Research Square

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

# R⁵-FRQR: Robust and Interaction-Aware Feature Selection Using Fuzzy Rough Sets

**Simon Peter Khabusi**

simonkhabusi@gmail.com

Uganda Christian University

**Godfrey Odongtto**

Busitema university

**Gilbert Gilibrays Ocen**

Busitema university

**Article**

**Keywords:** Fuzzy rough sets, feature selection, interaction-aware learning, stability selection, domain adaptation

**Posted Date:** March 27th, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-9145661/v1

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

**Additional Declarations:** No competing interests reported.

# $R^5$-FRQR: Robust and Interaction-Aware Feature Selection Using Fuzzy Rough Sets

Simon Peter Khabusi[1,2,*], Godfrey Odongtto[3], Gilbert Gilibrays Ocen[3]
[1]Department of Computing and Technology
Uganda Christian University
P.O Box 4, Mukono, Uganda
[2]Department of Computer Science and Engineering
Universal Technology and Management University
P.O Box 73307, Kampala, Uganda
[3]Department of Computer Engineering and Informatics
Busitema university
P.O Box 236, Tororo, Uganda
Email: simonkhabusi@gmail.com; godongtoo.eng@busitema.ac.ug; gocen@busitema.ac.ug
*Corresponding Author

***Abstract*—Traditional fuzzy rough quick reduct (FRQR) feature selection methods often rely on greedy, marginal relevance evaluation, which overlooks feature interactions. This leads to instability during resampling and consequently performs poorly under domain shifts. To overcome these limitations, this paper introduces $R^5$-FRQR, a fuzzy rough set-based feature selection framework that integrates five properties: reliability, relational awareness, robustness, reductiveness, and resilience ($R^5$). $R^5$-FRQR incorporates an interaction-aware selection mechanism that accounts for both feature redundancy and pairwise contribution to dependency scoring. A granular similarity approximation strategy is introduced to reduce computational cost while preserving selection quality. Moreover, feature stability is reinforced through bootstrap-based frequency scoring. A domain-adaptive dependency metric is proposed to maintain discriminative power across distribution shifts. Experimental results on 16 benchmark datasets demonstrate that $R^5$-FRQR consistently achieves higher performance across evaluation metrics compared to classical and state-of-the-art feature selection methods. The framework provides a scalable and interpretable solution for robust feature selection in high-dimensional, noisy, and dynamic data environments.**

***Index Terms*—Fuzzy rough sets, feature selection, interaction-aware learning, stability selection, domain adaptation.**

## I. INTRODUCTION

Feature selection is a crucial preprocessing step in machine learning especially when dealing with high-dimensional datasets where redundant, irrelevant and derogatory attributes can significantly degrade classification performance and increase the computational load (Zhang et al., 2023; Huang et al., 2024). An effective feature selection method must prioritize the selection of relevant features, eliminate detrimental and indifferent features, and carefully manage redundancy (Chakraborty & Pal, 2015). Traditional feature selection techniques such as filter (Hancer et al., 2024; Hu et al., 2023), wrapper (Xu et al., 2024), and embedded methods (Banerjee et al., 2024) have been extensively studied for identifying optimal subsets that improve model interpretability and reduce computational complexity. Among these, fuzzy rough set theory has emerged as a powerful approach due to its inherent ability to manage uncertainty and vagueness prevalent in real-world data. The integration of fuzzy set theory with rough set theory provides a robust foundation for modelling imprecise, noisy, or ambiguous information, thereby making the framework highly suitable for complex pattern recognition tasks (Zhang et al., 2023; Zabihimayvan & Doran, 2019; Yang et al., 2023).

Recent advancements in fuzzy rough set-based feature selection have aimed at enhancing both effectiveness and efficiency. For example, Xu et al. (2026) proposed a margin-aware fuzzy rough framework that considers both class compactness and separation, thereby reducing uncertainty in pattern classification. Gu et al. (2015) introduced a kernelized fuzzy rough set technique coupled with a memetic algorithm for transient stability

assessment in power systems, showing improved classification outcomes. Farahbakhshian and Ahvanooey (2020) applied fuzzy rough sets to gene selection for tumor classification, addressing the dimensionality challenges inherent in microarray gene expression data. Zabihimayvan and Doran (2019) applied fuzzy rough set-based feature selection to phishing attack detection, achieving a high F-score of 95% with a random forest classifier and demonstrating robustness to zero-day attacks. Building on this, Jensen and Shen (2004) introduced a semantics-preserving dimensionality reduction technique that retains structural information while improving analytical throughput. Similarly, Wang et al. (2017) utilized weighted fuzzy rough sets to reduce redundancy and clarify decision boundaries, thereby improving classification accuracy.

In other domains, Anaraki et al. (2018) combined the binary shuffled frog-leaping algorithm with fuzzy rough dependency metrics, while Xue et al. (2017) presented a method tailored for large-scale hierarchical classification tasks. Yang et al. (2023) introduced a significance-based weight assignment scheme using weighted fuzzy rough sets, and Zhang et al. (2023) proposed a localized density-based model that effectively captures underlying data structure. Other studies have explored instance selection and distributed optimization (Huang et al., 2023), self-adaptive weighting for monotonic classification (Sang et al., 2022), and hybrid selection-extraction strategies (Maji & Garai, 2013), further broadening the applicability of fuzzy rough sets.

In the application domain of remote sensing, Yin et al. (2018) enhanced segmentation accuracy by combining fuzzy region competition with Gaussian mixture models (GMMs), while Jebarani et al. (2021) integrated k-means with GMMs to improve clustering precision in semi-structured datasets. Other studies have used expectation-maximization with fuzzy logic (Uykan, 2023) and fuzzy machine learning frameworks (Lu et al., 2024) to enhance classification and feature selection performance in real-world applications.

To address multimodal and large-scale data challenges, Hu et al. (2018) proposed multi-kernel fuzzy rough sets. Evolutionary computation techniques have also been adopted to improve the global search capacity during feature selection (Xue et al., 2016). In the context of unsupervised learning, fuzzy mutual information-based methods have proven effective in evaluating feature relevance by quantifying mutual dependence with cluster assignments, thus enhancing clustering quality in the absence of labels (Antonelli et al., 2013). Naz et al. (2024) improved the clustering robustness with a fuzzy local-information k-means algorithm, while Rong et al. (2019) tackled big data scalability using hybrid models that balanced accuracy and computational demands.

Reduct computation within fuzzy rough sets has also improved. For example, a discernibility matrix-based method for identifying minimal reducts was proposed by Chen et al. (2012), while Wang et al. (2017) improved fuzzy decisions through fuzzy neighbourhood-based approximations. To preserve dataset semantics in noisy environments, Parthaláin et al. (2020) introduced fuzzy rough bireducts, which reduced both the dimensionality and the dataset size. Despite the theoretical strengths of these methods, they face several practical limitations. First, most existing FRQR-based methods evaluate features primarily using marginal relevance, implicitly assuming feature independence and thereby overlooking interaction effects that may reveal complementary or redundant relationships among attributes. Second, the greedy selection process can be unstable across different data samples, leading to inconsistent feature subsets under resampling. Third, the computation of fuzzy similarity relations requires quadratic pairwise comparisons $O(n^2)$, making traditional approaches computationally expensive for large datasets. Finally, many existing methods focus only on a single data distribution and do not consider domain shifts, which can significantly degrade the generalization ability of selected features across related datasets.

To address the limitations outlined above, we introduce R$^5$-FRQR: a framework based on fuzzy rough set theory and designed to identify a comprehensive set of feature subsets that are compact, stable, and generalizable in terms of maximizing class separability while remaining robust to noise and adaptable across data domains. R$^5$-FRQR integrates five underlying principles: reliability, relational awareness, robustness, reductiveness, and resilience. R$^5$-FRQR not only expands on the FRQR methodology for selecting features using the standard fuzzy dependency based greedy selection process but also introduces an interaction aware scoring strategy that determines both individual and pairwise scores to highlight which features exhibit complementarity (positive interaction) and which exhibit redundancy. R$^5$-FRQR employs a granular similarity approximation technique using clustering to provide a computationally efficient means for approximating the normally quadratic computations of feature similarity. Features are considered stable if they are consistently selected during the bootstrap sampling procedure from the original dataset. R$^5$-FRQR achieves robustness through the exclusion from the feature pool of features that are highly sensitive to label noise. R$^5$-FRQR also enables cross-domain resilience by establishing an adaptation score based on the mixed features from both source and target distributions to improve the confidence of feature selection in the presence of distributional change.

The main contributions of this work are as follows:

1) We propose R⁵-FRQR, a fuzzy rough set-based feature selection framework that integrates five complementary mechanisms: reliability, relational awareness, robustness, reductiveness, and resilience, to improve the stability, efficiency, and cross-domain generalization of selected feature subsets.
2) We introduce a novel interaction function that quantifies feature interdependence and redundancy based on deviations from additive fuzzy dependency, enabling the selection of features that provide complementary and non-redundant information.
3) To improve scalability, we introduce a clustering-based similarity approximation approach that reduces the computational cost of similarity evaluation from quadratic pairwise comparisons $O(n^2)$to cluster-level computations $O(c^2)$, while preserving the local structure of the data.
4) A resampling-driven mechanism is proposed that estimates the empirical stability of features across multiple data subsets, enhancing the robustness and reproducibility of the final selection.
5) The method incorporates a new domain-resilient scoring mechanism that evaluates feature utility across both source and target domains, enabling effective generalization under distributional shift.

The remainder of this paper is organized as follows. Section II presents the background for the proposed method. Section III details the proposed method including its mathematical formulation and implementation. Section IV provides experimental details. Section V presents the results and discussions. Section VI provides the conclusion and future research directions.

## II. Preliminaries

### *A. Problem Formulation*

Let $U = \{x_1, x_2, \dots, x_n\}$ denote a finite nonempty universe of instances, where each $x_i \in \mathbb{R}^m$ is described by a vector of real-valued attributes. The complete set of input features is denoted by $A = \{a_1, a_2, \dots, a_m\}$, and $Z = \{z_1, z_2, \dots, z_c\}$ represents the finite set of decision classes. Each instance $x_i \in U$ is associated with a decision label $Z(x_i) \in Z$.

Let $S \subseteq A$ denote a candidate subset of features to be evaluated. For any attribute $a \in A$, the value of attribute $a$ for instance $x_i$ is written as $x_i^a \in \mathbb{R}$. The feature selection task aims to find a compact subset $S^* \subseteq A$ such that the classification ability induced by $S^*$ approximates or maximizes that of the full feature set $A$, subject to interpretability and robustness constraints.

To handle uncertainty in decision labels, we consider a fuzzy modelling approach where each decision class $z \in Z$ is associated with a fuzzy set $\tilde{Z}_z \subseteq U$. A fuzzy membership function $\mu_z: U \to [0, 1]$ assigns to each instance $x \in U$ a degree $\mu_z(x)$ indicating its compatibility with class $z$. The values $\mu_z(x)$ are typically computed via probabilistic models such as kernel density estimation or Gaussian mixture models, satisfying the normalization condition $\sum_{z \in Z} \mu_z(x) = 1$ for all $x \in U$.

