MJFAS
Malaysian Journal of Fundamental and Applied Sciences

A. Karim *et al.* | Malaysian Journal of Fundamental and Applied Sciences, Vol. 21 (2025) 1832-1859

RESEARCH ARTICLE

Interaction-Based Feature Selection Technique Using Fuzzy Discretization and Class Association Rule Mining for Breast Cancer Classification

Shahiratul A. Karim[a]*, Ummul H. Mohamad[a,b], Puteri N. E. Nohuddin[a,b]

[a]Institute of Visual Informatics, Bangunan Akademia Siber Teknopolis, Universiti Kebangsaan Malaysia, 43600 Bangi, Malaysia; [b]iAI-UKM Research Group, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia

Abstract Breast cancer is a leading global cause of cancer-related deaths highlighting the need for an accurate diagnostic system. Up to now, computer-aided diagnosis (CAD) system plays an essential role in supporting pathologists with prompt and accurate classification. Feature selection within the CAD system is crucial as it helps identify the most relevant data for subsequent classification tasks. This paper proposed a novel method that focuses on fuzzy discretization in handling continuous features and selecting relevant and interactive features while eliminating redundancy using Class Association Rule Mining (CARM). The proposed method, FD-CARI was compared with other feature selection techniques including CFS, FCBF, Consistency, Relief-F, and mRMR using five different machine learning classifiers such as Decision Tree (DT), Random Forest (RF), Logistic Regression (LR), Naive Bayes (NB), and Support Vector Machine (SVM). Performance evaluation metrics such as Accuracy (ACC), Sensitivity (SEN), Specificity (SPE), Precision, F1-Score, and AUC were then utilized. Results: The experimental findings consistently showed that the proposed method achieved high performance with an ACC of 96.21%, SEN of 94.26%, and SPE of 97.38% on the SVM classifiers, and an ACC of 96.05%, SEN of 93.82%, and SPE of 97.38% on the LR classifiers. It demonstrated similar effectiveness to Relief-F for DT and RF classifiers. However, FCBF achieved the highest performance on NB with ACC, SEN, and SPE values of 96.21%, 92%, and 96.60%, respectively. The proposed method efficiently selects relevant and interactive features while enabling classifiers to achieve better classification accuracy.

Keywords: Breast cancer classification, class association rule, feature selection, feature interaction, fuzzy discretization.

***For correspondence:**
p105766@siswa.ukm.edu.my

Received: 16 Aug. 2024
Accepted: 25 Feb. 2025

Introduction

Breast cancer remains a global health challenge that significantly impacts populations worldwide. It has overtaken lung cancer as the most frequently diagnosed cancer globally, accounting for 1 in 8 cancer diagnoses with a combined total of 2.3 million new cases in both men and women annually [1]. According to the World Health Organization (WHO), there were 2.3 million women diagnosed with breast cancer and 670,000 deaths recorded in 2022. There is an estimate that 310,720 new cases of female breast cancer and 42,250 estimated deaths in 2024 [2]. Breast cancer tumors typically originate from the ductal or lobular regions of the breast and are defined by abnormal cell growth and the potential to spread to nearby tissues. Clinical manifestations include breast lumps, changes in breast size or shape, skin dimpling, redness, nipple alterations, and unusual discharge [3]. These tumors are categorized as benign or malignant where the malignant tumors present a higher risk due to their rapid growth and metastatic potential [4]. Early detection is crucial for timely intervention [5] with medical imaging playing a crucial role in facilitating this process. Various imaging modalities such as digital mammography (DM), ultrasound (US), computer tomography (CT) scan, and magnetic resonance imaging (MRI) are employed for breast cancer diagnosis [6]. Continuous advancements in medical diagnostics which are driven by

technological innovations enhance the precision of diagnoses thereby improving treatment outcomes and reducing mortality rates [7].

Computer-aided diagnosis (CAD) methods have become indispensable tools for radiologists especially in enhancing diagnostic accuracy [8]. In CAD, medical images often contain a large number of features or characteristics that can potentially be used for diagnosis. However, not all of these features are equally informative or relevant for distinguishing between normal and abnormal tissues. Feature selection algorithms are employed in CAD systems to identify the subset of the most relevant features [9] for accurate diagnosis. By selecting only the most important features, CAD systems can improve efficiency by reducing computational complexity and processing time. The feature selection procedure is structured into four key stages which involved: (1) creating subsets of features (2) evaluating these subsets (3) determining when to stop (4) validating the final results as depicted in Figure 1 [10]. During subset generation, a search strategy examined possible combinations of features. Each subset undergone evaluation against the current top-performing subset using predetermined standards. If enhancement was detected, the new subset replaced the existing one. This iterative process continued until a stopping criterion was fulfilled and the chosen optimal subset was subjected to validation.

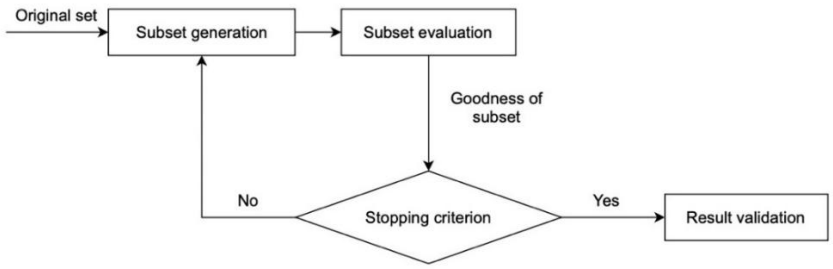

Figure 1. Basic steps for feature selection [10]

So far, most research on feature selection had focused on evaluating feature quality and developing search strategies [9], [11]. Assessing feature quality was essential to identify relevant features from the original dataset. The commonly used evaluation criteria include relevance and redundancy. Relevant features significantly benefit learning algorithms, while irrelevant and redundant features offer little to no useful information and may even reduce mining performance. Apart from these criteria, interaction between features is an essential yet frequently ignored [12], [13], [14]. Interactive features can greatly enhance classification accuracy even if they are individually irrelevant to the class. In recent years, numerous studies have concentrated on quantifying the interaction among features by employing information theory metrics such as mutual information [15], [16], [17]. Alternatively, the interaction between features can also be explored by uncovering interesting associations between them which can be facilitated by association rule mining [18].

This paper proposed a feature selection method designed to improve breast cancer diagnosis by focusing on relevant and interactive features while eliminating redundancy. This method known as FD-CARI where it applied Class Association Rule Mining (CARM) to discover meaningful feature associations with the target variable to ensure that the most informative features are selected for classification. In order to handle continuous data effectively, fuzzy discretization is utilized before CARM to preserve important information that might otherwise be lost with traditional discretization methods. This method enhances existing feature selection techniques like Relief-F and mRMR by not only filtering out redundant features but also identifying feature interactions that contribute to better classification accuracy. Relief-F [19] primarily ranks features based on their ability to distinguish between instances. While it can implicitly capture some dependencies, it does not explicitly search for or prioritize feature combinations that work together. In contrast, FD-CARI identifies groups of interacting features which can improve classification accuracy. Similarly, mRMR [20] focuses on reducing redundancy, but FD-CARI goes further by capturing meaningful feature associations so that the selected features are both non-redundant and highly relevant to the classification task.

The structure of this paper is as follows: Section 2: "Related Works," presented a literature review related to previous works on association rule mining in breast cancer diagnosis. Section 3: “Background Theory” discussed the background theories of fuzzy sets with the membership function definition and association rule mining which covers aspects such as relevancy, redundancy, and interactive features. including their definitions. This is followed by Section 4: "Methodology" which detailed the phases of the research

framework and also the proposed feature selection method. Section 5: "Experiment" revealed the experimental setup and empirical results of the proposed method. Finally, Section 6: “Conclusion and Future Works” provided the conclusion and suggested potential future works.

Related Works

The extensive research conducted since the 1970s in the field of feature selection (FS) emphasized the importance of evaluating feature quality to identify relevant attributes within datasets. FS methods can be divided into three categories [21]: i) those that exclusively target irrelevant features, ii) those that handle both irrelevant and redundant features, and iii) those that manage irrelevant and redundant features while also considering feature interactions. Most existing FS methods can effectively identify irrelevant features using various evaluation functions. However, not all of them are capable of eliminating redundant features while considering feature interactions. Feature weighting and ranking algorithms assess and prioritize features according to their relevance to the target concept on an individual basis. For instance, the Relief algorithm [22] assigns weights to features based on their capability to differentiate instances from various targets using a distance-based criterion function. Relief-F [19] is an extension of the Relief algorithm proposed to handle challenges like noisy and incomplete datasets as well as multi-class classification problems. However, despite these enhancements, Relief-F still encounters difficulties in addressing redundancy as it tends to assign high importance to two features that are predictive but highly correlated.

On the other hand, FS methods that consider redundant features include Correlation-based Feature Selection (CFS) [23], Fast Correlation-based Filter (FCBF) [24] and Consistency [25]. CFS and FCBF are two methods for selecting features that are highly correlated with class labels while avoiding redundancy. CFS calculates the merit of feature subsets using correlation-based heuristics while FCBF selects features based on a predefined threshold for correlation with class labels. Additionally, Consistency focuses on correctly recognizing samples using majority voting and efficiently removes redundant or irrelevant features. Although these algorithms effectively remove irrelevant and redundant features, they do not evaluate the interaction between features.

Feature interaction has been drawing more attention in recent years. Features that appear irrelevant individually may become highly relevant when combined with others which leads to two-way, three-way, or complex multi-way interactions among features. Jakulin and Bratko [12] introduced a method using interaction gain as a heuristic to detect feature interaction. Zhao and Liu [15] proposed an algorithm called INTERACT where feature interactions are implicitly handled by a carefully designed feature evaluation metric and a search strategy with a specially designed data structure. An interaction weight factor that can reflect whether a feature is redundant or interactive called IWFS is proposed by [16]. Wang, Jiang and Jiang [17] presented an FS method using the "maximum of the maximum" criterion to identify both highly relevant and maximally interactive features called MRMI.

Otherwise, association rule mining (ARM) also has been integrated into feature selection to evaluate the quality of features. Das and Nath [26] introduced a method combining genetic algorithms with ARM for dimensionality reduction which focuses on removing irrelevant and redundant data thereby improving accuracy and computation time. Xie, Wu and Qian [27] further developed a feature selection algorithm using ARM to identify features closely related to the class attribute which showed effectiveness in reducing feature sets while maintaining high classification performance. Karabatak and Ince [28] combined ARM with neural networks for diagnosing erythemato-squamous diseases and achieving high classification accuracy rates of 98.61% with reduced feature spaces. Sheikhan, Rad and Shirazi [29] developed a feature selection method using fuzzy association rules and fuzzy Adaptive Resonance Theory Mapping (ARTMAP) for intrusion detection. This approach effectively reduces features and improves detection rates and false alarm rates. Wang and Song [18] proposed an algorithm of Feature Subset Selection Algorithm based on Association Rule Mining (FEAST) using constraint association rules to identify relevant and redundant feature values using the FP-Growth method. They further refined the FEAST algorithm by incorporating partial least square regression for threshold prediction in association rule mining to improve feature selection and classification accuracy [21]. Waseem, Salman and Muhammad [30] combined JRip classifier with ARM to select relevant features. The approach is tested on various datasets which show higher classification accuracy with fewer features. Wang and Quo [31] integrated fuzzy clustering with ARM for soft-sensor applications, effectively selecting important variables and eliminating overlap. Harikumar, Dilipkumar and Kaimal [32] enhanced the apriori algorithm with information-theoretic methods and QR decomposition that improved computational efficiency and accuracy in high-dimensional data. Li et. [33] developed a method using equal probability discretization and ARM for fault diagnosis in rolling bearings successfully identifying relevant features and generating

effective association rules. Lin and Gao [34] proposed a method to select features and their interactions using ARM on various datasets and demonstrate computational efficiency and effectiveness. Liewlom [35] introduced a new measure for pruning class-association rules in imbalanced datasets, particularly in breast cancer. The prune rules show improved precision and conciseness. Farghaly and El-Hafeez [36] presented a feature selection technique for text classification based on frequent and correlated items achieving high accuracy with significantly reduced feature count. The final feature subset obtained retained frequent and significant features while removing redundant and unnecessary features and considering the interaction between features. Zhang [37] had employed three core metrics from information theory to enhance feature selection such as Mutual Information (MI) and Conditional Mutual Information (CMI) with additional dynamic weighting mechanism based on Information Gain (IG) to adjust feature importance. The proposed method achieved higher classification accuracy compared with six other feature selection algorithms. Lv *et al*. [38] presented Online Streaming Feature Selection via Feature Interaction (OFSI) algorithm where it dynamically selects features from continuous data streams while emphasizing feature interactions. This concept has been extended to Group-OFSI algorithm which handles batch-arriving features to enable a more detailed interaction analysis within feature groups. In multi-label classification, a graph theory-based algorithm called Online Streaming Feature Selection Method Based on Label Group Correlation and Feature Interaction (OSLGC) algorithm was introduced by [39]. The method was proposed to cluster labels and assigns weights using MI to capture correlations within label groups. OSLGC algorithm applied a sliding window approach to process data dynamically and achieved superior predictive accuracy and stability over traditional methods.

The previous experiments indicated that most FS methods efficiently identify irrelevant features through various evaluation functions. However, these methods often focus solely on removing redundant features and lack sufficient research on feature interactions based on ARM. Since the interaction between features may improve the performance of the machine learning classifiers, thus, we proposed a method that focuses on selecting relevant and interactive features while eliminating redundant ones using CARM in diagnosing the breast cancer. Table 1 provides a summary of various feature selection methods highlighting their objectives and results.

