Received: 14 May 2025 | Revised: 30 October 2025 | Accepted: 21 November 2025 | Published online: 30 January 2026

RESEARCH ARTICLE

Journal of Computational and Cognitive Engineering
2026, Vol. 00(00) 1–13
DOI: 10.47852/bonviewJCCE62026152

BON VIEW PUBLISHING

Intuitionistic Fuzzy Rough Mutual Information Aided Feature Subset Selection and Its Applications

Aneesh Kumar Mishra[1,*], Neelesh Kumar Jain[1] and Ravindra Kumar Singh[2]

[1]*Department of Computer Science and Engineering, Jaypee University of Engineering and Technology, India*

[2]*Department of Computer Science and Engineering, CHRIST University, India*

Abstract: Feature selection plays an essential role in solving the challenges of the "curse of dimensionality" in data analysis, aiming to increase the learning algorithms' performance. A key challenge in this field is achieving accurate attribute selection when handling both numerical and nominal attributes. To address this problem, we demonstrate a hybrid intuitionistic fuzzy (IF) similarity-based approach that flexibly handles mixed types of data for more precise attribute selection. The study shows an IF granular structure to manage noise in heterogeneous data and enhances the concepts of IF rough entropy, joint entropy, and conditional entropy to provide a comprehensive framework to deal with uncertainty. Moreover, IF rough mutual information is implemented to extract both uncertainty and the association between conditional attributes and the decision class, forming the basis of a novel attribute selection approach. The proposed algorithm contains intuitionistic fuzzification, evaluation of fuzzy mutual information to compute the significance of features, and recursive selection of the most important attributes, thus effectively reducing dimensionality. We set up a theoretical foundation using IF sets, IF information system, IF relation theory, and hybrid similarity relations, which ensures the robustness of the approach. The method is rigorously evaluated on seven benchmark datasets, showing superior performance on various metrics, including accuracy, sensitivity, and specificity, when compared with existing attribute selection approaches. Results demonstrate increased prediction of phospholipidosis-positive molecules with a sensitivity of 89.56%, a specificity of 92.63%, an accuracy of 95.98%, an AUC of 0.968, and an MCC of 0.908, which represents the strong class differentiation ability of the model. These findings underscore the effectiveness of the hybrid IF similarity-based attribute selection approach, which makes it a valuable tool for managing high-dimensional datasets and advancing the field of attribute selection.

Keywords: feature selection, intuitionistic fuzzy set, mutual information, rough set, similarity relation, granular structure

1. Introduction

In today's era, there is a rapid growth of data day by day at an unprecedented scale, which is larger in sample size and higher in dimensions. This reflects features that negatively impact the machine learning (ML) problem. It is problematic for traditional algorithms to manage high-dimensional data for effective model training. To mitigate the problem of redundant features from this sort of data, it is very critical to avoid the curse of dimensionality. Feature selection or reduction is emerging as a valuable method to mitigate such problems, which is an effective method to extract low-dimensional data from high-dimensional data on the basis of specific evaluation criteria. The primary goal is to reduce the data to make data analysis effective by extracting key characteristics from high-dimensional data by analyzing low low-dimensional subset of data. Currently, the major focus of feature selection is to recognize patterns to train ML algorithms effectively. It is essential to have appropriate evaluation criteria for selecting features, which directly impact the effectiveness of the process. Two approaches for making criteria functions for evaluation: the filter method and the wrapper method. The wrapper method utilizes a classifier for evaluating the selected subsets of features, but not all classifiers are suitable for this method. A classifier comfortable with wrapper methods must have the capability to handle high-dimensional features to produce accurate classification results even when the sample size is limited. While the filter method selects the features through an evaluation criterion like consistency [1], correlation [2], mutual information (MI) [3], and Euclidean distance [4], it is developed for the evaluation of the effectiveness of subsets of the features.

Rough set theory [5] is mainly used for selecting features. One of the approaches [6] consists of detecting the random relationships between conditional and decision features by leveraging a random probe-based strategy for identifying reducts. This approach confirms that attributes having common properties are not simultaneously selected, as identified through dependency analysis. Conventional rough set theory [7] relies on equivalence relations, which are best suited for data that is discrete in nature. However, there may be chances for loss of information by discretizing continuous data, which potentially affects the performance of learning methods. To mitigate this, a neighborhood rough set [8] was

*Corresponding author: Aneesh Kumar Mishra, Department of Computer Science and Engineering, Jaypee University of Engineering and Technology, India. Email: 193b001@juetguna.in; aneeshjuetphdcse2k19@gmail.com

proposed, which considers the neighbor instances and gives a feature reduction algorithm that eliminates redundant and irrelevant attributes effectively. Moreover, an enhanced neighborhood rough set model [9] has been introduced, with a corresponding feature selection, to avoid the issue due to discretization up to a certain extent. An improved rough set model has also been developed, also known as extension of knowledge based on the rough approximation (EKRA), to reduce boundary regions [10], resulting in more accurate feature selection. In fuzzy rough sets, the similarity between objects is assessed using fuzzy relations instead of crisp equivalence relations [11]. This method overcomes the limitations of conventional rough set theory to handle continuous data [12], which enables it to directly evaluate real-valued data. Fuzzy rough set theory has proved its flexibility in preprocessing the data and has been applied to select the features [13] successfully. To increase the precision of the measure of fuzzy relations, a Gaussian kernel rough set has been given, with some limitations, like the long-tail phenomenon in high-dimensional data. To improve this, a new Student-*t* kernelized fuzzy rough set model (SKFRS) has been proposed by Yang et al. [14], which consists of fuzzy rough feature selection algorithms. Uncertain information is evaluated by SKFRS by leveraging fuzzy divergence and is effective in noise reduction for discrete or real-valued data, without any requirement of user-supplied information. One of the approaches [15] that leverages the discernibility matrix for reducing the dimensionality, on the other hand, is critical reduction based on fuzzy rough sets [16], which identifies key conditional attributes for specified decision classes. Fuzzy rough feature selection has also been implemented within a graph theory framework [17] to show the time-consuming behavior of existing methods for large-scale datasets. For managing the multi-modal features in classification, a multi-kernel fuzzy rough set model has been proposed by Mondal and Singh [18] for dealing with fuzziness in this type of data. Another approach by An et al. [19] that is focused on selecting the features on the basis of fuzzy relationships between the attributes and decisions takes into account class densities that vary in the data distributions. Since fuzzy rough dependency functions do not give classification errors, the alternative approaches [20] rely on inner product dependency, and last, an adaptive weighted kernel density estimation method [21] has been given to mitigate the limitations of traditional feature selection approaches, especially in calculating correlations [22], redundancies, and noise in feature data. While these approaches mostly rely on dropping features that are redundant and irrelevant, moreover, processing such as extracting the features sometimes results in crucial information through these features.

