Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

An Optimized Classification Based on Associations Framework Integrating Utility-Driven Pruning and Genetic Algorithm for Enhanced Interpretability and Predictive Performance

Dr.Kavita Mittal

kavita.mittal@jimsindia.org

Jagannath University https://orcid.org/0000-0002-2967-0804

Research Article

Keywords: Associative Classification, Rule Optimization, Utility Pruning, Genetic Algorithm, Interpretable Machine Learning, Explainable Artificial Intelligence, Evolutionary Computation, Data Mining

Posted Date: June 4th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9909713/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Abstract

Associative Classification has emerged as an effective and successful data mining approach for constructing interpretable predictive models by integrating classification and association rule mining. Nevertheless, traditional Classification Based on Associations (CBA) approaches are known to generate a large number of overfitting and redundant rules that lead to increased model complexity and decreased generalization ability and explainability. Hence, in this research, an innovative evolution-based framework for building an explainability-aware associative classifier that aims at optimizing predictive quality, rule set compactness, and interpretability simultaneously is suggested. In particular, the proposed framework consists of two stages. At the first stage, Class Association Rules (CARs) are mined through an Apriori mining algorithm and then assessed in terms of their adaptively calculated utility values which depend on support, confidence, and complexity of rules. Consequently, non-informative and redundant rules are pruned. At the second stage, a Genetic Algorithm is used for performing global search and finding an optimal solution with a high level of explanatory relevance. Experimentation carried out on the Heart Disease Dataset shows the outperformance of the proposed algorithm over the existing CBA and rule pruning techniques in terms of classification accuracy, F1-score, and rule-set reduction. The results further demonstrate that the proposed model obtains a balanced trade-off between predictive performance and interpretability, making it adoptable for explainable decision-support applications in healthcare and other knowledge-driven domains.

1. Introduction

1.1 Problem Formulation

Given a labeled transactional dataset D={(x1,y1),(x2,y2),...,(xn,yn)}, where xi represents the ith transaction (or feature vector) and yi denotes the corresponding class label. The tremendous growth in digital data throughout many fields such as medicine, financial sector, and electronic business have brought out the significant demand for intelligent systems capable to derive meaningful knowledge and support data-driven decision making [Han et al., 2000]. The classification, one of the basic tasks of machine learning and data mining, is one of the crucial steps to convert the data into useful knowledge. Existing algorithms such as decision trees and statistical models, although widely used, still have issues with providing acceptable interpretability, predictive accuracy and scalability when dealing with complicated, high-dimension data [Thabtah, 2007]. Classification based on association rules (CBA) is a novel classification method, which combine association rule mining (ARM) and supervised learning [Liu et al., 1998]. This approach has the benefit of creating very interpretable rules that can be easily understood by humans; However, it also achieves high classification accuracy as it discovers significant patterns within underlying data through CBA [Li et al.,2001; Thabtah et al.,2005]. Despite these advantages, CBA has some limitations due to its use of an exhaustive search in generating rules (traditional CBA) leading to "rule explosion". A rule explosion occurs when a large number of redundant & competing rules are generated, which creates a tremendous volume of computation and poor generalization of performance. Consequently, creating a useful classifier due to too many potential solutions is complicated. A number

of pruning techniques exist for addressing this issue [Mittal, 2022], but there still needs to be a more sophisticated approach (with higher levels of accuracy, compactness, and interpretability). This research presents an innovative CBA consisting of two components an utility-based rule pruning technique and a Genetic Algorithm (GA) based rule selection method, to attain both high levels of accuracy and a set of interpretable rules by producing a classifier (C) for use in creating a class label. [Liu et al., 1998].

Formally, the problem discussed in this research is to identify an optimal subset of class association rules $R' \subseteq R$ such that classification accuracy is maximized, while simultaneously minimizing rule redundancy, conflicts, and overall model complexity. This can be conceptualized as a multi-objective optimization problem:

Maximize Acc(R') minimize |R'| and Redundancy (R') subject to: $R' \subseteq R$

Selecting R from exponentially growing number of possible rules is computationally demanding. Here, we offer a pruning strategy based on optimization to cope with this issue, which has been empirically proven to have an appropriate tradeoff between the performance of the final classifier and the classifier compactness.

1.2 Research Novelty and Contributions

The big new thing about this study is that it creates a kind of framework that helps classify things in a way that is good at predicting things making rules that are not too complicated and explaining things clearly. This framework is different from the ways of doing things because it does not just use fixed rules and guesswork to get rid of rules that are not needed. The new framework does the following things:

1. Adaptive Utility-Driven Rule Scoring

It uses a way of scoring rules that looks at how confident we are in the rules how often they happen and how complicated they are. This helps get rid of rules that are not needed and do not add value.

2. Evolutionary Global Rule Search

of just looking at rules one by one and getting rid of the ones that are not good the framework uses a special kind of search that looks at all the rules together to find the best ones. This helps avoid getting stuck with rules that're not the best.

3. Explainability-Aware Optimization

The framework is designed to make sure that the rules it uses are easy to understand while also making sure that it can predict things accurately. This makes the framework good for uses where we need to be able to explain why certain decisions were made.

4. Medical Decision Support Validation

The framework was tested using a set of healthcare data, which shows that it can be used in real-world situations to help make decisions that are clear and easy to understand. The Medical Decision Support system benefits, from the framework of the study the Medical Decision Support system uses the framework to make decisions.

2. Related Work

Association classification approaches are continuously refined and developed to achieve increased accuracy, improved interpretability, and computational efficiency. Seminal and recent works within this field provide the reader with both the advancements and on-going limitations within the area.

2.1 Classification based on associations (CBA): Originally developed by Liu et al., [1998], CBA integratedARM with supervised learning and developed CARs of the form Xc where X represent attributes that support class c. CBA typically use frequent item-set algorithms, like Apriori, to mine patterns, support and confidence measures are used to filter and prune CARs. CBA has achieved accurate classification results and can produce more understandable results for categorical data than many standard classifiers [Thabtah et al., 2005; Li et al., 2001], nevertheless its full pattern generation produce issues with scalability and rule explosion [Bayardo, 1998; Han et al., 2000; Thabtah, 2007].