Table 1. Summary of feature selection methods

Reference	Method	Objective	Results
[19]	Relief-F	Ranks individual features	Handles noisy & multi-class data
[23]	CFS	Removes irrelevant & redundant features	Uses correlation heuristics to select key features
[24]	FCBF	Fast filter-based FS	Reduces redundancy efficiently
[25]	Consistency	Majority voting-based FS	Ensures minimal redundancy
[15]	INTERACT	Feature interaction-aware FS	Uses data structures for implicit interaction detection
[16]	IWFS	Determines whether a feature is redundant or interactive	Identifies redundancy and interaction levels for better FS
[17]	MRMI	Interaction-aware FS	Selects highly relevant & interactive features
[26]	ARM + GA	Dimensionality reduction with ARM & GA	Removes irrelevant & redundant features, improving accuracy
[27]	ARM-based FS	ARM-based FS for class relevance	Identifies class-relevant features using ARM
[28]	ARM + Neural Networks	Disease diagnosis FS using ARM & Neural Networks	Achieves 98.61% accuracy with feature reduction
[29]	Fuzzy ARM + ARTMAP	Intrusion detection using fuzzy ARM & ARTMAP	Improves intrusion detection rates & reduces false alarms
[18]	FEAST	Uses constraint ARM and FP-Growth for FS	Effectively identifies relevant and redundant features
[30]	ARM + JRip Classifier	Combines ARM with JRip classifier for FS	Achieves high classification accuracy with fewer features
[31]	Fuzzy Clustering + ARM	Soft-sensor FS using fuzzy clustering & ARM	Effectively selects important variables & eliminates overlap
[32]	ARM + QR Decomposition	Enhances Apriori with QR decomposition	Improves efficiency and accuracy in high-dimensional data
[33]	ARM + Equal Probability	Uses ARM for fault diagnosis in rolling bearings	Successfully identifies relevant features & effective rules

Reference	Method	Objective	Results
	Discretization		
[34]	ARM for Feature Selection & Interaction	Uses ARM to select features & their interactions	Demonstrates computational efficiency and effectiveness
[36]	ARM	Uses ARM-based FS for text classification	Achieves high accuracy with significantly reduced features
[37]	Information Theory	Employs MI, CMI & IG-based weighting for FS	Outperforms six FS algorithms in classification accuracy
[38]	OFSI + Group OFSI	Dynamic FS for continuous data streams	Adapts dynamically to evolving data
[39]	OSLGC (Graph-Based FS)	Uses MI & graph theory for multi-label FS	Achieves superior predictive accuracy & stability

Background Theory

This section provides an overview of fuzzy sets and definition of association rule mining, class association rules and atomic association rules. Additionally, it discusses classic definitions of feature relevancy, redundancy, and interaction based on the theory of association rule mining.

Fuzzy Sets

A challenge with association rule mining is their dependence on categorical features which makes it difficult to directly handle numeric data which is frequently found in real-world applications. Data discretization is utilized on numeric data which involves dividing the range of a continuous attribute into several intervals where each is assigned an integer label. However, traditional discretization methods have a "sharp boundary problem," where values near the interval boundaries are treated the same as those in the middle which results in significant information loss and reduced classifier accuracy [40]. To tackle this issue, the paper adopts a fuzzification technique which is converting precise numeric values into fuzzy values hence enhancing the representation of continuous data for rule-based classification tasks.

A fuzzy set is a concept from fuzzy logic that was introduced by [41] which allows for the representation of uncertainty and vagueness in data. In classical set theory, an element is definitively either a member or not a member of set A that is $x \in A$ or $x \notin A$. This type of set is referred to as a crisp set. A fuzzy set is a set whose element has a degree of membership that can be expressed as:

$$A = \{(x, \mu_A(x)) | \ x \in X\} \tag{1}$$

where X is called the universe set and $\mu_A(x): X \rightarrow [0,1]$ is the membership function which maps the membership degree of each element x of X to a value between 0 to 1. This degree indicates how strongly the element belongs to the fuzzy set. If $\mu_A(x) = 0$, it indicates that x is not a member of A conversely if $\mu_A(x) = 1$, then x is considered a member of A, and values between 0 and 1 indicate partial membership.

Membership Functions

The membership function plays a crucial role in defining the fuzziness within a fuzzy set by defining the degree of membership to each element x in set X. There are various types of membership functions including triangular, Gaussian, trapezoidal, bell-shaped, and sigmoidal functions. In this study, the triangular membership function used to define the fuzzy sets which can be calculated as:

$$\mu_A(X;\ a,b,c) = \begin{cases} 0 & x \le a \\ \dfrac{x-a}{b-a} & a \le x \le b \\ \dfrac{c-x}{c-b} & b \le x \le c \\ 0 & x \ge c \end{cases} \tag{2}$$

where x is an element of a fuzzy set X, a and c represent the lower and upper boundary of fuzzy set X respectively and b represents the lower and the center of the fuzzy set X. An alternate expression can be presented by using min and max:

$$\mu_A(X;\ a,b,c) = max\left(min\left(\frac{x-a}{b-a}, \frac{c-x}{c-b}\right), 0\right) \tag{3}$$

Association Rule Mining

Association analysis is a methodology that is useful for discovering interesting relationships hidden in large datasets. Let D be a dataset, $I = \{i_1, i_2, \ldots, i_d\}$ be a set of items, and $T = \{t_1, t_2, \ldots, t_N\}$ is a set of transactions t from transaction database T. An itemset containing k items is termed a k-itemset. A crucial property of an itemset is its support count $\sigma(x)$ which denotes the number of instances containing a specific itemset can be expressed as:

$$\sigma(X) = |\{t_i | X \subseteq t_i, t_i \in T\}| \tag{4}$$

where the symbol |·| denotes the number of elements in a set.

An association rule is an implication expression of the form $X \Rightarrow Y$ where X and Y are disjoint itemset. The strength of the rule is measured by its support and confidence which can be calculated by: An Apriori algorithm pioneered method by [42] in association rule mining which effectively manages the exponential growth of candidate itemsets through support-based pruning. Its process involves two main steps: (i) iteratively finding frequent itemsets meeting a user-defined support threshold and (ii) constructing association rules that satisfy a user-defined minimum confidence level using these frequent itemsets. This procedure follows a level-wise approach that traverses the itemset lattice from frequent 1-itemsets to the maximum size of frequent itemsets and utilizes a generate-and-test strategy to discover frequent itemsets. Association rules are then generated iteratively until the antecedent becomes empty. There are various measures to evaluate the interestingness of association rules [43]. Interestingness refers to the quality and significance of an association rule in capturing meaningful relationships within the data. The interestingness measures such as lift, conviction, and certainty factors are utilized which can be described as follows:

$$supp(X \Rightarrow Y) = \frac{\sigma(X \cup Y)}{N} \tag{5}$$

$$conf(X \Rightarrow Y) = \frac{\sigma(X \cup Y)}{\sigma(X)} \tag{6}$$

Lift

Lift is a measure of how much more often the antecedent X and the consequent Y occur together than would be expected if they were statistically independent. Lift can be defined as

$$\text{lift}(X \Rightarrow Y) = \frac{conf(X \Rightarrow Y)}{supp(Y)} \tag{7}$$

If lift > 1, it indicates a positive correlation between X and Y (directly proportional), whereas lift < 1 shows a negative correlation between X and Y (inversely proportional). If lift =1, X and Y are uncorrelated to each other.

Conviction

Conviction is a measure of the strength of the implication of the rule $X \Rightarrow Y$. It quantifies how much more often X leads to Y than it would if Y occurred independently of X. Conviction can be expressed as:

$$\text{conviction}(X \Rightarrow Y) = \frac{1 - supp(Y)}{1 - conf(X \Rightarrow Y)} \tag{8}$$

Conviction values greater than 1 indicate that the X strongly implies the Y.

Certainty Factor

The certainty Factor is a measure to evaluate the degree to which the occurrence of an X increases the likelihood of the Y relative to the baseline probability of the Y. It helps quantify the strength of the association between X and Y. Certainty factor can be computed as:

$$\text{CF}(X \Rightarrow Y) = \begin{cases} \dfrac{conf(X \Rightarrow Y) - supp(Y)}{1 - supp(Y)} & conf(X \Rightarrow Y) > supp(Y) \\ \dfrac{conf(X \Rightarrow Y) - supp(Y)}{supp(Y)} & conf(X \Rightarrow Y) < supp(Y) \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

Class Association Rules and Atomic Association Rules

Class Association Rules (CARs) represent a specialized form of traditional association rules that explore relationships between items and a specified class label. On the other hand, Atomic Association Rules (AARs) are basic association rules that involve only one item in both the antecedent and consequent offering fundamental insights into item co-occurrences. Both CARs and AARs are derived from Strong Association Rules (SAR) to ensure they meet predefined thresholds of minimum support and minimum confidence. SAR, CAR, and AAR can be defined as follows [18]:

Definition 1. Strong association rule (SAR). A rule r of form $X \Rightarrow C$ is a SAR if and only if

$$supp(r) > minSup \cap conf(r) > minConf \quad (10)$$

Definition 2. Class association rule (CAR). A rule r of form $X \Rightarrow C$ is a CAR if and only if

$$r \in SAR \cap X \subseteq FVs \cap C \subseteq TVs \cap |C| = 1 \quad (11)$$

where FVs is the feature value itemset and TVs is the target value itemset.

Definition 3. Atomic association rule (AAR*).* A rule r of form $X \Rightarrow C$ is an AAR if and only if

$$\mathrm{r} \in SAR \cap |X| = 1 \cap |C| = 1 \quad (12)$$

Definition of Relevant, Redundant, and Interactive Features

The definition of relevancy, redundancy, and interaction of the features based on association rules [18] can be described as follows:

Definition 4. Relevant feature $(RelF)$. Feature F_i is relevant to the target concept Y if and only if:

$$\exists f_{ij} \in F_i, \{f_{ij} | \, fij \text{ is a } RelFV\} \neq \emptyset \quad (13)$$

Otherwise, F_i is an irrelevant feature $(iRelF)$. $RelFV$ is a feature value that can be described as:

A specific value f_{ij} is relevant to the target concept if and only if

$$\exists r \in CARs, f_{ij} \in r.Ante \quad (14)$$

where f_{ij} is denotes the jth-value *of* F_i *and* $r.Ante$ represents the antecedent of rule *r.* Otherwise, f_{ij} is an irrelevant feature value $(iRelFV)$.

Definition 5. Redundant feature (*RedF*). Feature F_i is redundant if and only if

$$\forall f_{ij} \in F_i, \{f_{ij} \text{ is a } RedFV \text{ or an } iRelFV\} \neq \emptyset \quad (15)$$

where $RedFV$ is a specific value f of a feature value set if and only if:

$$\exists r \in AARs, (\{f\} = r.\mathrm{Cons}) \cap (r.\mathrm{Ante} \subseteq FVs) \quad (16)$$

where *r.*Ante and *r.*Cons represent the antecedent and consequent of rule *r,* respectively.

Definition 6. Interactive features $(IntF)$. Let $F_s = \{F_1, F_2, \ldots, F_k\}$ be a feature subset with k features, and $VAset$ be its value-assignment sets. Features $F_1, F_2, \ldots, F_k$ are said to interact with each other if and only if:

$$\exists r \in AARs, (\{f\} = r.\mathrm{Cons}) \cap (r.\mathrm{Ante} \subseteq FVs) \quad (17)$$

where $FV_{sk}$is FV_s with k-th feature value interaction which can be defined as:

Suppose FV_s is a feature value set with *k* feature values. Let $(A \subset FV_s) \neq \emptyset$ and $B = FV_s - A$ and r_F, r_A, and r_B be the *CAR* of FV_s. The *k*-th feature value in FV_s are said to interact with each other if and only if:

$$conf(r_F) > conf(r_A) \cap conf(r_F) > conf(r_B) \quad (18)$$

Methodology

This section depicted the proposed research framework which includes data collection, data pre-processing, proposed feature selection model, classification using supervised machine learning algorithms, and performance evaluation procedure to assess the effectiveness and accuracy of the model in breast cancer diagnosis

Proposed Research Framework

The proposed research framework is illustrated in Figure 2 as follows:

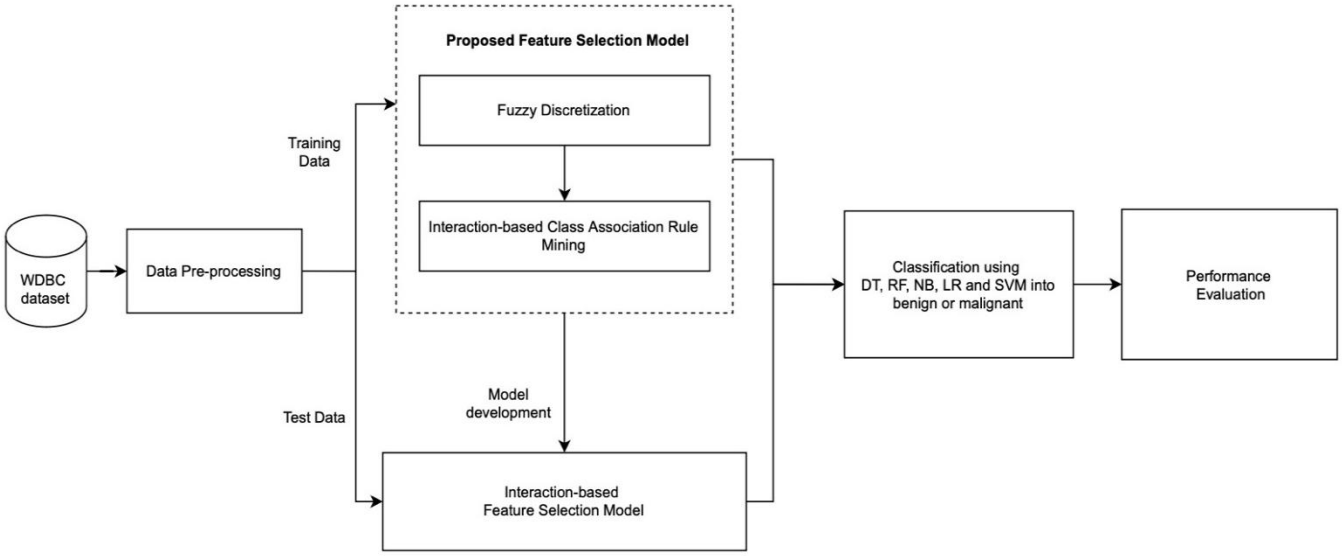

Figure 2. The proposed framework

Data Pre-Processing

Data pre-processing is crucial for refining and optimizing data to improve the accuracy and efficiency of machine learning models. It encompasses essential tasks aimed at cleaning and preparing the data. In this study, various data pre-processing methods have been adopted to ensure the quality of the dataset. These methods include:

a. Missing values checking: Missing values refer to any entries that are absent or undefined for certain attributes or features that may be due to various reasons such as human error during data entry, equipment malfunction during data collection, or simply because the information was not available.

b. Encode the categorical data: Encoding the data into categorical variables that can be understood by machine learning models is also an essential task. The categorical labels like 'diagnosis' which was originally expressed as 'B' for Benign and 'M' for Malignant were converted into numerical values which M=1 and B=0 to ensure compatibility with machine learning algorithms.

c. Outlier detection. Outliers are data points that significantly deviate from the rest of the dataset either being unusually high or low compared to the majority of observations. The Z-Score method is commonly used in outlier detection and removal. The Z-Score provides a method to assess the degree of divergence of an observation from the rest of the dataset which can be expressed as

$$z = \frac{(X - \mu)}{\sigma} \tag{19}$$

where X is the data point, μ is the mean of the dataset, and σ is the standard deviation. Data points with Z-Score exceeding a threshold of $z < -3, z > 3$ typically are considered outliers [44] and may be removed from the dataset.

d. Data transformation. Transforming the raw data into transactional format is the initial step required for association rule mining. This approach to treating patient records as 'transactions' draws inspiration from the methodology commonly employed in retail transactions. Each feature of the dataset became an ‘item' such as a patient's mean radius, texture, perimeter, and diagnosis in the analysis.

