Uncertainty Discourse and Applications

www.uda.reapress.com

Uncert. Disc. Appl. Vol. 3, No. 3 (2026) 212–224.

Paper Type: Original Article

Interval Fuzzy Rough Feature Selection under Simulated Missingness and Training Label Noise

Trayambak Rai[1,*], Kuan Mo[2]

[1] School of Computer Engineering, KIIT (Deemed to Be) University, Bhubaneswar-751024, Odisha, India; trayambakrai@research.edu.

[2] Department of Earth and Environmental Sciences, Brooklyn College, State University of New York, New York, USA; kuan.mo@brooklyn.cuny.edu.

Citation:

Received: 23 November 2025 *Revised: 17 February 2026* *Accepted: 13 April 2026*	Rai, T., & Mo, K. (2026). Interval fuzzy rough feature selection under simulated missingness and training label noise. *Uncertainty Discourse and Applications, 3*(3), 212-224.

Abstract

Feature selection on incomplete tables requires a criterion whose uncertainty representation is distinguished from the preprocessing used in its evaluation. We formulate an interval fuzzy-rough framework based on admissible completions and examine a simpler implementation, Interval Type-2 Fuzzy-Rough Feature Selection (IT2-FRFS), using interval Gaussian similarity scales after mean imputation. Reported experiments cover breast cancer, wine, an 800-row digits subsample, and an 800-by-20 synthetic table under 15% Missing Completely at Random (MCAR) and 10% label noise in downstream training folds, with ten corruption seeds. Selection uses clean labels. The tested midpoint-gain search reaches its feature cap on every dataset, selecting all features on wine and the synthetic table. Mean k-nearest-neighbour accuracy is 0.8738, compared with 0.8964 for minimum Redundancy Maximum Relevance (mRMR) and 0.9093 for ReliefF. A fixed-width ablation on two datasets produces small accuracy differences, without establishing an advantage of interval scales. Logistic regression improves relative to all features on breast cancer and digits, but mRMR remains stronger. The results characterize this capped, mean-imputed implementation; they do not evaluate selection from noisy labels, completion envelopes, adaptive interval widths, or missing-at-random and missing-not-at-random mechanisms. Reported pooled significance tests require reconsideration of seed dependence and validation nesting before they support general inference.

Keywords: Interval type-2 fuzzy sets, Fuzzy-rough feature selection, Missing data mechanisms, Label noise, Selection stability, Robust data science.

1 | Introduction

Clinical prediction, credit modelling, and survey analysis often combine missing attributes with uncertain outcomes. Selecting a compact panel in such settings requires separating uncertainty about unobserved values from uncertainty about labels, and evaluating whether the selected panel remains useful across perturbations.

Corresponding Author: trayambakrai@research.edu.ir

Licensee System Analytics. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0).

These applications motivate the present construction; the experimental evidence here comes from controlled tabular benchmarks rather than from operational deployments.

Feature selection includes filter, wrapper, and embedded approaches [1], and reproducibility across perturbations is a separate property from predictive accuracy [2]. Missing Completely at Random (MCAR), Missing at Random (MAR), and Missing Not at Random (MNAR) refer to different relations between missingness and observed or unobserved values [3]. Label noise introduces another source of uncertainty [4]. A selection procedure that uses only a completed table may conceal its sensitivity to alternative completions, motivating an explicit interval representation.

Rough-set reduction preserves decision information through approximation quality [5]. Fuzzy-rough approaches replace crisp indiscernibility with graded relations [6], [7]. Interval type-2 representations describe uncertainty in membership degrees through lower and upper membership functions [8], building on fuzzy sets [9]. Direct interval fuzzy-rough predecessors provide the starting point for the present construction [10], [11]; the empirical contribution is the diagnosis of a Gaussian-scale midpoint-gain variant.

Imputation can support valid analysis when its assumptions and uncertainty are handled appropriately [12]. For feature selection, however, a single completion need not reveal the range of dependency values compatible with missing observations. We therefore distinguish a completion-based framework from the mean-imputed similarity-scale implementation evaluated below. This distinction determines which research questions the reported experiment can answer.

The methodological contribution is a completion-based definition of interval lower approximations and dependency, together with a conservative gain criterion. The evaluated implementation instead varies Gaussian similarity scales on a mean-imputed table and uses midpoint gains.

The empirical contribution is a comparison of that implementation with information gain, ReliefF, minimum Redundancy Maximum Relevance (mRMR), and all features on four datasets. Reported observations include cap-limited selection, poor k-nearest-neighbour accuracy on digits, and small differences in a fixed-width ablation.

Complexity and evidential scope are stated explicitly. Neither greedy forward selection nor a nonnegative gain condition establishes a minimal reduct, and an interval enclosure need not be a tight range over jointly admissible completions.

The experiment provides a diagnostic case for improving interval construction and stopping rules. Claims about general accuracy, label-noise resistance in selection, or adaptive envelopes require additional evidence.

Section 2 reviews the relevant selection and uncertainty models. Sections 3 and 4 define the completion-based framework and its search rule. Section 5 describes the distinct implementation used for the reported experiments. Sections 6 and 7 discuss the implications and remaining research questions.

2 | Related Work

2.1 | Feature Selection Paradigms and Stability

Filters rank features independently of a downstream learner, wrappers evaluate candidate subsets through a learner, and embedded methods couple selection to model fitting [1]. The experimental baselines focus on filters, including mutual-information ranking and mRMR relevance-minus-redundancy selection [13]. Their tuned subset sizes and the proposed selector's fixed cap must be considered when interpreting accuracy differences.