However, misclassification of samples can result when there is significant overlap between several categories of data [23]. Additionally, traditional methods only reflect the membership of a sample to a set by limiting their ability to handle the latter uncertainty. This makes the requirement for the model to not only fit the data effectively but also handle later uncertainty that increases due to the presence of uncertainty not only in judgment but also in identification. This is where intuitionistic fuzzy (IF) sets [24] manage uncertainty by taking membership and non-membership of a sample. On the other hand, rough and IF sets both reflect the context of imprecision. It has been explained in research by Jensen and Shen [25] that the lower and upper approximations of IF rough sets are themselves IF sets. In recent years, various IF rough set models have been proposed: Jensen and Shen [25] proposed a method for attribute reduction using fuzzy rough method; Chakrabarty et al. [26] proposed concept of IF rough set; Zhan and Sun [27] also proposed IF rough sets with applications in multi-attribute for decision-making; Pandey et al. [28] proposed a method for selecting features using IF entropy method; De et al. [29] proposed a database for IF set; Zhang et al. [30] demonstrated a method using fuzzy rough sets based on overlap function for feature selection and edge extraction of images; Mahmood et al. [31] introduced a confidence level aggregation operators on the bases of IF rough sets; Mazarbhuiya and Shenify [32] proposed method for anomaly detection based on IF rough set; Xin et al. [33] proposed a novel method on for selecting attribute on the basis of IF three-way cognitive clustering; and Huang et al. [34] developed a risk evaluation model for effect analysis using IF rough method and implemented on various decision-making problems. It has also been examined by Tiwari et al. [35] that the relation between rough sets and IF sets shows that fuzzy rough sets can be taken into account as intuitionistic fuzzy sets (IFSs). Further research shows the various applications for IF rough set models, which include feature selection and reduction of attribute techniques: one such method, established by Hu et al. [36], uses genetic algorithms for attribute reduction using fuzzy kernel-induced relations; Esmail et al. [37] introduced rough set theory for IF information systems; Huang et al. [38] proposed a dominance approach on the basis of rough set in IF information systems; a method for attribute reduction in an categorical decision information is proposed by He et al. [39] on the basis of fuzzy rough set; Shreevastava et al. [40] proposed an IF neighborhood rough set model for selecting features; and Tiwari et al. [41] discussed new approaches for attribute reduction on the basis of IF rough set to reduce the attributes in the IF information system. Another model was developed by Huang et al. [38], leveraging a distance function, and was generalized for the reduction of attributes. Moreover, methods that were based on discernibility matrix [36, 42] have been proposed for the reduction of attributes, while others are the combination of fuzzy sets and information entropy to develop new reduction algorithms. The properties of IF rough sets are examined in several studies that present methods for the reduction of attributes and the extraction of rules. Other models have been implemented for the selection of a subset of attributes, and many feature subset selection [40, 43] have been proposed for supervised and semi-supervised datasets. Additionally, an IF clustering algorithm [23] has been developed for selecting the features for tracking the objects. Nowadays, a number of studies [41, 44–46] have introduced the IF rough set model, which has been implemented to select features. Despite these advancements, there is no single method that is capable of fully addressing both dataset [47] fitting and sample classification differentiation.

The IF set [24] is an enhanced generalization of the traditional fuzzy set, which incorporates both the membership degree and non-membership degree for classes in target data. A nuanced representation of inherent ambiguities is provided by this model in the real-world information compared to conventional fuzzy sets. The integration of IF set with rough set theory for the development of the IF rough set uses the approximations of rough set for describing the IF sets [27, 48, 49] by linking IF relations between the objects. Zhou and Wu [50] and Zhou et al. [51] explored the approximation operators for the IF rough set with the help of constructive and axiomatic approaches. Moreover, IF rough sets model is developed for interval-valued data by analyzing hierarchical structures and uncertainty measures. For IF rough models, a generalized framework has been provided by Ghosh et al. [52] on the basis of binary relations between two universes, and the reduction of attributes has been studied by Xin et al. [33] and Hesameddini et al. [53] with structures of reductions that are constructed with the help of discernibility matrices. These studies reflect the increasing attention of IF rough set in the research community. However, there is a great focus on generalization [54–56] of models, exploration of property [57–59], and descriptions of measures [35, 60] with limited application in areas such as granular computing and feature selection.

A hybrid IF similarity relation is an algebraic structure formulated to evaluate the similarity between data objects with both numerical and categorical features, using the concepts of IFSs [61]. In IFSs, every component has a degree of membership (μ), a degree of non-membership (ν), and a hesitation degree $(\pi) = (1 - \mu - \nu)$. The hybrid component is about integrating various similarity measures specifically suited to address different types of data within a single IF paradigm. For example, numerical features could be compared based on their distance-based measures, and categorical features could implement set-theoretic or probabilistic measures. This hybridization allows a comprehensive similarity measurement that can cater to the heterogeneity of real-world data.