Proposed Feature Selection Method

The proposed feature selection model consists of six key phases. Initially, candidate feature subsets are generated through feature ranking using multiple ranker methods followed by combining them into an ensemble feature subset using a union combination method. The second phase involves incorporating a fuzzy discretization technique to handle continuous features. In the third phase, an *Apriori* algorithm is applied to mine association rules based on minimum support and minimum confidence and the fourth phase involves extracting the class association rules (CAR) and atomic association rules (AAR). The fifth

phase focuses on the discovery of interactive relevant features (RFs) on the pruned rules of CAR (pCAR) based on the minimum lift, minimum conviction, and minimum certainty factor thresholds. The sixth phase focuses on the elimination of redundant features through the evaluation of AAR. This process concludes in the identification of an optimal feature subset S. The proposed feature selection algorithm, termed Fuzzy Discretization-Class Association Rule Mining based on Interaction (FD-CARI), summarizes these phases and is illustrated as in Figure 3.

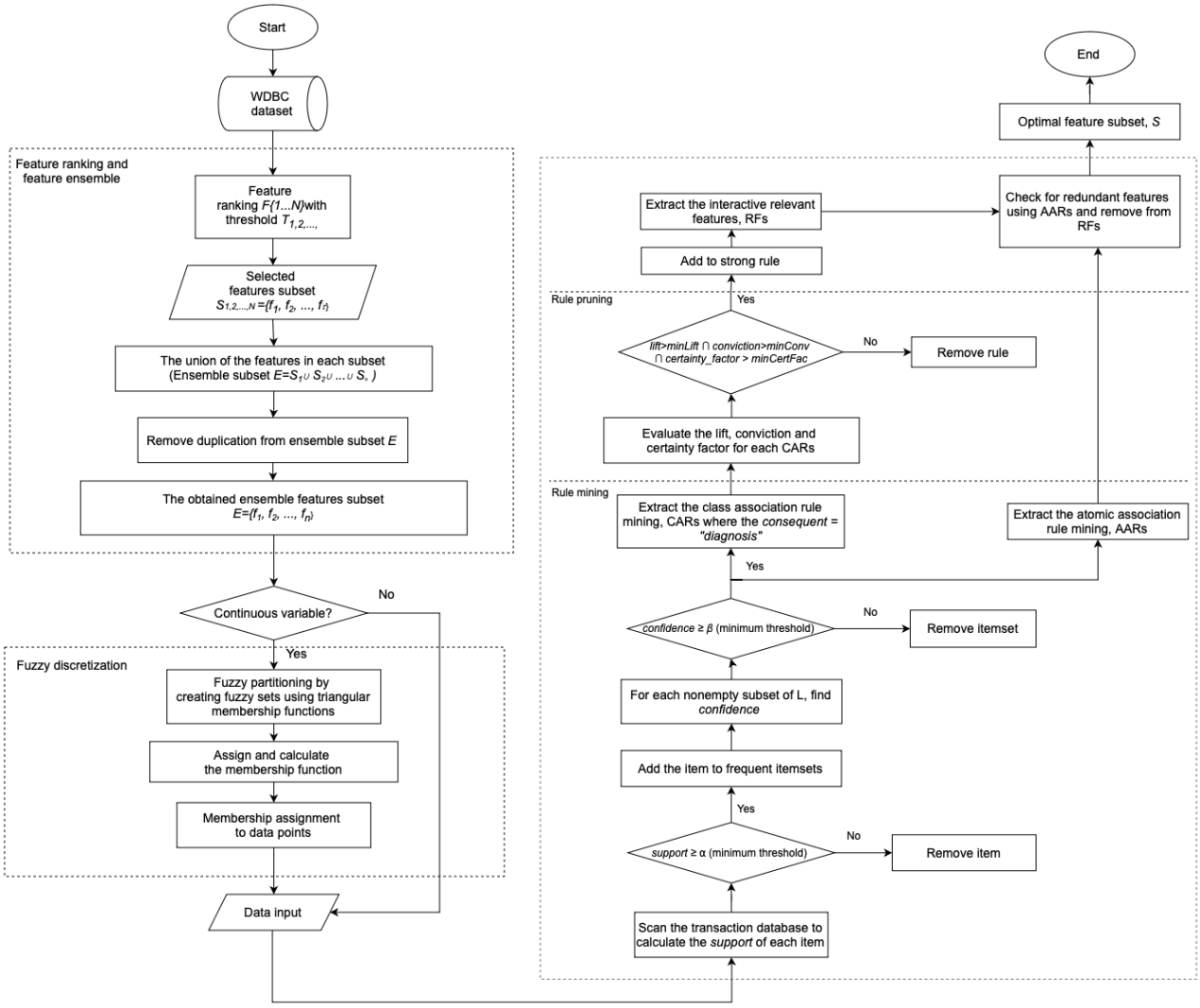

Figure 3. Proposed FD-CARI workflow phases

The proposed FD-CARI workflow phases are detailed as follows:

a. Evaluation of candidate feature subsets using Ranker methods

Phase I involves the evaluation of features through a systematic approach. Initially, features are evaluated using the ranker search method by various attribute evaluator methods. In this study, the features undergo evaluation using three distinct ranker methods: correlation, mutual information, and feature importance.

Pearson Correlation Coefficient (PCC). PCC measures the linear relationship between pairs of features. Features with high correlation to the target variable or each other are often considered more relevant. PCC between features X and Y can be evaluated as:

$$\mathrm{PCC}(X,Y)=\frac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\overline{x})^2(y_i-\overline{y})^2}} \tag{20}$$

where x_i and y_i are the individual data points for features X and Y respectively, $\overline{x}$ and $\overline{y}$ are the mean values of features X and Y respectively and n is the number of data points.

Mutual Information (MI): MI quantifies the amount of information obtained about one feature through the observation of another. It is a measure of the mutual dependence between features. MI between features X and Y is calculated as:

$$I(X,Y)=\sum_{x\in X}\sum_{y\in Y}p(x,y)\log\frac{p(x,y)}{p(x)p(y)} \tag{21}$$

where $p(x,y)$ is the joint probability distribution of X and Y, $p(x)$ and $p(y)$ are the marginal probability distributions of X and Y, respectively.

Feature Importance (FI): FI evaluates the importance of features in predicting the target variable. Commonly, it employs methodologies like decision trees or ensemble techniques such as random forests to rank features according to their impact on predictive accuracy. The Gini Index is used to assess impurity in decision tree splits for a node t with K classes can be expressed as:

$$G(t)=1-\sum_{i=1}^{K}p(i)^2 \tag{22}$$

where $p(i)$ is the probability of randomly selecting a sample with class i in node t.

Setting a threshold T is an essential procedure in order to obtain a subset of candidate-relevant features after employing the ranking methods. Standard deviation (SD) is used as the threshold in this study which quantifies the disparity between measured data and the average value [45]. SD can be expressed as:

$$SD=\sqrt{\frac{\sum_{i=1}^{n}(x_i-\overline{x})^2}{n}} \tag{23}$$

where x_i is the individual measure value of each i, $\overline{x}$ is the average of measure values of all features and n is the total number of features.

Generation of an ensemble feature subset

An ensemble subset which refers to a collection of features selected from multiple subsets is generated using a union method. It involves merging the candidate feature subsets generated by individual ranker methods previously to create a single ensemble feature subset. The features selected at least one by the ranker method will be retained in the ensemble feature subset, while any duplicated features from the multiple ranker methods will be eliminated. The ensemble feature subset E from the combination of individual candidate feature subset $S_1, S_2, \ldots, S_N$ can be expressed as [46]:

$$E=S_1\cup S_2\cup\ldots\cup S_N \tag{24}$$

b. Fuzzy discretization on continuous features

Phase II occupies data discretization which involves the transformation of continuous-valued features into a series of discrete intervals. This methodology is crucial as it significantly influences the efficacy of predictors and classifiers [47], [48], [49]. Fuzzy discretization is implemented in this study through a four-step procedure as follows:

- Step 1: Identification of each continuous feature presents in the Wisconsin Dataset Breast Cancer (WDBC) dataset and the execution of fuzzy partitioning wherein each continuous feature's range is segmented into the specified number of intervals. This study uses three intervals which are 'Low', 'Medium', and 'High' for each continuous feature. Each interval represents a different magnitude level of the feature's values. Specifically, 'Low' corresponds to values below a certain threshold, 'Medium' is moderate values within a middle range, and 'High' indicates values above a certain threshold.
- Step 2: Generating fuzzy sets $A_1, A_2, \ldots, A_m$ where m denotes the number of intervals for each continuous feature in the dataset. Triangular membership functions are utilized to construct these fuzzy sets, with parameters (a, b, c) derived from the lower and upper bounds of the intervals. Specifically, parameters 'a' and 'c' correspond to the lower and upper boundaries of the fuzzy set respectively while parameter 'b' denotes the center of the fuzzy set. Prior to assigning ranges of features, outlier detection is performed to ensure the normal distribution of the data using the Z-Score method.
- Step 3: Assigning membership values to data points within each interval to define discrete boundaries.

- Step 4: Deriving the discrete attribute values by selecting the interval with the highest membership value for each data point. This iterative process is repeated for each continuous attribute within the dataset. This process results in the creation of a finalized dataset containing fuzzy discretized attribute values.

c. Association rule mining

Phase III emphasizes on association rule mining which is a process aimed at discovering patterns and relationships within the dataset. This involves extracting two sets of rules: the Classification Association Rule set (CARs) and the Atomic Association Rule set (AARs). These sets are generated using a widely used method for mining association rules which is the Apriori algorithm. The algorithm operates on dataset D while considering parameters such as minimum support ($minSup$), and minimum confidence ($minConf$) to control the quality and significance of the discovered rules.

In the FD-CARI framework, the $minSup$ threshold was set at 0.2 to ensure that only frequently occurring feature associations are considered. A lower $minSup$ value would generate more rules, but it could also introduce noise and making it harder to distinguish truly relevant patterns. On the other hand, a higher $minSup$ might eliminate useful associations and reducing the effectiveness of the feature selection process. After empirical analysis, we found that setting $minSup$ within the range of [0.2, 0.5] balances rule generation and computational efficiency. Similarly, the $minConf$ threshold was set at 0.7 to ensure that only strong and reliable association rules are retained. A higher $minConf$ value helps maintain strong dependencies among selected features while reducing the likelihood of weak or misleading correlations. However, setting $minConf$ too high could result in the loss of valuable feature interactions. Based on experimental evaluations, we determined that an optimal $minConf$ range of [0.7, 1.0] provides the best balance between rule quality and feature selection effectiveness.

d. Relevant interactive features discovery

Phase IV represents a critical juncture in the research framework. During this stage, minimum lift ($minlift$), minimum conviction ($minConv$), and minimum certainty factor ($minCertFac$) thresholds are applied to obtain the pruned CAR (pCAR). These thresholds help in ensuring that only the most significant interactions are retained in the RFs. Building upon this framework, this study evaluates the lift, conviction, and certainty factor levels of the associated rules. This evaluation helps determine whether combinations of feature value sets effectively describe the target concept.

e. Feature redundant elimination

Phase V involves the removal of redundant features from the Relevant Feature value set (RFs). Redundancy occurs when a feature value contains information already present in another. Inspired by [18] which utilized higher confidence to detect redundancy, this study adopted atomic association rules set ($AARs$) with higher certainty factor levels to signify stronger implications for identifying and eliminating redundant values.

f. Optimal feature subset identification

Final phase VI focuses on identifying the optimal feature subset S. RFs now comprise solely relevant and non-duplicative values. Phase III guarantees that RFs encompass all-important feature value interactions necessary for defining feature interactions (refer to Definition 6). The final subset S effectively preserves relevant features while discarding irrelevant and redundant ones and considering feature interaction thoroughly.

Classification Using Supervised Machine Learning Models

In this study, supervised machine learning methods were employed to classify breast cancer tumors. The utilized models include Decision Trees (DT), Random Forest (RF), Naïve Bayes (NB), Logistic Regression (LR), and Support Vector Machine (SVM).