### *B. Fuzzy Rough Set Model*

The fuzzy rough set model combines the fuzzy theory of graded memberships with the rough set theory of approximations based on indiscernibility or similarity (Wang et al., 2024; Wang et al., 2025). A crucial component is the definition of a fuzzy similarity relation between instances with respect to selected attributes. For a single attribute $a \in A$, the fuzzy similarity between two instances $x_i$ and $x_j$ is given by the Gaussian kernel:

$$\ell_a(x_i, x_j) = \exp\left(-\frac{\left(a(x_i) - a(x_j)\right)^2}{2\sigma_a^2}\right) \tag{1}$$

where $\sigma_a > 0$ is a bandwidth parameter controlling the sensitivity of similarity decay for attribute $a$. For a subset of attributes $K \subseteq A$, the aggregated fuzzy similarity relation is defined as the average:

$$\ell_K(x_i, x_j) = \frac{1}{|K|} \sum_{a \in K} \ell_a(x_i, x_j), \tag{2}$$

yielding a symmetric, normalized similarity measure in the interval [0,1]. Given a fuzzy similarity relation $\ell_K$ and fuzzy decision membership function $\mu_z$, the fuzzy lower and upper approximations of a decision class $z \in Z$ are defined as follows:

$$\underline{K}Z(x) = \inf_{y \in U} \max(1 - \ell_K(x, y), \mu_Z(y)) \tag{3}$$

$$\overline{K}Z(x) = \sup_{y \in U} \min(\ell_K(x, y), \mu_Z(y)) \tag{4}$$

for each instance $x \in U$. The lower approximation quantifies the degree to which $x$ can be certainly classified into class $z$, while the upper approximation measures the possibility that $x$ may belong to $z$. The fuzzy positive region induced by feature subset $K$ is defined as:

$$POS_K(Z)(x) = \underline{KZ}(x) \tag{5}$$

which maps each instance to a degree of certainty in class assignment. The global quality of the feature subset $K$ is quantified using the fuzzy dependency degree:

$$\gamma_K(Z) = \frac{1}{|U|}\sum_{x\in U} POS_K(Z)(x) \tag{6}$$

This measure reflects the average confidence with which instances in the universe can be assigned to their respective classes using the selected features. A higher value of $\gamma_K(Z)$ implies stronger explanatory power of the feature subset.

*C. Greedy Forward Selection with Fuzzy Dependency Degree*

The fuzzy rough quick reduct (FRQR) algorithm forms the foundational baseline for our proposed framework. It operates under a greedy forward selection strategy, where features are selected iteratively based on their marginal contribution to the fuzzy dependency degree (FDD), denoted by $\gamma_K$. At each iteration, the candidate feature $a \in A\backslash S$ is evaluated relative to the current subset $S \subseteq A$ using the marginal gain:

$$\Delta\gamma(a|S) = \gamma_{S\cup\{a\}} - \gamma_S \tag{7}$$

The feature that yields the highest marginal gain is selected:

$$a^* = \arg\max_{a\in A\backslash S} \Delta\gamma(a|S) \tag{8}$$

The marginal gain $\Delta\gamma$ reflects the additional explanatory power a feature contributes when added to the current subset.

## III. The Proposed R⁵-FRQR Feature Selection Method

The objective of R⁵-FRQR is to identify a minimal, stable, and generalizable subset of features that maximizes discriminative power while ensuring computational efficiency and resilience to noise and domain shifts. The method integrates five complementary components: reliability, relational awareness, robustness, reductiveness, and resilience into a unified fuzzy rough set-based framework. Each of these components is detailed in the subsequent subsections. Figure 1 illustrates the complete feature selection process. The critical aspects of this framework include measuring inter-selection compatibility (interaction-aware scoring), estimating similarity (granular similarity approximation), evaluating consistency (bootstrap-based stability aggregation) and developing Robustness metrics, and utilizing a domain-specific methodology (adaptivity-dependent validation) for the establishment of a compact and generalizable feature set.

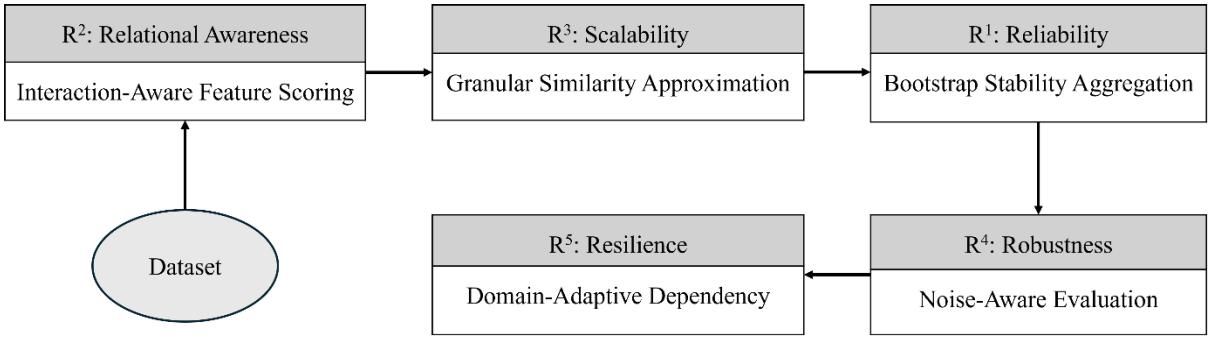



**Fig. 1.** Workflow of the proposed R⁵-FRQR feature selection framework.

*A. Interaction-Aware Selection: Relational Reduct Enhancement*

The FRQR method adopts a greedy forward selection strategy that evaluates candidate features solely based on their individual contribution to the fuzzy dependency degree (Dai et al., 2018). It implicitly assumes that features contribute independently and additively. However, in many real-world scenarios, such assumptions do not hold. Certain features may be uninformative in isolation but become highly predictive when used in conjunction with others. Similarly, some features may appear individually useful yet are redundant when jointly included. We therefore introduce a pairwise interaction mechanism that captures the relational influence between features. Given two features $a, b \in A$, the interaction score $\wp(a, b)$ is defined by measuring the deviation of their joint contribution from the average of their individual contributions:

$$\wp(a,b) = \gamma_{\{a,b\}}(Z) - \frac{1}{2}\left[\gamma_{\{a\}}(Z) + \gamma_{\{b\}}(Z)\right] \tag{9}$$

This score quantifies the extent of constructive interaction or redundancy. A positive interaction score, that is $\wp(a,b) > 0$, indicates the features jointly provide more discriminative power than expected from their individual effects, suggesting complementarity. A score of zero ($\wp(a,b) = 0$) implies independent, additive behavior, while a negative score ($\wp(a,b) < 0$) reflects redundancy, where the combination offers little additional benefit over either feature alone.

To provide an intuitive interpretation of the interaction score, Fig. 2 illustrates the relationship between the expected additive contribution of two features and their observed joint contribution. The horizontal axis represents the expected additive dependency $\frac{1}{2}\left[\gamma_{\{a\}}(Z) + \gamma_{\{b\}}(Z)\right]$, while the vertical axis represents the joint dependency $\gamma_{\{a,b\}}(Z)$. The diagonal line corresponds to additive independence. Points above this boundary indicate complementary interactions where the joint contribution exceeds the expected additive contribution, whereas points below the boundary indicate redundancy between features.

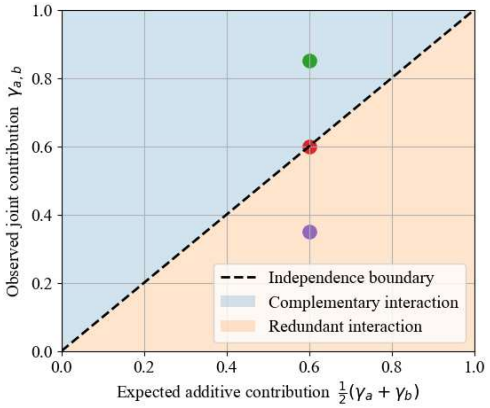



**Fig. 2**. Interpretation of the feature interaction score $\mathcal{P}(a,b)$.

The diagonal line $y = x$represents additive independence between two features. Points above the line indicate complementary interactions where the joint contribution exceeds the expected additive contribution, while points below the line indicate redundancy between features.

The interaction function $\wp$ is symmetric, $\wp(a,b) = \wp(b,a)$ and bounded within [-1, 1]. This is because it is computed using fuzzy dependency degrees, which themselves lie in the normalized range [0, 1]. Although the framework can be extended to higher-order interactions or Shapley-value approximations, such methods introduce prohibitive computational overhead. For tractability and interpretability, we restrict attention to pairwise interactions, which capture a large portion of relational structure while maintaining computational feasibility. To incorporate interaction into the feature selection process, we define a unified scoring function that balances relevance and pairwise interaction.

Let $S \subseteq A$ denote the current set of selected features, and $a \in A \backslash S$ be a candidate feature. The selection score for $a$ is given by:

$$score(a) = \Delta\gamma(a|S) + \psi \sum_{b \in S} \wp(a,b) \tag{10}$$

where the first term $\Delta\gamma(a|S) = \gamma_{S\cup\{a\}}(Z) - \gamma_S(Z)$ represents the marginal gain in dependency, and the second term captures the cumulative interaction between $a$ and the features in $S$. The parameter $\psi \in [-1,1]$ serves as a trade-off coefficient controlling the influence of interaction. A positive $\psi$ encourages the selection of features with positive interaction effect, whereas a negative $\psi$ discourages redundancy by penalizing overlapping information. Based on this formulation, the updated interaction-aware selection rule becomes:

$$a^* = \arg\max_{a \in A \backslash S} \left[\Delta\gamma(a|S) + \psi \sum_{b \in S} \wp(a,b)\right] \tag{11}$$

For each candidate feature $a$ not yet selected, the algorithm computes the marginal gain in fuzzy dependency $\Delta\gamma(a|S)$, which quantifies the improvement in classifiability when $a$ is added to the current subset $S$. It then evaluates how feature $a$ interacts with already selected features using a pairwise interaction score $\wp(a,b)$, summing over all $b \in S$. This accounts for redundancy (negative interaction) and complementarity (positive interaction). The total score for feature $a$ is calculated as:

$$\text{total score} = \Delta\gamma(a|S) + \psi \cdot \sum_{b \in S} \wp(a, b) \tag{12}$$

The interaction-aware feature scoring algorithm (Algorithm 1) is called inside the R⁵-FRQR (Algorithm 2) loop to select the next feature. It enhances classical greedy selection by incorporating a joint relevance criterion that balances individual contribution and inter-feature interaction, improving subset quality and informativeness.

**Algorithm 1** Interaction-Aware Feature Scoring.

**Input:**
$S \leftarrow$ current selected feature set
$A \leftarrow$ full feature set
$\psi \leftarrow$ interaction weight
$\gamma(S) \leftarrow$ fuzzy dependency degree of set $S$
$\gamma(S \cup \{a\}) \leftarrow$ fuzzy dependency degree after adding $a$
$\wp(a, b) \leftarrow$ interaction score between features $a$ and $b$
**Output:**
$a^* \leftarrow$ feature with highest total score
**Initialize:**
best score$\leftarrow -\infty$
$a^* \leftarrow \varphi$
**for** each candidate feature $a \in A \backslash S$:
Compute $\Delta\gamma \leftarrow \gamma(S \cup \{a\}) - \gamma(S)$
Interaction sum $\leftarrow 0$
**for** each feature $b$ in $S$:
interaction sum $\leftarrow$ interaction sum $+\wp(a, b)$
**End for**
Total score $\leftarrow \Delta\gamma + \psi \cdot$ interaction sum.
**if** total score $>$ best score
Best score $\leftarrow$ total score
$a^* \leftarrow a$
**End if**
**End for**
**Return** $a^*$

The parameter $\psi$ provides flexibility in tuning the selection strategy. Setting $\psi = 0$ reduces the method to the original FRQR algorithm based solely on marginal dependency. When $\psi > 0$, the algorithm promotes complementarity among features, and when $\psi < 0$, it penalizes redundancy more aggressively, thereby favoring feature diversity.

*B. Granular Similarity Approximation for Scalable Computation*

FRQR repetitively evaluates the fuzzy similarity relation $\ell_K(x_i, x_j)$ across all pairs of instances. This operation requires $O(n^2)$ pairwise computations for a dataset of $n$ instances and must be recomputed during each iteration of the feature selection process. As $n$ grows, this quadratic complexity becomes increasingly prohibitive, particularly for high-dimensional datasets. To mitigate this issue, we introduce a granular approximation strategy that replaces computationally intensive instance-level feature comparisons with a more efficient, cluster-level estimation. The idea is to partition the universe $U$ into a smaller number of coherent groups, referred to as granular balls, and compute similarities at the granule level.