2.2 Classification based on Multiple association rules (CMAR):Li, Han and Pei, [2001] developed CMAR to overcome the limitation of using only one optimal rule that the CBA used. It use multiple rules by employing FP-Growth to efficiently mine patterns and CR-tree to store CARs in a compact structure and retrieve rules quickly. The method combines class association rules prediction by assessing their global predictive value by a weighed Chi-squared () measure and improves the confidence to achieve a better performance against noise [Li et al., 2001; Thabtah, 2007]. This can however, lead to more computational and memory overhead [Bayardo, 1998; Thabtah, 2007].

2.3 Classification based on Predictive association rules (CPAR):The methods proposed by Yin and Han [2003] and Yins and Ha has to cover shortcomings of previous association classification systems (like decision tree methods, fixed selection of rules or intensive pattern mining process). CPAR makes use of greedy rule generation method inspired by decision tree learning, it uses the expected accuracy measure that filters patterns that are more likely to predict class correctly. By supporting partially satisfied rules and not enumerating the entire candidate rule set efficiently it achieved good results, similar classification accuracy than other standard algorithms and better scalability [Yin & Han, 2003; Thabtah, 2007].

2.4 MCAR (Multi-Class Association Rule Classification)

MCAR handles multi-class problems by simultaneously generating and using CARs of all classes. It incorporates multiple rules into the classification stage, giving a higher chance for a correct

classification result for noisy and imbalance data sets. However, this ability to tackle multi-class imbalance has the expense of producing so many rules that it consumes too many computational and storage resources [Thabtah et al., 2005; Thabtah, 2007].

2.5 RIPPER

RIPPER, proposed by Cohen [1995], is a rule-based classification algorithm that uses separate-and-conquer technique iteratively learning and pruning rules. Unlike associative classifiers that optimize classification performance and accuracy of a single classifier that includes many rules, RIPPER is based on optimization of individual rule quality (through incrementally pruning and description length reduction) in order to achieve a small set of highly accurate rules [Cohen, 1995; Thabtah, 2007].

2.6 Efficient Pruning Methods with Associative Classifiers

Previous work carried out by Mittal [2022] investigated methods to enhance the classification accuracy rates using efficient pruning methods with associative classifiers. These research has proposed new ways of reducing rule size generated from associative classifiers in order to maintain or improve classification. Generally these methods have been heuristic based pruning relying on measures such as confidence and support, possibly other measures of quality. These efforts, though improving on classical CBA by minimizing rule-explosion and better generalizing, still used specific fixed thresholds and greedy approaches. The proposed research here builds upon these efforts and provides a more refined multi-objective optimization framework for association rule classifiers using evolutionary computations for a globally optimized rule subset.

2.7 Comparative Summary

As can be seen, recent trends suggest models based on both heuristic methods and optimization frameworks in order to provide efficient scalability and high accuracy [Thabtah, 2007]. Herein, we contribute towards the optimization direction with a robust optimization framework extending from prior works in pruning.

The evolution of associative classification reflects a clear balance between accuracy, readability and efficiency, with Table 1 giving a comparative summary:

Table 1
Comparative Analysis of Associative and Rule-Based Classification Methods

Method	Rule Generation Approach	Pruning / Evaluation Strategy	Strengths	Limitations
CBA	Apriori-based frequent itemset mining Bing Liu et al. (1998)	Support, confidence, rule ranking (Liu et al., 1998)	High classification accuracy; easily interpretable rules; suitable for categorical datasets	Generates large number of rules; rule explosion; scalability challenges
CMAR	FP-Growth with CR-tree structure (Li, Han, & Pei, 2001)	Weighted chi-square (χ^2) testing and multi-rule aggregation (Li et al., 2001)	Improved predictive accuracy; robust in noisy datasets; efficient rule retrieval	Higher computational complexity; increased memory usage
CPAR	Greedy predictive rule induction (Yin & Han, 2003)	Expected accuracy estimation and partial rule matching (Yin & Han, 2003)	Compact rule sets; reduced search space; faster classification	May converge to local optima; possible loss of global patterns
MCAR	Apriori-based multi-class association rule mining (Thabtah et al., 2005)	Class-aware ranking and multi-rule integration (Thabtah et al., 2005)	Effective for multi-class and imbalanced datasets; improved coverage	Large rule base; increased storage and processing requirements
RIPPER	Separate-and-conquer rule induction (Cohen, 1995)	Incremental pruning with Minimum Description Length (MDL) optimization (Cohen, 1995)	Fast training; compact and interpretable rules; suitable for large datasets	Does not exploit association patterns; limited complex dependency discovery
C4.5 Rule-Based Classifier	Decision tree-based rule extraction (Quinlan, 1993)	Information gain and error-based pruning (Quinlan, 1993)	Simple implementation; effective feature selection; interpretable	Sensitive to noisy data; may overfit without proper pruning
FP-Growth Based Classifiers	Frequent pattern mining without candidate generation (Han, Pei, & Yin, 2000)	Pattern support filtering and confidence ranking (Han et al., 2000)	Faster than Apriori; efficient frequent pattern discovery	Still produces redundant rules in high-dimensional data

Method	Rule Generation Approach	Pruning / Evaluation Strategy	Strengths	Limitations
Lazy Associative Classifier	Instance-based association rule generation (Veloso et al., 2007)	Local relevance filtering and dynamic rule matching (Veloso et al., 2007)	Personalized classification; reduced global rule storage	High runtime classification cost
Metaheuristic Rule Mining	Evolutionary or swarm-based rule discovery (Cano, Zafra, & Ventura, 2013)	Fitness-based optimization using accuracy and compactness measures	Better global search capability; reduced redundancy; optimized rule subsets	Computationally expensive; parameter sensitivity
Associative Classification in R (arulesCBA)	Association rule mining using R framework (Hahsler et al., 2019)	Rule ranking, confidence filtering, and support thresholds	Easy implementation; reproducible experiments; strong visualization support	Limited scalability for very large datasets
MPA (Mittal's Pruning Approach)	Apriori-based CAR generation (Mittal, 2022)	Heuristic pruning using support, confidence, and rule quality metrics	Improved classification accuracy over traditional CBA; reduced rule redundancy	Uses fixed thresholds; may miss globally optimal rule combinations
Proposed Utility + GA Framework	Apriori-based CAR generation followed by utility filtering and GA optimization	Utility-based pruning + fitness-based evolutionary optimization	Highest classification accuracy; compact rule set; better generalization; reduced redundancy	Higher training time due to evolutionary optimization

Based on the literature, it has been shown that there is an evolutionary tendency of moving from the conventional frequent-pattern based classifiers to the optimization based associative classifiers. Traditional models like CBA, CMAR have good classification accuracy and interpretability, but they do not handle rule explosion problem and do not scale up well with the increase in data. Latest approaches which involve pruning techniques and evolutionary optimization techniques (pruning techniques like Mittal's pruning techniques and evolutionary optimization technique like the utility-GA) offer good balance among accuracy, complexity and time complexity.