The hybrid IF similarity relation is more suitable for heterogeneous data since it effectively works on mixed types of data by utilizing specific measures of similarity for multiple types of features, that is, numerical and categorical data. Integrating the principles of IF set and rough set, it captures membership, non-membership, and hesitation degrees, thus allowing uncertainty and vagueness present in real-world data. This fine-grained approach supports a better similarity evaluation compared to classical crisp approaches. In turn, this improves decision-making in multiple applications, such as medical diagnosis, pattern recognition, and feature selection, by suitably extracting the similarity between sophisticated data objects.

The recursive selection of features begins by training a random forest (RF) model on the set of whole features to compute the relative importance (RI) of every feature, based on information gain across all trees. Features are ranked according to RI, and the features with less importance are excluded in each iteration. A new RF is then trained on the reduced set of features, and the importance of the feature is computed. This process repeats recursively. The criterion for removing features at every step is their RI ranking. The process does not internal stopping criterion; instead, multiple subset sizes of features are evaluated eventually on the basis of the performance of the model on the validation set. The optimal number of features is selected on the basis of the subset that results in high validation accuracy.

The study offers multiple notable contributions. It demonstrates a new hybrid IF similarity relation, which is capable of handling both nominal and numerical features, giving a more comprehensive approach for the feature analysis. To handle noise within mixed data, an IF granular structure is proposed, which enhances the resilience of the data. The study further demonstrates the IF rough entropy concepts, joint entropy, and conditional entropy that allow a deeper understanding of uncertainty in information entropy. Moreover, it gives IF rough MI to use uncertainty and the correlation between conditional features and decision classes. Utilizing this approach, the study proposes a novel feature selection approach that prioritizes relevant features while reducing dimensionality. Finally, a framework built on these proposed methods is provided specifically to improve the predictive accuracy for finding phospholipidosis-positive molecules, which showcase the practical effectiveness of the method.

2. Theoretical Background

2.1. Illustration: IF set

If set A in V is a storage of well-defined objects that is given as follows:

$$A = \{\langle v, \alpha_A(v), \beta_A(v)\rangle | v \in V\} \tag{1}$$

Let V represent the set of data points. Moreover, $\alpha_A : V \to [0, 1]$and $\beta_A : V \to [0, 1]$ satisfy the condition $0 \leq \alpha_A(v) + \beta_A(v) \leq 1$ for $v \in V$. Here, $\alpha_A(v)$ and $\beta_A(v)$ denote the membership and non-membership degrees of an element $v \in V$, respectively.

Additionally, $\pi_A(v) = 1 - \alpha_A(v) - \beta_A(v)$ shows the degree of hesitancy of the given element $v \in V$. Hence, A is referred to as the IF set.

2.2. Illustration: Intuitionistic fuzzy information system (IFIS)

An intuitionistic fuzzy information system (IFIS) is described as a structure $S = (W, A \cup B, I, \text{IF})$, where W represents a non-empty finite set of objects referred to as the universe. The set A shows the conditional features, while $B = \{b\}$ is a singleton set consisting of the decision feature, with $A \cap B = \phi$. The set I shows the collection of all IF values and is composed of two subsets I_1 and I_2, corresponding to the domains of conditional and decision attributes, respectively. The information function IF maps each object-attribute pair to an IF value, that is, $IF : W \times (A \cup B) \to I$, where $IF(x, a) \in I_1 \; \forall a \in A$ and $IF(x, b) \in I_2$ are for the decision attribute $b \in B$. Each value $IF(x, a)$ or $IF(x, b)$ is denoted as a pair $(\alpha(x), \beta(x))$, where $\alpha(x)$ and $\beta(x)$ show the membership and non-membership degrees of the object x under the given attribute. When $\alpha(x) + \beta(x) = 1$ for all objects and attributes, the IFIS minimizes to a standard fuzzy information system. Hence, a fuzzy information system is considered a special case of an IFIS [62].

2.3. Illustration: IF relation

Consider $R(b_k, b_l) = (\alpha_A(b_k, b_l), \beta_A(b_k, b_l)$, where $k, l \in 1, 2, \ldots, N$, is a binary relation of IF, which is added to the system. $R(b_k, b_l)$ be an IF relation for the condition [63]:

1) For any k and l,

$$\alpha_R(b_k, b_l) = 1 \; and \; \beta_R(b_k, b_l) = 0 \tag{2}$$

2) Symmetry: For any k and l,

$$\begin{aligned} \alpha_R(b_k, b_l) &= \alpha_R(b_l, b_k) \; and \; \beta_R(b_k, b_l) \\ &= \beta_R(b_l, b_k) \forall \, b_k \quad b_l \in W \end{aligned} \tag{3}$$

2.4. Illustration: MI

MI can be represented in terms of the broadly defined concepts of entropy and conditional entropy, which is represented in the provided equation:

$$I(E; F) = H(F) - H(F \mid E) \tag{4}$$

here, $E \subseteq Z$, $H(F)$ represents the entropy of information, and $H(F \mid E)$ is the entropy of the condition. MI evaluates the reduction in uncertainty about F due to E, and similarly, its information is evaluated. It simplifies the amount of information that E provides about F or vice versa. Moreover, $H(E)$ denotes the amount of information that is included E itsef, which represents that $I(E; E) = H(E)$.