Let $U \subset \mathbb{R}^m$ be partitioned into $c \ll n$ clusters $C = \{C_1, C_2, \ldots, C_c\}$ using a lightweight algorithm such as MiniBatch $k$-Means. Each cluster $C_k$ is then represented by its centroid $\mu_k \in \mathbb{R}^m$. For each instance $x_i \in U$, we define a mapping $c(x_i) \in \{\mu_1, .., \mu_c\}$ that assigns the instance to the centroid of its cluster. Using this representation, we define an approximate similarity measure between instances $x_i$ and $x_j$ as the similarity between their associated centroids. Specifically, the granular similarity is given by:

$$\tilde{\ell}_K(x_i, x_j) = \ell_K\left(c(x_i), c(x_j)\right) = \frac{1}{|K|} \sum_{a \in K} \exp\left(-\frac{\left(c(x_i)^a - c(x_j)^a\right)^2}{2\sigma_a^2}\right) \tag{13}$$

This formulation significantly reduces the number of computations required, from $O(n^2)$ to $O(c^2)$, by shifting the similarity estimation from individual instances to their respective cluster representatives. The Gaussian kernel remains the similarity basis, but the inputs are now the attributes of the centroids, rather than the raw instance-level features. The granularity parameter $c$ introduces a trade-off between speed and fidelity. A smaller $c$ accelerates computation but results in coarser approximations, while a larger $c$ provides finer resolution at the cost of increased complexity. To ensure the quality of approximation, it is essential that the clustering process preserves local neighbourhood structure, as these underlie the semantic relationships captured by the fuzzy similarity measure.

The approximate similarity $\tilde{\ell}_K(x_i, x_j)$ is used in place of $\ell_K(x_i, x_j)$ throughout the feature selection process, including the computation of lower and upper approximations, defined in (3) and (4), the fuzzy positive region, defined in (5), the dependency degree $\gamma_K(Z)$, defined in (6) the marginal gain $\Delta\gamma(a|S)$, defined in (7) and the interaction score $\wp(a, b)$, defined in (9). This approximation enables the proposed method to scale effectively to larger datasets without compromising the structural integrity of the fuzzy rough framework.

### *C. Stability-Aware Feature Consolidation Via Bootstrap Aggregation*

While relevance and efficiency are vital, stability is equally crucial, especially in sensitive domains like healthcare or finance where consistency across data splits matters. Classical methods, including fuzzy rough approaches, can be unstable under sampling noise. To address this, we incorporated a bootstrap-based aggregation mechanism that identifies features consistently selected across resampled datasets, enhancing the reliability and robustness of the final subset.

Let $B$ denote the number of bootstrap iterations and let $\{U^{(1)}, U^{(1)}, \dots, U^{(B)}\}$ represent resampled datasets drawn with replacement from the original universe $U$. For each bootstrapped sample $U^{(B)}$, the interaction-aware selection rule defined in (11) is applied, yielding a corresponding feature subset $S^{(b)} \subseteq A$. Due to the stochasticity of resampling and selection, these subsets may differ across iterations. For each feature $a \in A$, we define a stability score $\varphi(a) \in [0,1]$ that reflects the proportion of bootstrap runs in which $a$ is selected. This score is computed as

$$\varphi(a) = \frac{1}{B}\sum_{b=1}^{B} \mathbb{I}\left(a \in S^{(b)}\right) \tag{14}$$

where $\mathbb{I}(\cdot)$ denotes the indicator function, returning 1 if $a$ appears in $S^{(b)}$, and 0 otherwise. A higher stability score implies that the feature is consistently selected under different sampling conditions, suggesting higher reliability and robustness. To extract the final subset of stable features, a stability threshold $\wp \in [0,1]$ is applied to the scores $\varphi(a)$. The final set of retained features is given by

$$S_{\text{final}} = \{a \in A | \varphi(a) \geq \wp\}, \tag{15}$$

thus, including only those features that are selected in at least a proportion $\wp$ of the bootstrap runs. The threshold $\wp$ offers a tunable trade-off: $\wp$ = 1.0 yields a conservative selection consisting only of universally chosen features, while $\wp$=0.5 includes those that are moderately stable. Intermediate values such as $\wp$=0.7 or 0.8 are often adopted to balance reliability and representativeness.

### *D. Domain-Adaptive Dependency for Cross-Domain Generalization*

While stability ensures consistency within a domain, real-world applications demand generalization across domains with varying data distributions. To meet this need, we introduce a domain-adaptive dependency score, enabling feature subsets to remain discriminative even under distribution shifts.

Let $D_s$ and $D_t$ denote the source and target domains, respectively. For a feature subset $K \subseteq A$, the fuzzy dependency degree is computed separately for each domain using (6) resulting in:

$$\gamma_K^{(s)}(Z) = \frac{1}{|D_s|}\sum_{x \in D_s} \underline{KZ}(x), \gamma_K^{(t)}(Z) = \frac{1}{|D_t|}\sum_{x \in D_t} \underline{KZ}(x) \tag{16}$$

where $\underline{KZ}(x)$ denotes the fuzzy lower approximation value of instance $x$ with respect to decision attribute $Z$ and feature subset $K$. These domain-specific dependency scores reflect how well the feature subset explains the decision attribute in each respective environment. To blend these two measures into a unified, domain-adaptive criterion, we define the resilient dependency score $\wp_K(Z)$ as a convex combination of the source and target domain scores:

$$\wp_K(Z) = \alpha \cdot \gamma_K^{(s)}(Z) + (1 - \alpha) \cdot \gamma_K^{(t)}(Z) \tag{17}$$

where α ∈ [0.5,1.0] is a tunable parameter representing the degree of trust placed in the source domain. A value of α =1.0 implies full reliance on the source domain, reducing (17) to the standard dependency score. When α = 0.5 gives equal weight to both domains, promoting balanced generalization.

**Algorithm 2** R⁵-FRQR Feature Selection Algorithm.

**Inputs:**

Dataset $U \in \mathbb{R}^{n\times m}$, decision labels $Z \in \{z_1, z_2, \dots, z_C\}$

Parameters:

$B$: Number of bootstrap iterations

$\wp \in [0,1]$: stability threshold

$\psi \in [0,1]$: Interaction weight

$c$: Number of clusters

$\varepsilon > 0$: Improvement threshold

$\alpha \in [0.5,1]$: Domain blending weight

**Output:**

Selected feature subset: $S_{final} \subseteq A$

**Step 1: Preprocessing**

Normalize features to [0, 1]

Estimate fuzzy memberships $\mu_z(x)$

Compute $\sigma_a$ for each feature $a \in A$

**Step 2: Initialize stability scores**

Set $\varphi(a) = 0$ for all $a \in A$

**Step 3: Bootstrap-based selection (repeat for $b = 1$ to $B$)**

Resample data to form $U^{(b)}$

Cluster $U^{(b)}$ into $c$ clusters → centroids $\{\mu_1, \dots, \mu_c\}$

Assign each instance $x \in U^{(b)}$ to a cluster center $c(x)$

Initialize selected feature set $S \leftarrow \emptyset$

**Iterative feature selection**

While improvement $\Delta\gamma > \varepsilon$

For each candidate feature $a \in A \setminus S$:

Use **Algorithm 1** to compute:

$\text{score}(a) = \Delta\gamma(a|S) + \psi \sum_{b\in S} \wp(a,b)$

Select $a^* = \text{argmax}_a \text{Score}(a)$

Update: S ← S ∪ {a*}

Record selected subset $S^{(b)}$, update stability scores:

$\varphi(a) \leftarrow \varphi(a) + 1$ for $a \in S^{(b)}$

**Step 4: Final feature selection**

Normalize $\varphi(a) \leftarrow \varphi(a)/B$

Define final set: $S_{final} \leftarrow \{a \in A | \varphi(a) \geq \wp\}$

**Step 5: Domain-resilient validation**

Compute FDD in source $\gamma^{(s)}$ and target $\gamma^{(t)}$ domains

Compute domain-adaptive score

$$\gamma_{\text{blend}} \leftarrow \alpha \cdot \gamma^{(s)} + (1 - \alpha) \cdot \gamma^{(t)}$$

Prune $S_{final}$ to maximize $\gamma_{\text{blend}}$

**Return:** Final feature subset $S_{\text{final}}$

The proposed R⁵-FRQR framework presents five mechanisms to enhance the reliability, efficiency, and generalization ability of the proposed fuzzy rough feature selection method. First, the reliability of the proposed method is enhanced through the application of a stability aggregation mechanism based on the bootstrap method. Second, the relational awareness of the proposed method is achieved through the application of an interaction-aware feature scoring mechanism, which evaluates the relationships between the features to be selected and the existing features. Third, the robustness of the proposed method is achieved through the evaluation of the robustness of the features to label noise. Fourth, the reductiveness of the proposed method is achieved through the application of a greedy forward feature selection mechanism based on the maximization of the fuzzy dependency

measure. Finally, the proposed method is enhanced to incorporate the resilience of the proposed method through the application of a dependency evaluation mechanism based on the source and target domains.

## IV. EXPERIMENTATION

This section presents the implementation and evaluation details of the proposed R⁵-FRQR framework.

### *A. Implementation*

The proposed method was implemented in Python 3.10, and the experiments were run on an Intel Core i9 workstation with 64 GB RAM, with an NVIDIA GeForce RTX 2060 GPU. The procedure consists of four interlinked modules: The interaction-aware quick reduct in which features are incrementally selected using the marginal gain in fuzzy dependency $\Delta\gamma$ and interaction score $\wp(a,b)$, the granular similarity approximation to address the $O(n^2)$ complexity of pairwise fuzzy similarity computations, instance similarities are approximated using centroids from Minibatch k-means clustering, bootstrapped stability aggregation where feature selection is repeated over $B$=10 bootstrap iterations; stability scores $\varphi(a)$ are computed and final features are filtered by a threshold $\wp$, and domain-adaptive dependency in which FDD is computed across source and target domains and blended using a tunable trust parameter $\alpha \in$[0.5,1.0].

### *B. Experimental Datasets*

Experiments were conducted on 16 diverse sets of benchmark datasets spanning multiple biomedical, sensor-based, and pattern recognition applications accessed from the repository[1]. Each dataset was first randomly shuffled and then partitioned into a 60% source domain and 40% target domain, ensuring stratified class distribution. To simulate a realistic domain shift, we applied a mild feature perturbation to the target domain by introducing small Gaussian noise (mean zero, standard deviation proportional to each feature's standard deviation). These mimics variations arising from differences in acquisition conditions, instrumentation, or environmental factors. The dataset class counts, sample sizes, and feature dimensions, are listed in Table I.

TABLE I. EXPERIMENTAL DATASETS AND FEATURE SET SIZES.

| ID | Dataset | Classes | Instances | Features |
|---|---|---|---|---|
| 1 | CNS | 5 | 42 | 7129 |
| 2 | Colon | 2 | 62 | 2000 |
| 3 | HCV for Egyptian patients | 4 | 1385 | 28 |
| 4 | Hill Valley | 2 | 1212 | 100 |
| 5 | Leaf | 30 | 340 | 15 |
| 6 | Leukemia-2 | 2 | 72 | 7129 |
| 7 | Leukemia-3 | 3 | 72 | 7129 |
| 8 | Leukemia-4 | 4 | 72 | 7129 |
| 9 | Libras Movement | 15 | 360 | 90 |
| 10 | Lung Cancer | 2 | 181 | 12533 |
| 11 | Lymphoma | 3 | 62 | 4026 |
| 12 | MEU-Mobile KSD | 56 | 2856 | 71 |
| 13 | MLL | 3 | 72 | 12582 |
| 14 | Ovarian Cancer | 2 | 253 | 15154 |
| 15 | SRBCT | 4 | 83 | 2308 |
| 16 | Waveform Database Generator | 3 | 5000 | 21 |

### *E. Evaluation*

We evaluated the predictive quality of the selected feature subsets using three complementary classifiers:

**Random Forest (RF)** (Ren et al., 2023) was chosen for its strong ensemble capabilities and robustness to noise and overfitting, particularly in high-dimensional spaces. It also provides an effective baseline for benchmarking the quality of selected features due to its inherent feature bagging and averaging across multiple trees. In all experiments, we used the standard implementation with 100 trees (estimators) and random state set to 42 to ensure reproducibility.