Decision Trees (DT). DT represents a non-parametric supervised learning method utilized for both classification and regression tasks. Its hierarchical structure includes a root node, branches, internal nodes, and leaf nodes. Several approaches exist for choosing the optimal attribute at each node in a decision tree. One of them is Information Gain which is calculated using the entropy measure [50]. Entropy measures the impurity or uncertainty in a set of data. The formula for entropy $H(S)$ of a set S with K classes is given by:

$$H(S) = -\sum_{i=1}^{K} p_i \log_2(p_i) \tag{25}$$

where p_i is the proportion of samples in class i in set S.

Random Forest (RF). RF is a type of ensemble model that utilizes multiple decision trees for making predictions [51]. Each decision tree within the ensemble generates a prediction for the input data and

subsequently, RF combines these predictions and chooses the most frequently appearing outcome as the final prediction. RF promotes diversity among its constituent trees by selecting the best feature from a random subset of features thereby decreasing the likelihood of overfitting [52]. The prediction of RF can be represented as:

$$RF(x) = \text{mode}(\{f_1(x), f_2(x), \ldots, f_T(x)\}) \quad (26)$$

where T is the number of decision trees in the RF and $f_T(x)$ is the prediction of the T-th decision tree for the input data.

Naïve Bayes (NB). NB is a probabilistic machine learning algorithm that relies on Bayes' theorem with an assumption of independence between features. It is used to calculate the conditional probability of a class given a set of features [53]. Mathematically, Bayes' theorem is expressed as:

$$P(C|X) = \frac{P(X|C) \times P(C)}{P(X)} \quad (27)$$

where $P(C|X)$ is the posterior probability of class C given features X, $P(X|C)$ is the likelihood of observing features X given class C, $P(C)$ is the probability of class C, and $P(X)$ is the probability of observing features X.

Logistic Regression (LR). LR stands as a frequently employed method specifically designed for binary classification tasks. Its primary objective is to establish a clear decision boundary within the input feature space and effectively segregate data instances into positive and negative classes. The specialized activation function called the logistic response function [54] is expressed by Equation 3 with its respective parameters described in Equation 4.

$$h_\theta(x) = \frac{1}{1 + e^{-\theta^T X}} \quad (28)$$

$$\theta^T X = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \ldots + \theta_n x_n \quad (29)$$

where $h_\theta(x)$ is the predicted probability of the positive class while $\theta^T X$ is the weighted sum of input features.

Support Vector Machine (SVM). SVM is a powerful algorithm for binary classification tasks and is capable of handling linearly and non-linearly separable data by finding the optimal hyperplane in a high-dimensional space [55]. In its basic form, a Linear SVM establishes a linear decision boundary. The mathematical representation of the decision function for a linear SVM is:

$$f(x) = sign(w^T x + b) \quad (30)$$

where x denotes the input vector, w represents the weight vector perpendicular to the hyperplane, T is the transpose operation, b is the bias term, and the function sign determines the class label.

Performance Evaluation Measures

Evaluation measures are essential in assessing the effectiveness and proficiency of a model or system which offers quantitative insights into its performance in achieving objectives. Several evaluation metrics were employed in this study such as Accuracy (Acc), Sensitivity (Sen), Specificity (Spe), Precision (Pre) and F1-Score, and Area Under the Curve (AUC). The formula of these measures can be observed in Table 2 [56].

Table 2. Performance evaluation metrics

Evaluation Metric	Formula
Accuracy (Acc)	$\frac{TP + TN}{TP + TN + FN + FP}$
Sensitivity (Sen)	$\frac{TP}{TP + FN}$
Specificity (Spe)	$\frac{TN}{TN + FP}$
Precision (Pre)	$\frac{TP}{TP + FP}$
F1 Score	$2 \times \frac{Pre \times Sen}{Pre + Sen}$
True Positive Rate (TPR)	Sen
False Positive Rate (TPR)	$\frac{FP}{FP + TN}$

*TP (True Positive), TN (True Negative), FN (False Negative), FP (False Positive)

Experiment

In this section, we firstly designed some experiments to evaluate the performance of the FD-CARI algorithm by making a comparison with other representative feature selection algorithms and then reporting the empirical results.

Experimental Setup

A breast cancer dataset derived from the WDBC dataset obtained from the UC Irvine (UCI) Machine Learning Repository [57] was employed. This dataset encompasses a range of features computed from digitized Fine Needle Aspiration (FNA) images of breast masses which elucidate the cellular composition of breast cancer specimens. The diagnosis variable comprises 357 benign and 212 malignant samples with 31 features. Table 3 provides a list of these features along with their brief descriptions with their range values (rounded to 2 d.p.).

Table 3. List of features in the WDBC dataset

Feature Name.	Feature Description	Mean (μ)	Standard Error	Max
Radius	Average distance from the center to points on the circumference	6.98-28.11	0.11-2.87	7.93-36.04
Texture	Variations gray-scale levels	9.71-39.28	0.36-4.9	12.02-49.54
Perimeter	Length of the tumor boundary	43.79-188.5	0.76-21.86	50.41-251.2
Area	Area occupied by the tumor	143.5-2501	6.8-542.2	185.2-4254
Smoothness	Local variation in radius lengths	0.05-0.16	0-0.03	0.07-0.22
Compactness	Compactness of the tumor shape	0.02-0.35	0-0.1	0.02-1.06
Concavity	Severity of concave portions of the tumor contour	0-0.43	0-0.39	0-1.25
Concave points	Number of concave portions of the tumor contour	0-0.2	0-0.05	0.0-0.29
Symmetry	Symmetry of the cell nucleus	0.11-0.3	0-0.08	0.16-0.66
Fractal dimension	Complexity of the tumor boundary	0.05-0.1	0-0.03	0.06-0.21

The following experiments are conducted by comparing our method with other widely used feature selection algorithms including CFS, FCBF, Consistency, Relief-F, and mRMR. Among them, CFS, Consistency, FCBF, and Relief-F can be found in the WEKA environment [58]while Relief-F and mRMR are utilized using MATLAB R2023a software [59]. Among them, CFS, FCBF, and Consistency can select optimal feature subsets by their own stop criteria but mRMR and Relief-F select features based on a ranking procedure. The experiments were carried out using Python programming with Spyder version 5.4.3 on a PC running 64-bit Windows 11 Pro at a speed of 2.5 GHz with 32 GB of RAM. The experiments were conducted using holdout cross-validation where 2/3 of the data are used for training while 1/3 is used for validation. In order to ensure the results are not biased by the sample sequences, each dataset is tested over 10 trials with different random initializations [17]. Classification tasks utilized machine learning classifiers such as Decision Tree (DT), Random Forest (RF), Naive Bayes (NB), Logistic Regression (LR), and Support Vector Machine (SVM).

Results and Discussion

The evaluation of the proposed methodology is presented in this section.

Generating an Ensemble Candidate Feature Subset

In this phase, features were ranked based on their corresponding values derived from a standard deviation threshold. Table 4 presented the features selected using different ranking methods and a union combination of these methods. The features are ordered in descending values of their corresponding ranking criteria.

Table 4. List of selected features with the corresponding ranking methods

PCC	MI	FI	Union Combination
concave points_worst	perimeter_worst	area_worst	radius_mean
perimeter_worst	area_worst	concave points_worst	texture_mean
concave points_mean	radius_worst	concave points_mean	perimeter_mean
radius_worst	concave points_worst	radius_worst	area_mean
perimeter_mean	concave points_mean	perimeter_worst	smoothness_mean
area_worst	perimeter_mean	perimeter_mean	compactness_mean
radius_mean	concavity_mean	concavity_mean	concavity_mean
area_mean	radius_mean	area_mean	concave points_mean
concavity_mean	area_mean		symmetry_mean
concavity_worst	area_se		radius_se
compactness_mean	concavity_worst		perimeter_se
compactness_worst	perimeter_se		area_se
radius_se	radius_se		compactness_se
perimeter_se	compactness_worst		concave points_se
area_se	compactness_mean		radius_worst
texture_worst	concave points_se		texture_worst
smoothness_worst			perimeter_worst
symmetry_worst			area_worst
texture_mean			smoothness_worst
concave points_se			compactness_worst
smoothness_mean			concavity_worst
symmetry_mean			concave points_worst
fractal_dimension_worst			symmetry_worst
compactness_se			fractal_dimension_worst

Based on Table 3, features such as *'concave points_worst*', *'perimeter_worst*', *'radius_worst*', and *'area_worst'* consistently appeared across different ranking methods which indicated their significance in the dataset. In addition, it is crucial to highlight the irrelevant or weakly relevant features even though they are not strong predictors on their own, they may provide valuable interactions when combined with other features. From the table, features such as *'fractal_dimension_worst*', *'symmetry_mean'*, *'smoothness_mean'*, *'concave points_se*', and *'texture_mean'* appear less frequently and have lower rankings across different methods. These features are considered weakly relevant or irrelevant as they do not consistently contribute to the effectiveness of the predictive model across multiple ranking methods. However, despite their individual lack of significance, these features can still be valuable when used to find interactions between features through class association rule mining. The ensemble candidate feature subset consisted of 24 features out of the original 30 features excluding the target variable '*diagnosis*'.

Fuzzy Discretization of the Continuous Features

The process of fuzzy discretization transforms continuous features into a set of linguistic terms to facilitate better interpretability and improve performance in machine learning models. This phase includes the removal of outliers, generation of fuzzy sets using triangular membership functions, and assignment of membership values to data points within each interval to define discrete boundaries. For instance, consider the feature *'radius_mean'* in the dataset which originally ranged from 6.981 to 28.11 narrowed to 6.981 to 24.63 after removing outliers. Similarly, the original range of the '*texture_mean'* feature ranged from 9.71 to 39.28 reduced to 9.71 to 31.12 after the outliers' removal. This refinement ensures that the feature values are within a more reasonable and relevant range for analysis. The boxplots in Figure 4(a) and 4(b) illustrated the distribution of '*radius_mean'* and *'texture_mean'* before and after the removal of outliers, respectively.

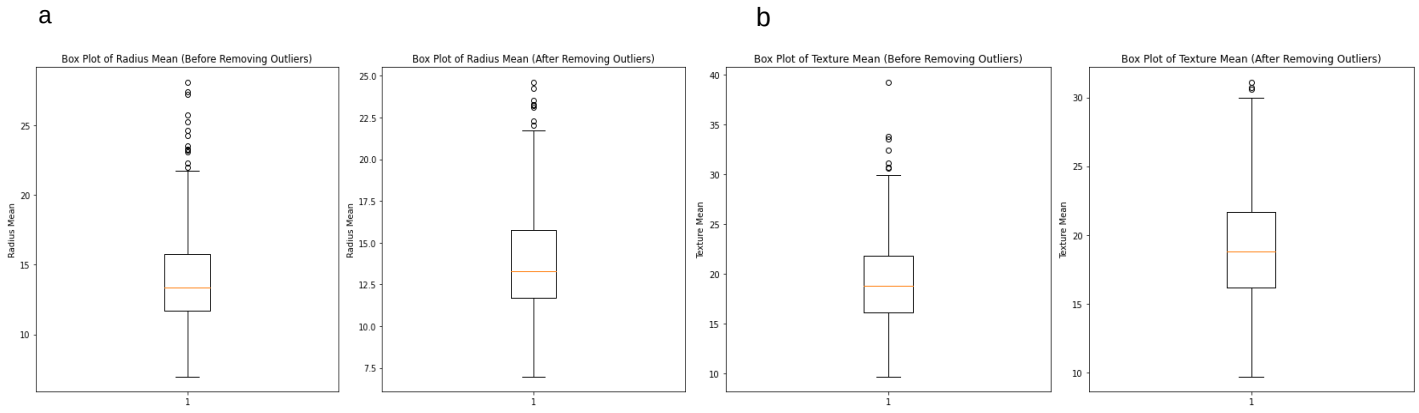

Figure 4. Box-plot of (a) *radius_mean* (b) *texture_mean* before and after removal of outliers

Once the outliers are removed, fuzzy sets for each interval of Low, Medium, and High are derived from the minimum and maximum values of each feature. These fuzzy sets are created using evenly distributed membership functions to ensure a smooth and continuous transition between different degrees of membership. For example, the fuzzy sets for the 'radius_mean' are defined as follows:

$$\begin{gathered}\text{Low}_{\text{radius\_mean}} \leq 14.33 \rightarrow \text{Low}_{\text{radius\_mean}} = \{(-0.37;0),(6.98;1),(14.33;0)\};\\ 8.45 \leq \text{Medium}_{\text{radius\_mean}} \leq 23.16 \rightarrow \text{Medium}_{\text{radius\_mean}} = \{(8.45;0),(15.81;1),(23.16;0)\};\\ \text{High}_{\text{radius\_mean}} \geq 17.28 \rightarrow \text{High}_{\text{radius\_mean}} = \{(17.28;0),(24.63;1),(31.98;0)\}.\end{gathered}$$

while the fuzzy sets for the '$texture\_mean$' are defined as follows:

$$\begin{gathered}\text{Low}_{\text{texture\_mean}} \leq 18.63 \rightarrow \text{Low}_{\text{texture\_mean}} = \{(0.78;0),(9.71;1),(18.63;0)\};\\ 11.49 \leq \text{Medium}_{\text{texture\_mean}} \leq 29.34 \rightarrow \text{Medium}_{\text{texture\_mean}} = \{(11.49;0),(20.42;1),(29.34;0)\};\\ \text{High}_{\text{texture\_mean}} \geq 22.2 \rightarrow \text{High}_{\text{texture\_mean}} = \{(22.2;0),(31.12;1),(40.04;0)\}.\end{gathered}$$

The triangular membership functions in Figure 5(a) and 5(b) depicted how the continuous values are transformed into fuzzy sets for these features.