2.5. Illustration: Impact of conditional feature

For an IFIS and $C \subseteq Y$, if a conditional feature that is randomly chosen $c \in (Y - C)$, its significance can be represented by the equation, which is given by:

$$\begin{aligned} \mathrm{SGF}(\mathrm{c}, \mathrm{C}, \mathrm{E}) &= \mathrm{I}(\mathrm{C} \cup \{\mathrm{c}\}; \mathrm{E}) - \mathrm{I}(\mathrm{C}; \mathrm{E}) = \mathrm{H}(\mathrm{E} \mid \mathrm{C}) \\ &\quad - \mathrm{H}(\mathrm{E} \mid \mathrm{C} \cup \{\mathrm{c}\}) \end{aligned} \tag{5}$$

and $\mathrm{C} = \emptyset$, $\mathrm{SGF}(\mathrm{c}, \mathrm{C}, \mathrm{E}) = \mathrm{H}(\mathrm{E}) - \mathrm{H}(\mathrm{E} \mid \mathrm{c}) = \mathrm{I}(\mathrm{c}; \mathrm{E})$, which is the MI between the condition attribute c and decision attribute E. If the evaluated value of $SGF(c, C, E)$ is high, then it implies that, given the known subset of features C, the dimension c is found to be more valuable for the decision feature that is available.

3. Proposed Method

Assume P and Q are two IFSs in U. Then:

1) If $P \subseteq Q \; iff \; \mu_P(u_i) \leq \mu_Q(u_i) \; and \; \nu_P(u_i) \geq \nu_Q(u_i)$ for any $u_i \in U$ [24].
2) $P \preccurlyeq Q$ that represents P is less fuzzy than Q; that is, for any $u_i \in U$, the following statements hold,
   a. If $\mu_Q(u_i) \leq \nu_Q(u_i)$ *then* $\mu_P(u_i) \leq \mu_Q(u_i)$ *and* $\nu_P(u_i) \geq \nu_Q(u_i)$
   b. If $\mu_Q(u_i) \geq \nu_Q(u_i)$ *then* $\mu_P(u_i) \geq \mu_Q(u_i)$ *and* $\nu_P(u_i) \leq \nu_Q(u_i)$

3.1. IF relation

An IF relation is defined by $(W, Z, V_{\{IF\}}, IF)$, $\forall z \in Z$ and w_i, w_j in W, the hybrid similarity denoted by $S_z^h(w_i, w_j)$, between w_i, w_j w.r.t any given z can be represented as:

$$S_z^h(w_i, w_j) = \begin{cases} 1, \; if \; z(w_i) = z(w_j) \text{ and if z is nominal} \\ 0, \; if \; z(w_i) \neq z(w_j) \text{ and if z is nominal} \\ 1 - \frac{1}{2n} \sum_{i=1}^{n} \; if \; \begin{array}{c} |\mu_P(u_i) - \mu_Q(u_i)| \\ + |\nu_P(u_i) - \nu_Q(u_i)|; \\ |\mu_P(u_i) - \mu_Q(u_i)| \leq \zeta_z \text{ and} \\ |\nu_P(u_i) - \nu_Q(u_i)| \geq \zeta_z \text{ and if } z \textit{ is} \\ \textit{numerical} \end{array} \\ 0; \; if \; \begin{array}{c} |\mu_P(u_i) - \mu_Q(u_i)| \geq \zeta_z \text{ and} \\ |\nu_P(u_i) - \nu_Q(u_i)| \leq \zeta_z \text{ and if } z \textit{ is} \\ \textit{numerical} \end{array} \end{cases} \tag{6}$$

4. Granular Structure

A fuzzy granular structure organizes data into overlapping granules with varying membership levels, which improves the robustness of noise. It handles noise using feature-based similarity averaging, decreases the impact of outliers with the help of gradual membership, and maintains stability utilizing fuzzy granular structure distance. This allows reliable analysis even with noisy or slightly varied data.

The IF granule $\forall u_i \in W$ is represented by Z_1, which is given as:

$$\mu^{\epsilon}_{[\mathrm{u_i}]^{\epsilon}_{\mathrm{P}}} = \begin{cases} 0, \text{ if } \mu_{S^{\mathrm{h}}_{\mathrm{Z}}}(\mathrm{w_i}, \mathrm{w_j}) \leq \epsilon \\ \mu_{S^{\mathrm{h}}_{\mathrm{Z}}}(\mathrm{w_i}, \mathrm{w_j}), \text{ if } \mu_{S^{\mathrm{h}}_{\mathrm{Z}}}(\mathrm{w_i}, \mathrm{w_j}) > \epsilon \end{cases} \tag{7}$$

Further:

$$\nu^{\epsilon}_{[\mathrm{u_i}]^{\epsilon}_{\mathrm{P}}} = \begin{cases} 0, \text{ if } \nu_{S^{\mathrm{h}}_{\mathrm{Z}}}(\mathrm{w_i}, \mathrm{w_j}) \leq \epsilon \\ \nu_{S^{\mathrm{h}}_{\mathrm{Z}}}(\mathrm{w_i}, \mathrm{w_j}), \text{ if } \nu_{S^{\mathrm{h}}_{\mathrm{Z}}}(\mathrm{w_i}, \mathrm{w_j}) > \epsilon \end{cases} \tag{8}$$

This holds for all $\mathrm{p} \in \mathrm{P} \in \mathrm{Z}$ and $\varepsilon \in [0, 1]$.

With a granulation structure IF is present, rough entropy is defined within the IF rough framework, which allows the IF rough entropy of a feature to be represented as follows:

Entropy: Entropy serves as a fundamental measure of uncertainty in information systems; on the other hand, conventional Shannon entropy extracts the nuances of uncertainty in datasets characterized by vagueness and hesitation. To address this, IF entropy extends the concept by using membership degrees, non-membership, and hesitation, which provides a more comprehensive assessment of uncertainty.

