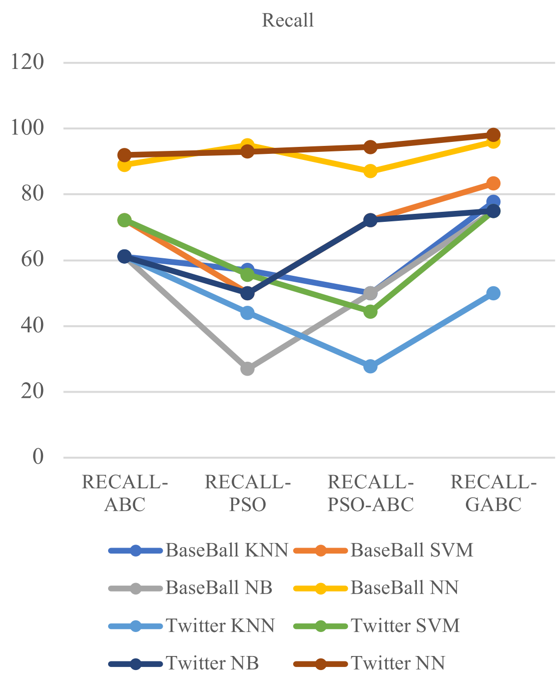


FIGURE 8. Recall comparison between algorithms.

it's important to note that these rules are hypothetical and may not accurately reflect the complexity of sentiment analysis. In practice, sentiment analysis using neural networks involves training the model on labeled data, learning the patterns and relationships between features and sentiment classes, and making predictions based on the learned model.

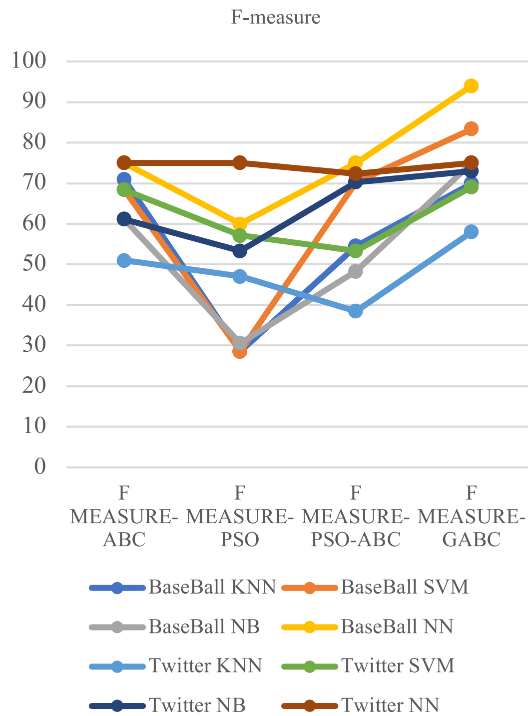


FIGURE 9. F-measure comparison between algorithms.

Figure 6 illustrates the accuracy values, showcasing that the proposed algorithms achieve around 98 percent accuracy across both datasets, outperforming the existing algorithms. Moving to Figure 7, the precision values are depicted, revealing that the proposed algorithm attains approximately 68 percent precision for the Twitter dataset and around 76 percent for the Baseball dataset, both surpassing the performance of existing algorithms. In Figure 8, the recall values are shown, with approximately 98 percent and 96 percent for the Twitter dataset and around 94 percent for the Baseball dataset. Finally, Figure 9 presents a comparison of the F-measure values, indicating that the proposed method consistently achieves superior F-measure results compared to other algorithms. This outcome suggests its heightened efficacy in mining substantial datasets, owing to the efficient processing enabled by mean-variance rule mining.

Experimental results show that a neural network classifier is best suited for identifying classes of similar items that are close together in an unlabeled dataset.

VI. CONCLUSION AND FUTURE WORK

The research study has presented an advanced data mining framework based on G-ABC with associated rule mining and mean-variance optimization followed by neural network architecture. The objective of the study was to explore how Artificial Intelligence can be applied to rule mining while involving nature-inspired optimization approaches and rule-based data mining. The research comprises twin evaluation that is performed at the feature selection level as well as classification level. The different optimization techniques such as PSO, ABC, PSO+ABC, and G-ABC have been compared using the Twitter and baseball datasets. The performance metrics such as Precision, Sensitivity, F-measure, and Execution Time has been computed. G-ABC technique performs better in comparison to other techniques. The execution time using the G-ABC technique for the NN classifier is 2.8s and 2.6s respectively which is comparatively less than other techniques. The performance of the G-ABC technique is superior in comparison to other optimization techniques as well.