Kuncheva's [2] index measures chance-corrected overlap for fixed-size subsets. The experiment below reports mean pairwise Jaccard overlap instead. These metrics are not interchangeable, and the size of a selected subset affects their interpretation. Selecting the full feature set yields perfect Jaccard stability without demonstrating an informative reduction. Accuracy, subset size, and stability must therefore be read together.

2.2 | Rough and Fuzzy Rough Feature Selection

Rough-set reduction preserves a specified decision structure [5]. Fuzzy-rough methods permit graded relations and class approximations [6], [7], [14]. Jensen and Shen [10] explicitly extend fuzzy-rough selection to missing values using interval-valued fuzzy sets and interval dependency. Maji and Garai [11] develop interval type-2 fuzzy-rough attribute selection with a max-relevance/max-significance criterion. Interval representation within fuzzy-rough selection therefore predates this study. The present empirical object is a Gaussian-scale midpoint-gain variant evaluated after mean imputation, rather than a first interval fuzzy-rough criterion.

Incomplete information systems have been treated using relations that accommodate unknown values without requiring a unique imputation [15]. Such constructions motivate the use of admissible completion sets here. A specific completion-based dependency definition must state both its similarity relation and the decision classes it approximates. *Table 1* compares representative design choices; its entries describe the approaches considered, rather than an exhaustive classification of the literature.

Table 1. Representative selection-criterion families and their handling of missing values and label ambiguity (qualitative comparison of the approaches considered).

Family	Representative Method	Missingness Handling	Label Ambiguity Handling	Stability Profile
Rough-set reducts [5]	Discernibility-based reducts	Tolerance relations or complete-case	Crisp labels only	Sensitive to discretization
Type-1 fuzzy-rough [7], [14]	Fuzzy-rough dependency, greedy search	Imputation or deletion beforehand	Fuzzy decisions possible, point-valued	Moderate; imputation-sensitive
Information-theoretic filters [13]	Mutual information; mRMR	Imputation beforehand	Point labels expected	Moderate; redundancy-aware
Embedded sparsity	L1-regularized linear models	Imputation beforehand	Point labels expected	Depends on learner and tuning
Impute-then-select	mRMR after single imputation	Single-point imputation	Point labels expected	Inherits imputation variance
Proposed IT2-FRFS	Interval type-2 fuzzy-rough dependency	Completion-based enclosure in general framework; mean imputation in tested variant	Decision bounds proposed; clean selection labels tested	Stability evaluated, no general guarantee

2.3 | Missing Data and Label Noise

Missing-data theory supplies the MCAR, MAR, and MNAR distinction and explains the assumptions needed for incomplete-data analysis [3], [12]. Flexible imputation and chained equations provide alternative ways to represent missing-value uncertainty [16], [17]. The reported experiment uses uniform random masking only. Conclusions about informative missingness require a mechanism-specific design rather than extrapolation from the MCAR condition.

Label-noise models distinguish class-independent, class-dependent, and instance-dependent corruption [4]. Angluin and Laird [18] analyse learning from noisy examples, while Natarajan et al. [19] develop loss-based learning with noisy labels. The present experiment applies label noise only to downstream training folds; the selectors themselves use clean labels. It consequently evaluates prediction after selection under training noise, rather than robustness of the selection criterion to corrupted labels.

2.4 | Type 2 Uncertainty Modelling and Gap Synthesis

Interval type-2 (IT2) fuzzy sets use lower and upper membership functions to represent a footprint of uncertainty [8]. This representation is distinct from a calibrated probability interval. In feature selection, interval dependency and IT2 fuzzy-rough attribute criteria already appear in the direct predecessors above [10], [11]. The present construction and diagnostic experiment should be assessed against those methods, with explicit similarities, aggregation, and uncertainty assumptions.

The reviewed work supplies complementary tools for feature selection, incomplete-data analysis, and graded uncertainty. We combine these tools in a completion-based dependency framework, then evaluate a restricted similarity-scale implementation. This organization makes the theoretical construction explicit while keeping the empirical claim tied to the method actually described in the experimental setup.

3 | Problem Formulation and Research Questions

3.1 | Research Questions and Hypotheses

RQ1 asks how incomplete feature values and uncertain decision memberships can be represented in a well-defined selection criterion. RQ2 asks whether the criterion improves subset stability and downstream accuracy relative to type-1 selection and classical filters. RQ3 asks how interval width and a missingness-adaptive envelope affect selection. The reported midpoint-scale experiment addresses only a restricted version of these questions.

H1 concerns whether interval selection improves subset stability over type-1 selection and mRMR. H2 concerns slower degradation as label-noise rates increase. H3 concerns whether missingness-adaptive intervals improve stability over fixed widths. These are research hypotheses, without a verified preregistration record. The reported experiment offers only a limited comparison relevant to H1; it has no noise-rate sweep for H2 and implements no adaptive envelope for H3.

3.2 | System Model and Notation

Let U be a finite set of n objects and F a set of m features. The indicator M_{ij} equals one when x_{ij} is observed. A completion specifies all unobserved values within nonempty admissible sets and, when needed, class memberships $d_{i(c)}$ in [0,1]. Denote the joint admissible set by Ω. Fixed observed values are retained in every completion. A supplied soft-label vector alone does not determine an uncertainty interval; label bounds require an explicit calibration or elicitation rule. *Table 2* lists the notation

Table 2. Notation used throughout the manuscript.