Burillo and Bustince [64] defined entropy as a real-value function $E: IFS(Y) \rightarrow \mathbb{R}^+$ is referred to as an entropy measure on the collection of IFSs $IFS(Y)$ if it satisfies the following properties:

1) (EP1) $E(B) = 0$ if and only if B is a classical fuzzy set on Y.
2) (EP2) $E(B) = |Y| = M$ if and only if the membership and non-membership functions of B, denoted by $\alpha_B(y)$ and $\beta_B(y)$, are both $0 \; \forall y \in Y$.
3) (EP3) $E(B) = E(B') \forall B \in IFS(Y)$, where B' denotes the complement of B.
4) (EP4) If $B_1 \subseteq B_2$, then $E(B_1) \geq E(B_2)$.

In the context of fuzzy rough sets, information entropy computes uncertainty by assuming both inner-class and outer-class information, which extracts the granularity of effective partitions of data, and provides a more nuanced understanding of uncertainty in environments of fuzzy.

Furthermore, based on hybrid similarity relation, MI frameworks have been demonstrated to handle high-dimensional datasets with inherent uncertainty; these frameworks use IF rough entropy, joint entropy, and conditional entropy to calculate uncertainty and relevancy of features, which enhances the performance of learning algorithms in complex data environments.

These advancements in the measure of entropy in IF and fuzzy rough set frameworks provide more robust tools for uncertainty quantification, particularly in systems where ambiguity of data and hesitation are prevalent.

4.1. Illustration: IF rough entropy

The rough entropy of IF of Z_1 is described as:

$$RE(Z_1) = RE\left(S_Z^h\right) = -\frac{1}{n} \sum_{i=1}^{n} log_2 \frac{1}{|[u_i]_{S^h_{Z_1}}|} \tag{9}$$

It is evident that $0 \leq RE(Z_1) \leq log_{2n}$ iff $\forall u_i \in W$, $S^h_{Z_1} = 1$, $\left|[u_i]_{S^h_{Z_1}}\right| = n$. In this case, all pairs of samples that are observed distinctly result in the space of granulation being at its maximum. Conversely, $\forall u_i \neq u_j$, $S^h_{Z_1}(u_i, u_j) = 0$, which implies that $\left|[u_i]_{S^h_{Z_1}}\right| = 1$. Thus, $RE(Z_1) = log_{2n} = 0$. At this point, the granulation space is created as low as possible.

4.2. Illustration: IF joint rough entropy

The joint rough entropy Z_1 *or* Z_2 of the IF system can be represented by:

$$RE(Z_1, Z_2) = RE\left(S^h_{\{Z_1 \cup Z_2\}}\right) = -\frac{1}{n}\sum_{i=1}^{n} \log_2 \frac{1}{|[u_i]_{S^h_{Z_1}} \cap [u_i]_{S^h_{Z_2}}|} \quad (10)$$

4.3. Illustration: IF rough conditional entropy

The IF rough Condition entropy of Z_1 w.r.t Z_2 is given as:

$$RE(Z_1|Z_2) = -\frac{1}{n}\sum_{i=1}^{n} \frac{\left|[u_i]_{S^h_{Z_1}}\right|}{|[u_i]_{S^h_{Z_1}} \cap [u_i]_{S^h_{Z_2}}|} \quad (11)$$

Intuitionistic fuzzy rough mutual information (IFRMI): It extracts the correlation between conditional features and decision classes by integrating the concepts of IF and rough set theory. This approach assumes the membership, non-membership, and hesitation, which enables for nuanced assessment of uncertainty and similarity in datasets consisting of both numerical and nominal features.

In the aspect of rough, it represents the use of lower and upper approximation in rough set theory, which describes the data granules' boundaries based on similarity relations. In the framework of IFRMI, these approximations are built using a hybrid similarity relation that uses mixed types of data. By calculating the MI between these approximations and the decision classes, IFRMI quantifies the amount of information shared, effectively calculating each feature's relevance. This method allows the identification of significant features that contribute the accurate decision-making, even in the presence of uncertainty and imprecision inherent in mixed datasets.

4.4. Illustration: IF rough mutual information

The IF rough MI of Z_1 and Z_2 is evaluated as:

$$I(Z_1; Z_2) = -\frac{1}{n}\sum_{i}^{n} \log_2 \frac{\left|[u_i]_{S^h_{Z_1}}\right| \cap \left|[u_i]_{S^h_{Z_2}}\right|}{\left|[u_i]_{S^h_{Z_1}}\right| \left|[u_i]_{S^h_{Z_2}}\right|} \quad (12)$$

The IF rough MI of Z_1 and D is evaluated as:

$$I(D; Z_1) = -\frac{1}{n}\sum_{i}^{n} \log_2 \frac{\left|[u_i]_{S^h_{Z_1}}\right| \cap |[u_i]_D|}{\left|[u_i]_{S^h_{Z_1}}\right| |[u_i]_D|} \quad (13)$$

With the use of this, IF rough MI $I(d; Z_1)$ evaluates the correlation between Z_1 and D, which is the decision feature. IF rough MI between D and Z_2 is greater, and it reflects a strong correlation between Z_1 and Z_2.

In a given IFIS, consider a subset C of condition feature Z. Then, for any $K \in (Z - C)$, the impact of Y compared to P and D (decision features) is expressed as $\sigma(K,\ C, F)$ that is represented by the equation:

$$\sigma(K, C, F) = I(C \cup K; F) - I(K, F) \quad (14)$$

$K = \phi, \sigma(G, C, F)$, and can be expressed as $\sigma(K, F) = RE(F) - RE(F|K) = I(K, F)$, which shows the MI of IF conditional feature set G and decision features D. If $\sigma(G, C, F)$ increases, it shows that the IF conditional feature G becomes more relevant to the feature D.