**CART** (**classification and regression trees)** (Loh, 2011) was included as a lightweight, interpretable model that is highly sensitive to feature quality and structure. Unlike ensemble or margin-based methods, CART directly reflects the impact of selected features on decision boundaries and branching patterns. We used default Gini-based impurity for splitting and unrestricted depth to allow full tree development.

[1] 1https://archive.ics.uci.edu/datasets

**HOPK-SVM support vector machines with Hermite orthogonal polynomial kernels** (Moghaddam & Hamidzadeh, 2016) were adopted to evaluate the ability of the selected features to support complex nonlinear decision boundaries. The Hermite kernel is defined as:

$$K_{Hermite}(x, x') = \exp\left(-\frac{\|x - x'\|^2}{2\sigma^2}\right) \cdot H_d(x \cdot x') \tag{18}$$

where $H_d(\cdot)$ denotes the Hermite polynomial of degree $d$, and $\sigma$ is the kernel bandwidth. The polynomial degree $d \in \{1,2,3\}$ and bandwidth $\sigma \in \{0.1,1.0,10\}$ were selected through nested 3-fold cross-validation performed within each training fold. The combined kernel variant uses:

$$K_c(x, x') = \sigma \cdot K_{RBF}(x, x') + (1 - \sigma) \cdot K_{Hermite}(x, x') \tag{19}$$

To further evaluate the generalization behaviour, noise robustness, and feature selection stability, we performed three additional diagnostic evaluations: noisy label perturbation to assess robustness, 10% of the class labels were randomly flipped, and classification performance was compared to that on the original labels, bootstrap-based resampling ($B$ = 20) to quantify feature stability, selection was repeated across 20 resampled datasets, and consistency was measured using Jaccard and Kuncheva indices, and cross-domain generalization for domain-resilience analysis, source and target domain accuracies were computed separately, and the drop in performance (ΔDomain) was used to quantify transferability.

*F. Parameter Configuration and Justification*

All hyperparameters were selected based on empirical sensitivity analysis, guided by best practices in stability selection and fuzzy rough modelling (Baek & Kim, 2017; Sysoev, 2023). All parameter values were held fixed across datasets to ensure fairness and reproducibility, avoiding dataset-specific tuning. The number of clusters used in the granular similarity approximation was set to $c$ = 20. This setting was chosen via a grid search over $c$ ∈ {10, 20, 30, 50, 100}, tracking both approximation fidelity and runtime. Larger values of $c$ yielded marginal improvements in dependency scores (1-2%) but incurred substantial runtime costs especially on datasets with >5000 instances. Conversely, $c$ < 20 led to oversimplified similarity approximations that degraded classification accuracy. The selected value of $c$ = 20 consistently provided >95% approximation fidelity relative to full pairwise similarity, with approximately 3× computational speedup. Bootstrap resampling was performed $B$=5 times in all experiments to compute the stability score $\varphi(a)$. The choice of $B$ was guided by both literature on stability selection (Meinshausen & Bühlmann, 2010) and convergence analysis. Specifically, we observed that the Kuncheva index and Jaccard index of selected subsets stabilized between $B$=4 and $B$=7, with negligible changes beyond this range.

The interaction-aware selection process was capped at a maximum of 50 iterations to avoid overfitting. In our trials, we observed that for most datasets, the FDD, $\gamma_S$ saturated within 35-45 iterations, with further additions yielding marginal gains (often $\Delta\gamma$ <0.005). Beyond 50 iterations, classification performance of selected features plateaued/dropped, thus, setting maximum iteration to 50 serves as a conservative upper bound while preventing over-selection. The interaction parameter $\psi \in [0,1]$ controls the trade-off between feature relevance and interaction complementarity. In all reported experiments, we set $\psi$=1.0 to fully activate the interaction term. To justify this, we conducted ablation experiments with $\psi \in \{0.0,0.25,0.5,0.75,1.0\}$. When $\psi$ = 0, the method reverted to classical FRQR and often selected redundant features due to a lack of interaction modeling. Settings like $\psi$ = 0.25 or 0.5 yielded moderate improvements but still missed mutual combinations, particularly in high-dimensional data. The best classification accuracy and lowest redundancy (as measured by pairwise cosine similarity among selected features) consistently occurred at $\psi$ =1.0.

The stability filtering step used a threshold of $\wp$=0.6, meaning a feature must appear in at least 60% of the bootstrap runs to be retained. At lower thresholds (for example, $\wp$=0.4, unstable features with spurious correlations were admitted, leading to high variance in classification accuracy. At stricter thresholds, for example, $\wp$=0.8, the final subsets became too small (in some cases, <3 features), compromising classification quality. The value $\wp$=0.6 retained moderately stable features while maintaining classification robustness, as evidenced by consistent F1-scores and robustness indices across noisy data experiments.

## V. Results And Discussion

*A. Feature Subset Size and Classification Performance*

To evaluate the discriminative strength of selected features, we report the number of features selected along with accuracy, precision, recall, and F1-score using three classifiers: RF, HKSVM, and CART. Results, averaged over 5-fold cross-validation (Table II), demonstrate that R⁵-FRQR achieves high classification performance with remarkably compact subsets. In datasets like CNS, Colon, HCV, Leukemia, Lung Cancer, and MLL, fewer than

six features yielded F1-scores above 90% across all classifiers, underscoring the strong relevance of the selected features. On high-dimensional datasets such as Lymphoma and Ovarian, more features were retained, yet performance especially for HKSVM remained exceptional, exceeding 99% in several cases. RF maintained consistent performance across datasets, while HKSVM excelled in noisy or nonlinear domains. CART also performed competitively, benefitting from the quality of selected features.

To further interpret the selection dynamics, we tracked the FDD (γ) and gain across 50 iterations over 5 bootstrap runs, as visualized in Fig. 3. In most datasets, γ initially had a high value which rapidly dropped, suggesting early dominance by a few highly informative features. The diminishing marginal gain justifies the algorithm's early stopping mechanism. These trends demonstrate that R⁵-FRQR efficiently prioritizes the most impactful features early in the selection process.

*B. Runtime, Stability, and Generalization Analysis*

To provide a holistic evaluation of the proposed R⁵-FRQR framework, we categorize the performance results under four themes assessed using various metrics on 10 bootstrapped selection trials across 16 datasets. Table III presents detailed results.

TABLE II. SELECTED FEATURE SIZE AND CLASSIFICATION PERFORMANCE RESULTS OF THE SELECTED FEATURES.

| | | RF | | | | H-SVM | | | | CART | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ID | F | Acc. (%) | P (%) | R (%) | F (%) | Acc. (%) | P (%) | R (%) | F (%) | Acc. (%) | P (%) | R (%) | F (%) |
| 1 | 4 | 89.6±14 | 95.0±21 | 91.3±16 | 93.1±13 | 96.8±06 | 83.4±03 | 97.0±00 | 90.2±02 | 92.5±15 | 92.0±10 | 96.2±08 | 94.2±10 |
| 2 | 5 | 86.1±24 | 86.6±28 | 89.0±22 | 87.8±18 | 75.7±06 | 82.5±22 | 93.3±07 | 87.9±11 | 68.9±05 | 64.0±06 | 76.0±06 | 70.0±04 |
| 3 | 2 | 76.9±03 | 68.4±03 | 77.4±03 | 72.9±02 | 75.5±03 | 63.2±02 | 64.9±03 | 64.1±02 | 67.4±01 | 66.2±02 | 65.6±01 | 65.9±01 |
| 4 | 99 | 95.8±04 | 98.3±04 | 92.5±04 | 95.4±03 | 94.6±02 | 92.8±07 | 88.7±02 | 90.8±04 | 96.3±02 | 96.2±02 | 95.8±02 | 96.0±02 |
| 5 | 15 | 92.7±06 | 94.5±04 | 88.5±04 | 91.5±03 | 82.3±00 | 86.5±01 | 81.8±01 | 84.2±01 | 92±00 | 90.6±02 | 90.4±03 | 91.3±02 |
| 6 | 45 | 95.3±06 | 96.9±04 | 93.3±08 | 94.3±05 | 71.4±05 | 82.9±03 | 85.0±00 | 84.0±02 | 86.4±08 | 85.8±09 | 85.5±09 | 85.1±09 |
| 7 | 4 | 95.6±19 | 94.2±25 | 94.8±22 | 94.5±17 | 90.2±08 | 91.9±13 | 95.9±04 | 93.9±10 | 95.0±14 | 90.5±23 | 95.6±22 | 93.1±21 |
| 8 | 5 | 94.2±21 | 96.4±29 | 91.2±26 | 93.8±19 | 88.3±07 | 93.7±14 | 85.5±12 | 89.6±13 | 95.5±13 | 95.2±20 | 90.6±19 | 92.9±18 |
| 9 | 79 | 89.7±05 | 90.1±07 | 88.6±06 | 89.4±07 | 86.8±05 | 88.8±05 | 89.1±06 | 89.0±06 | 85.2±11 | 85.0±14 | 88.1±12 | 86.6±13 |
| 10 | 4 | 93.7±05 | 91.8±11 | 89.8±09 | 90.8±09 | 92.9±04 | 87.4±09 | 91.5±08 | 89.5±08 | 93.9±05 | 94.1±17 | 94.8±14 | 94.5±14 |
| 11 | 218 | 95.0±06 | 82.4±22 | 86.7±16 | 83.9±20 | 99.9±00 | 98.9±00 | 99.5±00 | 99.2±00 | 92.5±10 | 82.2±22 | 85.3±18 | 83.5±21 |
| 12 | 58 | 92.7±03 | 91.3±03 | 89.5±03 | 89.4±03 | 62.5±01 | 69.4±01 | 72.0±01 | 70.7±01 | 77.4±02 | 79.9±03 | 77.4±02 | 78.7±02 |
| 13 | 5 | 99.9±00 | 99.9±00 | 99.8±00 | 99.9±00 | 95.3±06 | 96.9±04 | 93.3±08 | 94.3±07 | 97.8±04 | 98.6±03 | 96.7±07 | 97.2±06 |
| 14 | 110 | 98.0±03 | 98.2±02 | 97.7±04 | 97.8±03 | 98.7±03 | 99.1±02 | 98.2±04 | 98.5±03 | 95.4±03 | 95.9±02 | 94.4±04 | 94.9±03 |
| 15 | 230 | 98.0±04 | 98.8±03 | 97.5±05 | 97.6±05 | 96.0±05 | 97.5±03 | 95.8±05 | 95.9±05 | 92.0±10 | 94.2±07 | 93.3±08 | 92.7±09 |
| 16 | 21 | 84.1±01 | 84.2±01 | 84.1±01 | 84.1±01 | 86.5±01 | 86.6±01 | 86.5±01 | 86.4±01 | 74.3±02 | 74.3±02 | 74.3±02 | 74.3±02 |

*1) Runtime, Stability, and Selection Dynamics*

The computational efficiency and internal consistency of the feature selection process are assessed using several metrics across all datasets. The average runtime (ART) represents the mean execution time in seconds required to complete the feature selection process over multiple bootstrap iterations. Table III shows how runtimes vary significantly based on dataset size and complexity, ranging from as low as 11.30s in HCV for Egyptian patients' dataset to as high as 7674.8s in Lymphoma. To quantify selection consistency across bootstrap samples, the stability score (SS) is computed as the average selection frequency of features retained in the final subset. Higher scores suggest that the selected features are robust to sampling variations; for instance, Leaf and Waveform Database Generator datasets yielded maximum stability scores of 1.00. Further insights into selection dynamics are provided by the Kuncheva index (KI) and the average Jaccard index (AJD) which measure the degree of overlap among feature subsets across different bootstrap runs. While Jaccard captures raw intersection-over-union, the Kuncheva index accounts for random chance agreement. Across the datasets, KI values range from 0.68 to 0.85 and AJD values from 0.59 to 1.00, indicating a generally consistent selection behavior, especially in datasets with well-defined structure or class separation.