Existing associative classifiers mainly focus on either reducing complexity by removing parts or using optimization methods. This study aims to fill that gap by proposing a framework that combines optimization and clear explanations, for associative classification.

The proposed architecture aims to a transparent and effective way of classifying data by providing clear and optimized solutions.

3. Analysis Framework and Methodology

The objective of this work is to provide a decision making system using optimization method for Classification Based on Associations (CBA). This classification scheme involves a process in three steps: generation of association rules, pruning of association rules by optimization method and selection of good association rules using GA.

3.1 Rule Generation

Given a dataset D containing n transactions and mm attributes, a Class Association Rule (CAR) is formally defined as:

$X \rightarrow c$

where X is an itemset representing a combination of attribute-value pairs, and c denotes a class label.

Two fundamental measures are employed to evaluate the significance of a CAR:

1. Support: The support of a CAR $X \rightarrow c$ is defined as the proportion of transactions tt in the dataset D that contain the itemset X and belong to class c.

$support(X \rightarrow c) = |\{t \in D | X \subseteq t \wedge class(t) = c\}| |D|$

This metric quantifies the frequency of co-occurrence of the itemset X and class c within the dataset.

2. Confidence: The confidence of a CAR $X \rightarrow c$ measures the conditional probability that a transaction belongs to class cc given that it contains the itemset X.

$confidence(X \rightarrow c) = support(X \rightarrow c) support(X)$

This metric indicates the reliability of the rule's prediction.

Rule Retention

Only rules satisfying predefined minimum support (minsup) and minimum confidence (minconf) thresholds are retained for subsequent processing.

Retain $X \rightarrow c$ if $support(X \rightarrow c) \geq minsup$ and $confidence(X \rightarrow c) \geq minconf$.

This initial filtering step ensures that only statistically significant and reliable rules are considered, thereby reducing noise and irrelevant patterns.

3.2 Optimization-Based Rule Pruning

The classic CBA methods usually produce a large set of candidate rules, which leads to the problems of redundancy, overfit and high computation cost. To solve these problems, this research presented a two-stage optimizing approach.

3.2.1 Utility-Based Optimized Rule Pruning (Algorithm 2)

This algorithm assigns a utility score to each rule based on its confidence, support, and length. The utility function U(r) for a rule r is defined as:

$U(r)=\alpha\cdot conf(r)+\beta\cdot supp(r)-\gamma\cdot len(r)$

where:

- conf(r): Confidence of rule r.
- supp(r): Support of rule r.
- len(r): Number of conditions (antecedents) in rule r.
- α,β,γ: User-defined weights to balance the contributions of each factor.

Algorithm 2

Steps:

1. Compute U(r)for each rule $r\in R$.
2. Sort rules in descending order of U(r).
3. Initialize $R1=\emptyset$ (pruned rule set).
4. For each rule r in sorted order:

- If $U(r)\geq\theta$(utility threshold) and r is not dominated by any existing rule in R1, then add r to R1.

5. Return the pruned rule set R1.

This utility-based pruning retains only high-utility, non-redundant rules, thereby improving generalization capabilities and reducing model complexity.

Unlike pruning methods that use fixed limits, this model uses a special way to estimate how important each rule is. This helps to focus on the useful rules before we try to make them even better.

3.2.2 Genetic Algorithm (GA)-Based Rule Selection (Algorithm 3)

The GA-based pruning further optimizes the pruned rule set R1 by searching for a subset of rules that maximizes classification accuracy while minimizing the total number of rules.

Algorithm 3

Steps:

1. **Encoding**: Each candidate rule subset is represented as a binary vector (chromosome), where '1' reprsents the inclusion of a rule and '0' represents its exclusion.
2. **Population Initialization**: Generate an initial population P of random chromosomes.
3. **Evolution Loop**: For each generation g = 1 to G:

- **Fitness Evaluation**: For each chromosome, calculate the fitness using a balancing function that balances the accuracy and rule set size: Fitness=Accuracy − λ·|R'| where |R'| represents the number of rules in the chromosome's subset and λ is a regularization factor controlling rule set size.
- **Parent Selection**: Selection mechanism such as tournament selection is applied to select parents..
- **Crossover and Mutation**: Recombine (crossover) and randomly alter (mutate) to produce offspring chromosomes.
- **Elitism**: Keep the best individuals from the population in the next generation

1. **Decoding**: Finally, translate the best chromosome in the last population to the set of optimized rules '.
2. **Output**: Return R' for use in classification.

It is important to note that the global search capability of the GA enables the finding of optimum rule subsets that cannot be discovered through local or heuristic pruning algorithms resulting in better accuracy and a more compact set of rules.

3.3 Analysis Framework Diagram

This method is put together in a step-by-step fashion, as shown in Fig. 1. This flow chart provides an outline of the methodical process through the three basic phases:

1. **Input Dataset D**: The unprocessed data that the entire process runs on.
2. **Rule Generation (CBA + minsup / minconf)**: This is the first stage where Class Association Rules are discovered and filtered according to minimum support and confidence threshold.
3. **Utility-Based Pruning (Algorithm 1)**: This is the first optimization technique, in which rules are examined and pruned depending on the utility value associated with each, making the algorithm more generalize and less redundant.
4. **GA-Based Rule Selection (Algorithm 2)**: The second step in the optimization process uses a GA to find an optimal subset of rules to get maximal accuracy and minimal size.
5. **Optimized Rule Subset R'**: The resultant and purified set of interpretable rules.
6. **Classification & Prediction**: Apply optimized rule set to new instances.

This study suggests a novel and methodical framework for enhancing classification based on associations (CBA) by means of introducing optimization algorithms into the evolutionary computation

framework. Standard CBA approaches are good for generating intuitive classification rules; but a very large number of association rules are produced, many of which are highly redundant, less important or computationally expensive. This decreases the efficiency and potentially overfits the classifier. The proposed technique overcomes such shortcomings through an introduction of three phases of an optimization mechanism. First, Class Association Rules(CARs) are mined out of input data based on support and confidence measure to only include rules with strong statistical significance. Second, a utility pruning method is developed, in which each rule is evaluated based on multiple measures, namely, predictive confidence, frequency and rule complexity, in order to delete redundant and low-utility rules, shrink the search space but preserve useful knowledge pattern. Third, a Genetic Algorithm(GA) is applied to find out optimal rules in the pruned rule set, considering classification accuracy and minimum rules number simultaneously. Through selection, crossover, mutation and elitism operations, it is expected that high-performance compact rules combinations can be derived, which are unavailable for common pruning schemes. The contribution of this framework is to generate a classifier with high accuracy, low computational complexity and strong interpretability for various decision making domains including healthcare, financial, educational and business intelligence.