5. Algorithm: Feature Selection Using IFIS and Hybrid Similarity Relation

Step 1: Given a real-valued information system ⟨K, C ∪ F⟩ and adaptive radius ζ_a.
Step 2: Implement intuitionistic fuzzification for ⟨K, C⟩.
Step 3: Evaluate mutual information (MI) I(F; U) for the IFIS based on hybrid similarity relation S_h, where U ⊆ C.
Step 4: Initialize S ← ∅.
Step 5: Loop Until Convergence: Repeat the following steps until I(U; F) = I(C; F):
- Set K ← S.
- For each feature x ∈ (C − U):
  • Calculate the conditional feature significance Ω(x, U, F).
- Select the attribute that gives the maximum value of Ω(x, U, F).
- If multiple features have the maximum value, select the one that, when combined with U, gives the smallest output value.
- Update S ← U ∪ {x}.
Step 6: Return the set S.

Figure 1 shows the flow chart for the proposed feature selection method using IFIS and hybrid similarity relation.

6. Experiment

In this experiment, the proposed method's performance is evaluated by comparing it with earlier fuzzy approaches and IFS-based techniques. The steps of preprocessing are done using MATLAB [65], while the learning algorithms are performed in WEKA [65]. Initially, fuzzification and intuitionistic fuzzification from real-time data are implemented with the methods that are proposed by Jensen and Shen [25] and Tan et al. [66], respectively. Before performing the reduction, the datasets are transformed into IF decision systems. For simplicity, we use a straightforward method by calculating the similarity degree between objects. Various threshold parameter values are adjusted in the proposed approach to find a reduct. The reduced datasets are created by minimizing the noise to the greatest extent which is possible. The reduct is evaluated by varying the adaptive radius between 0.1 and 0.8 in small intervals. The final reduct is computed on the basis of that value of the adaptive radius, which produces the reduced dataset by minimizing the redundancy and maximizing the relevancy. The whole experiment follows this setup to produce the comprehensive results.

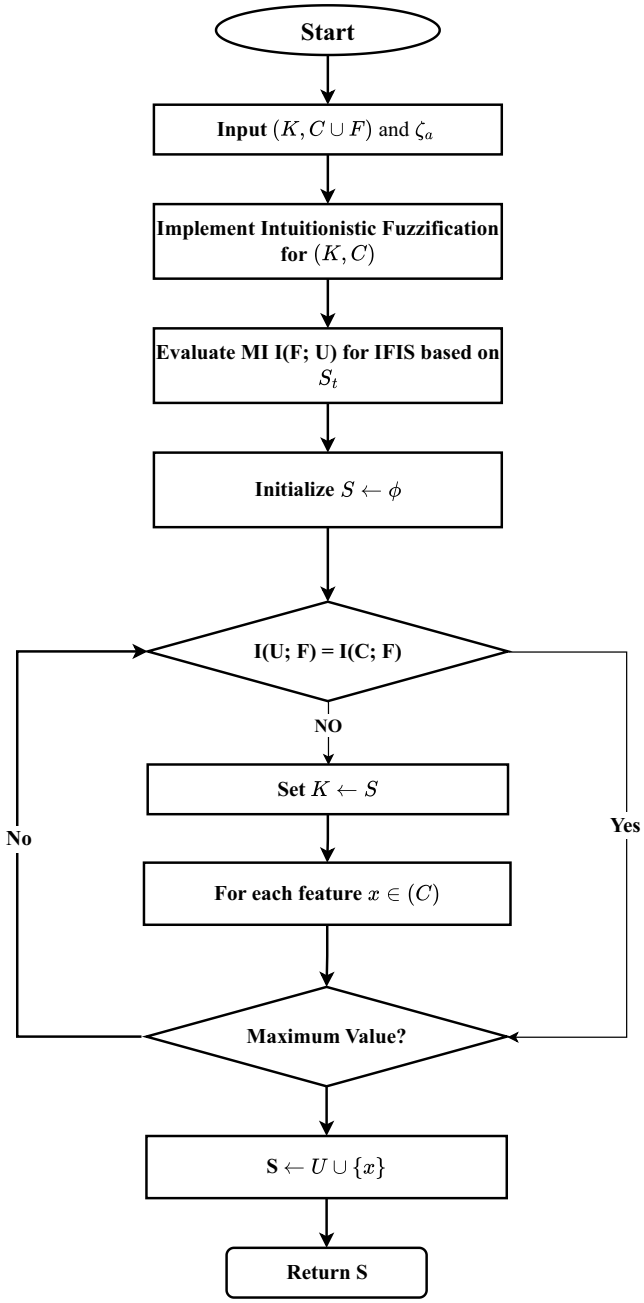
7. Dataset

Seven benchmark datasets were chosen from the University of California, Irvine (UCI) Machine Learning Repository by Kelly et al. [67] to carry out the experiments. The dataset details are discussed in Table 1. The shape of the datasets shows that the data ranges from small to large, with the number of rows ranging from 64 to 1941 and the number of features ranging from 15 and 10000.

8. Classifiers

Three different learning techniques [68] were leveraged in this experiment to evaluate the performance of the reduced datasets generated from various feature selection approaches. RARF and IBK

Figure 1
Flow chart for feature selection using intuitionistic fuzzy information system (IFIS) and hybrid similarity relation



were utilized to evaluate the overall accuracy of the classification and standard deviations by using multiple validation approaches across the seven datasets. Furthermore, J48 and PART methods were also implemented to calculate the performance by using various performance metrics on the reduced datasets [69]. This evaluation aims to demonstrate the effectiveness of the proposed method compared with existing methods for differentiating between positive and negative classes.