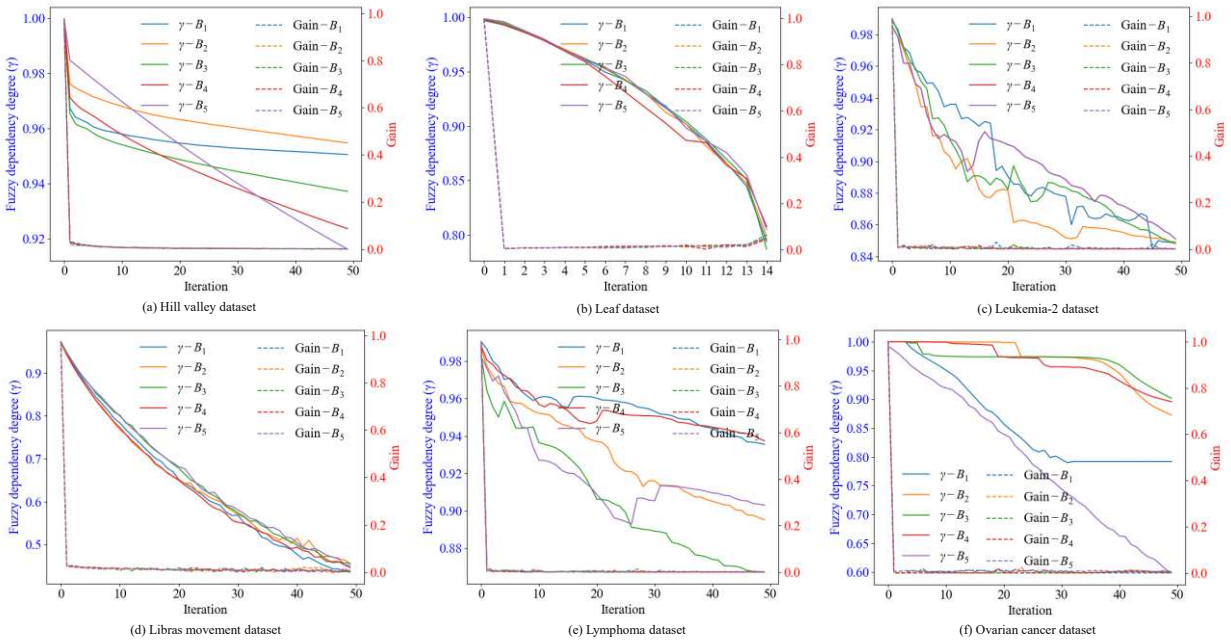



**Fig. 3.** Evolution of FDD (γ) and interaction gain across 50 selection iterations for five bootstrap runs. Each curve represents one bootstrap. High initial γ values followed by declining trends suggest early selection of highly informative features, with diminishing gain indicating redundancy or saturated dependency.

*2) Redundancy Analysis*

We assess the uniqueness of selected features by measuring the redundancy score (RS) computed as the mean pairwise absolute Pearson correlation among selected features. Low scores such as 0.001 for HCV for Egyptian patients' dataset, 0.032 for Leaf dataset indicate that selected features contribute non-overlapping information, critical for interpretability and generalization. Most datasets maintain RS < 0.15, indicating strong anti-redundancy behavior.

TABLE III. RUNTIME, STABILITY, AND GENERALIZATION ANALYSIS RESULTS OF THE PROPOSED $R^5$-FRQR FEATURE SELECTION METHOD.

| Dataset | CA | NA | RI↑ | SA | TA | Δ↓ | γ | RS↓ | KI↑ | AJD↑ | SS↑ | ART |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CNS | 89.6±14 | 82.9±11 | 0.925 | 100.0 | 92.5 | 7.5 | 0.928 | 0.090 | 0.72 | 0.63 | 0.62 | 1969.9 |
| Colon | 86.1±24 | 86.0±17 | 0.999 | 100.0 | 95.8 | 4.2 | 0.875 | 0.080 | 0.68 | 0.59 | 0.55 | 564.1 |
| HCV | 76.9±03 | 74.5±04 | 0.969 | 84.0 | 78.3 | 5.7 | 0.862 | 0.001 | 0.81 | 0.73 | 0.58 | 11.3 |
| Hill Valley | 95.8±04 | 95.3±05 | 0.995 | 100.0 | 92.7 | 7.3 | 0.871 | 0.094 | 0.77 | 0.70 | 0.76 | 766.5 |
| Leaf | 92.7±06 | 89.2±01 | 0.962 | 100.0 | 96.8 | 3.2 | 0.905 | 0.032 | 0.78 | 1.00 | 1.00 | 17.2 |
| Leukemia-2 | 95.3±06 | 83.6±07 | 0.877 | 100.0 | 97.9 | 2.1 | 0.979 | 0.154 | 0.75 | 0.69 | 0.42 | 7362.1 |
| Leukemia-3 | 95.6±19 | 80.3±12 | 0.840 | 100.0 | 98.2 | 1.8 | 0.946 | 0.109 | 0.73 | 0.66 | 0.38 | 1657.5 |
| Leukemia-4 | 94.2±21 | 84.3±14 | 0.895 | 100.0 | 97.1 | 2.9 | 0.914 | 0.071 | 0.79 | 0.71 | 0.44 | 1298.5 |
| Libras Movement | 89.7±05 | 82.1±08 | 0.915 | 100.0 | 90.7 | 9.3 | 0.900 | 0.270 | 0.74 | 0.68 | 0.70 | 568.2 |
| Lung Cancer | 93.7±05 | 84.8±02 | 0.905 | 100.0 | 91.3 | 8.7 | 0.926 | 0.079 | 0.85 | 0.78 | 0.44 | 3105.1 |
| Lymphoma | 95.0±06 | 91.5±08 | 0.963 | 100.0 | 92.3 | 7.7 | 0.977 | 0.269 | 0.69 | 0.60 | 0.38 | 7674.8 |
| MEU-Mobile KSD | 92.7±03 | 85.1±02 | 0.918 | 100.0 | 91.2 | 8.8 | 0.974 | 0.148 | 0.70 | 0.82 | 0.80 | 874.1 |
| MLL | 99.9±00 | 91.4±04 | 0.915 | 100.0 | 99.3 | 0.7 | 0.968 | 0.170 | 0.68 | 0.71 | 0.43 | 1507.7 |
| Ovarian Cancer | 98.0±03 | 86.9±05 | 0.887 | 100.0 | 100.0 | 0.0 | 1.000 | 0.117 | 0.76 | 0.69 | 0.49 | 7128.7 |
| SRBCT | 98.0±04 | 97.0±16 | 0.990 | 100.0 | 100.0 | 0.0 | 1.000 | 0.164 | 0.81 | 0.72 | 0.43 | 6111.4 |
| Waveform Database | 84.1±01 | 76.9±01 | 0.914 | 100.0 | 86.5 | 13.5 | 0.960 | 0.071 | 0.84 | 1.00 | 1.00 | 148.7 |

*3) Robustness to Label Noise*

To simulate realistic conditions where mislabeled data may occur, 10% of class labels were randomly flipped, and the performance of random forest was reevaluated on both the original and noise-injected datasets. The results are presented using two metrics: clean accuracy (CA) and noisy accuracy (NA), which quantify model performance before and after label corruption, respectively. From these, the robustness index (RI) is computed as the ratio of NA to CA, indicating the degree to which performance is preserved under noisy conditions. High robustness values, such as 0.999 for Colon dataset and 0.995 for Hill valley dataset, confirm that the $R^5$-FRQR method retains strong predictive accuracy even in the presence of significant label noise.

*4) Cross-Domain Generalization* (Δ*Domain*)

To evaluate the generalizability of selected features across distribution shifts, we assess the performance of a random forest classifier trained on the source domain and tested on an unseen target domain. Source accuracy (SA) and target accuracy (TA) are used to measure predictive performance before and after a domain shift, respectively. The difference between them, denoted as Δdomain (Δ), captures the performance degradation due to domain mismatch, with lower values indicating better generalization. Notably, Ovarian Cancer and SRBCT datasets exhibit perfect transferability with $\Delta = 0$, highlighting the domain-robustness of the selected features. Additionally, we report the domain-adaptive gamma ($\gamma$), which reflects the average fuzzy dependency degree across source and target domains. Most datasets show $\gamma$ values above 0.9, confirming that the semantic significance of selected features is preserved even under substantial distributional variation.

*C. Performance Comparison*

To validate the performance of $R^5$-FRQR, we compared it against 11 widely adopted feature selection methods using four core metrics: classification accuracy (Acc.), F1-score (F1), robustness to label noise (RI), and the number of selected features (F). These metrics jointly assess predictive strength, stability under noise, and subset compactness.

The comparison includes methods spanning graph-based, information-theoretic, sparsity-driven, fuzzy, kernel-based, ensemble, and hybrid paradigms. Specifically, we considered SPEC (Spielman, 2007), ReliefF (Spolaôr et al., 2013), LASSO (Demir-Kavuk et al., 2011), mRMR (Wang et al., 2018), FRQR (Anaraki & Eftekhari, 2011), FMI (Antonelli et al., 2013), Sparse Group Lasso (SGL) (Rao et al., 2016), HSIC-Lasso (HL) (He et al., 2024), Boruta (Kursa & Rudnicki, 2010), DeepFS (Taherkhani et al., 2018), and a ReliefF + SVM-RFE hybrid (Fu et al., 2014). These methods were selected because they span filter, wrapper, embedded, fuzzy, and deep learning-based techniques, allowing us to compare $R^5$-FRQR against both traditional and state-of-the-art alternatives. All methods were applied to the source domain using consistent data splits. Ranking-based methods used the top $k = 50$ features, while others reported the number of features selected intrinsically. Results are presented in Table IV.

$R^5$-FRQR consistently achieved higher performances than all baselines across most datasets, with particularly strong margins on the Hill Valley, Leaf, Leukemia-2, Leukemia-3, MLL, Ovarian Cancer, and SRBCT datasets. For instance, on the MLL dataset, $R^5$-FRQR achieved near-perfect performance with 99.9% accuracy and F1, performing better than even deep learning-based DeepFS and the Hybrid methods. Similarly, on SRBCT dataset, it reached 98.0% accuracy and 97.6% F1 with a robustness index of 0.99, demonstrating strong generalization and resilience. While methods such as DeepFS, Boruta, and Hybrid also performed well in certain cases, their feature counts were generally higher, for example, 36-45 features on average, making them less interpretable and computationally demanding. In contrast, $R^5$-FRQR often required fewer features, as low as 2-7 on some datasets, to achieve comparable or stronger results, highlighting its discriminative power and compactness.