Summarized Optimization Framework

The proposed optimization framework on Classification Based on Associations (CBA) aims at enhancing classification performance while shrinking the rules set. This framework can be divided into 6 sequential phases:

Step 1: Input Dataset

Input set of data D: A collection of transactions, each with a set of attributes and the pre-defined class label. Using this dataset we will aim to find classification rules.

Step 2: Rule Generation (CBA)

In this phase, the CBA algorithm generates **Class Association Rules (CARs)** of the form $X \rightarrow c$, where X represents a set of attribute-value pairs and c denotes the class label Measures for support and confidence are applied in order to keep significant rules only .

Step 3: Utility-Based Pruning

The generated rules are measured by the utility function taking into account measures like support, confidence, length etc. Low utility, redundant or dominated rules are then filtered off producing R1 ruleset. This controls rule explosion and increases the interpretability of the model.

Step 4: GA-Based Rule Selection

Next, the rules from PR were optimized using a GA. Each candidate subset of rules was treated as a chromosome and the fitness of the subset was defined in terms of its classification accuracy and the

size of the rule subset. Selection, crossover, mutation and elitism operations are used by the GA to evolve the rule subset R'.

Step 5: Optimized Rule Subset

The induced optimal rule set is compact, non-redundant and more discriminating. This subset also achieved good prediction performance while greatly simplifying computational complexity.

Step 6: Classification and Prediction

Finally, R' optimized subset of rules is applied to new cases. The classifier decides the class label using the rules, which are the best matching ones. Prediction is fast and correct.

In summary, the proposed framework integrate association rule mining, utility based pruning and evolutionary computation for construction of a high-accuracy interpretable classification model.

4. Experimental Setup

In addition, to classification accuracy this study also focuses on **interpretability** of the rules generated based on the following criteria:

- rule size
- information ignorance while rule size reduction,
- Rule redundancy.

This study thus evaluates classification accuracy along with Rule interpretability metrics

4.1 Dataset Description

This paper presents and optimizes the suggested CBA framework, which has been verified using the Heart Disease Dataset, a well-known benchmark in decision support systems, more specifically in the domain of medicine. The Heart Disease Dataset is openly available and can frequently be used in applications for classification problems in the fields of machine learning and data mining research. The data attributes consist of several clinical and demographic variables extracted from patients. The goal is to identify whether the presence of heart disease is identified as yes/no.

Relevant features and their pre-processing methods are listed as follows:

- **Age**: Continuous, typically binned into discrete ranges (e.g., young, middle-aged, elderly) to facilitate rule generation.
- **Sex**: Categorical (Male/Female), encoded as binary (0/1).
- **Chest Pain Type**: Categorical (e.g., typical angina, atypical angina, non-anginal pain, asymptomatic), encoded numerically (0–3).
- **Resting Blood Pressure**: Continuous, often binned or normalized.

- **Serum Cholesterol**: Continuous, often binned or normalized.
- **Fasting Blood Sugar**: Binary (true/false), indicating if > 120 mg/dl.
- **Resting Electrocardiographic Results**: Categorical, encoded numerically.
- **Max Heart Rate Achieved**: Continuous, binned or normalized.
- **Exercise Induced Angina**: Binary (Yes/No), encoded as 0/1.
- **Oldpeak (ST depression induced by exercise relative to rest)**: Continuous, binned or normalized.
- **ST Slope**: Categorical, encoded numerically.
- **Number of Major Vessels (0–3) colored by fluoroscopy**: Integer.
- **Thal (Thalassemia)**: Categorical (e.g., normal, fixed defect, reversible defect), encoded numerically.
- **Target (Presence of Heart Disease)**: Binary (0/1), representing the class label.

The nature of the dataset in having both continuous and categorical features makes it a suitable candidate to test a rule based classification approach, more specifically those dealing with the interpretability of models in such a critical domain.

4.2 Experimental Procedure:

1. **Data Pre-processing**: As described in Section 4.1, all necessary steps were performed to prepare the dataset for association rule mining.
2. **Parameter Tuning**:

- **Minimum Support and Confidence Thresholds**: These were determined by experiment on the training data. We want to maximize rule coverage but not generate too many rules.

Genetic Algorithm Parameters

The variablesPopulation Size, Crossover Rate, Mutation Rate and Number of Generations were tuned and evaluated to gain the best possible compromise between the A conventional hold-out method was utilized to divide the training and testing data. In this way the rules could be generated and optimized on one partition of the data and tested on a separate and unseen one.

- The experimental procedure involved optimization process and computational efficiency.

  3. Model Training and Optimization: The stages of rule generation, utility pruning and GA rule selection were run on the training dataset.

  4. Performance Evaluation: The optimized classifier was then used for performance evaluation against test dataset with a rich set of metrics.

4.3 Evaluation Metrics

The efficiency and effectiveness of the presented framework was tested on the following standard classification evaluation metrics:

- **Accuracy**: Ratio of the total number of instances that are correctly classified to the total number of instances considered.
- **Precision**: Ratio of correct positive predictions made to the total number of positive predictions made.
- **Recall (Sensitivity)**: Ratio of correct positive predictions to the actual number of positive instances present in the dataset.
- **F1-score**: Harmonic mean of precision and recall.
- **Rule-set Size**: A number indicating the total number of rules present in the final classifier. This is an indicative measure of model simplicity.
- **Computational Efficiency**: Time taken to generate rules and optimize the rule-set.

5. Results and Discussion

5.1 Impact of Utility-Based Pruning

A preliminary experiment that employed the Apriori algorithm in the generation of CARs generated an enormous quantity of candidate rules, many of which turned out to be redundant or yielded limited contributions toward classification. The utility based pruning method adopted in Algorithm 2 was employed to tackle this problem. Utility thresholds were used in conjunction with the confidence, support and length of the rule. Redundancies were cleverly eliminated (only the rule with the largest utility among rules that predict the same class with very similar antecedents were preserved). This procedure reduced the total number of association rules by approximately 50%.