The different classifiers such as NB, KNN, SVM, and NN have been compared for different performance metrics. The analysis evaluation shows that G-ABC is found to be more suitable at the pre-processing stage for the feature selection. The complete evaluation showed that the neural network proved to be the best classifier in the present framework.

The simulation study is conducted using a number of simulations varied from 200 to 1000 for both datasets used in the evaluation. The major idea here is to evaluate the effect of an increase in the simulation rounds and the neural layers on the performance of the proposed G-ABC+NN. The analysis using multiple simulations shows that an increase in the number of neural layers in the proposed G-ABC+NN considerably increases the overall execution time. In addition to this, it has been observed that G-ABC+NN with 5 to 10 neural layers performed better than those using 15 to 30 neural layers. Therefore, the number of neural layers is restricted to 5 to 10 so as to lower the overall execution time without compromising the performance of the proposed framework.

In the present work, G-ABC with NN has been proposed for advanced data mining based on association rule mining and mean-variance optimization. In the future, more datasets will be involved to justify the effectiveness of the designed framework over a large number of datasets. Further, more metaheuristic algorithms can be evaluated to resolve the challenges associated with prediction accuracy and selection problems. In addition to this, deep learning will be involved in the presented research work to further improve the feature extraction accuracy to reduce the overall execution time of the process.

ACKNOWLEDGMENT

The authors would like to acknowledge the support of Prince Sultan University for paying the Article Processing Charges (APC) of this publication.

REFERENCES

[1] Y. Jiang, G. Yang, H. Li, and T. Zhang, "Knowledge driven approach for smart bridge maintenance using big data mining," *Autom. Construction*, vol. 146, Feb. 2023, Art. no. 104673.

[2] X. Shu and Y. Ye, "Knowledge discovery: Methods from data mining and machine learning," *Social Sci. Res.*, vol. 110, Feb. 2023, Art. no. 102817.

[3] M. U. Rehman, A. Shafique, S. Khalid, M. Driss, and S. Rubaiee, "Future forecasting of COVID-19: A supervised learning approach," *Sensors*, vol. 21, no. 10, p. 3322, May 2021.

[4] S. Ben Atitallah, M. Driss, and I. Almomani, "A novel detection and multi-classification approach for IoT-malware using random forest voting of fine-tuning convolutional neural networks," *Sensors*, vol. 22, no. 11, p. 4302, Jun. 2022.

[5] M. A. Khan, M. A. K. Khattk, S. Latif, A. A. Shah, M. U. Rehman, W. Boulila, M. Driss, and J. Ahmad, "Voting classifier-based intrusion detection for IoT networks," in *Advances on Smart and Soft Computing*. Singapore: Springer, 2021, pp. 313–328.

[6] S. Batool, J. Rashid, M. W. Nisar, J. Kim, H.-Y. Kwon, and A. Hussain, "Educational data mining to predict students' academic performance: A survey study," *Educ. Inf. Technol.*, vol. 28, no. 1, pp. 905–971, Jan. 2023.

[7] M. Driss, I. Almomani, Z. E. Huma, and J. Ahmad, "A federated learning framework for cyberattack detection in vehicular sensor networks," *Complex Intell. Syst.*, vol. 8, no. 5, pp. 4221–4235, Oct. 2022.

[8] W. Boulila, I. R. Farah, K. S. Ettabaa, B. Solaiman, and H. B. Ghézala, "Spatio-temporal modeling for knowledge discovery in satellite image databases," in *Proc. CORIA*, 2010, pp. 35–49.

[9] U. Maulik and S. Bandyopadhyay, "Genetic algorithm-based clustering technique," *Pattern Recognit.*, vol. 33, no. 9, pp. 1455–1465, Sep. 2000.