SGL and HL, though nonlinear and strong in kernel settings, were less robust to noise, often yielding lower robustness indices than $R^5$-FRQR. For example, on Leukemia-2 dataset, SGL and HL had RI values of 0.848 and 0.847, compared to 0.877 for $R^5$-FRQR. Traditional methods such as mRMR, ReliefF, and FMI showed moderate performance but lacked consistency across datasets. Their feature counts fluctuated between 8 and 30, and their robustness indices were generally lower, for example, ReliefF averaged ~0.85 RI. LASSO, while embedded and fast, underperformed in most cases due to its linear assumptions and sparse biases, especially on complex or high-dimensional datasets. Notably, $R^5$-FRQR maintained high robustness ($RI \geq 0.90$) on nearly all datasets, indicating stable performance.

*D. Ablation Study*

To evaluate the individual and cumulative contributions of each component in the proposed $R^5$-FRQR framework, we performed an ablation study across six diverse datasets. Starting from the baseline FRQR which relies solely on the marginal fuzzy dependency gain, we progressively integrated the five enhancements: ($R^2$) granular similarity for efficiency, ($R^3$) interaction-aware scoring, ($R^1$) bootstrap-based stability selection, ($R^4$) robustness to label noise, and ($R^5$) domain-adaptive dependency for cross-domain generalization. Each variant was assessed using Random Forest and Hermite-kernel SVM classifiers and the highest results were reported. Evaluation metrics included classification accuracy and F1-score (Acc, F), redundancy (RS), robustness index (RI), domain shift (Δ), domain-adaptive gamma ($\gamma$), stability score (SS), and selection consistency (Kuncheva index KI, Jaccard index AJD), as well as average runtime (ART). Feature selection was repeated across bootstrap iterations to capture reliability.

The results of the ablation study detailed in Table V confirm that each enhancement in the $R^5$-FRQR framework contributed meaningfully to performance, robustness, and generalizability. Starting from the baseline FRQR, we observed significant improvements in performance metrics as components were progressively integrated. For

example, on the HCV dataset, the baseline FRQR achieved 71.8% accuracy and 67.2% F1-score with a runtime of 25.6 seconds. Adding granular clustering ($R^2$) improved accuracy to 73.4% and F1 to 69.5%, while maintaining a low redundancy score of 0.004. With interaction-awareness ($R^3$), performance increased further (Acc: 74.8%, F1: 71.0%) and continued to improve with stability bootstrapping ($R^1$) (Acc: 75.9%, F1: 72.1%). The full $R^5$-FRQR yields the best performance (Acc: 76.9%, F1: 72.9%, RI: 0.969), validating the cumulative effect of each module. While the runtime increased to 11.3s, the selection became more consistent (SS: 0.58), with improved generalization (ΔDomain ↓ to 5.7 and γ ↑ to 0.862).

A similar trend was observed with the Hill Valley dataset, where the baseline achieved 89.4% accuracy and 88.2% F1. Each module contributed progressively culminating in 95.8% accuracy and 95.4% F1 under $R^5$-FRQR. The stability score also improved from 0.51 (baseline) to 0.76, and Kuncheva index from 0.45 to 0.77. Notably, the domain gap dropped from 10.4% to 7.3%, highlighting better cross-domain resilience. Although the runtime increased from 685.1s to 766.5s with the full $R^5$-FRQR model, it remained substantially lower than the baseline FRQR runtime of 1228.3s. This improvement was primarily due to the efficient feature selection strategy in $R^5$-FRQR, which leverages clustering to reduce the feature search space, in contrast to baseline FRQR's exhaustive comparison across all feature subsets. Similar upward trends were observed across the Libras Movement, MEU-Mobile KSD, Ovarian Cancer, and Waveform datasets, where the full model consistently outperformed its ablated variants in predictive performance, robustness, and stability. While runtime did increase in the addition of various components, the increases were not arbitrary but directly corresponded to greater selection consistency, reduced noise sensitivity, and better improved generalization.

TABLE IV. PERFORMANCE COMPARISON RESULTS OF $R^5$-FRQR WITH CLASSICAL AND STATE-OF-THE-ART FEATURE SELECTION METHODS.

| Dataset | Metrics | SPEC | ReliefF | LASSO | mRMR | FRQR | FMI | SGL | HL | Boruta | DeepFS | Hybrid | Ours |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CNS | Acc. (%) | 80.4 | 85.3 | 88.2 | 85.1 | 82.6 | 86.0 | 86.5 | 86.9 | 87.4 | 90.8 | 88.7 | 89.6 |
| | F1 (%) | 86.6 | 87.0 | 84.0 | 82.9 | 89.4 | 89.8 | 90.3 | 90.8 | 91.2 | 92.0 | 92.2 | 93.1 |
| | RI | 0.896 | 0.891 | 0.794 | 0.833 | 0.736 | 0.837 | 0.858 | 0.808 | 0.831 | 0.859 | 0.904 | 0.925 |
| | F | 8 | 25 | 6 | 17 | 24 | 45 | 26 | 17 | 54 | 18 | 13 | 4 |
| Colon | Acc. (%) | 79.2 | 80.1 | 80.9 | 81.8 | 82.2 | 82.7 | 83.1 | 83.5 | 83.9 | 84.4 | 85.2 | 86.1 |
| | F1 (%) | 81.7 | 82.1 | 83.0 | 83.8 | 84.3 | 84.7 | 85.2 | 85.6 | 86.0 | 86.5 | 86.9 | 87.8 |
| | RI | 0.972 | 0.958 | 0.938 | 0.831 | 0.974 | 0.953 | 0.953 | 0.934 | 0.906 | 0.944 | 0.915 | 0.999 |
| | F | 9 | 8 | 10 | 12 | 15 | 15 | 30 | 10 | 58 | 19 | 7 | 5 |
| HCV for Egyptian patients | Acc. (%) | 70.7 | 72.5 | 72.3 | 73.1 | 73.4 | 73.8 | 74.2 | 74.6 | 75.0 | | 75.0 | 76.9 |
| | F1 (%) | 67.8 | 68.2 | 68.9 | 69.6 | 70.0 | 70.3 | 76.0 | 74.4 | 71.4 | 71.8 | 76.2 | 72.9 |
| | RI | 0.815 | 0.814 | 0.811 | 0.691 | 0.804 | 0.885 | 0.788 | 0.836 | 0.908 | 0.796 | 0.813 | 0.969 |
| | F | 19 | 20 | 11 | 11 | 9 | 33 | 28 | 18 | 57 | 20 | 14 | 2 |
| Hill Valley | Acc. (%) | 85.6 | 86.6 | 89.0 | 90.5 | 89.8 | 90.9 | 90.6 | 90.4 | 92.5 | 93.0 | 93.2 | 95.8 |
| | F1 (%) | 88.7 | 89.2 | 90.2 | 91.1 | 91.6 | 92.1 | 92.5 | 93.0 | 93.5 | 94.0 | 94.4 | 95.4 |
| | RI | 0.970 | 0.972 | 0.961 | 0.930 | 0.969 | 0.960 | 0.970 | 0.976 | 0.962 | 0.968 | 0.975 | 0.995 |
| | F | 6 | 24 | 6 | 14 | 24 | 39 | 45 | 30 | 59 | 12 | 10 | 99 |
| Leaf | Acc. (%) | 85.3 | 86.2 | 87.1 | 88.1 | 88.5 | 89.0 | 89.5 | 89.9 | 88.4 | 94.8 | 91.8 | 92.7 |
| | F1 (%) | 85.1 | 85.6 | 86.5 | 89.4 | 86.8 | 88.3 | 88.8 | 89.2 | 90.0 | 90.2 | 90.6 | 91.5 |
| | RI | 0.729 | 0.934 | 0.921 | 0.944 | 0.933 | 0.937 | 0.939 | 0.960 | 0.957 | 0.937 | 0.955 | 0.962 |
| | F | 18 | 6 | 25 | 20 | 5 | 29 | 20 | 20 | 45 | 39 | 8 | 15 |
| Leukemia-2 | Acc. (%) | 87.7 | 88.6 | 89.6 | 90.5 | 91.0 | 91.5 | 92.4 | 92.4 | 92.9 | 93.4 | 94.3 | 95.3 |
| | F1 (%) | 89.7 | 88.2 | 90.1 | 90.1 | 90.5 | 90.0 | 92.0 | 91.9 | 92.4 | 92.9 | 93.4 | 94.3 |
| | RI | 0.684 | 0.822 | 0.731 | 0.807 | 0.719 | 0.738 | 0.848 | 0.847 | 0.845 | 0.862 | 0.852 | 0.877 |
| | F | 14 | 23 | 18 | 20 | 6 | 46 | 43 | 18 | 46 | 18 | 15 | 45 |
| Leukemia-3 | Acc. (%) | 88.0 | 88.9 | 89.9 | 90.8 | 89.3 | 91.8 | 90.3 | 92.5 | 93.2 | 90.7 | 90.6 | 95.6 |
| | F1 (%) | 87.9 | 90.4 | 88.3 | 90.2 | 90.7 | 89.2 | 91.5 | 86.1 | 88.6 | 92.0 | 89.0 | 94.5 |
| | RI | 0.739 | 0.717 | 0.759 | 0.801 | 0.753 | 0.769 | 0.816 | 0.822 | 0.821 | 0.816 | 0.795 | 0.840 |
| | F | 5 | 24 | 8 | 11 | 25 | 13 | 44 | 17 | 45 | 32 | 8 | 4 |
| Leukemia-4 | Acc. (%) | 86.0 | 90.6 | 88.5 | 89.5 | 90.0 | 90.0 | 90.0 | 89.0 | 88.8 | 90.3 | 90.3 | 94.2 |
| | F1 (%) | 88.2 | 87.7 | 88.6 | 89.0 | 88.0 | 90.8 | 89.6 | 92.8 | 92.0 | 89.4 | 91.0 | 93.8 |
| | RI | 0.622 | 0.703 | 0.611 | 0.745 | 0.628 | 0.734 | 0.743 | 0.791 | 0.856 | 0.866 | 0.874 | 0.895 |
| | F | 14 | 20 | 12 | 11 | 5 | 20 | 39 | 18 | 24 | 34 | 13 | 5 |
| Libras Movement | Acc. (%) | 82.5 | 83.4 | 86.3 | 85.0 | 85.7 | 86.1 | 86.6 | 87.0 | 87.5 | 87.9 | 88.0 | 89.7 |
| | F1 (%) | 83.0 | 85.0 | 84.5 | 85.4 | 85.8 | 84.0 | 86.7 | 89.2 | 87.6 | 88.1 | 87.2 | 89.4 |
| | RI | 0.710 | 0.753 | 0.753 | 0.847 | 0.842 | 0.736 | 0.721 | 0.848 | 0.727 | 0.851 | 0.844 | 0.915 |
| | F | 18 | 9 | 13 | 17 | 24 | 40 | 42 | 25 | 38 | 10 | 6 | 79 |
| Lung Cancer | Acc. (%) | 86.2 | 88.5 | 88.5 | 89.0 | 89.5 | 90.0 | 90.4 | 90.0 | 90.4 | 88.8 | 92.8 | 93.7 |
| | F1 (%) | 84.4 | 87.0 | 84.8 | 86.7 | 90.2 | 87.6 | 88.2 | 88.5 | 89.0 | 90.1 | 88.0 | 90.8 |
| | RI | 0.841 | 0.847 | 0.844 | 0.746 | 0.731 | 0.733 | 0.853 | 0.868 | 0.853 | 0.876 | 0.858 | 0.905 |
| | F | 17 | 5 | 8 | 15 | 12 | 44 | 29 | 13 | 24 | 25 | 6 | 4 |
| | Acc. (%) | 87.4 | 88.3 | 89.3 | 90.2 | 90.5 | 91.2 | 87.7 | 92.1 | 92.6 | 93.1 | 94.0 | 95.0 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Lymphoma | F1 (%) | 78.0 | 78.4 | 93.3 | 88.5 | 90.0 | 91.0 | 82.4 | 81.8 | 82.2 | 82.6 | 83.1 | 83.9 |
| | RI | 0.898 | 0.851 | 0.926 | 0.866 | 0.959 | 0.884 | 0.876 | 0.932 | 0.868 | 0.924 | 0.904 | 0.963 |
| | F | 7 | 21 | 19 | 11 | 7 | 22 | 29 | 14 | 42 | 35 | 8 | 218 |
| MEU-Mobile KSD | Acc. (%) | 85.3 | 86.2 | 87.1 | 88.1 | 88.5 | 89.0 | 89.5 | 89.9 | 90.4 | 90.8 | 91.8 | 92.7 |
| | F1 (%) | 83.1 | 87.6 | 84.5 | 85.4 | 85.8 | 86.3 | 90.7 | 90.2 | 89.6 | 88.1 | 88.5 | 89.4 |
| | RI | 0.889 | 0.874 | 0.827 | 0.931 | 0.908 | 0.883 | 0.849 | 0.882 | 0.909 | 0.959 | 0.893 | 0.918 |
| | F | 58 | 6 | 27 | 51 | 29 | 22 | 10 | 18 | 17 | 24 | 5 | 7 |
| MLL | Acc. (%) | 91.6 | 92.9 | 88.5 | 94.0 | 95.4 | 95.9 | 96.4 | 96.9 | 97.4 | 97.9 | 98.0 | 99.9 |
| | F1 (%) | 92.0 | 93.4 | 84.4 | 95.4 | 90.9 | 97.4 | 96.0 | 97.4 | 95.0 | 98.4 | 99.0 | 99.9 |
| | RI | 0.856 | 0.867 | 0.791 | 0.910 | 0.945 | 0.813 | 0.905 | 0.815 | 0.842 | 0.894 | 0.888 | 0.915 |
| | F | 7 | 5 | 5 | 14 | 20 | 38 | 20 | 30 | 39 | 20 | 9 | 5 |
| Ovarian Cancer | Acc. (%) | 90.2 | 92.1 | 92.1 | 93.4 | 93.6 | 94.4 | 94.6 | 95.5 | 95.5 | 96.0 | 97.0 | 98.0 |
| | F1 (%) | 91.0 | 89.4 | 88.4 | 90.5 | 89.9 | 95.0 | 92.9 | 90.0 | 97.0 | 96.3 | 94.8 | 97.8 |
| | RI | 0.840 | 0.838 | 0.857 | 0.857 | 0.850 | 0.872 | 0.853 | 0.904 | 0.852 | 0.854 | 0.868 | 0.887 |
| | F | 16 | 22 | 6 | 8 | 25 | 14 | 40 | 22 | 38 | 14 | 7 | 110 |
| SRBCT | Acc. (%) | 90.2 | 95.6 | 97.0 | 93.0 | 93.6 | 94.2 | 89.6 | 90.0 | 94.5 | 94.0 | 98.0 | 98.0 |
| | F1 (%) | 94.8 | 92.1 | 90.2 | 86.2 | 93.0 | 79.0 | 78.0 | 88.4 | 95.0 | 89.5 | 95.6 | 97.6 |
| | RI | 0.957 | 0.929 | 0.800 | 0.843 | 0.854 | 0.703 | 0.935 | 0.781 | 0.750 | 0.676 | 0.662 | 0.990 |
| | F | 19 | 15 | 13 | 16 | 6 | 30 | 18 | 20 | 36 | 36 | 10 | 230 |
| Waveform Database Generator | Acc. (%) | 80.4 | 79.4 | 79.1 | 79.9 | 80.3 | 86.7 | 85.2 | 81.6 | 82.4 | 83.6 | 83.3 | 84.1 |
| | F1 (%) | 83.2 | 77.6 | 82.5 | 80.3 | 83.7 | 81.2 | 79.6 | 83.0 | 69.9 | 82.8 | 84.0 | 84.1 |
| | RI | 0.896 | 0.849 | 0.819 | 0.778 | 0.833 | 0.693 | 0.802 | 0.869 | 0.824 | 0.816 | 0.954 | 0.914 |
| | F | 7 | 6 | 12 | 10 | 24 | 10 | 41 | 21 | 20 | 37 | 7 | 21 |