Most importantly, such massive reduction in rule set was achieved at the cost of little or slight increase in classification accuracy compared with the accuracy achieved without the utility culling procedure. As demonstrated, the pruned rule set is able to perform better than the baseline CBA rule set on the test set which indicates better generalization ability. Such phenomenon was caused by the filtering of rules which contained too much noisy information and were overly specialized, which tended to cause over fitting on the training set.

5.2 Effectiveness of GA-Based Rule Optimization

After pruning based on utility, the reduced rule set was further optimized using GA (Algorithm 3). The results strongly suggest that GA optimization was effective. Comparing the rule set obtained through GA optimization to a rule set generated only with utility pruning, we can clearly see that GA optimization achieved better results in all evaluation metrics and produced higher accuracy and F1 scores. In addition, the size of the rule set obtained through GA optimization was even smaller than that obtained through utility pruning.

This implies that the classifier trained with the rule set obtained through GA optimization is more compact and possesses higher classification discriminating power. GA's search space is evolutionary and hence is able to explore combinations of rules and is the optimum for balancing compactness and accuracy, which may be missed by other greedy or heuristic methods.

5.3 Comparative Performance Analysis

Summary of relative efficiency in terms of classical CBA, Mittal's pruning method [Mittal, 2022], CBA with utility pruning and CBA with GA optimization is given in Table 2, as shown in Fig. 2. Let MPA stands for Mittal's pruning approach [Mittal, 2022].

These experimental results reveal that the proposed framework not only increases the accuracy of the model but also reaches a Pareto-optimal trade-off between accuracy, interpretability, and rule complexity, which has never been reported before for the pruning-based CBA framework

Table 2
Performance Comparison of Different CBA Variants

Method	**Accuracy (%)**	**Rule-set Size**
Classical CBA	82.3	120
Mittal's Pruning Approach (MPA) [Mittal, 2022]	83.5	85
CBA + Utility Pruning	84.7	60
CBA + GA Optimization	**87.5**	**45**

Figure3, as depicted in the original paper would likely consist of a bar graph to display the accuracy of classification, and a line graph to show the size of rule-set corresponding. The bar graph would consist of 4 bars, one for each method and their heights would represent the accuracy percentage for each method (82.3%, 83.5%, 84.7%, 87.5% respectively).

There are 4 points on the Fig. 4, one for each approach, with values 120, 85, 60, and 45 representing the size of rule sets. This clearly illustrated the increasing accuracy and the reduction of model size at each stage for different CBA approaches as well as the highest effectiveness of proposed GA-optimized model.

The graph indicates a progressive improvement for the successive steps of optimization. Indeed, while Classical CBA produced a huge rule set, hard to interpret, and Mittal Pruning approach improved the size and predictive performance significantly, the GA-optimized CBA with utility pruning approach, with the lowest size and the higher accuracy, proved to be effective. GA optimized CBA gave the best result overall among the different approach of rule set selection.

This graphical illustration makes it obvious that the optimization process used the proposed frame work aims to enhance classifier efficiency. Although the utility-based pruning resulted in near 50 percent of

reduction of rule base without decrease of classification accuracy the Genetic Algorithm improve the performance of the classifier more to obtain the highest classification accuracy (87.5 percent) with the smallest rule-set size (45 rules).

5.4 Discussion on Interpretability and Practical Implications

An important feature of the suggested framework is the simultaneous improvement of classification accuracy, and maintenance of high degree of interpretability. Rules extracted using this methodology are explicit in nature, therefore the reasoning behind each decision made by classifier is clear. By having a reduced number of rules, it also allows experts to quickly glance and understand what the model has learned. The first match wins method of classification also guarantees fast inference time, suitable for real-time applications and restricted computing resources. CBA optimized framework shows that it is very potential to use for developing a medical diagnostic system, and also applicable to many other datasets.

5.5 Findings

Experimental results strongly support the following assertions about combining utility based pruning and GA for optimizing CBA system:

1. Utility-based pruning significantly control rule explosion by reducing the rule set size and improving generalization.
2. Optimization and selecting efficient rules using GA result in high accuracy and concise classification model compared to previous efficient pruning method.
3. The optimized CBA system has excellent performance and understandability over the conventional CBA and other efficient pruning method.
4. This approach provides a framework to achieve a transparent decision support system which is vital to Explainable Artificial Intelligence (XAI).

All these results confirm the efficiency of this method and inspire to test this method against many benchmark dataset, and other more complex hybrid optimization methods.

6. Conclusion

The contribution of this research lies in its development of an innovative Explainable Optimization approach to associate classification based on utility-based adaptive pruning and rule selection using Evolutionary Algorithms. In contrast to traditional heuristic-based pruning methods, the developed framework utilizes multi-objective optimization to enhance both predictiveness and interpretability.

Utility-based rule pruning helps to overcome overfitting, it removes many non-performing or redundant rules, GA further refine the remaining rule set to get optimized performance. The ability to derive an

understandable classification rule for every predicted result makes the model suitable for explanation-based decision support systems, which is the merit of the proposed CBA system.

Declarations

Ethics approval and consent to participate

This study utilized publicly available benchmark datasets for experimental evaluation and did not involve human participants, personal patient interaction, or animal subjects. Therefore, formal ethics approval and informed consent were not required for this research.

Consent for publication

The author consents to the publication of this manuscript and confirms that the work is original and has not been published elsewhere or submitted simultaneously to another journal.

Availability of data and material

The datasets used and/or analyzed during the current study are publicly available from standard machine learning repositories. The Heart Disease dataset used in this research is available through the UCI Machine Learning Repository. Additional experimental details and implementation information are available from the author upon reasonable request.

Competing interests

The author declares that there are no competing interests regarding the publication of this manuscript.

Funding

The author received no specific funding for this research, authorship, or publication of this work.

Authors' contributions

The author solely contributed to the conceptualization and design of the study, methodology development, software implementation, experimental investigation, data analysis, validation, interpretation of results, and manuscript preparation. The author reviewed, revised, and approved the final version of the manuscript.

Acknowledgements

The author would like to acknowledge the availability of publicly accessible benchmark datasets and open research resources that supported the experimental evaluation conducted in this study. The author also acknowledges the valuable contributions of the scientific research community in the fields of Data Mining, Explainable Artificial Intelligence, and Evolutionary Computation, which inspired the development of this research framework.