Symbol	Meaning
U, n	Object set and its cardinality
F, m	Feature set and its cardinality
x_{ij}, M_{ij}	Cell value and missingness indicator (one if observed)
$\mu_{[L,U],ij}$	Interval type-2 membership bounds for cell (i, j)
$C_{ij}(\omega)$)	Admissible-completion envelope of a missing cell, with width parameter omega
$\rho_P(x_i, x_k)$	Fuzzy similarity between objects under feature subset P
$d_i(c)$	Soft-label probability of class c for object i
$\mu_{D,L,U]}(x_i)$	Interval-valued decision memberships
$\mathrm{LOW}(\rho_P, \mu_D)$	Fuzzy lower approximation of the decision, Eq. (2)
$POS_P(D)$	Maximum of class-specific lower approximations
$\gamma_P^{112}(D)$	Conservative dependency enclosure, Eq. (4)
$\Delta_{[L,U]}(f \mid P)$	Interval-valued dependency gain of adding feature f
λ, η	Injected missingness rate and label-noise rate

3.3 | Formal Problem Statement

For each cell, *Eq. (1)* bounds a chosen membership function over its admissible completion set. Infima and suprema are used because endpoint completions need not attain the extrema, and a nonlinear membership function need not have its extrema at the endpoints of the value interval. For a completion ω in Ω, define a per-feature similarity in [0,1] and aggregate it by the minimum over features in P, with similarity one for the empty subset. *Eq. (2)* defines the Kleene–Dienes lower approximation for each decision class c. Its positive region $\mathrm{iPOS_P(D)(x_i;\omega)s} = \max_c \mathrm{LOW}(\rho_P, d_c)(x_i;\omega)$. *Eqs. (3)* and *(4)* form conservative endpoint enclosures. The resulting Gaussian relation is a reflexive symmetric fuzzy tolerance; transitivity is not assumed, and equivalence-based reduct properties do not follow automatically.

$$\mu_A(x_{ij}) = [\mu_{L,ij}, \mu_{U,ij}], \mu_{L,ij} = \inf_{v \subset C_{ij}} \mu_A(v), \mu_{U,ij} = \sup_{v \subset C_{ij}} \mu_A(v). \quad (1)$$

$$\mathrm{LOW}(\rho_P, d_c)(x_i;\omega) = \min_{x_k \in U} \max\{1 - \rho_P(x_i, x_k;\omega), d_k(c;\omega)\}. \quad (2)$$

$$\mathrm{LOW}_{L,P,c}(x_i) = \inf_{\omega \in \Omega} \mathrm{LOW}(\rho_P, d_c)(x_i;\omega),\ \mathrm{LOW}_{U,P,c}(x_i) = \sup_{\omega \in \Omega} \mathrm{LOW}(\rho_P, d_c)(x_i;\omega). \quad (3)$$

$$\gamma_P(D;\omega) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{POS_P(D)(x_i;\omega)}, \underline{\gamma}_P = \frac{1}{n}\sum_{i=1}^{n} \inf_{\omega \in \Omega} \mathrm{POS_P(D)(x_i;\omega)}, \bar{\gamma}_P = \frac{1}{n}\sum_{i=1}^{n} \sup_{\omega \in \Omega} \mathrm{POS_P(D)(x_i;\omega)}. \quad (4)$$

The interval $[\underline{\gamma}_P, \bar{\gamma}_P]$ encloses the dependency at every joint completion, but the sum of objectwise extrema can be wider than the attainable range of γ_P. Define the gain by paired completions: $\Delta_L(f|P) = \inf_{\{\omega \in \Omega\}} [\gamma_{P\cup\{f\}}(D;\omega) - \gamma_P(D;\omega)]$ and Δ_U analogously with a supremum. If only dependency enclosures are available, $[\underline{\gamma}_{\{P\cup\{f\}\}} - \bar{\gamma}_P, \bar{\gamma}_{\{P\cup\{f\}\}} - \underline{\gamma}_P]$ is a conservative gain enclosure, not generally a tight range. The conservative search admits a candidate only when its lower gain bound exceeds a positive tolerance ε. It is a heuristic and does not solve a simultaneous maximum-dependency and minimum-cardinality problem.

f is admissible given P if $\Delta_L(f \mid P) \geq \varepsilon > 0$; f interval-dominates f′ if $\Delta_L(f \mid P) > \Delta_U(f' \mid P)$. **(5)**

4 | Proposed Methodology

4.1 | Architecture Overview

The completion-based framework constructs value and decision bounds, derives similarity relations, evaluates dependency enclosures, and performs a conservative forward search. Each stage requires an explicit parameterization and a common admissible set Ω. *Algorithm 1* states this conceptual search. The experiment in Section 5 evaluates a different midpoint-gain rule on imputed values and uncertain similarity scales.

Completion envelopes are estimated from observed training information and any defensible domain constraints. Decision-membership bounds require their own evidence or elicitation; they do not follow automatically from a soft-label vector.

Similarity bounds are propagated using operations that preserve enclosures. Sharing completion variables across objects can make independently propagated bounds conservative.

Class-specific lower approximations are combined into positive regions and normalized dependency values. Diagnostics report interval width and whether the enclosure remains informative.