9. Data Splitting

The method for attribute selection is employed on the whole dataset. After getting the reduced data, each algorithm is used with a 66:34 split of data and k-fold cross-validation. In the proportional split method, the data is grouped into two groups: 66% random data for training the learning algorithm and 34% kept for testing the algorithm. In the k-fold cross-validation, the dataset is split into random k subsets, where k-1 subsets are utilized for training the algorithm and the remaining subsets are kept for testing. This step is executed k-times, with a different subset that serves as the test set each time. The final performance is computed by averaging the evaluation metrics across each iteration. In this study, the value of k is set to 10.

10. Evaluation Metrics

The prediction evaluated from the three learning algorithms is computed from multiple categories, which is evaluated with threshold-independent evaluation metrics. These metrics are evaluated on the basis of true positive values (*TRPve*), true negative (*TRNve*), false positive (*FLPve*), and false negative (*FLNve*). *TRPve* denotes the number of positive classes that are correctly predicted as positive, while *TRNve* shows the number of negative classes that are predicted correctly as negative. *FLNve* represents the sample counts that are predicted positives as incorrect, and *FLPve* shows the incorrectly predicted negative samples. To compute the overall performance of each algorithm, multiple parameters are employed that include specificity (S_p), sensitivity (S_n), Area Under the Curve (AUC), accuracy (Ac), and Matthew's correlation coefficient (MCC). These performance metrics can be described as follows:

S_n measures the percentage of positive classes that are correctly classified and described as:

$$S_n = \frac{TRPve}{TRPve + FLNve} * 100 \tag{15}$$

S_p evaluates the effective percentage of negative classes with a correctly classified and is directed by:

$$S_n = \frac{TRNve}{TRNve + FLPve} * 100 \tag{16}$$

Table 1
Characteristics of datasets and their reduced sizes

Dataset	*Instances*	*Attributes*	*MIFRFS*	*LIFRFS*	*DIFRFS*	*IFRFS*	*Proposed Method*
Arrhythmia	452	262	159	145	178	129	84
Leaf	340	15	12	10	11	9	8
Arcene	200	10000	225	165	271	188	86
Db-world-bodies-steamed	64	3721	25	47	52	36	29
Movement Libras	270	90	37	54	61	43	39
Parkinsons	195	22	14	11	16	10	12
Steel Plates Faults	1941	33	25	22	15	12	14

Ac measures the amount of required correctly classified positive and negative classes, and it is given by:

$$Ac = \frac{TRPve + TRNve}{TRPve + FLNve + TRNve + FLPve} * 100 \quad (17)$$

AUC is used to evaluate the critical and relevant area under the receiver operating characteristic (ROC) curve. The closer the value is to 1 for AUC, the better the predictor performs.

Matthews correlation coefficient is a highly valuable and anticipated parameter that is calculated by the equation:

$$MCC = \frac{TRPve * TRNve - FLNve * FLPve}{\sqrt{(TRPve + FLPve)(TRPve + FLNve)}} * 100 \quad (18)$$

Matthews correlation coefficient is not only used for demonstrating the effectiveness of binary classification but also used for validating the efficiency. Matthews correlation coefficient value tending to 1 reflects the high reliability of the reflector.

11. Result Discussion

The seven benchmark datasets details along with their reducts obtained from four earlier methods and the proposed approach are shown in Table 1. The real-time data was converted into fuzzy and IF values with the help of well-established methods introduced by Jensen and Shen [25] and Kelly et al. [67]. The process of reduction was implemented on the whole dataset using fuzzy and IF-assisted methods. Mutual Information based IFRFS (MIFRFS), Different classes ratio based IFRFS (DIFRFS), Laplace summation operator based IFRFS (LIFRFS), and Intuitionistic fuzzy rough feature selection (IFRFS) were utilized for the analysis of comparison as shown in Table 2. The proposed approach decreased the reduct size to a range from 8 to 84, with smaller reduct sizes in comparison with previous approaches. For the bank marking datasets, Fuzzy set-based feature reduction method (FSFrM) showed a smaller reduction size; on the other hand, Feature selection using fuzzy rough mutual information (FSFrMI) [25, 70] output smaller sizes for thyroid-hypothyroid and fertility diagnosis datasets in comparison with IFRFSMI. For the breast cancer datasets, FSFrM and FRFS resulted in similar reduction sizes, and FRFS outperformed a comparable size to the proposed approach for the fertility diagnosis data. On the basis of the reduction in Table 1, it is ensured that the proposed approach consistently generated a more reduced dimension throughout the various datasets in comparison to previous methods. The process of reduction, visualized on various methods, reflects that the proposed approach gives a higher overall percentage attribute elimination as the number of conditional

Table 2
Performance of classifiers using reduced datasets from state-of-the-art and proposed methods (10-fold CV)