TABLE V. ABLATION STUDY RESULTS OF $R^5$-FRQR ACROSS DIVERSE DATASETS

| ID | Method Variant | Acc. | F1 | RS↓ | RI↑ | Δ↓ | γ | KI↑ | AJD↑ | SS↑ | ART (s) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| HCV for Egyptian Patients | Baseline FRQR | 71.8 | 67.2 | 0.005 | 0.944 | 8.0 | 0.827 | 0.69 | 0.65 | 0.52 | 25.6 |
| | +$R^2$ | 73.4 | 69.5 | 0.004 | 0.954 | 6.4 | 0.842 | 0.72 | 0.68 | 0.54 | 8.1 |
| | +$R^2$ + $R^3$ | 74.8 | 71.0 | 0.003 | 0.962 | 5.4 | 0.852 | 0.76 | 0.71 | 0.56 | 8.5 |
| | +$R^2$ + $R^3$ + $R^1$ | 75.9 | 72.1 | 0.002 | 0.968 | 5.9 | 0.859 | 0.78 | 0.72 | 0.57 | 10.2 |
| | +$R^2$ + $R^3$ + $R^1$ + $R^4$ | 76.7 | 72.6 | 0.002 | 0.968 | 5.8 | 0.861 | 0.80 | 0.73 | 0.58 | 11.0 |
| | $R^5$-FRQR (Ours) | 76.9 | 72.9 | 0.001 | 0.969 | 5.7 | 0.862 | 0.81 | 0.73 | 0.58 | 11.30 |
| Hill Valley | Baseline FRQR | 89.4 | 88.2 | 0.132 | 0.923 | 10.4 | 0.861 | 0.45 | 0.41 | 0.51 | 1228.3 |
| | +$R^2$ | 91.8 | 90.5 | 0.104 | 0.936 | 9.1 | 0.869 | 0.61 | 0.54 | 0.60 | 685.1 |
| | +$R^2$ + $R^3$ | 93.5 | 92.7 | 0.094 | 0.944 | 8.5 | 0.872 | 0.70 | 0.63 | 0.66 | 689.0 |
| | +$R^2$ + $R^3$ + $R^1$ | 94.7 | 94.0 | 0.087 | 0.951 | 7.9 | 0.875 | 0.74 | 0.67 | 0.72 | 755.3 |
| | +$R^2$ + $R^3$ + $R^1$ + $R^4$ | 95.3 | 94.9 | 0.080 | 0.983 | 7.5 | 0.877 | 0.75 | 0.70 | 0.74 | 766.3 |
| | $R^5$-FRQR (Ours) | 95.8 | 95.4 | 0.094 | 0.995 | 7.3 | 0.871 | 0.77 | 0.70 | 0.76 | 766.5 |
| Libras Movement | Baseline FRQR | 87.3 | 85.9 | 0.290 | 0.882 | 12.1 | 0.864 | 0.61 | 0.56 | 0.64 | 551.2 |
| | +$R^2$ | 88.5 | 86.7 | 0.280 | 0.898 | 9.6 | 0.881 | 0.67 | 0.61 | 0.67 | 478.4 |
| | +$R^2$ + $R^3$ | 89.0 | 87.1 | 0.280 | 0.905 | 7.9 | 0.890 | 0.70 | 0.65 | 0.68 | 489.0 |
| | +$R^2$ + $R^3$ + $R^1$ | 89.3 | 87.5 | 0.270 | 0.912 | 6.4 | 0.897 | 0.72 | 0.67 | 0.69 | 533.9 |
| | +$R^2$ + $R^3$ + $R^1$ + $R^4$ | 89.6 | 88.4 | 0.270 | 0.915 | 5.1 | 0.900 | 0.74 | 0.68 | 0.69 | 529.0 |
| | $R^5$-FRQR (Ours) | 89.7 | 89.4 | 0.270 | 0.915 | 9.3 | 0.900 | 0.74 | 0.68 | 0.70 | 568.16 |
| MEU-Mobile KSD | Baseline FRQR | 90.3 | 89.2 | 0.190 | 0.890 | 10.5 | 0.959 | 0.61 | 0.78 | 0.70 | 1470.3 |
| | +$R^2$ | 91.5 | 89.8 | 0.170 | 0.905 | 9.2 | 0.962 | 0.65 | 0.79 | 0.74 | 620.5 |
| | +$R^2$ + $R^3$ | 91.9 | 90.3 | 0.160 | 0.910 | 9.0 | 0.968 | 0.68 | 0.81 | 0.76 | 766.2 |
| | +$R^2$ + $R^3$ + $R^1$ | 92.3 | 90.8 | 0.150 | 0.912 | 8.8 | 0.970 | 0.69 | 0.82 | 0.78 | 823.0 |
| | +$R^2$ + $R^3$ + $R^1$ + $R^4$ | 92.6 | 91.1 | 0.140 | 0.916 | 8.5 | 0.973 | 0.70 | 0.82 | 0.79 | 825.5 |
| | $R^5$-FRQR (Ours) | 92.7 | 89.4 | 0.148 | 0.918 | 8.8 | 0.974 | 0.70 | 0.82 | 0.80 | 874.14 |
| Ovarian Cancer | Baseline FRQR | 96.7 | 95.4 | 0.190 | 0.863 | 1.8 | 0.983 | 0.70 | 0.67 | 0.43 | 9283.7 |
| | +$R^2$ | 97.3 | 96.0 | 0.170 | 0.880 | 1.2 | 0.990 | 0.72 | 0.68 | 0.42 | 6115.1 |
| | +$R^2$ + $R^3$ | 97.8 | 96.9 | 0.150 | 0.884 | 0.6 | 0.996 | 0.75 | 0.69 | 0.44 | 6452.5 |
| | +$R^2$ + $R^3$ + $R^1$ | 98.0 | 97.5 | 0.130 | 0.886 | 0.2 | 0.999 | 0.76 | 0.69 | 0.46 | 6937.4 |
| | +$R^2$ + $R^3$ + $R^1$ + $R^4$ | 98.0 | 97.6 | 0.120 | 0.887 | 0.0 | 1.000 | 0.76 | 0.69 | 0.48 | 7098.3 |
| | $R^5$-FRQR (Ours) | 98.0 | 97.8 | 0.117 | 0.887 | 0.0 | 1.000 | 0.76 | 0.69 | 0.49 | 7128.71 |
| Waveform Database Generator | Baseline FRQR | 80.1 | 78.2 | 0.090 | 0.881 | 17.5 | 0.922 | 0.78 | 0.92 | 0.88 | 893.4 |
| | +$R^2$ | 82.0 | 80.1 | 0.080 | 0.899 | 15.2 | 0.934 | 0.81 | 0.96 | 0.89 | 143.6 |
| | +$R^2$ + $R^3$ | 83.5 | 81.5 | 0.070 | 0.906 | 14.3 | 0.942 | 0.82 | 0.98 | 0.90 | 144.9 |
| | +$R^2$ + $R^3$ + $R^1$ | 83.9 | 83.0 | 0.070 | 0.910 | 13.8 | 0.949 | 0.83 | 0.99 | 0.91 | 144.0 |
| | +$R^2$ + $R^3$ + $R^1$ + $R^4$ | 84.1 | 83.7 | 0.070 | 0.914 | 13.6 | 0.959 | 0.84 | 0.99 | 0.92 | 148.0 |
| | $R^5$-FRQR (Ours) | 84.1 | 84.1 | 0.071 | 0.914 | 13.5 | 0.960 | 0.84 | 1.00 | 1.00 | 148.70 |

## VI. Conclusion and Future work

This paper introduced R⁵-FRQR, a fuzzy rough set-based feature selection framework that integrates five complementary principles; reliability, relational awareness, robustness, reductiveness, and resilience to address some of the limitations of classical FRQR in high-dimensional, noisy, and cross-domain scenarios. Unlike marginally greedy FRQR variants that overlook feature interdependencies and are prone to instability and domain overfitting, R⁵-FRQR proposes an interaction-aware scoring function that captures both individual relevance and pairwise complementarity, a granular similarity approximation that reduces fuzzy relation complexity via clustering, bootstrap-based stability filtering for reproducible selection, and a domain-adaptive dependency metric that promotes cross-domain generalization. Experiments on 16 diverse benchmark datasets demonstrate that R⁵-FRQR consistently achieves higher performance than 11 state-of-the-art methods across classification accuracy, robustness, stability, and generalization.