[10] G. D'Angelo, D. Della-Morte, D. Pastore, G. Donadel, A. De Stefano, and F. Palmieri, "Identifying patterns in multiple biomarkers to diagnose diabetic foot using an explainable genetic programming-based approach," *Future Gener. Comput. Syst.*, vol. 140, pp. 138–150, Mar. 2023.

[11] J. Mata, J. L. Alvarez, and J. C. Riquelme, "Discovering numeric association rules via evolutionary algorithm," in *Proc. Pacific-Asia Conf. Knowl. Discovery Data Mining*. Berlin, Germany: Springer, 2002, pp. 40–51.

[12] M. Jemmali, M. Denden, W. Boulila, G. Srivastava, R. H. Jhaveri, and T. R. Gadekallu, "A novel model based on window-pass preferences for data emergency aware scheduling in computer networks," *IEEE Trans. Ind. Informat.*, vol. 18, no. 11, pp. 7880–7888, Nov. 2022.

[13] V. Beiranvand, M. Mobasher-Kashani, and A. Abu Bakar, "Multi-objective PSO algorithm for mining numerical association rules without a priori discretization," *Exp. Syst. Appl.*, vol. 41, no. 9, pp. 4259–4273, Jul. 2014.

[14] R. J. Kuo, M. Gosumolo, and F. E. Zulvia, "Multi-objective particle swarm optimization algorithm using adaptive archive grid for numerical association rule mining," *Neural Comput. Appl.*, vol. 31, no. 8, pp. 3559–3572, Aug. 2019.

[15] F. Moslehi and A. Haeri, "A novel hybrid wrapper–filter approach based on genetic algorithm, particle swarm optimization for feature subset selection," *J. Ambient Intell. Humanized Comput.*, vol. 11, no. 3, pp. 1105–1127, Mar. 2020.

[16] Q. Pu, J. Gan, L. Qiu, J. Duan, and H. Wang, "An efficient hybrid approach based on PSO, ABC and K-means for cluster analysis," *Multimedia Tools Appl.*, vol. 81, no. 14, pp. 19321–19339, Jun. 2022.

[17] J. Agrawal, S. Agrawal, A. Singhai, and S. Sharma, "SET-PSO-based approach for mining positive and negative association rules," *Knowl. Inf. Syst.*, vol. 45, no. 2, pp. 453–471, Nov. 2015.

[18] M. Karunyalakshmi and N. Tajunisha, "Classification of cancer datasets using artificial bee colony and deep feed forward neural networks," *Int. J. Adv. Res. Comput. Commun. Eng.*, vol. 62, no. 2, pp. 33–41, 2017.

[19] H. Ishibuchi and T. Yamamoto, "Fuzzy rule selection by multi-objective genetic local search algorithms and rule evaluation measures in data mining," *Fuzzy Sets Syst.*, vol. 141, no. 1, pp. 59–88, Jan. 2004.

[20] C. Creighton and S. Hanash, "Mining gene expression databases for association rules," *Bioinformatics*, vol. 19, no. 1, pp. 79–86, Jan. 2003.

[21] A. Ghosh and B. Nath, "Multi-objective rule mining using genetic algorithms," *Inf. Sci.*, vol. 163, nos. 1–3, pp. 123–133, Jun. 2004.

[22] V. D. Badal, P. J. Kundrotas, and I. A. Vakser, "Natural language processing in text mining for structural modeling of protein complexes," *BMC Bioinf.*, vol. 19, no. 1, pp. 1–10, Dec. 2018.

[23] B. Alatas and E. Akin, "An efficient genetic algorithm for automated mining of both positive and negative quantitative association rules," *Soft Comput.*, vol. 10, no. 3, pp. 230–237, Feb. 2006.

[24] W. Huo, X. Feng, and Z. Zhang, "An efficient approach for incremental mining fuzzy frequent itemsets with FP-tree," *Int. J. Uncertainty, Fuzziness Knowl.-Based Syst.*, vol. 24, no. 3, pp. 367–386, Jun. 2016.

[25] A. Pal and M. Kumar, "Distributed synthesized association mining for big transactional data," *Sādhanā*, vol. 45, no. 1, pp. 1–13, Dec. 2020.

[26] F. Bao, L. Mao, Y. Zhu, C. Xiao, and C. Xu, "An improved evaluation methodology for mining association rules," *Axioms*, vol. 11, no. 1, p. 17, Dec. 2021.