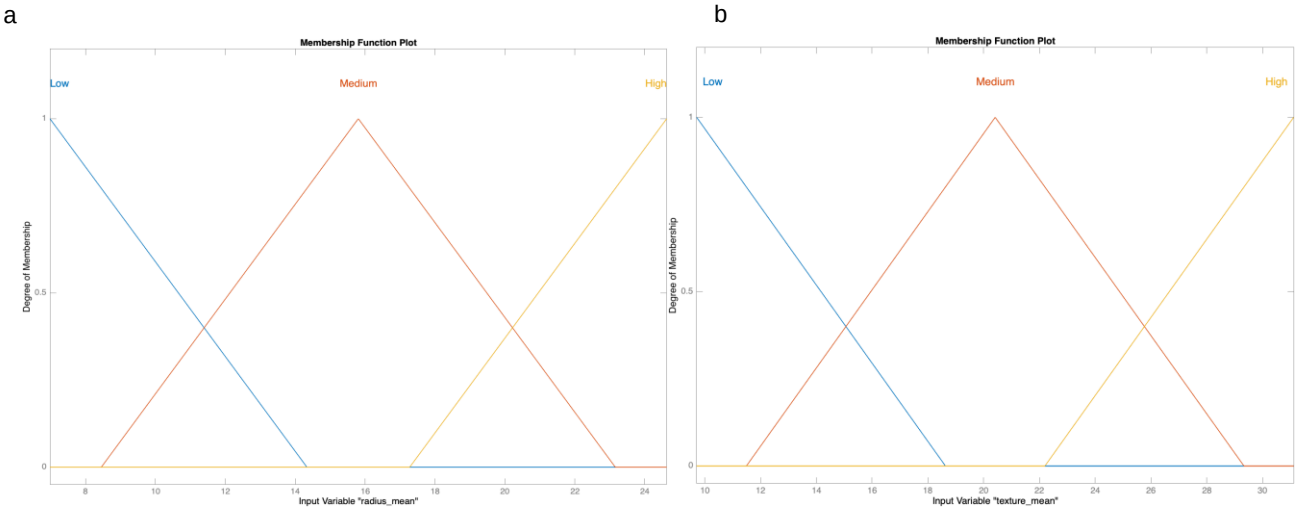

Figure 5. Triangular membership function of (a) *radius_mean* (b) *texture_mean*

Discovering Insightful Rules from the Association Rule Mining

The discovery of insightful rules from association rule mining is a pivotal step in understanding the underlying patterns and relationships within the dataset. In the process of performing association rule

mining, selecting the appropriate minimum support (*minsup*) threshold is crucial because it directly affects the number of frequent itemsets generated and the computational feasibility of the analysis. In this study, the number of association rules generated was analyzed based on different *minsup* (α) and *minconf* (*β*) thresholds. These thresholds were varied to assess their impact on the quantity and quality of the rules produced. Initially, setting the minsup to 0.15 resulted in the generation of 751,230 frequent itemsets, which was beyond the processing capacity of our system. This excessive number of frequent itemsets led to high memory usage and long processing times, making the analysis impractical on the available hardware. To address these challenges, we tested higher minsup of 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, and 0.5. By increasing the number of frequent itemsets was significantly reduced to 141,735, 33,439, 10,284, 3332, 1269, 513, and 219, respectively. This eased the computational burden and ensured that the analysis could be completed efficiently. For each minsup value, we also tested different minimum confidence (*minconf*) levels of 0.7, 0.8, 0.9, and 1.0. This helped in identifying strong association rules to provide valuable insights while keeping the computational requirements manageable. Table 5 presents the number of association rules (AR), class association rules (CAR), and atomic association rules (AAR) identified at various *minsup* and *minconf* combinations.

Table 5. Number of association rules based on certain minsup and minconf thresholds

α	*β*	AR	CAR	AAR
0.20	0.7	6,478,440	40,328	570
	0.8	3,622,430	38,349	313
	0.9	1,384,161	35,491	89
	1.0	178,666	8,866	8
0.25	0.7	922,022	9,078	521
	0.8	536,232	8,772	296
	0.9	214,887	8,071	78
	1.0	24,236	1,080	5
0.30	0.7	178,694	2,502	454
	0.8	106,204	2,417	268
	0.9	44,852	2,171	75
	1.0	4,601	120	5
0.35	0.7	36,832	692	358
	0.8	22,198	660	200
	0.9	9,598	550	60
	1.0	939	16	0
0.40	0.7	9,789	205	291
	0.8	6,042	192	162
	0.9	2,721	147	46
	1.0	239	0	1
0.45	0.7	2,702	68	225
	0.8	1,709	63	126
	0.9	780	44	39
	1.0	64	0	1
0.50	0.7	828	15	156
	0.8	522	13	89
	0.9	241	6	27
	1.0	16	0	1

Based on Table 4, at a *minsup* of 0.2 and *minconf* of 0.7, a substantial 6,478,440 *AR* were identified. This number drastically drops to 178,666 when the *minconf* is raised to 1.0. Similarly, the number of *CAR* decreases from 40,328 at a *minsup* of 0.2 and *minconf* of 0.7 to 8,866 at a *minconf* of 1.0. For *AAR*, the reduction is even more obvious from 570 at *minsup* of 0.2 and *minconf* of 0.7 to just 8 at *minconf* of 1.0. The pattern was consistent across different *minsup* and *minsconf* thresholds. These findings indicated that lower *minsup* and *minconf* values yielded a higher number of rules which may include many less significant ones, while higher thresholds produced fewer but potentially more meaningful and robust rules. Figures 6(a), 6(b), and 6(c) visually represent the number of *AR*, *CAR,* and *AAR* respectively across the different *minsup* and *minconf* thresholds. The graphs clearly illustrate the steep decline in the number of rules as the thresholds increase.

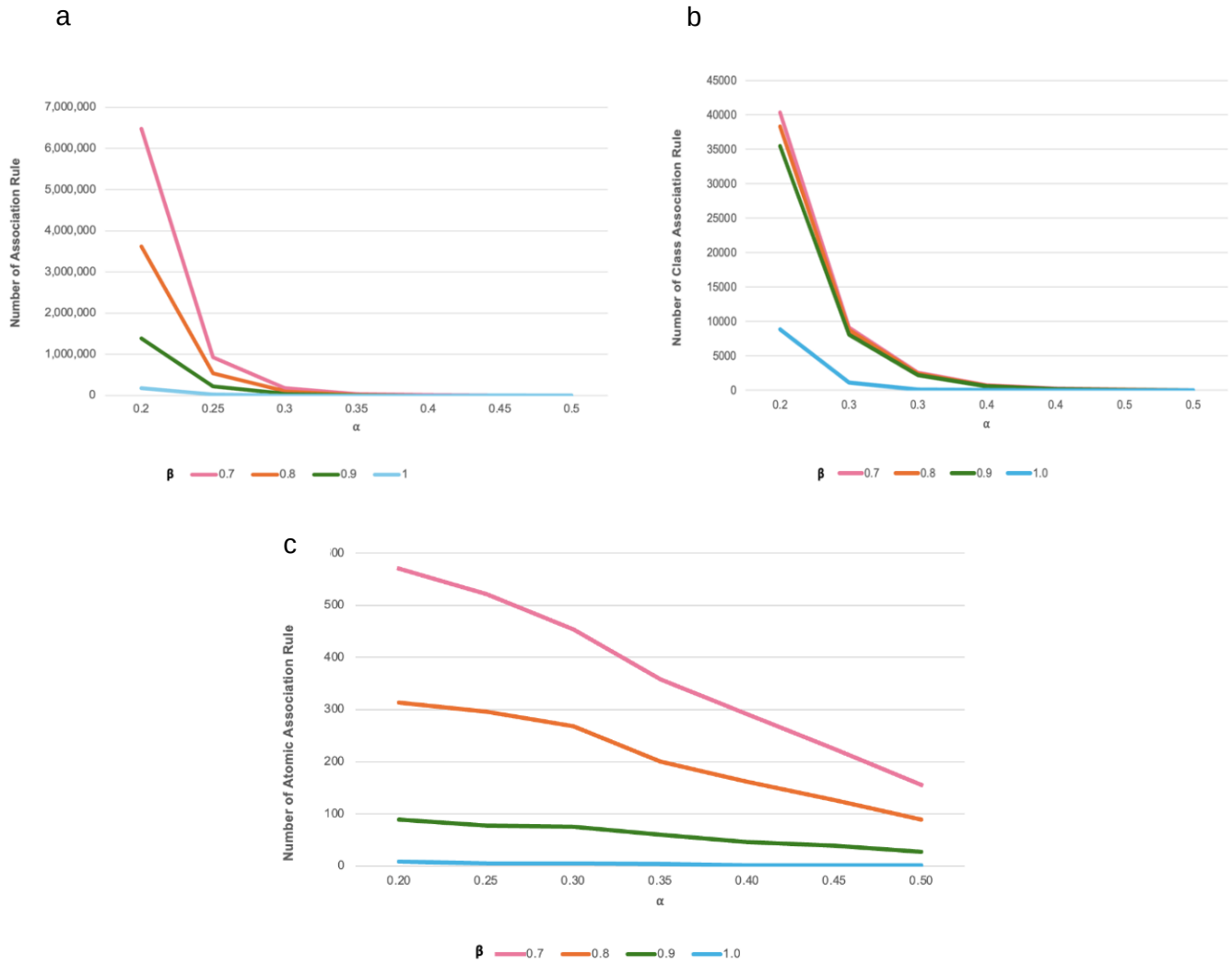

Figure 6. Number of (a) *AR* (b) *CAR* (c) *AAR* with different *minsupp* and *minconf* thresholds

Interactive features from class association rule mining

The process of identifying interactive features from class association rule mining provides insights into the relationships between different features and their influence on the target variable. Tables 5 and 6 provide an analysis of the identification of initial relevant features (RFs) using pruned class association rules (pCAR) and optimal RFs after the removal of redundant features using Atomic Association Rules (AAR).

Table 6. List of RFs corresponds to pCAR and AAR

α	*β*	*pCAR*	*AAR*	Before remove redundant	After remove redundant	
					No. of RF	List of RFs
0.20	0.7	1,785	264	22	2	['concave points_worst=Low', 'perimeter_mean=Medium']
	0.8	1,785	141	22	4	['texture_worst=Medium', 'concave points_worst=Low', 'radius_worst=Low', 'perimeter_mean=Medium']
	0.9	1,785	45	22	9	['symmetry_worst=Medium', 'radius_worst=Low', 'texture_worst=Medium', 'concave points_worst=Low', 'compactness_se=Low', 'radius_se=Low', 'compactness_mean=Low', 'smoothness_mean=Medium', 'perimeter_mean=Medium']
	1.0	486	5	18	14	['concavity_mean=Low', 'symmetry_worst=Medium', 'concavity_worst=Low', 'radius_worst=Low', 'perimeter_se=Low', 'perimeter_worst=Low', 'compactness_worst=Low', 'smoothness_worst=Medium', 'symmetry_mean=Medium', 'concave points_worst=Low', 'compactness_se=Low', 'radius_se=Low', 'compactness_mean=Low', 'smoothness_mean=Medium']
0.25	0.7	1,785	258	22	3	['concave points_worst=Low', 'radius_worst=Low', 'perimeter_mean=Medium']
	0.8	1,785	141	22	4	['texture_worst=Medium', 'concave points_worst=Low', 'radius_worst=Low', 'perimeter_mean=Medium']
	0.9	1,785	45	22	9	['symmetry_worst=Medium', 'radius_worst=Low', 'texture_worst=Medium', 'concave points_worst=Low', 'compactness_se=Low', 'radius_se=Low', 'compactness_mean=Low', 'smoothness_mean=Medium', 'perimeter_mean=Medium']
	1.0	486	5	18	14	['concavity_mean=Low', 'symmetry_worst=Medium', 'concavity_worst=Low', 'radius_worst=Low', 'perimeter_se=Low', 'perimeter_worst=Low', 'compactness_worst=Low', 'smoothness_worst=Medium', 'symmetry_mean=Medium', 'concave points_worst=Low', 'compactness_se=Low', 'radius_se=Low', 'compactness_mean=Low', 'smoothness_mean=Medium']
0.30	0.7	640	198	19	3	['radius_worst=Low', 'texture_worst=Medium', 'compactness_mean=Low']
	0.8	640	120	19	3	['radius_worst=Low', 'texture_worst=Medium', 'compactness_mean=Low']
	0.9	640	37	19	9	['symmetry_worst=Medium', 'concavity_worst=Low', 'radius_worst=Low', 'compactness_worst=Low', 'texture_worst=Medium', 'compactness_se=Low', 'radius_se=Low', 'compactness_mean=Low', 'smoothness_mean=Medium']

α	*β*	*pCAR*	*AAR*	Before remove redundant	After remove redundant	
					No. of RF	List of RFs
	1.0	120	2	14	12	['area_mean=Low', 'concavity_mean=Low', 'symmetry_worst=Medium', 'concavity_worst=Low', 'perimeter_se=Low', 'perimeter_worst=Low', 'compactness_worst=Low', 'smoothness_worst=Medium', 'symmetry_mean=Medium', 'concave points_mean=Low', 'compactness_se=Low', 'radius_se=Low']
0.35	0.7	98	90	12	2	['area_mean=Low', 'compactness_worst=Low']
	0.8	98	66	12	3	['area_mean=Low', 'symmetry_worst=Medium', 'compactness_worst=Low']
	0.9	98	18	12	7	['area_mean=Low', 'symmetry_worst=Medium', 'concavity_worst=Low', 'compactness_worst=Low', 'smoothness_worst=Medium', 'concave points_mean=Low', 'radius_se=Low']
	1.0	16	1	7	6	['area_worst=Low', 'concavity_mean=Low', 'concavity_worst=Low', 'concave points_mean=Low', 'perimeter_se=Low', 'radius_se=Low']
0.40	0.7	22	38	7	2	['concavity_worst=Low', 'radius_se=Low']
	0.8	22	29	7	2	['concavity_worst=Low', 'radius_se=Low']
	0.9	22	13	7	3	['concave points_mean=Low', 'concavity_worst=Low', 'radius_se=Low']
	1.0	0	0	0	0	[]
0.45	0.7	2	12	4	1	['concave points_mean=Low']
	0.8	2	9	4	1	['concave points_mean=Low']
	0.9	2	5	4	1	['concave points_mean=Low']
	1.0	0	0	0	0	[]
0.50	0.7	0	0	0	0	[]
	0.8	0	0	0	0	[]
	0.9	0	0	0	0	[]
	1.0	0	0	0	0	[]

Table 5 detailed the influence of varying α and β threshold parameters on the identification of RFs and removal of redundant features. The data show a clear pattern that the number of optimal RFs decreases significantly across different α values. However, increasing the β threshold alongside each α threshold resulted in an increase in the number of RFs. This happens because higher β values lead to fewer AARs and result in fewer redundant features. For instance, at α = 0.2 and β = 0.7, the number of RFs drops from 22 to 2 after redundancy removal while at α = 0.2 and β = 1.0, the number of RFs drops from 18 to 14. This is because the number of AAR reduced from 264 to 5 resulting in fewer redundant features. Similarly, at α = 0.35 and β = 0.7, the number of RFs reduced from 12 to 2 while at α = 0.35 and β = 1.0, the number of RFs reduced from 7 to 6 with the number of AAR of 90 to 1 only. These reductions indicate the effectiveness of the AAR in refining the feature set to ensure it retains only the most significant and non-redundant features without excessively removing redundant ones.