The conservative search adds a feature with a lower gain bound above ε and stops when none meets that condition or a feature cap is reached. No minimal-reduct guarantee follows from this rule.

4.2 | Constructing Interval Memberships under Missingness and Label Ambiguity

For a missing cell, C_{ij} may be based on local observed values, domain knowledge, or a missingness model. Restricting the set improves informativeness only if the restriction is defensible: A narrow but misspecified envelope can exclude the true value. Envelope construction must use training information during predictive evaluation. The proposed local and mechanism-aware construction is not implemented in the reported experiment.

If annotator or adjudication information supplies uncertainty sets over class memberships, lower and upper decision bounds can be obtained by optimizing over those sets. Membership uncertainty does not specify a probability-to-interval conversion for labels. The chosen class-membership bounds, their dependence, and their calibration must therefore be stated before the framework can be used with ambiguous labels.

4.3 | Interval Type 2 Fuzzy Rough Dependency

The dependency criterion quantifies how objects similar under P belong to the same decision classes [6], [7]. Its completion-based enclosure separates the attained dependency at a fixed completion from uncertainty across completions. High lower bounds support stronger conservative statements than high midpoints alone; a wide interval indicates sensitivity or conservative bound propagation, rather than establishing that a point estimate is an imputation artefact.

Single imputation supplies one completed table, while multiple imputation explicitly propagates uncertainty under a statistical model [16], [17]. The present enclosure represents a different uncertainty set and is valid only relative to that set. It provides a conservative admission rule when the gain bound is computed correctly; it does not establish robustness to values or mechanisms excluded from Ω.

4.4 | Dominance Guided Greedy Search

Algorithm 1 starts with the empty subset and ranks admissible candidates by their lower gain bound, breaking ties by the upper bound and then a fixed feature order. With minimum aggregation of feature similarities and the Kleene–Dienes implication, dependency is nondecreasing as features are added at each fixed completion. Nonnegative gains alone therefore cannot identify a reduct. A positive tolerance limits small additions but does not ensure inclusion minimality or globally optimal cardinality.

A backward pass could test whether deleting any retained feature preserves a stated dependency target; such a pass and its target must be specified before calling the result a reduct. *Algorithm 1* returns a conservative subset and diagnostics. In Section 5, the tested implementation uses midpoint dependency gains, $\varepsilon = 0.002$, and a cap of 20 features; it does not implement the lower-bound admission rule or completion-envelope memberships.

Algorithm 1. Conceptual completion based conservative forward search.

Input: joint admissible set Ω, feature set F, decision classes, tolerance ε > 0, optional feature cap k_max

1: specify admissible value and decision-membership bounds
2: construct per-feature similarities for each admissible completion
3: use enclosure-preserving operations when exact completion optimization is unavailable
4: retain the common completion variables in paired gain definitions
5: initialize P ← ∅
6: while unselected features remain and |P| < k_max do
7: compute valid gain bounds Δ_L(f|P), Δ_U(f|P) for every f ∈ F \ P
8: form A ← {f ∈ F \ P : Δ_L(f|P) ≥ ε}
9: if A is empty, stop
10: choose f* maximizing Δ_L; break ties by Δ_U and fixed feature order
11: update P ← P ∪ {f*}
12: end while
13: record dependency widths and enclosure diagnostics
14: report whether the tolerance or feature cap ended the search
Output: conservative subset P and diagnostic report; no minimality certificate

4.5 | Complexity Assumptions and Failure Modes

A dense pairwise relation requires $O(n^2)$ storage. Evaluating a subset of p features directly costs $O(n^2p)$, including the lower approximations over a fixed number of decision classes. A forward search that evaluates all remaining features at k steps costs $O(n^2 \Sigma_{t = 0}^{k - 1}(m - t)(t + 1))$ without reuse; cached minimum relations can reduce repeated work but add storage and implementation choices. Exact optimization over coupled completions may introduce further cost. Restricting the infimum to sampled neighbours changes the lower approximation and generally raises it, so it is an approximation requiring validation. Calibrated envelopes and explicit label bounds are assumptions, not guarantees supplied by interval arithmetic.

5 | Experimental Results and Analysis

5.1 | Setup

The manuscript reports a four-dataset experiment: Breast cancer (569 objects, 30 features, two classes), wine (178, 13, three), an 800-row digits subsample (64 features, ten classes), and a synthetic make_classification table (800 objects, 20 features, eight informative and four redundant, class separation 0.8). The first three use scikit-learn datasets of UCI origin[1] [20]. Ten seeds inject 15% MCAR missingness. All selectors receive mean-imputed values and clean labels; 10% label noise is applied only to downstream training folds. Folds and corruption realizations are reported as shared across selectors. It remains necessary to verify whether imputation, scaling, ranking, and subset-size tuning were refitted inside each outer training fold, so the reported cross-validation scores require verification of that nesting.