Future work will explore higher-order interaction modeling such as Shapley values, adaptive clustering granularity for improved approximation fidelity, and multi-objective optimization to better balance performance metrics. The model code can be accessed here[2].

## References


1. Zhang, X., Mei, C., Li, J., Yang, Y., & Qian, T. (2023). Instance and feature selection using fuzzy rough sets: A bi-selection approach for data reduction. *IEEE Transactions on Fuzzy Systems, 31*(6), 1981-1994.
2. Huang, Z., Li, J., & Wang, C. (2024). Robust feature selection using multigranulation variable-precision distinguishing indicators for fuzzy covering decision systems. *IEEE Transactions on Systems, Man, and Cybernetics: Systems, 54*(2), 903-914.
3. Chakraborty, R., & Pal, N. R. (2015). Feature selection using a neural framework with controlled redundancy. *IEEE Transactions on Neural Networks and Learning Systems, 26*(1), 35-50.
4. Hancer, E., Xue, B., & Zhang, M. (2024). A multimodal multiobjective evolutionary algorithm for filter feature selection in multilabel classification. *IEEE Transactions on Artificial Intelligence, 5*(9), 4428-4442.
5. Hu, Y., Zhang, Y., Gong, D., & Sun, X. (2023). Multiparticipant federated feature selection algorithm with particle swarm optimization for imbalanced data under privacy protection. *IEEE Transactions on Artificial Intelligence, 4*(5), 1002-1016.
6. Xu, W., Yuan, Z., & Liu, Z. (2024). Feature selection for unbalanced distribution hybrid data based on k-nearest neighbourhood rough set. *IEEE Transactions on Artificial Intelligence, 5*(1), 229-243.
7. Banerjee, S., Bhuyan, D., Elmroth, E., & Bhuyan, M. (2024). Cost-efficient feature selection for horizontal federated learning. *IEEE Transactions on Artificial Intelligence, 5*(12), 6551-6565.
8. Zabihimayvan, M., & Doran, D. (2019). Fuzzy rough set feature selection to enhance phishing attack detection. In *Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)* (pp. 1-6).
9. Yang, X., Chen, H., Wang, H., et al. (2023). Feature selection with local density-based fuzzy rough set model for noisy data. *IEEE Transactions on Fuzzy Systems, 31*(5), 1614-1627.
10. Xu, S., Shang, L., Liu, K., Ju, H., Yang, X., & Pedrycz, W. (2026). Margin-aware fuzzy rough feature selection: Bridging uncertainty characterization and pattern classification. *IEEE Transactions on Fuzzy Systems, 34*(1), 138-151.
11. Gu, X., Li, Y., & Jia, J. (2015). Feature selection for transient stability assessment based on kernelized fuzzy rough sets and memetic algorithm. *International Journal of Electrical Power & Energy Systems, 64*, 664-670.
12. Farahbakhshian, S. F., & Ahvanooey, M. T. (2020). A new gene selection algorithm using fuzzy-rough set theory for tumor classification. *Control Engineering and Applied Informatics, 22*(1), 14-23.
13. Jensen, R., & Shen, Q. (2004). Semantics-preserving dimensionality reduction: Rough and fuzzy-rough-based approaches. *IEEE Transactions on Knowledge and Data Engineering, 16*(12), 1457-1471.
14. Jensen, R., & Shen, Q. (2009). New approaches to fuzzy-rough feature selection. *IEEE Transactions on Fuzzy Systems, 17*(4), 824-838.
15. Anaraki, J. R., Samet, S., Eftekhari, M., & Ahn, C. W. (2018). A fuzzy-rough feature selection based on binary shuffled frog leaping algorithm. *International Journal of Computer and Information Engineering, 12*(9), 722-729.
16. Xue, Y., Chen, Y., Hu, Y., & Chen, H. (2017). Fuzzy rough set algorithm with binary shuffled frog-leaping (BSFL-FRSA): An innovative approach for identifying main drivers of carbon exchange in temperate deciduous forests. *Ecological Indicators, 83*, 41-52.
17. Huang, W., She, Y., He, X., & Ding, W. (2023). Fuzzy rough sets-based incremental feature selection for hierarchical classification. *IEEE Transactions on Fuzzy Systems, 31*(10), 3721-3733.
18. Sang, B., Chen, H., Wan, J., et al. (2022). Self-adaptive weighted interaction feature selection based on robust fuzzy dominance rough sets for monotonic classification. *Knowledge-Based Systems, 253*, 1-14.
19. Maji, P., & Garai, P. (2013). Fuzzy-rough simultaneous attribute selection and feature extraction algorithm. *IEEE Transactions on Cybernetics, 43*(4), 1166-1177.
20. Yin, S., Zhang, Y., & Karim, S. (2018). Large scale remote sensing image segmentation based on fuzzy region competition and Gaussian mixture model. *IEEE Access, 6*, 26069-26080.
21. Jebarani, P. E., Umadevi, N., Dang, H., & Pomplun, M. (2021). A novel hybrid k-means and GMM machine learning model for breast cancer detection. *IEEE Access, 9*, 146153-146162.
22. Uykan, Z. (2023). Fusion of centroid-based clustering with graph clustering: An expectation-maximization-based hybrid clustering. *IEEE Transactions on Neural Networks and Learning Systems, 34*(8), 4068-4082.
23. Lu, J., Ma, G., & Zhang, G. (2024). Fuzzy machine learning: A comprehensive framework and systematic review. *IEEE Transactions on Fuzzy Systems, 32*(7), 3861-3878.
24. Hu, Q., Zhang, L., Zhou, Y., & Pedrycz, W. (2018). Large-scale multimodality attribute reduction with multi-kernel fuzzy rough sets. *IEEE Transactions on Fuzzy Systems, 26*(1), 226-238.
25. Xue, B., Zhang, M., Browne, W. N., & Yao, X. (2016). A survey on evolutionary computation approaches to feature selection. *IEEE Transactions on Evolutionary Computation, 20*(4), 606-626.


[2] https://github.com/simo2020-arch/R5-FRQR/blob/main/r5_frqr.py

26. Liang, C. (2024). Construction of soft computing anomaly detection model based on independence criteria and maximum entropy clustering algorithm. *IEEE Access, 12*, 111111-111125.
27. Naz, H., Saba, T., Alamri, F. S., Almasoud, A. S., & Rehman, A. (2024). An improved robust fuzzy local information k-means clustering algorithm for diabetic retinopathy detection. *IEEE Access, 12*, 78611-78623.
28. Rong, M., Gong, D., & Gao, X. (2019). Feature selection and its use in big data: Challenges, methods, and trends. *IEEE Access, 7*, 19709-19725.
29. Chen, D., Zhang, L., Zhao, S., Hu, Q., & Zhu, P. (2012). A novel algorithm for finding reducts with fuzzy rough sets. *IEEE Transactions on Fuzzy Systems, 20*(2), 385-389.
30. Wang, C., Qi, Y., & Shao, M. (2017). A fitting model for feature selection with fuzzy rough sets. *IEEE Transactions on Fuzzy Systems, 25*(4), 741-753.
31. Parthaláin, N. M., Jensen, R., & Diao, R. (2020). Fuzzy-rough set bireducts for data reduction. *IEEE Transactions on Fuzzy Systems, 28*(8), 1840-1850.
32. Wang, C., Wang, C., Qian, Y., & Leng, Q. (2024). Feature selection based on weighted fuzzy rough sets. *IEEE Transactions on Fuzzy Systems, 32*(7), 4027-4037.
33. Wang, C., Wang, C., An, S., Ding, W., & Qian, Y. (2025). Feature selection and classification based on directed fuzzy rough sets. *IEEE Transactions on Systems, Man, and Cybernetics: Systems, 55*(1), 699-711.
34. Dai, J., Hu, Q., Hu, H., & Huang, D. (2018). Neighbour inconsistent pair selection for attribute reduction by rough set approach. *IEEE Transactions on Fuzzy Systems, 26*(2), 937-950.
35. Ren, Y., Zhu, X., Bai, K., & Zhang, R. (2023). A new random forest ensemble of intuitionistic fuzzy decision trees. *IEEE Transactions on Fuzzy Systems, 31*(5), 1729-1741.
36. Loh, Y.-W. (2011). Classification and regression trees. *WIREs Data Mining and Knowledge Discovery, 1*, 14-23.
37. Moghaddam, V. H., & Hamidzadeh, J. (2016). New Hermite orthogonal polynomial kernel and combined kernels in support vector machine classifier. *Pattern Recognition, 60*, 921-935.
38. Baek, S., & Kim, D. Y. (2017). Empirical sensitivity analysis of discretization parameters for fault pattern extraction from multivariate time series data. *IEEE Transactions on Cybernetics, 47*(5), 1198-1209.
39. Sysoev, A. (2023). Sensitivity analysis of mathematical models. *Computation, 11*(8), 1-13.
40. Meinshausen, N., & Bühlmann, P. (2010). Stability selection. *Journal of the Royal Statistical Society, 72*(4), 417-473.
41. Spielman, D. A. (2007). Spectral graph theory and its applications. In *Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science (FOCS'07)* (pp. 29-38).
42. Spolaôr, N., Cherman, E. A., Monard, M. C., & Lee, H. D. (2013). ReliefF for multi-label feature selection. In *Proceedings of the Brazilian Conference on Intelligent Systems* (pp. 6-11).
43. Demir-Kavuk, O., Kamada, M., Akutsu, T., & Knapp, E. W. (2011). Prediction using step-wise L1, L2 regularization and feature selection for small data sets with large number of features. *BMC Bioinformatics, 12*, 1-10.
44. Wang, X., Tao, Y., & Zheng, K. (2018). Feature selection methods in the framework of mRMR. In *Proceedings of the International Conference on Instrumentation & Measurement, Computer, Communication and Control (IMCCC)* (pp. 1490-1495).
45. Anaraki, J. R., & Eftekhari, M. (2011). Improving fuzzy-rough quick reduct for feature selection. In *Proceedings of the 19th Iranian Conference on Electrical Engineering* (pp. 1-1).
46. Antonelli, M., Ducange, P., & Marcelloni, F. (2013). Feature selection based on fuzzy mutual information. In *Proceedings of the 10th International Workshop on Fuzzy Logic and Applications* (pp. 36-43).
47. Rao, N., Nowak, R., Cox, C., & Rogers, T. (2016). Classification with the sparse group Lasso. *IEEE Transactions on Signal Processing, 64*(2), 448-463.
48. He, D., Sun, S., & Xie, L. (2024). Multi-target HSIC-Lasso via exploiting target correlations. *Expert Systems with Applications, 242*, 1-19.
49. Kursa, M. B., & Rudnicki, W. R. (2010). Feature selection with the Boruta package. *Journal of Statistical Software, 36*(11), 1-13.
50. Taherkhani, A., Cosma, G., & McGinnity, T. M. (2018). Deep-FS: A feature selection algorithm for deep Boltzmann machines. *Neurocomputing, 322*, 22-37.
51. Fu, R., Wang, P., Gao, Y., & Hua, X. (2014). A new feature selection method based on relief and SVM-RFE. In *Proceedings of the 12th International Conference on Signal Processing (ICSP)* (pp. 1363-1366).



## Funding Declaration

This research received no external funding.