Table 7 lists the top 10 pCAR with their corresponding minimum confidence, lift, conviction, and certainty factor values. In order to ensure a balanced approach between capturing informative associations and removing noise, the pruning criteria were set at thresholds designed to filter out weaker associations while retaining potentially meaningful patterns. The thresholds such as minLift = 1.5, minConv = 50, and

minCertFac = 0.8 were chosen. These criteria allowed for the exclusion of noisy or less relevant rules while still permitting the inclusion of associations that include the weakly relevant features individually but may contribute significantly when combined with other features.

Table 7. 10 *pCAR* correspond to confidence, lift, conviction, and certainty factor values

No.	Antecedents	Consequents	α	β	Lift	Conviction	Certainty Factor
1	{concavity_worst=Low', 'area_mean=Low', 'perimeter_se=Low'}	{'diagnosis=B'}	0.30053	1.00000	1.59384	inf	1.0000
2	{'concavity_worst=Low', 'area_se=Low', 'area_mean=Low', 'perimeter_se=Low'}	{'diagnosis=B'}	0.30053	1.00000	1.59384	inf	1.0000
3	{'concavity_worst=Low', 'area_worst=Low', 'perimeter_se=Low', 'area_mean=Low'}	{'diagnosis=B'}	0.30053	1.00000	1.59384	inf	1.0000
4	{'concavity_mean=Low', 'area_mean=Low', 'perimeter_se=Low', 'concavity_worst=Low'}	{'diagnosis=B'}	0.30053	1.00000	1.59384	inf	1.0000
5	{'concavity_mean=Low', 'perimeter_worst=Low', 'perimeter_se=Low', 'area_mean=Low'}	{'diagnosis=B'}	0.30053	1.00000	1.59384	inf	1.0000
6	{'concavity_worst=Low', 'area_se=Low', 'area_mean=Low', 'perimeter_se=Low', 'area_worst=Low'}	{'diagnosis=B'}	0.30053	1.00000	1.59384	inf	1.0000
7	{'concavity_worst=Low', 'area_se=Low', 'area_mean=Low', 'concavity_mean=Low', 'perimeter_se=Low'}	{'diagnosis=B'}	0.30053	1.00000	1.59384	inf	1.0000
8	{'area_se=Low', 'area_mean=Low', 'concavity_mean=Low', 'perimeter_worst=Low', 'perimeter_se=Low'}	{'diagnosis=B'}	0.30053	1.00000	1.59384	inf	1.0000
9	{'concavity_worst=Low', 'area_mean=Low', 'concavity_mean=Low', 'perimeter_se=Low', 'area_worst=Low'}	{'diagnosis=B'}	0.30053	1.00000	1.59384	inf	1.0000
10	{'area_mean=Low', 'concavity_mean=Low', 'perimeter_worst=Low', 'perimeter_se=Low', 'area_worst=Low'}	{'diagnosis=B'}	0.30053	1.00000	1.59384	inf	1.0000

As pCAR represent associations between features and the diagnosis, they provide potential interactions among features that might be relevant for precise classification. For instance, the first rule indicates that when features like *'concavity_worst=Low'*, *'perimeter_se=Low'*, and *'area_mean=Low'* are present, there's a 100% confidence that the diagnosis will be 'B'. This suggests a strong association between these features and the diagnosis outcome. Interestingly, certain combinations feature seemingly irrelevant features like *'area_se=Low'* alongside more relevant ones in the second rule highlighting the potential for such seemingly insignificant features to enhance classification performance when paired with relevant ones. This potential is further supported by the absence of *'area_se'* in the FI ranking as well as its low ranking on measures like PCC and MI might initially suggest that this feature has limited individual predictive power or relevance to the target variable. However, its inclusion in some of the top pCAR in Table 5 alongside other features indicated that while *'area_se'* might not be informative on its own, it can still contribute to predictive performance when combined with other relevant features.

Comparison of the Number of Selected Features by Several Feature Selection Methods

A comparison of the number of selected features by several FS methods presents in this section. Table 8 provided the features selected by various FS methods based on their unique criteria.

Table 8. List of selected features of feature selection methods

CFS (11)	FCBF (7)	Consistency (8)	Relief-F (9)	mRMR (9)	FD-CARI (9)
texture_mean	*perimeter_worst*	*texture_mean*	*radius_worst*	*perimeter_worst*	*concavity_worst*
concavity_mean	*concave*	*radius_se*	*concave*	*fractal_dimension*	*radius_se*
concave	*points_worst*	*perimeter_se*	*points_worst*	*_se*	*radius_worst*
points_mean	*radius_se*	*radius_worst*	*perimeter_worst*	*texture_mean*	*smoothness_mean*
area_se	*texture_mean*	*texture_worst*	*texture_worst*	*symmetry_worst*	*compactness_mean*
symmetry_se	*smoothness_worst*	*concavity_worst*	*radius_mean*	*radius_se*	*compactness_worst*
radius_worst	*symmetry_worst*	*concave*	*perimeter_mean*	*concave*	*compactness_se*
perimeter_worst	*symmetry_se*	*points_worst*	*concave*	*points_worst*	*texture_worst*
area_worst		*symmetry_worst*	*points_mean*	*smoothness_worst*	*symmetry_worst*
smoothness_worst			*area_worst*	*concavity_mean*	
concavity_worst			*area_mean*	*area_mean*	
concave					
points_worst					

According to Table 8, FCBF and Consistency selected the fewest features with 7 and 8 respectively while CFS chose 11 features. Since Relief-F and mRMR employ a ranking procedure to select features, it was decided to select the same number of features as our proposed method FD-CARI. This approach ensures that each method is evaluated under the same conditions and with an equal number of features eliminating any potential bias due to different feature set sizes.

In order to obtain the optimal feature subset, the performance of each feature subset obtained through different *minsup* and *minconf* in Table 5 was evaluated. After rigorous testing, the subset identified with min p=0.3 and *minconf*=0.9 with 9 features consistently exhibited the highest performance compared to other feature subsets. The proposed method chose '*concavity_worst*', '*radius_se'*, *'radius_worst'*, '*smoothness_worst*', '*perimeter_se'*, '*concavity_mean'*, '*concave points_mean'*, '*area_mean'*, '*smoothness_mean'*, '*compactness_mean'*, '*compactness_worst'*, '*compactness_se', 'texture_worst'* and '*symmetry_mean'*. Notably, FD-CARI included *'compactness_mean',* '*compactness_worst'* and '*symmetry_worst'* which were absent in the selections of CFS, FCBF, Consistency, Relief-F, and mRMR. These features were initially unselected in FI and had low ranking in MI. Their inclusion in the optimal feature subset of FD-CARI suggests that each feature might have limited individual predictive power or relevance to the target variable. Yet, it can still contribute to classification performance when combined with other relevant features. Thus, this validated the capability of FD-CARI to select relevant features and remove redundant ones while considering feature interactions.

Comparison of Performance Evaluation of Feature Selection Methods with Supervised Machine Learning Classifiers

The evaluation of different feature selection methods alongside various supervised machine learning classifiers provides valuable insights into their effectiveness in enhancing model performance. Table 8 shows the performance of the features chosen by six FS algorithms including the average and standard deviation (following the ± symbol) across 10 trials. Bolded values indicate the highest among the six feature selection algorithms. Additionally, each performance evaluation measures that higher, equal, or lower value with respect to FD-CARI is represented by W/T/L (win/tie/loss) respectively.

Table 9. Comparison of performance evaluation of feature selection methods with machine learning classifiers

Evaluation measures	Feature selection methods	DT	RF	NB	LR	SVM	W/T/L
Accuracy	CFS	93.26 ± 1.41	95.63 ± 1.08	94.95 ± 0.59	94.79 ± 1.23	95.32 ± 1.28	1/0/5
	FCBF	93.68 ± 1.56	95.74 ± 0.89	**96.21 ± 0.07**	95.00 ± 1.53	95.26 ± 1.60	1/1/3
	Consistency	94.00 ± 1.67	96.21 ± 0.74	94.95 ± 1.00	95.16 ± 1.79	95.79 ± 1.05	2/0/3
	Relief-F	**94.63 ± 1.26**	**96.26 ± 1.36**	94.05 ± 1.49	95.16 ± 1.77	95.58 ± 1.20	3/0/2
	mRMR	93.37 ± 1.63	95.68 ± 1.56	95.47 ± 0.92	95.11 ± 1.41	94.84 ± 1.35	1/0/4
	FD-CARI	94.47 ± 2.44	95.74 ± 1.09	92.58 ± 1.57	**96.05 ± 1.6**	**96.21 ± 1.07**	
Sensitivity	CFS	90.72 ± 4.34	93.30 ± 2.03	91.88 ± 2.85	91.85 ± 2.52	92.48 ± 1.97	1/0/4
	FCBF	90.66 ± 3.64	92.88 ± 2.13	92.00 ± 2.34	91.55 ± 3.55	92.15 ± 3.31	1/0/4
	Consistency	92.58 ± 3.18	94.09 ± 2.46	**91.98 ± 3.05**	92.72 ± 3.50	94.16 ± 2.45	2/0/3
	Relief-F	92.78 ± 2.31	93.56 ± 1.36	88.82 ± 2.46	92.11 ± 2.75	92.44 ± 2.40	0/0/5
	mRMR	91.78 ± 2.76	**94.18 ± 2.46**	91.34 ± 2.47	92.60 ± 3.22	90.82 ± 3.46	2/0/3
	FD-CARI	**93.33 ± 4.01**	93.68 ± 2.37	90.11 ± 2.64	**93.82 ± 3.39**	**94.26 ± 2.33**	
Specificity	CFS	94.65 ± 1.86	96.97 ± 1.57	96.72 ± 1.95	96.47 ± 1.92	96.98 ± 1.91	2/0/3
	FCBF	95.38 ± 1.61	97.39 ± 1.55	**98.60 ± 1.18**	96.98 ± 1.77	97.05 ± 1.51	3/0/2
	Consistency	94.87 ± 2.98	97.48 ± 2.10	96.65 ± 1.60	96.65 ± 2.11	96.81 ± 1.73	2/0/3
	Relief-F	**95.67 ± 2.46**	**97.78 ± 1.71**	97.06 ± 2.24	96.91 ±2.15	97.4 ± 1.76	3/0/2
	mRMR	94.26 ± 1.96	96.55 ± 1.95	97.88 ± 1.54	96.58 ± 1.93	97.15 ± 1.86	1/0/4
	FD-CARI	95.20 ± 2.65	96.91 ± 1.74	94.00 ± 2.68	**97.38 ± 1.41**	**97.38 ± 1.26**	
Precision	CFS	90.57 ± 3.25	94.53 ± 3.04	94.17 ± 3.61	93.68 ± 3.68	94.5 ± 3.66	2/0/3
	FCBF	91.73 ± 3.00	95.22 ± 2.95	**97.44 ± 2.10**	94.43 ± 3.49	94.58 ± 2.97	3/0/2
	Consistency	91.09 ± 5.11	95.50 ± 3.76	93.88 ± 3.00	93.99 ± 4.07	94.35 ± 3.23	2/0/3
	Relief-F	**92.71 ± 3.95**	**95.99 ± 3.16**	94.49 ± 4.41	94.32 ± 4.30	95.18 ± 3.40	3/0/2
	mRMR	89.96 ± 3.9	93.92 ± 3.51	96.06 ± 3.02	93.78 ± 3.67	94.69 ± 3.50	1/0/4
	FD-CARI	91.52 ± 4.94	94.40 ± 3.36	89.56 ± 4.18	**95.24 ± 2.74**	**95.28 ± 2.43**	
F1-Score	CFS	90.55 ± 2.64	93.87 ± 1.66	92.91 ± 0.93	92.69 ± 1.92	93.43 ± 1.95	1/0/4
	FCBF	91.14 ± 2.50	93.99 ± 1.45	**94.60 ± 1.02**	92.90 ± 2.41	93.31 ± 2.43	2/0/3
	Consistency	91.70 ± 2.61	94.7 ± 1.12	92.85 ± 1.80	93.26 ± 2.58	94.19 ± 1.46	2/0/3
	Relief-F	**92.65 ± 1.44**	**94.74 ± 1.99**	91.49 ± 2.31	93.14 ± 2.85	93.74 ± 1.96	3/0/2
	mRMR	90.81 ± 2.67	94.00 ± 2.21	93.58 ± 1.34	93.11 ± 2.28	92.64 ± 2.22	2/0/3
	FD-CARI	92.34 ± 3.69	93.98 ± 1.82	89.75 ± 2.19	**94.48 ± 2.33**	**94.73 ± 1.51**	
AUC	CFS	92.68 ± 2.02	98.97 ± 0.53	98.98 ± 0.46	98.69 ± 0.63	99.00 ± 0.46	1/0/4
	FCBF	93.02 ± 1.93	99.33 ± 0.26	**99.33 ± 0.25**	98.41 ± 0.77	98.69 ± 0.66	2/0/3
	Consistency	93.72 ± 1.57	**99.37 ± 0.26**	99.01 ± 0.41	**99.14 ± 0.49**	**99.36 ± 0.35**	4/0/1
	Relief-F	94.23 ± 1.00	99.03 ± 0.65	98.58 ± 0.59	98.66 ± 0.66	98.84 ± 0.54	0/0/5
	mRMR	93.02 ± 1.73	98.93 ± 0.57	99.00 ± 0.41	98.84 ± 0.51	98.93 ± 0.48	1/0/4
	FD-CARI	**94.26 ± 2.66**	99.17 ± 0.41	97.56 ± 1.01	99.06 ± 0.53	99.29 ± 0.40	