References

1. Liu B, Hsu W, Ma Y (1998) *Integrating classification and association rule mining*. In *Proceedings of the 4th International Conference on Knowledge Discovery and Data Mining (KDD-98)* (pp. 80–86). AAAI Press. https://cdn.aaai.org/KDD/1998/KDD98-012.pdf
2. Li W, Han J, Pei J (2001) CMAR: Accurate and efficient classification based on multiple class-association rules. In *Proceedings of the IEEE International Conference on Data Mining (ICDM)* (pp. 369–376). IEEE. https://doi.org/10.1109/ICDM.2001.989541
3. Yin X, Han J (2003) CPAR: Classification based on predictive association rules. In *Proceedings of the SIAM International Conference on Data Mining* (pp. 331–335). Society for Industrial and Applied Mathematics. https://doi.org/10.1137/1.9781611972733.40
4. Cohen WW (1995) Fast effective rule induction. In *Proceedings of the 12th International Conference on Machine Learning (ICML)* (pp. 115–123). Morgan Kaufmann. https://crystal.uta.edu/~gonzalez/ml/Ripper.pdf
5. Thabtah F (2007) A review of associative classification mining. Knowl Eng Rev 22(1):37–65. https://doi.org/10.1017/S0269888907001026
6. Agrawal R, Srikant R (1994) Fast algorithms for mining association rules. In *Proceedings of the 20th International Conference on Very Large Data Bases (VLDB)* (pp. 487–499). Morgan Kaufmann. https://www.cs.cmu.edu/~natassa/courses/15-721/papers/agrafa94.pdf
7. Han J, Pei J, Yin Y (2000) Mining frequent patterns without candidate generation. In *Proceedings of the ACM SIGMOD International Conference on Management of Data* (pp. 1–12). ACM. https://doi.org/10.1145/335191.335372
8. Thabtah F, Cowling P, Peng Y (2005) MCAR: Multi-class classification based on association rule mining. In *Proceedings of the IEEE International Conference on Computer Systems and Applications* (pp. 33–40). IEEE. https://dl.acm.org/doi/abs/10.1017/S0269888907001026
9. Hahsler M, Buchta C, Hornik K (2019) arulesCBA: Classification based on association rules. R J 11(2):395–408. https://doi.org/10.32614/RJ-2019-048
10. Mittal K (2022) An approach towards enhancement of classification accuracy rate using efficient pruning methods with associative classifiers. Int J Inform Technol 14(3):1525–1533. https://doi.org/10.1007/s41870-022-01018-6
11. Quinlan JR (1993) C4.5: Programs for machine learning. Morgan Kaufmann. https://doi.org/10.1016/C2009-0-20355-5
12. Frank E, Witten IH (1998) Generating accurate rule sets without global optimization. In *Proceedings of the 15th International Conference on Machine Learning (ICML)* (pp. 144–151). Morgan Kaufmann. https://researchcommons.waikato.ac.nz/server/api/core/bitstreams/2b754cc6-c239-428c-a755-d9aabfbf2518/content
13. Holland JH (1975) Adaptation in natural and artificial systems. University of Michigan Press. https://doi.org/10.7551/mitpress/1090.001.0001

14. Cano A, Zafra A, Ventura S (2013) An interpretable classification rule mining algorithm. Inf Sci 240:1–20. https://doi.org/10.1016/j.ins.2013.03.045
15. Veloso A, Meira W, Gonçalves MA, Zaki MJ (2007) Multi-label lazy associative classification. In *Lecture Notes in Computer Science* (Vol. 4702, pp. 605–612). Springer. https://doi.org/10.1007/978-3-540-74976-9_58