The baselines are all features, mutual-information ranking, a ReliefF implementation with 200 sampled points and five nearest hits and misses [21], and relevance-minus-redundancy mRMR using quantile-binned mutual information [13]. The tested Interval Type-2 Fuzzy-Rough Feature Selection (IT2-FRFS) uses Gaussian similarities with scale bounds [0.8,1.4] times the feature standard deviation, dependency-interval midpoints for ranking and gain, ε = 0.002, and a cap of 20 features. This is an interval-scale implementation on completed data, rather than *Algorithm 1*. Baseline subset sizes are tuned using five-fold k-Nearest Neighbors (kNN) accuracy on clean labels; the IT2 count follows its stopping rule. Downstream evaluation uses stratified five-fold cross-validation with kNN (k = 5) and logistic regression. Stability is pairwise Jaccard overlap across ten

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

seeds. Two-dataset ablations use a point scale of 1.0 and wider scale bounds [0.6,1.8]. Reported Friedman, Wilcoxon, and Holm analyses treat 40 dataset–seed combinations as blocks [22–24]; their dependence is addressed in Section 5.5. The reported software is Python 3.12 with NumPy, SciPy, scikit-learn, statsmodels, and matplotlib on one CPU core.

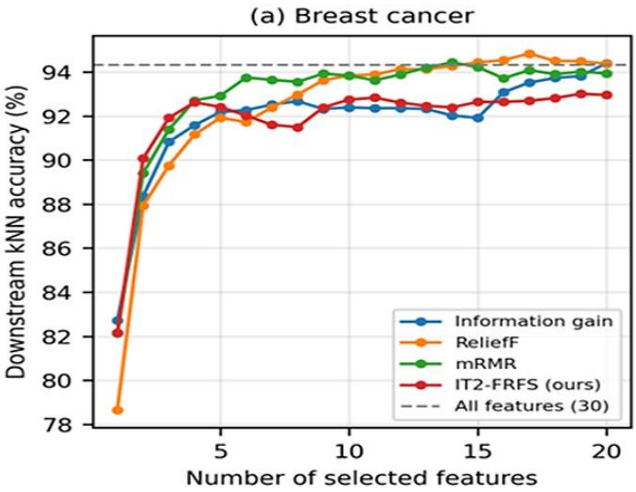
5.2|Main Comparison RQ1

Table 3 reports the principal comparison. Mean kNN accuracy across the four dataset means is 0.8738 for IT2-FRFS, 0.8964 for mRMR, 0.9093 for ReliefF, and 0.8951 for all features. The selector reaches min(20,m) features on every dataset: 20 on breast cancer and digits, 13 on wine, and 20 on the synthetic table. Its wine and synthetic scores therefore coincide with all features. Relative to all features, the displayed means imply deficits of 1.36 percentage points on breast cancer and 7.18 points on digits. ReliefF selects an average of 14.15 features across the four dataset means and has the highest kNN accuracy on breast cancer and the synthetic table; mRMR leads on wine and all features leads on digits. These descriptive differences characterize the reported pipeline.

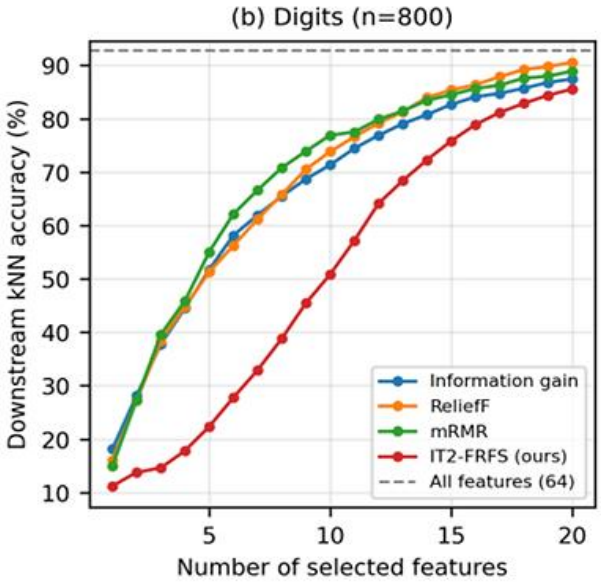
Fig. 1 shows that ranking behaviour varies by dataset. IT2-FRFS reaches approximately 91.9% on breast cancer with three features, but the mRMR curve is higher at that size. On digits its early ordering is substantially weaker than the other ranking selectors. Logistic-regression results in *Table 3* are more favourable to IT2-FRFS relative to all features: The displayed means imply gains of 1.17 percentage points on digits and 0.66 points on breast cancer. mRMR remains higher on both datasets. These observations are compatible with learner-specific effects of compact subsets; they do not isolate an interval-specific regularization mechanism.

Table 3. Main comparison over ten seeds (mean ± std): Downstream 5-fold CV accuracy, number of selected features, and selection stability.