Based on Table 9, there is a noticeable variation in classifier performance across feature selection methods. According to the evaluation, FD-CARI consistently achieved the highest average accuracy with LR (96.05%), and SVM (96.21%). FCBF also outperformed NB with 97.37% while Relief-F excelled on DT and RF with average accuracy of 94.63% and 96.26% respectively. Besides, high sensitivity is crucial in applications like medical diagnostics where detecting false negatives is crucial. FD-CARI obtained the highest average sensitivity with DT, LR and SVM while mRMR and Consistency acquired the highest performance on RF and NB respectively. In addition, high specificity is important to minimize false positives. FD-CARI achieved the highest specificity with LR and SVM (both 97.38%). For DT and RF, Relief-F achieved the highest specificity of 95.67% and 97.78% respectively while FCBF obtained highest specificity of 98.60%. Furthermore, high precision indicates that the classifier has a low false positive rate. FD-CARI again performed well, achieving the highest precision with LR and SVM with 95.24% and 95.28% respectively. This highlights FD-CARI ability to select features correctly for positive predictions. Even so, Relief-F had the highest precision (97.97%) for DT (92.71%) and RF (95.99%) while FCBF outperfomed on NB (97.44%). In the aspect of balancing precision and sensitivity, FD-CARI achieved the highest F1-scores of both LR and SVM with 94.48% and 94.73% respectively while again Relief-F had the outperformed on DT (92.65%) and RF (94.74%) while FCBF outperfomed for NB with 94.60%. In terms of the effectiveness of classifiers with superior discriminative ability, FD-CARI achieved the highest AUC values on RF (94.26%) while second highest AUC values on RF, LR, and SVM with 99.17%, 99.06% and 99.29% respectively, following Consistency as the highest. Despite this, FD-CARI still outperfomed other methods like CFS, FCBF, Relief-F, and mRMR for these classifiers. Yet, FCBF

achieved the highest AUC with 99.33% for NB. These results indicated that FD-CARI has a strong capability in feature selection.This can be attributed to three key factors which are its ability to capture feature interactions, preserve information through fuzzy discretization and optimize feature selection with CARM-based rule filtering. The performance evaluation measures of accuracy, sensitivity, specificity, precision, F1-score, and AUC for each feature selection method with several classifiers are visualized in Figures 7(a) through 7(f), respectively.

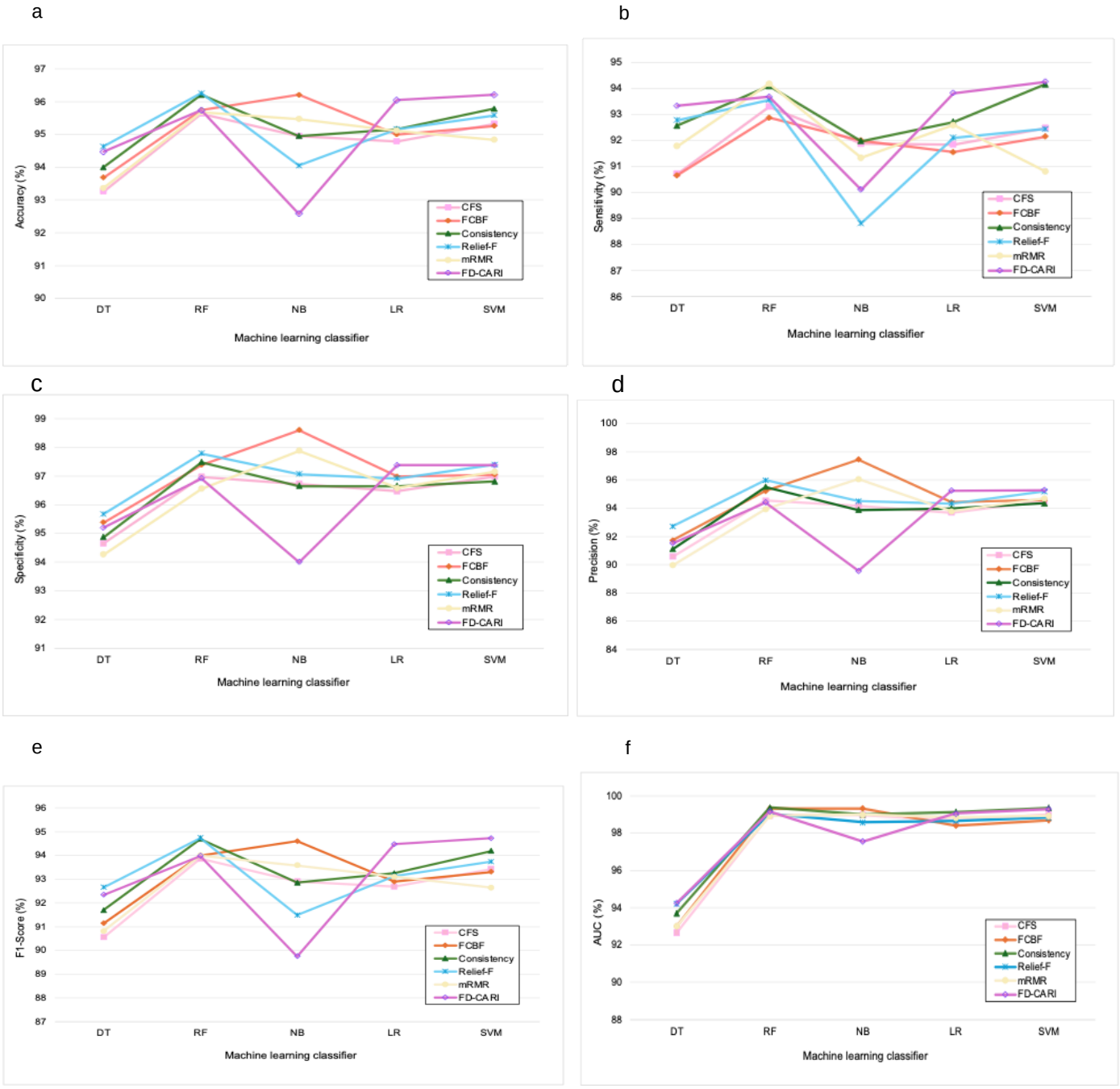

Figure 7. Classification (a) accuracy (b) sensitivity (c) specificity (d) precision (e) F1-score (f) AUC for each feature selection method

The AUC curve can be further visualized within ROC curves illustrated in Figures 8(a) through 8(f) for CFS, FCBF, Consistency, Relief-F, mRMR, and FD-CARI, respectively.

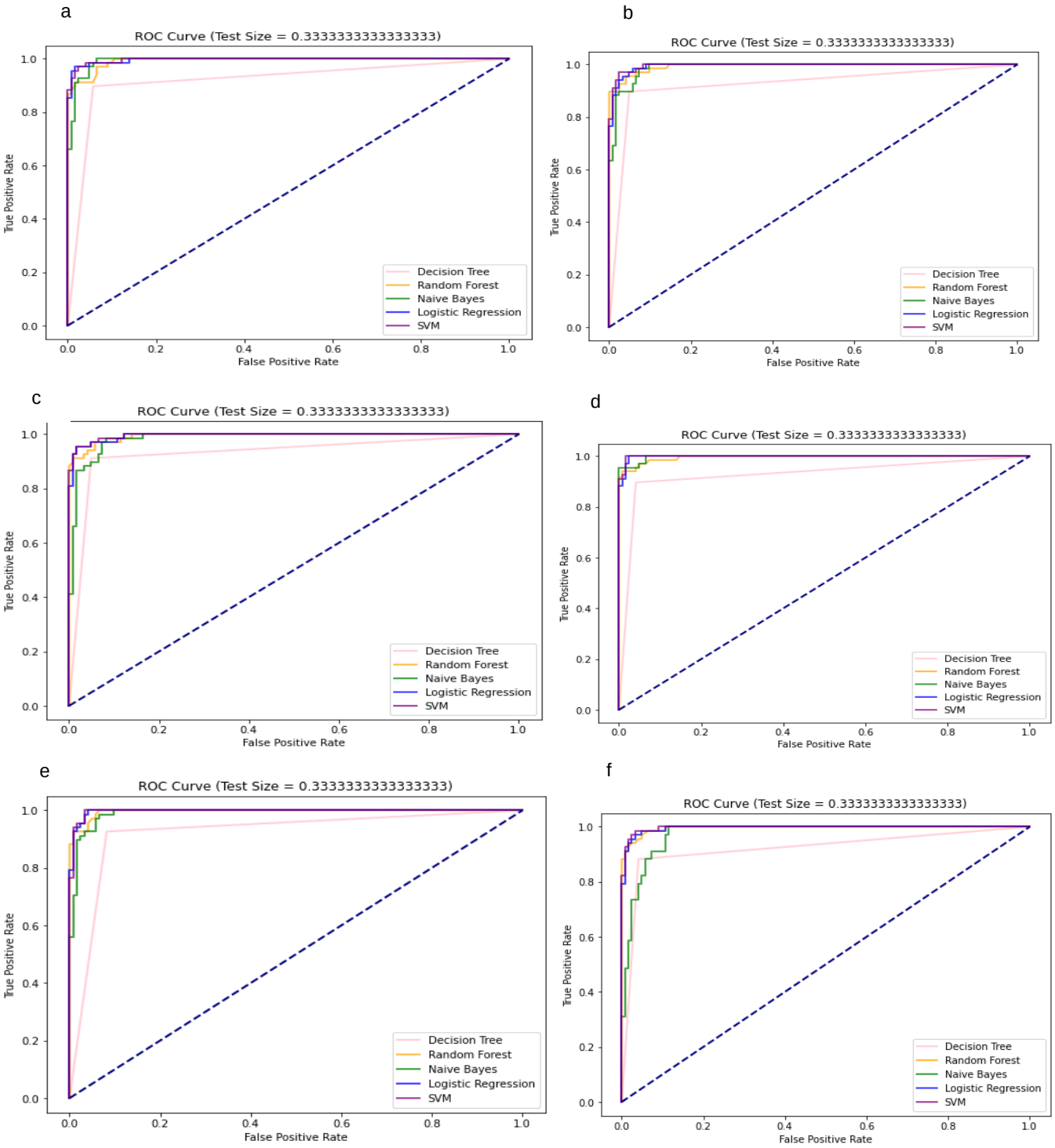

Figure 8. ROC curves for (a) CFS (b) FCBF (c) Consistency (d) Relief-F (e) mRMR (f) FD-CARI

From Figure 8(f), it can be seen that the ROC curves for FD-CARI and Relief-F with RF, LR and SVM classifiers very close to the top-left corner indicated that they had better overall performance compared to others. Besides, Relief-F also had the highest overall performance for NB while FCBF and Consistency had similar overall performance for DT.

These findings align with the performance trends observed in Table 10 which summarizes FD-CARI's accuracy across classifiers with the best-performing feature selection method in each case and explaining the factors influencing FD-CARI's performance.

Table 10. Summary of the performance of proposed method across classifiers

Classifier	Best FS Method	FD-CARI's Accuracy (%)	Performance Level	Reason for Performance
SVM	FD-CARI	96.21	High	Handles interaction-based feature selection effectively aligning well with SVM's margin-based classification.
LR	FD-CARI	96.05	High	Excels due to well-selected feature subsets benefiting LR's linear classification approach.
DT	Relief-F	94.47	Moderate	DT naturally manages feature redundancy, making Relief-F's ranking-based FS more effective.
RF	Relief-F	95.74	Moderate	RF internally handles feature dependencies reducing FD-CARI's advantage over other methods.
NB	FCBF	92.58	Lower	NB assumes feature independence making simpler correlation-based selection methods more effective than FD-CARI.

Table 10 shows that FD-CARI performs best with LR and SVM because these models benefit from well-selected, relevant, and interactive features. However, for DT and RF, Relief-F outperforms FD-CARI since these classifiers naturally manage feature redundancy on their own. Similarly, FCBF is more effective as NB assumes feature independence which making simpler correlation-based selection methods a better fit.

Conclusion and Future Work

Recent studies have focused on evaluating the interaction between features using information theory metrics like mutual information. Otherwise, the interaction between features can also be explored by uncovering interesting associations between them using association rule mining. Thus, this paper presented a novel feature selection method focusing on mining relevant and interactive features using pruned classification association rules (CAR) while removing redundant features using atomic association rules (AAR) to enhance the classification performance of machine learning models for breast cancer diagnosis. The proposed method first generated an ensemble candidate feature subset by applying several ranking methods to identify potentially relevant features. Then, fuzzy discretization was used to convert continuous features into meaningful categories. Subsequently, pruned CAR was applied using measures like lift, conviction, and certainty factors to extract strong rules and mine interactive relevant features. Finally, AAR was employed to eliminate redundant features and hence the optimal feature subset was obtained. A comparative analysis was conducted between FD-CARI with other well-known feature selection algorithms employing different supervised machine learning classifiers using the WDBC dataset. The experimental findings indicated that FD-CARI consistently outperformed other methods across LR and SVM classifiers. These findings validated its capability in selecting interactive relevant features while removing redundant features and enhancing the performance of classifier models.