[27] L. Liu, J. Wen, Z. Zheng, and H. Su, "An improved approach for mining association rules in parallel using spark streaming," *Int. J. Circuit Theory Appl.*, vol. 49, no. 4, pp. 1028–1039, Apr. 2021.

[28] M. Shawkat, M. Badawi, S. El-Ghamrawy, R. Arnous, and A. El-Desoky, "An optimized FP-growth algorithm for discovery of association rules," *J. Supercomput.*, vol. 78, no. 4, pp. 5479–5506, Mar. 2022.

[29] R. Rastogi and M. Bansal, "Diabetes prediction model using data mining techniques," *Measurement, Sensors*, vol. 25, Feb. 2023, Art. no. 100605.

[30] S. H. Basha, A. M. Anter, A. E. Hassanien, and A. Abdalla, "Hybrid intelligent model for classifying chest X-ray images of COVID-19 patients using genetic algorithm and neutrosophic logic," *Soft Comput.*, vol. 27, no. 6, pp. 3427–3442, 2023.

[31] M. Rana and J. Singla, "A framework for selecting features using various soft computing algorithms," in *Proc. 6th Int. Conf. Trends Electron. Informat. (ICOEI)*, Apr. 2022, pp. 1743–1748, doi: 10.1109/ICOEI53556.2022.9776721.

[32] R. Cioffi, M. Travaglioni, G. Piscitelli, A. Petrillo, and F. De Felice, "Artificial intelligence and machine learning applications in smart production: Progress, trends, and directions," *Sustainability*, vol. 12, no. 2, p. 492, Jan. 2020, doi: 10.3390/SU12020492.

[33] A. Holzinger, P. K. Andreas, W. Edgar, and A. M. Tjoa, "Current advances, trends and challenges of machine learning and knowledge extraction: From machine learning to explainable AI," in *Proc. Mach. Learn. Knowl. Extraction (IFIP)*, vol. 2. Hamburg, Germany: Springer, Aug. 2018, pp. 1–8.

[34] S. Makridakis, "The forthcoming artificial intelligence (AI) revolution: Its impact on society and firms," *Futures*, vol. 90, pp. 46–60, Jun. 2017.

[35] S. Dey, S. Wasif, D. S. Tonmoy, S. Sultana, J. Sarkar, and M. Dey, "A comparative study of support vector machine and naive Bayes classifier for sentiment analysis on Amazon product reviews," in *Proc. Int. Conf. Contemp. Comput. Appl. (ICA)*, Feb. 2020, pp. 217–220.

[36] M. Jemmali, "Intelligent algorithms and complex system for a smart parking for vaccine delivery center of COVID-19," *Complex Intell. Syst.*, vol. 8, no. 1, pp. 597–609, Feb. 2022.

[37] M. Al-Sarem, A. Alsaeedi, F. Saeed, W. Boulila, and O. AmeerBakhsh, "A novel hybrid deep learning model for detecting COVID-19-related rumors on social media based on LSTM and concatenated parallel CNNs," *Appl. Sci.*, vol. 11, no. 17, p. 7940, Aug. 2021.

[38] I. K. C. U. Perera and H. A. Caldera, "Aspect based opinion mining on restaurant reviews," in *Proc. 2nd IEEE Int. Conf. Comput. Intell. Appl. (ICCIA)*, Sep. 2017, pp. 542–546.

[39] Z. Jianqiang and G. Xiaolin, "Comparison research on text preprocessing methods on Twitter sentiment analysis," *IEEE Access*, vol. 5, pp. 2870–2879, 2017.

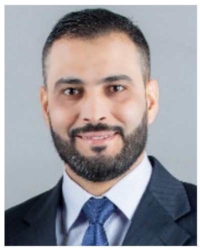
[40] N. M. Ali, M. M. A. El Hamid, and A. Youssif, "Sentiment analysis for movies reviews dataset using deep learning models," *Int. J. Data Mining Knowl. Manage. Process*, vol. 9, no. 3, pp. 19–27, May 2019.

[41] A. Madasu and S. Elango, "Efficient feature selection techniques for sentiment analysis," *Multimedia Tools Appl.*, vol. 79, nos. 9–10, pp. 6313–6335, Mar. 2020.