Dataset	Selector	kNN accuracy	LR accuracy	Features	Jaccard
Breast cancer	All features	0.9431 ± 0.0101	0.9301 ± 0.0089	30.0 ± 0.0	1.000
Breast cancer	Information gain	0.9413 ± 0.0101	0.9357 ± 0.0121	19.2 ± 1.1	0.881
Breast cancer	ReliefF	0.9453 ± 0.0085	0.9385 ± 0.0089	17.1 ± 1.3	0.863
Breast cancer	mRMR	0.9427 ± 0.0087	0.9425 ± 0.0085	14.3 ± 3.4	0.650
Breast cancer	IT2-FRFS (ours)	0.9295 ± 0.0125	0.9367 ± 0.0118	20.0 ± 0.0	0.734
Wine	All features	0.9140 ± 0.0182	0.9128 ± 0.0160	13.0 ± 0.0	1.000
Wine	Information gain	0.9308 ± 0.0142	0.9257 ± 0.0230	9.2 ± 3.0	0.678
Wine	ReliefF	0.9319 ± 0.0157	0.9285 ± 0.0165	8.5 ± 2.5	0.631
Wine	mRMR	0.9399 ± 0.0173	0.9274 ± 0.0145	8.2 ± 2.5	0.689
Wine	IT2-FRFS (ours)	0.9140 ± 0.0182	0.9128 ± 0.0160	13.0 ± 0.0	1.000
Digits (n=800)	All features	0.9274 ± 0.0071	0.8024 ± 0.0107	64.0 ± 0.0	1.000
Digits (n=800)	Information gain	0.8741 ± 0.0089	0.8177 ± 0.0144	19.9 ± 0.3	0.837
Digits (n=800)	ReliefF	0.9037 ± 0.0084	0.8226 ± 0.0185	19.7 ± 0.7	0.723
Digits (n=800)	mRMR	0.8891 ± 0.0098	0.8271 ± 0.0122	19.9 ± 0.3	0.801
Digits (n=800)	IT2-FRFS (ours)	0.8556 ± 0.0165	0.8141 ± 0.0140	20.0 ± 0.0	0.381
Synthetic (800x20)	All features	0.7961 ± 0.0152	0.8317 ± 0.0096	20.0 ± 0.0	1.000
Synthetic (800x20)	Information gain	0.8277 ± 0.0180	0.8376 ± 0.0105	10.9 ± 2.4	0.724
Synthetic (800x20)	ReliefF	0.8562 ± 0.0157	0.8401 ± 0.0095	11.3 ± 1.3	0.897
Synthetic (800x20)	mRMR	0.8140 ± 0.0162	0.8364 ± 0.0076	8.6 ± 4.5	0.703
Synthetic (800x20)	IT2-FRFS (ours)	0.7961 ± 0.0152	0.8317 ± 0.0096	20.0 ± 0.0	1.000



a.



b.

Fig. 1. Downstream kNN accuracy versus the number of selected features for the four ranking-based selectors (mean over ten seeds). Dashed line: All-features accuracy at full dimensionality; a. Breast cancer (30 features), and b. Digits subsample (64 features).

5.3 | Ablation and Interval Diagnostics RQ2

Table 4 compares fixed point and interval scales on two datasets. On breast cancer all variants select 20 features, but equal counts do not imply identical subsets. The displayed kNN means range from 0.9283 to 0.9309, a spread of 0.26 percentage points, and Jaccard values span 0.017. On the synthetic table every variant selects all features and the scores coincide. Reported dependency widths of 0.324–0.361 are much larger than $\varepsilon = 0.002$, but width alone does not determine the midpoint gain or prove why stopping failed. The ablation offers no clear interval-scale benefit at this operating point and does not test the adaptive missingness envelopes of H3.

Table 4. Interval-width ablation on breast cancer and the synthetic table (mean ± std over ten seeds).

Dataset	Variant	kNN accuracy	LR accuracy	Features	Jaccard
Breast cancer	Type-1 (point scale)	0.9283 ± 0.0130	0.9369 ± 0.0118	20.0	0.739
Breast cancer	Interval [0.8, 1.4] (full)	0.9295 ± 0.0125	0.9367 ± 0.0118	20.0 ± 0.0	0.734
Breast cancer	Wide interval [0.6, 1.8]	0.9309 ± 0.0137	0.9382 ± 0.0115	20.0	0.751
Synthetic (800x20)	Type-1 (point scale)	0.7961 ± 0.0152	0.8317 ± 0.0096	20.0	1.000
Synthetic (800x20)	Interval [0.8, 1.4] (full)	0.7961 ± 0.0152	0.8317 ± 0.0096	20.0 ± 0.0	1.000
Synthetic (800x20)	Wide interval [0.6, 1.8]	0.7961 ± 0.0152	0.8317 ± 0.0096	20.0	1.000

5.4 | Robustness and Stability under the Joint Corruption RQ3

At the sole corruption setting, 15% MCAR plus 10% noise in downstream training labels, H2 cannot be evaluated as a degradation curve. *Table 3* gives ReliefF improvements over all features of 6.01 percentage points on the synthetic table and 1.79 points on wine; mRMR gains 2.59 points on wine. Every selected subset trails all features on digits. IT2-FRFS has higher displayed Jaccard overlap than mRMR on breast cancer (0.734 versus 0.650), lower overlap on digits (0.381 versus 0.801), and trivial perfect overlap where it selects all features. Pooled Jaccard means are approximately 0.779 for both IT2-FRFS and ReliefF and 0.711 for mRMR. These comparisons do not establish a general stability advantage, particularly because subset sizes differ.