While this study has confirmed the effectiveness of the FD-CARI method in feature selection, there are several areas for further exploration and refinement. Expanding the evaluation of FD-CARI to various datasets beyond the WDBC dataset would enable a comprehensive assessment of its applicability across diverse data domains. Besides, exploring hybrid feature selection techniques that integrate FD-CARI with other methods such as bio-inspired algorithms could enhance its ability to select relevant features and improve classifier performance. Furthermore, the scalability and computational efficiency of FD-CARI on large-scale datasets also should be examined since it is crucial for its practical use in big data scenarios

Conflicts of Interest

The author(s) declare(s) that there is no conflict of interest regarding the publication of this paper.

Acknowledgement

This work was supported by Universiti Kebangsaan Malaysia, through Grant TAP-K021917. High appreciation goes to above sponsor.

References

[1] Sung, H., *et al.* (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians, 71*(3), 209–249. https://doi.org/10.3322/caac.21660

[2] Siegel, R. L., Giaquinto, A. N., & Ahmedin, J. (2024). Cancer statistics, 2024. https://doi.org/10.3322/caac.21820

[3] Yadav, R. K., Singh, P., & Kashtriya, P. (2023). Diagnosis of breast cancer using machine learning techniques: A survey. *Procedia Computer Science, 218*, 1434–1443. https://doi.org/10.1016/j.procs.2023.01.122

[4] Oskouei, R. J., Kor, N. M., & Maleki, S. A. (2017). Data mining and medical world: Breast cancers' diagnosis, treatment, prognosis, and challenges. *American Journal of Cancer Research, 7*(3), 610–627.

[5] Khamparia, A., *et al.* (2021). Diagnosis of breast cancer based on modern mammography using hybrid transfer learning. *Multidimensional Systems and Signal Processing, 32*(2), 747–765. https://doi.org/10.1007/s11045-020-00756-7

[6] Kuhl, C. K., *et al.* (2005). Mammography, breast ultrasound, and magnetic resonance imaging for surveillance of women at high familial risk for breast cancer. *Journal of Clinical Oncology, 23*(33), 8469–8476. https://doi.org/10.1200/JCO.2004.00.4960

[7] Pulumati, A., Pulumati, A., Dwarakanath, B. S., Verma, A., & Papineni, R. V. L. (2023). Technological advancements in cancer diagnostics: Improvements and limitations. *Cancer Reports, 6*(2), 1–17. https://doi.org/10.1002/cnr2.1764

[8] Drukker, K., Sennett, C. A., & Giger, M. L. (2009). Automated method for improving system performance of computer-aided diagnosis in breast ultrasound. *IEEE Transactions on Medical Imaging, 28*(1), 122–128. https://doi.org/10.1109/TMI.2008.928178

[9] Li, J., *et al.* (2017). Feature selection: A data perspective. *ACM Computing Surveys, 50*(6). https://doi.org/10.1145/3136625

[10] Dash, M., & Liu, H. (1997). Feature selection for classification. *Intelligent Data Analysis, 1*(3), 131–156. https://doi.org/10.3233/IDA-1997-1302

[11] Vergara, J. R., & Estévez, P. A. (2014). A review of feature selection methods based on mutual information. *Neural Computing and Applications, 24*(1), 175–186. https://doi.org/10.1007/s00521-013-1368-0

[12] Jakulin, A., & Bratko, I. (2004). Testing the significance of attribute interactions. In *Proceedings of the Twenty-First International Conference on Machine Learning* (pp. 409–416). https://doi.org/10.1145/1015330.1015377

[13] Deisy, C., Baskar, S., Ramraj, N., Koori, J. S., & Jeevanandam, P. (2010). A novel information theoretic-interact algorithm (IT-IN) for feature selection using three machine learning algorithms. *Expert Systems with Applications, 37*(12), 7589–7597. https://doi.org/10.1016/j.eswa.2010.04.084

[14] Gu, X., Guo, J., Wei, H., & He, Y. (2020). Spatial-domain steganalytic feature selection based on three-way interaction information and KS test. *Soft Computing, 24*(1), 333–340. https://doi.org/10.1007/s00500-019-03910-x

[15] Zhao, Z., & Liu, H. (2007). Searching for interacting features. In *Proceedings of the International Joint Conference on Artificial Intelligence* (pp. 1156–1161).

[16] Zeng, Z., Zhang, H., Zhang, R., & Yin, C. (2015). A novel feature selection method considering feature interaction. *Pattern Recognition, 48*(8), 2656–2666. https://doi.org/10.1016/j.patcog.2015.02.025

[17] Wang, L., Jiang, S., & Jiang, S. (2021). A feature selection method via analysis of relevance, redundancy, and interaction. *Expert Systems with Applications, 183*, 115365. https://doi.org/10.1016/j.eswa.2021.115365

[18] Wang, G., & Song, Q. (2012). Selecting feature subset via constraint association rules. In *Lecture Notes in Computer Science: Vol. 7301. Advances in Knowledge Discovery and Data Mining* (pp. 304–321). https://doi.org/10.1007/978-3-642-30220-6_26

[19] Šikonja, M.-R., & Kononenko, I. (2003). Theoretical and empirical analysis of ReliefF and RReliefF. *Machine Learning, 53*, 23–69.

[20] Peng, H., Long, F., & Ding, C. (2005). Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 27*(8), 1226–1238. https://doi.org/10.1109/TPAMI.2005.159

[21] Wang, G., & Song, Q. (2013). A novel feature subset selection algorithm based on association rule mining. *Intelligent Data Analysis, 17*(5), 803–835. https://doi.org/10.3233/IDA-130608

[22] Kira, K., & Rendell, L. A. (1992). The feature selection problem: Traditional methods and a new algorithm. In *Proceedings of the National Conference on Artificial Intelligence* (pp. 129–134). AAAI Press. https://doi.org/10.1201/9781003402848-6

[23] Hall, M. A., & Smith, L. A. (1995). Feature selection for machine learning: Comparing a correlation-based filter approach to the wrapper. In *Proceedings of the Florida Artificial Intelligence Research Society Conference* (pp. 235–239).

[24] Yu, L., & Liu, H. (2003). Feature selection for high-dimensional data: A fast correlation-based filter solution. In *Proceedings of the Twentieth International Conference on Machine Learning* (pp. 856–863).

[25] Dash, M., & Liu, H. (2003). Consistency-based search in feature selection. *Artificial Intelligence, 151*(1–2), 155–176. https://doi.org/10.1016/S0004-3702(03)00079-1

[26] Das, S., & Nath, B. (2008). Dimensionality reduction using association rule mining. In *Proceedings of the IEEE*

Region 10 Colloquium and 3rd International Conference on Industrial and Information Systems (pp. 1–6). https://doi.org/10.1109/ICIINFS.2008.4798351

[27] Xie, J., Wu, J., & Qian, Q. (2009). Feature selection algorithm based on association rules mining method. In *Proceedings of the 8th IEEE/ACIS International Conference on Computer and Information Science* (pp. 357–362). https://doi.org/10.1109/ICIS.2009.103

[28] Karabatak, M., & Ince, M. C. (2009). A new feature selection method based on association rules for diagnosis of erythemato-squamous diseases. *Expert Systems with Applications, 36*(10), 12500–12505. https://doi.org/10.1016/j.eswa.2009.04.073

[29] Sheikhan, M., Rad, M. S., & Shirazi, H. M. (2011). Application of fuzzy association rules-based feature selection and fuzzy ARTMAP to intrusion detection. *Majlesi Journal of Electrical Engineering, 5*(4). https://doi.org/10.1109/ICSMC.2006.385078

[30] Waseem, S., Salman, A., & Muhammad, A. K. (2013). Feature subset selection using association rule mining and JRip classifier. *International Journal of Physical Sciences, 8*(18), 885–896. https://doi.org/10.5897/ijps2013.3842

[31] Wang, L., & Guo, H. (2014). Feature selection based on fuzzy clustering analysis and association rule mining for soft-sensor. In *Proceedings of the 33rd Chinese Control Conference* (pp. 5162–5166). https://doi.org/10.1109/ChiCC.2014.6895819

[32] Harikumar, S., Dilipkumar, D. U., & Kaimal, M. R. (2017). Efficient attribute selection strategies for association rule mining in high dimensional data. *International Journal of Computational Science and Engineering, 15*(3–4), 201–213. https://doi.org/10.1504/IJCSE.2017.087416

[33] Li, Y., *et al.* (2019). Association rule-based feature mining for automated fault diagnosis of rolling bearing. *Shock and Vibration, 2019*. https://doi.org/10.1155/2019/1518246

[34] Lin, Q., & Gao, C. (2023). Discovering categorical main and interaction effects based on association rule mining. *IEEE Transactions on Knowledge and Data Engineering, 35*(2), 1379–1390. https://doi.org/10.1109/TKDE.2021.3087343

[35] Liewlom, P. (2021). Class-association-rules pruning by the profitability-of-interestingness measure: Case study of an imbalanced class ratio in a breast cancer dataset. *Journal of Advances in Information Technology, 12*(3), 246–252. https://doi.org/10.12720/jait.12.3.246-252

[36] Farghaly, H. M., & Abd El-Hafeez, T. (2023). A high-quality feature selection method based on frequent and correlated items for text classification. *Soft Computing, 27*(16), 11259–11274. https://doi.org/10.1007/s00500-023-08587-x

[37] Zhang, L. (2021). A feature selection algorithm integrating maximum classification information and minimum interaction feature dependency information. *Computational Intelligence and Neuroscience, 2021*. https://doi.org/10.1155/2021/3569632

[38] Lv, Y., Lin, Y., Chen, X., Wang, C., & Li, S. (2021). Feature interaction based online streaming feature selection via buffer mechanism. *Concurrency and Computation: Practice and Experience, 33*(21), e6435. https://doi.org/10.1002/cpe.6435

[39] Liu, J., Yang, S., Zhang, H., Sun, Z., & Du, J. (2023). Online multi-label streaming feature selection based on label group correlation and feature interaction. *Entropy, 25*(7), Article 1071. https://doi.org/10.3390/e25071071

[40] Kianmehr, K., Alshalalfa, M., & Alhajj, R. (2010). Fuzzy clustering-based discretization for gene expression classification. *Knowledge and Information Systems, 24*(3), 441—465. https://doi.org/10.1007/s10115-009-0214-2

[41] Zadeh, L. A. (1965). Fuzzy sets. *Information and Control, 8*, 338–353.

[42] Agrawal, R., Imieliński, T., & Swami, A. (1993). Mining association rules between sets of items in large databases. *ACM SIGMOD Record, 22*(2), 207–216. https://doi.org/10.1145/170036.170072

[43] Akbas, K. E., *et al.* (2022). Assessment of association rule mining using interest measures on the gene data. *Medical Records, 4*(3), 286–292. https://doi.org/10.37990/medr.1088631

[44] Anusha, P. V., Anuradha, C., Chandra Murty, P. S. R., & Kiran, C. S. (2019). Detecting outliers in high dimensional data sets using Z-score methodology. *International Journal of Innovative Technology and Exploring Engineering, 9*(1), 48–53. https://doi.org/10.35940/ijitee.A3910.119119

[45] Prasetiyowati, M. I., Maulidevi, N. U., & Surendro, K. (2021). Determining threshold value on information gain feature selection to increase speed and prediction accuracy of random forest. *Journal of Big Data, 8*(1). https://doi.org/10.1186/s40537-021-00472-4

[46] Damtew, Y. G., Chen, H., & Yuan, Z. (2023). Heterogeneous ensemble feature selection for network intrusion detection system. *International Journal of Computational Intelligence Systems, 16*(1). https://doi.org/10.1007/s44196-022-00174-6

[47] Ishibuchi, H., Yamamoto, T., & Nakashima, T. (2001). Fuzzy data mining: Effect of fuzzy discretization. In *Proceedings of the IEEE International Conference on Data Mining* (pp. 241–248). https://doi.org/10.1109/icdm.2001.989525

[48] Kianmehr, K., Alshalalfa, M., & Alhajj, R. (2008). Effectiveness of fuzzy discretization for class association rule-based classification. In *Lecture Notes in Computer Science: Vol. 4994. Advances in Artificial Intelligence* (pp. 298–308). https://doi.org/10.1007/978-3-540-68123-6_33

[49] Shanmugapriya, M., Nehemiah, H. K., Bhuvaneswaran, R. S., Arputharaj, K., & Sweetlin, J. D. (2017). Fuzzy discretization based classification of medical data. *Research Journal of Applied Sciences, Engineering and Technology, 14*(8), 291–298. https://doi.org/10.19026/rjaset.14.4953

[50] Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning, 1*(1), 81–106. https://doi.org/10.1023/A:1022643204877

[51] Breiman, L. (2001). Random forests. *Machine Learning, 45*, 5–32.

[52] Han, S., & Kim, H. (2019). On the optimal size of candidate feature set in random forest. *Applied Sciences, 9*(5). https://doi.org/10.3390/app9050898

[53] Shobha, G., & Rangaswamy, S. (2018). Machine learning. In *Handbook of Statistics: Vol. 38. Advances in Machine Learning and Data Science* (1st ed.). Elsevier B.V. https://doi.org/10.1016/bs.host.2018.07.004

[54] Berkson, J. (1944). Application of the logistic function to bio-assay. *Journal of the American Statistical Association, 39*(227), 357–365.
[55] Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 297*(20), 273–297.
[56] Tharwat, A. (2018). Classification assessment methods. *Applied Computing and Informatics, 17*(1), 168–192. https://doi.org/10.1016/j.aci.2018.08.003
[57] Wolberg, W., Mangasarian, O., Street, N., & Street, W. (1995). Breast Cancer Wisconsin (Diagnostic). *UCI Machine Learning Repository*. https://doi.org/10.24432/C5DW2B
[58] Garner, S. R. (1995). WEKA: The Waikato Environment for Knowledge Analysis. In *Proceedings of the New Zealand Computer Science Research Students Conference* (pp. 57–64).
[59] The MathWorks. (2023). MATLAB (Version 9.14.0 R2023a) [Software]. Natick, MA: The MathWorks Inc.