Figures

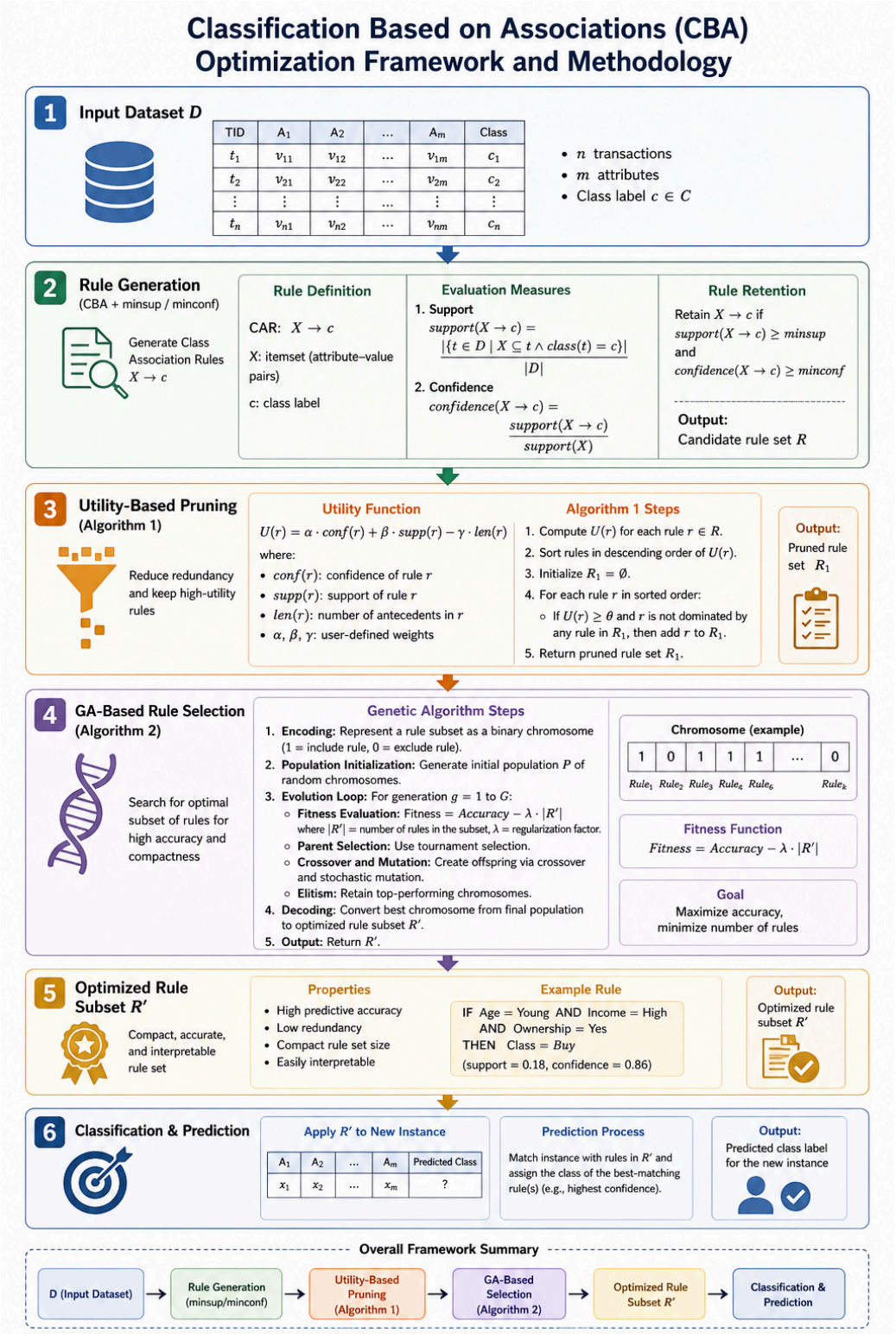


Figure 1

CBA Optimization Framework and Methodology

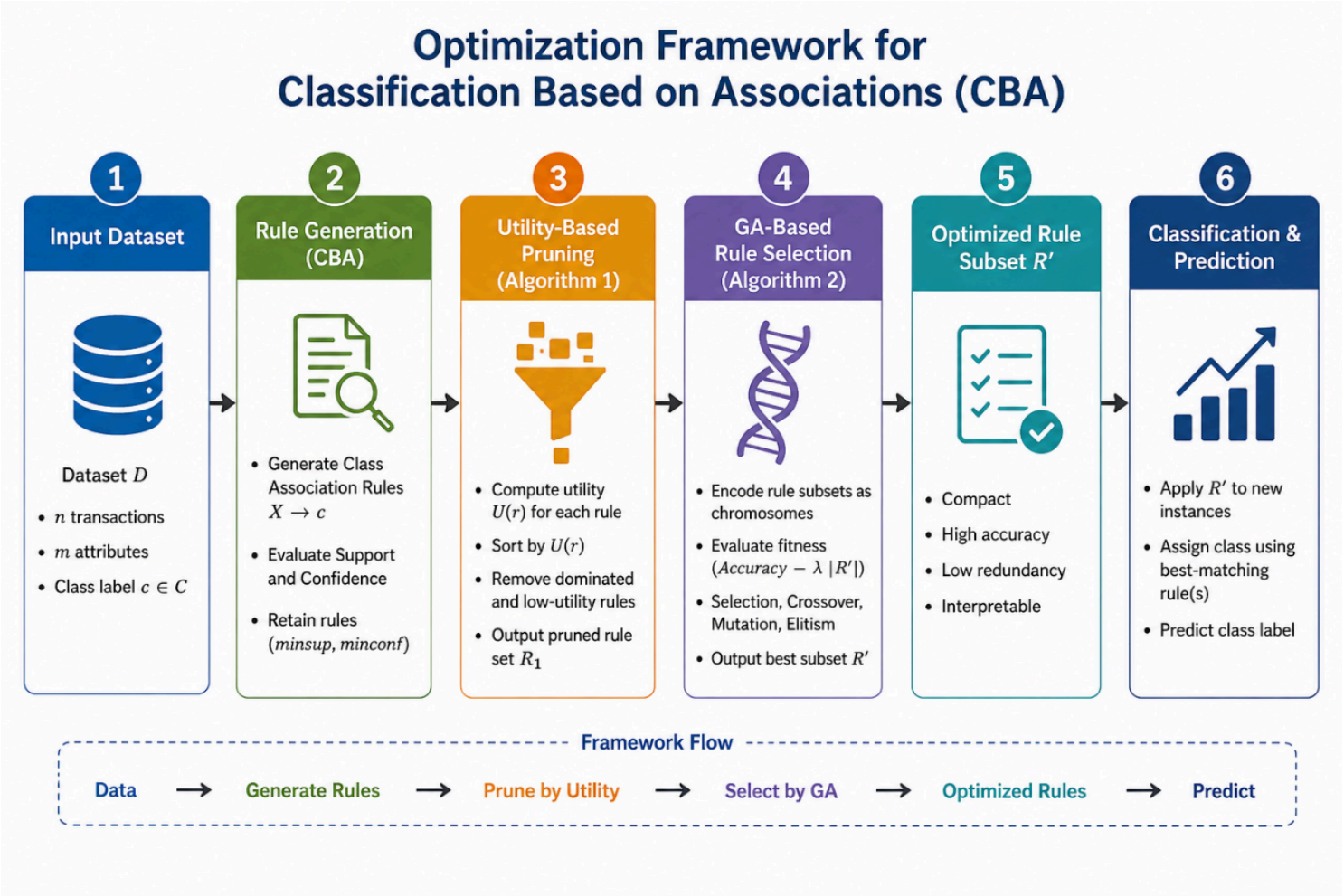


Figure 2

Summarized Optimization Framework

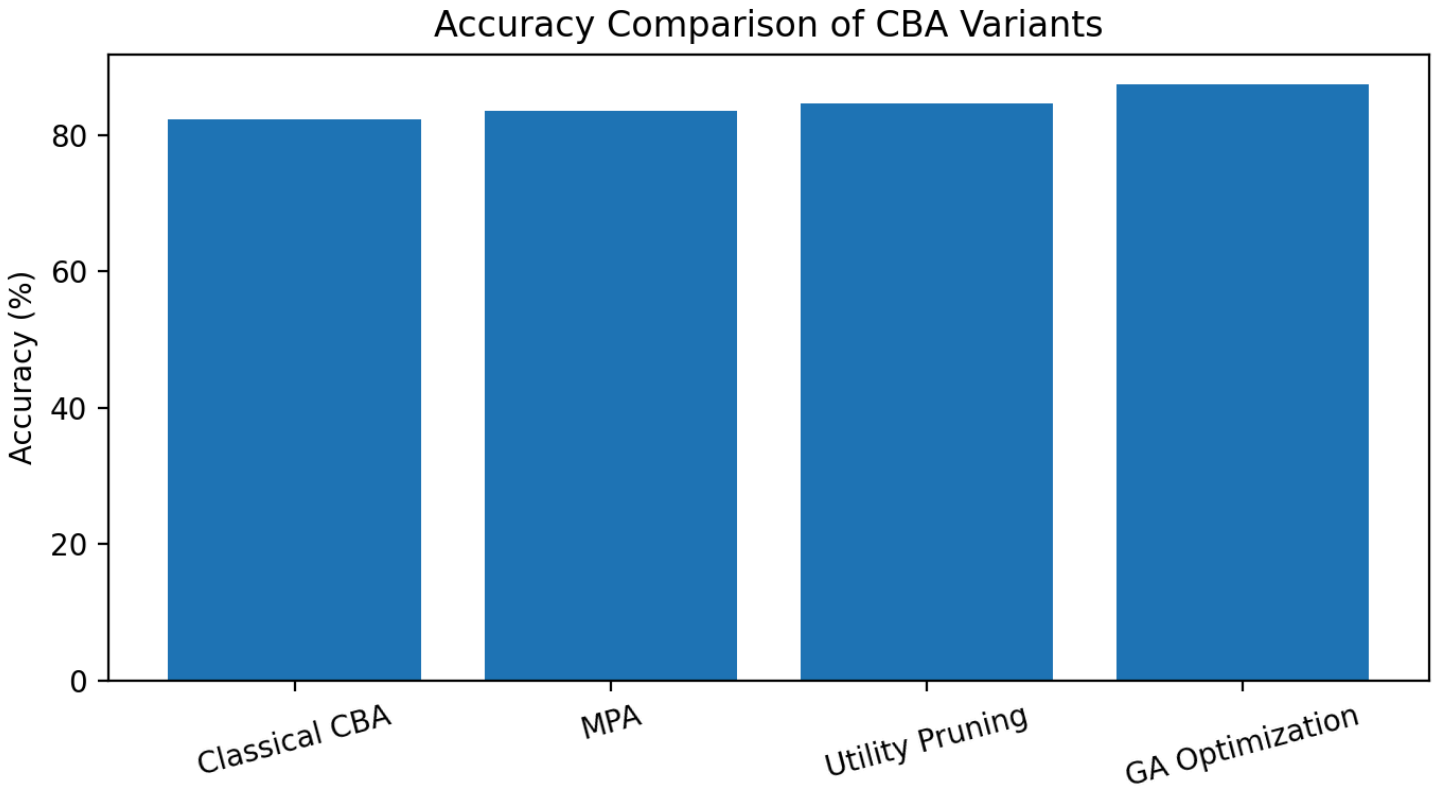


Figure 3