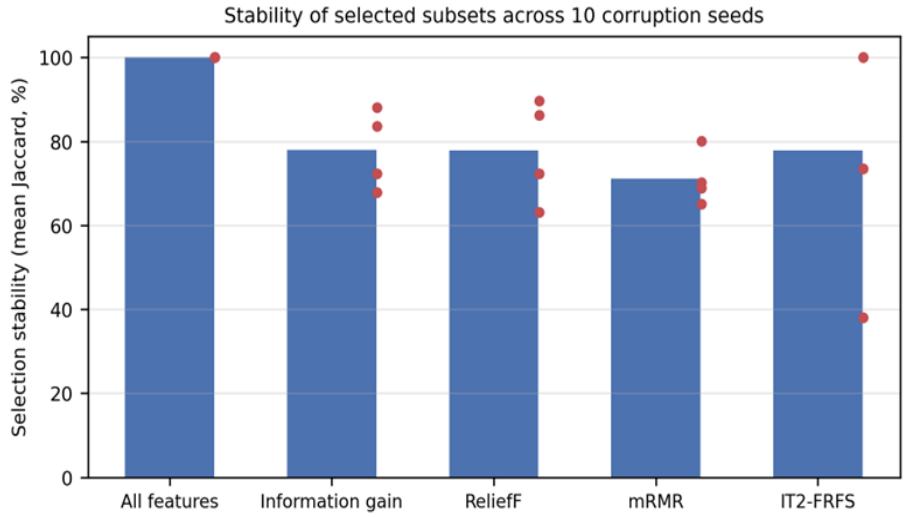


Fig. 2. Selection stability by selector: Mean pairwise Jaccard overlap of the selected sets across the ten corruption seeds. Bars: Mean over the four datasets; points: Per-dataset values.

5.5 | Statistical Analysis

Table 5 preserves the reported Friedman statistic ($\chi^2(4) = 56.73, p = 1.4 \times 10^{-11}$) and paired Wilcoxon results with Holm correction. The signs of the displayed pooled differences favour all four baselines over IT2-FRFS. However, repeated corruption seeds on each of only four datasets are not 40 independent sampled datasets. Treating them as interchangeable inferential blocks can overstate precision for cross-dataset claims, and raw paired outcomes are needed to verify ties, ranks, and p values. The tests are therefore reported as calculations from the original campaign, while the present conclusions rely on descriptive dataset-level differences rather than declaring a broadly established significance result [23].

Table 5. Paired Wilcoxon signed-rank tests on downstream kNN accuracy, IT2-FRFS versus each baseline, over 40 (dataset, seed) blocks, with Holm correction.

Comparison	W	p (raw)	p (Holm)	Mean diff (points)	Reject at 0.05
IT2-FRFS vs all features	1.0	1.03e-04	1.03e-04	-2.13	Yes
IT2-FRFS vs information gain	48.0	1.82e-06	3.63e-06	-1.97	Yes
IT2-FRFS vs ReliefF	17.0	3.77e-10	1.51e-09	-3.55	Yes
IT2-FRFS vs mRMR (primary)	21.0	2.61e-07	7.84e-07	-2.26	Yes

5.6 | Discussion Findings Losses and Scope

The reported implementation reaches its cap on every dataset and offers no consistent predictive or stability advantage over the established filters. Its poorest kNN result occurs on digits; full-set selection explains identical wine and synthetic outcomes. The fixed-width ablation changes breast-cancer kNN accuracy by at most 0.26 percentage points and is degenerate on the synthetic table. Logistic-regression improvements over all features on two datasets warrant learner-specific investigation, without establishing that interval scales caused the gains.

Mean imputation and midpoint gains distinguish the tested implementation from the completion-based framework. Missing values filled with column means can alter pairwise similarities, while wide scale intervals may change ordering without preserving the lower-gain admission criterion. These are plausible explanations for the observed cap behaviour, rather than identified causal mechanisms. Testing them requires traceable gain trajectories, fold-specific preprocessing, comparable subset budgets, and an implemented completion-envelope variant. Scaling ε relative to dependency or width is a candidate design change whose effect remains to be tested.

The evidence is limited to four datasets, one MCAR rate, one downstream training-noise rate, a 20-feature cap, kNN and logistic regression, and fixed-width ablations on two datasets. H2 and H3 remain untested in their stated form. H1 receives mixed descriptive evidence, partly confounded by full-set selection and differing subset sizes. The absence of confirmed validation nesting limits claims of out-of-sample prediction.

The numerical results are reported at proof-of-concept scale. Independent regeneration requires the experiment code, raw paired outcomes, fold indices, and execution record; a verified release is not specified.

6 | Discussion Limitations and Counterarguments

6.1 | Limitations

Completion-based bounds may be conservative because dependent uncertainties are propagated independently. Their validity depends on defensible completion sets, while their usefulness depends on informativeness. Dense similarity computation limits scale, and approximate neighbour relations change the criterion. A lower-gain rule can reject all candidates, whereas midpoint selection can keep adding features until a cap; these are different failure modes of different search rules. The reported experiment evaluates only the latter implementation.

6.2 | Alternative Viewpoints and Responses

Probabilistic completion models and multiple imputation offer alternatives for representing missing-data uncertainty [12]. Their validity rests on assumptions different from an admissible-set enclosure. Comparing these approaches should match training information and subset budgets. Likewise, missingness-aware prediction may reduce the value of selection for accuracy while leaving measurement cost and interpretability relevant. Stability and predictive accuracy should remain separate reported outcomes, especially where perfect overlap is obtained by retaining every feature.

7 | Conclusion and Future Directions

We defined a completion-based interval fuzzy-rough framework and examined reported results for a simpler mean-imputed, interval-scale selector. The tested IT2-FRFS consistently reaches its feature cap and has lower mean kNN accuracy than the four baselines. Its logistic-regression gains over all features on breast cancer and digits are modest and do not exceed mRMR. Fixed-width ablations show small differences on breast cancer and identical full-set outcomes on the synthetic table. These findings diagnose the tested implementation; they do not refute or validate the unimplemented completion-envelope and ambiguous-label construction.