MRINALINI RANA received the B.Tech. and M.Tech. degrees in computer science and engineering from Punjab Technical University, Jalandhar. She is currently an Assistant Professor with the School of Computer Science and Engineering, Lovely Professional University, Phagwara, Punjab, India. She has more than eight years of experience in teaching. She has guided many UG/PG students. Her current research interests include data mining, soft computing, databases, and machine learning techniques.

OMDEV DAHIYA received the Ph.D. degree in computer science and engineering from Maharshi Dayanand University, Rohtak, Haryana, India. He is currently an Assistant Professor with the School of Computer Science and Engineering, Lovely Professional University, Phagwara, Punjab, India. He has more than five years of teaching and research experience in various fields of computer science and engineering. He has published more than 35 research articles duly indexed in Scopus and Web of Science databases. He has guided several UG/PG students and currently three Ph.D. scholars are working under his supervision. His current research interests include various domains of computer engineering, including data mining, learning analytics, cloud computing, and soft computing approaches involving advanced AI/ML techniques. He has been the session chair and a technical program committee member at various national and international conferences. Apart from that, he is a potential reviewer of several international journals and conferences of repute.

PARMINDER SINGH (Member, IEEE) received the B.Tech. degree from Punjabi University, Patiala, the M.Tech. degree from Punjab Technical University, Jalandhar, and the Ph.D. degree from Lovely Professional University, India, in 2019. In 2020, he was a Postdoctoral Researcher with Mohammed VI Polytechnic University (UM6P), Ben Guerir, Morocco. He has been an Associate Professor with the School of Computer Science and Engineering, Lovely Professional University. Currently, five students are pursuing the Ph.D. degree under his supervision. 15 master's dissertations and two Ph.D. thesis have been completed under his supervision. He has published more than 50 papers in reputed journals, conferences, and book chapters, including IEEE Transactions. His current research interests include cyber-security, machine learning, and cloud/fog/edge computing. He is an active member of ACM. He won the Best Smart City Project Award from the Government of India, in 2019. He has been awarded research excellence awards in the A+ category from LPU, in 2019, 2020, and 2021. He has also been awarded a Teaching Excellence Award from LPU, in 2021. He has been the session chair and a technical program committee member at various national and international conferences.



WADII BOULILA (Senior Member, IEEE) received the B.Eng. degree (Hons.) in computer science from the Aviation School of Borj El Amri, in 2005, the M.Sc. degree in computer science from the National School of Computer Science (ENSI), University of Manouba, Tunisia, in 2007, and the Ph.D. degree in computer science jointly from the ENSI and Telecom-Bretagne, University of Rennes 1, France, in 2012. He is currently an Associate Professor of computer science with Prince Sultan University, Saudi Arabia. He is also a Senior Researcher with the RIOTU Laboratory, Prince Sultan University, and a Senior Researcher with the RIADI Laboratory, University of Manouba. Previously, he was a Senior Research Fellow with the ITI Department, University of Rennes 1. He has participated in numerous research and industrial-funded projects. His current research interests include data science, computer vision, big data analytics, deep learning, cybersecurity, artificial intelligence, and uncertainty modeling. He is an ACM Member and a Senior Fellow of the Higher Education Academy (SFHEA), U.K. He received the Award of the Young Researcher in computer science from Beit El-Hikma, Tunisia, in 2021; the Award of Best Researcher from the University of Manouba, in 2021; and the Award of Most Cited Researcher at the University of Manouba, in 2022. He has served as the chair, a reviewer, and a TPC member for many leading international conferences and journals. His work has gained global recognition and he has been nominated as one of the top 2% of scientists in his field by Stanford University.

ADEL AMMAR received the engineering degree from Ecole des Mines de Nancy, Nancy, France, in 2003, the master's degree from the University of Versailles St-Quentin-en-Yvelines, Versailles, France, in 2005, and the Ph.D. degree in applied data processing from Université Paul Sabatier, Toulouse, France, in 2008. He is currently an Associate Professor with the College of Computer Science and Information Systems and a Researcher with the Robotics and Internet of Things Laboratory, Prince Sultan University, Riyadh, Saudi Arabia. His current research interests include machine learning, deep learning, pattern recognition, and image processing.