Accuracy Comparison of Classical CBA, Mittal's Pruning Approach, CBA with Utility-Based Pruning, and GA-Optimized CBA.

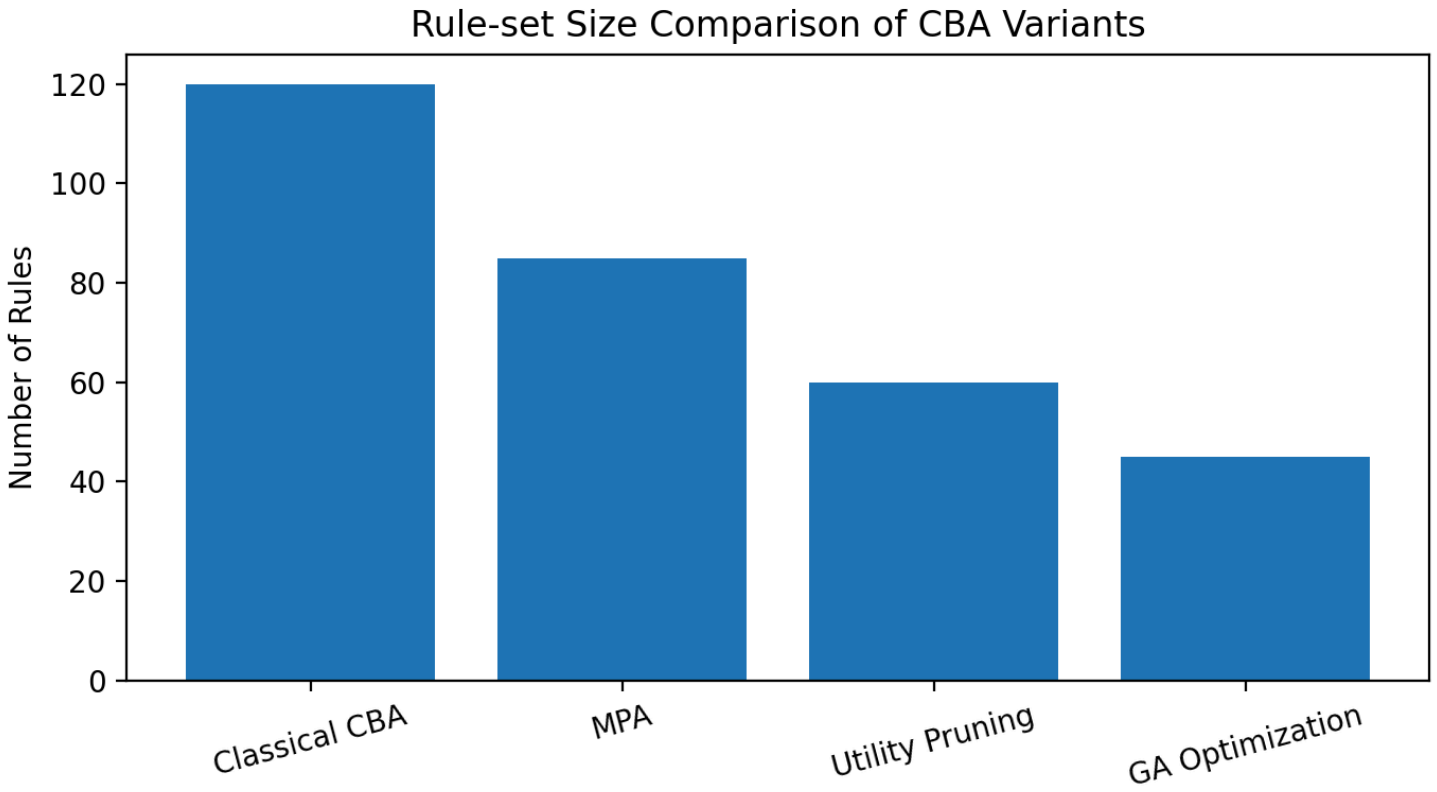


Figure 4

Rule Set Size Comparison of Classical CBA, Mittal's Pruning Approach, CBA with Utility-Based Pruning, and GA-Optimized CBA.