The next experiment should implement the stated lower-gain search and completion envelopes, document their calibration, and fit every preprocessing and selection stage within training folds. Equal subset or computation budgets would separate ranking quality from tuning freedom. Noise-rate sweeps must corrupt the labels visible to selection if they are to test H2, and adaptive envelopes must be implemented to test H3. Larger and mechanism-structured datasets, together with raw outcomes and dependency-gain traces, would determine whether the proposed uncertainty representation adds practical value.

Author Contributions

T. R.: Conceptualization, study design, methodology, supervision, and project administration. K. M.: Data collection, formal analysis, interpretation of results, and original manuscript drafting.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data Availability

The reported public datasets are the breast-cancer Wisconsin diagnostic, wine, and digits collections distributed with scikit-learn, together with a synthetic make_classification table. The exact digits subsample, synthetic-generation settings, and corruption realizations require the experiment records.

Code availability

A verified public release of the experiment code and raw paired outcomes is not specified.

Conflicts of Interest

The authors declare no conflicts of interest.

Generative AI Disclosure

AI-assisted tools were used in drafting and revising the manuscript text. The authors retain responsibility for verification of the scientific content and the applicable disclosure

References

[1] Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of machine learning research*, *3*(3), 1157–1182. https://dl.acm.org/doi/pdf/10.5555/944919.944968

[2] Kuncheva, L. I. (2007). A stability index for feature selection. *Artificial intelligence and applications* (pp. 390–395). ACTA Press. https://lucykuncheva.co.uk/papers/lkAIA07.pdf

[3] Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*(3), 581–592. https://doi.org/10.1093/biomet/63.3.581

[4] Frénay, B., & Verleysen, M. (2013). Classification in the presence of label noise: A survey. *IEEE transactions on neural networks and learning systems*, *25*(5), 845–869. https://doi.org/10.1109/TNNLS.2013.2292894

[5] Pawlak, Z. (1982). Rough sets. *International journal of computer & information sciences*, *11*(5), 341–356. https://doi.org/10.1007/BF01001956

[6] Dubois, D., & Prade, H. (1990). Rough fuzzy sets and fuzzy rough sets. *International journal of general system*, *17*(2–3), 191–209. https://doi.org/10.1080/03081079008935107

[7] 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. https://doi.org/10.1109/TKDE.2004.96

[8] Mendel, J. M., & John, R. I. B. (2002). Type-2 fuzzy sets made simple. *IEEE transactions on fuzzy systems*, *10*(2), 117–127. https://doi.org/10.1109/91.995115

[9] Zadeh, L. A. (1965). Fuzzy sets. *Information and control*, *8*(3), 338–353. https://doi.org/10.1016/S0019-9958(65)90241-X

[10] Jensen, R., & Shen, Q. (2009). Interval-valued fuzzy-rough feature selection in datasets with missing values. *2009 IEEE international conference on fuzzy systems* (pp. 610–615). IEEE. https://doi.org/10.1109/FUZZY.2009.5277289

[11] Maji, P., & Garai, P. (2014). IT2 fuzzy-rough sets and max relevance-max significance criterion for attribute selection. *IEEE transactions on cybernetics*, *45*(8), 1657–1668. https://doi.org/10.1109/TCYB.2014.2357892

[12] Little, R. J. A., & Rubin, D. B. (2019). *Statistical analysis with missing data*. John Wiley & Sons. https://doi.org/10.1002/9781119482260

[13] 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

[14] Jensen, R., & Shen, Q. (2008). New approaches to fuzzy-rough feature selection. *IEEE transactions on fuzzy systems*, *17*(4), 824–838. https://doi.org/10.1109/TFUZZ.2008.924209

[15] Kryszkiewicz, M. (1998). Rough set approach to incomplete information systems. *Information sciences*, *112*(1–4), 39–49. https://doi.org/10.1016/S0020-0255(98)10019-1

[16] Van Buuren, S. (2018). *Flexible imputation of missing data*. Chapman and hall/CRC. https://doi.org/10.1201/9780429492259

[17] Van Buuren, S., & Groothuis-Oudshoorn, K. (2011). MICE: Multivariate imputation by chained equations in R. *Journal of statistical software*, *45*(3), 1–67. https://doi.org/10.18637/jss.v045.i03

[18] Angluin, D., & Laird, P. (1988). Learning from noisy examples. *Machine learning*, *2*(4), 343–370. https://doi.org/10.1007/BF00116829

[19] Natarajan, N., Dhillon, I. S., Ravikumar, P. K., & Tewari, A. (2013). Learning with noisy labels. *Advances in neural information processing systems* (pp. 1196–1204). Curran Associates, Inc. https://papers.nips.cc/paper/5073-learning-with-noisy-labels

[20] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., ...& Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of machine learning research*, *12*, 2825–2830. https://jmlr.org/papers/v12/pedregosa11a.html

[21] Kononenko, I. (1994). Estimating attributes: Analysis and extensions of relief. *European conference on machine learning* (pp. 171–182). Springer, Berlin, Heidelberg. https://doi.org/10.1007/3-540-57868-4_57

[22] Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics bulletin*, *1*(6), 80–83. https://doi.org/10.2307/3001968

[23] Demšar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *Journal of machine learning research*, *7*(1), 1–30. https://jmlr.org/papers/v7/demsar06a.html

[24] Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, *7*(2), 65–70. https://www.jstor.org/stable/4615733