Dataset	Classifier	MIFRFS	LIFRFS	DIFRFS	IFRFS	Proposed method
Arrhythmia	J48	82.33 ± 4.62^{4}	86.31 ± 6.59^{3}	91.8 ± 4.79^{1}	79.21 ± 6.98^{5}	90.32 ± 4.98^{2}
	PART	78.43 ± 6.23^{4}	81.43 ± 5.95^{2}	81.12 ± 6.56^{3}	77.83 ± 3.93^{5}	86.35 ± 3.21^{1}
Leaf	J48	71.52 ± 7.21^{5}	74.18 ± 5.87^{3}	72.45 ± 4.53^{4}	77.81 ± 5.65^{2}	79.12 ± 4.16^{1}
	PART	68.21 ± 8.99^{4}	69.53 ± 8.65^{3}	66.33 ± 8.18^{5}	71.29 ± 7.42^{2}	73.26 ± 5.77^{1}
Arcene	J48	84.44 ± 6.66^{3}	88.36 ± 5.29^{2}	83.21 ± 8.36^{4}	82.11 ± 8.91^{5}	91.59 ± 6.21^{1}
	PART	78.53 ± 7.43^{4}	80.83 ± 4.18^{2}	79.48 ± 9.53^{3}	77.15 ± 6.26^{5}	84.64 ± 5.18^{1}
Db-world-bodies-steamed	J48	78.32 ± 9.88^{5}	83.93 ± 10.25^{3}	86.73 ± 8.53^{2}	81.33 ± 8.51^{4}	89.32 ± 7.17^{1}
	PART	69.43 ± 11.53^{4}	68.31 ± 8.44^{5}	73.44 ± 8.73^{2}	70.83 ± 10.41^{3}	74.31 ± 8.63^{1}
Movement Libras	J48	72.31 ± 8.92^{4}	74.93 ± 12.62^{3}	77.53 ± 9.55^{1}	72.05 ± 11.22^{5}	75.52 ± 9.33^{2}
	PART	56.21 ± 7.65^{5}	61.23 ± 10.82^{3}	65.56 ± 12.43^{2}	59.29 ± 9.63^{4}	68.21 ± 10.71^{1}
Parkinsons	J48	87.38 ± 8.38^{3}	84.61 ± 9.09^{4}	82.46 ± 8.77^{5}	90.78 ± 7.28^{1}	89.23 ± 7.93^{2}
	PART	78.41 ± 9.59^{3}	75.65 ± 0.52^{4}	71.66 ± 10.53^{5}	79.31 ± 8.66^{2}	80.06 ± 7.53^{1}
Steel Plates Faults	J48	78.31 ± 8.47^{3}	75.72 ± 6.88^{4}	78.91 ± 8.32^{2}	74.48 ± 5.92^{5}	83.53 ± 3.21^{1}
	PART	71.79 ± 7.54^{1}	69.32 ± 8.77^{4}	70.46 ± 7.84^{3}	68.87 ± 8.93^{5}	70.97 ± 8.11^{2}
Average Rank	J48	3.85	3.14	2.71	3.85	1.57
	PART	3.57	3.85	3.28	0.71	1.14
F-statistics	J48	3.7801				
	PART	4.5734				
F-tabular	F(4,24)	2.776				

features increases. To calculate the learning performance, J48 and PART algorithms were implemented, which measure the standard deviation and overall accuracy. The datasets, which were reduced with existing and proposed approaches, were assessed using 10-fold cross-validation to prevent overfitting. The results with rankings are summarized in Table 2. As a result, it can be concluded that our proposed method outperforms other attribute selection techniques throughout the various datasets.

On the Arrhythmia dataset, J48 gives a high accuracy of 90.32% with the proposed method, which is slightly less than DIFRFS (91.89%), but outperforms better than MIFRFS (82.33%) and IFRFS (79.21%). Similarly, PART describes the improved performance with the proposed method by leveraging the accuracy of 86.35% in comparison to other methods. On the Leaf dataset, the proposed method gives a notable improvement in both J48 (79.12%) and PART (73.26%) compared to the other methods. On the Arcene dataset, the proposed method shows the highest accuracy using RF with an accuracy of 91.59%, and PART has an accuracy of 84.64%, which gives better performance compared to other methods, especially IFRFS and DIFRFS. In the Db-world-bodies-steamed dataset, both J48 and PART classifiers achieved an improved accuracy with the proposed method by achieving an accuracy of 89.32% and 74.31%, respectively. On the Movement Libras dataset, the proposed method did not perform well but still improved classifier performance compared to MIFRFS and IFRFS. The same situation occurred in the case of the Parkinson's dataset, where J48 (89.23%) and PART (80.06%) gave better results using the proposed method. Last for the Steel Plates Faults dataset, J48 posted the best results and a high accuracy rate of 83.53% outcompeting all the other algorithms, while PART followed behind with moderate performance of 70.97%. Collectively, these results assert the efficiency of the proposed method in increasing the level of classification accuracy of PART and J48 algorithms, as shown in Figures 2 and 3, respectively, in a number of datasets and classifiers higher than the utilized MIFRFS and IFRFS heuristics. The complete results are visualized in Figure 4, which demonstrates that the proposed approach is highly effective for both low- and high-dimensional data, as the reduced dataset output by the approach gives consistent accuracies through various learning algorithms. The assumptions that we have made for validating the impact of the proposed approach are as follows:

Null assumptions: All approaches that are used are equivalent.

Alternative assumptions: There is a difference between the methods that are employed.

Two major methods, which are used for testing: the Friedman test [71] and the Bonferroni–Dunn test [72], are implemented for validating the proposed method's effectiveness. The Friedmann test is used for conducting the comparative analysis of various models. The Bonferroni–Dunn test is implemented for the determination of methods that are effectively different from the proposed method. The null assumptions are rejected at an α level of significance if there is a difference between the mean ranks that exceeds the critical distance value. In this method, the average ranks for RARF and IBK methods on the basis of the proposed approach are recorded as minimum values in Table 2, which clearly shows the superiority of our methods that are established. Moreover, the evaluated F-statistics on the basis of IFRFSMI are greater for J48 and PART compared to the F-tabular value. Specifically, for J48 and PART, the F-statistic values are 3.7801 and 4.5734, while the tabulated value is 2.776 at a significant level of 5%. Since the calculated F-statistic is greater than the tabulated (critical) value, we reject the null hypothesis, indicating that the model is statistically significant. Thus, according to the Dunn test, our proposed approach is effectively different.

Case study: An application for the discriminator between positive and negative classes.

The one main application of the ML approach in cheminformatics is the reduction of a large chemical space that concerns specific properties of interest. This reduces the chemical space that can subsequently be confirmed through experiments, which highlights the importance of the accuracy of ML techniques. A notable point is the accumulation of phospholipids in several tissues, such as the kidneys and eyes, mainly because of cationic amphiphilic

Figure 2
Mean accuracy comparison of PART on reduced datasets

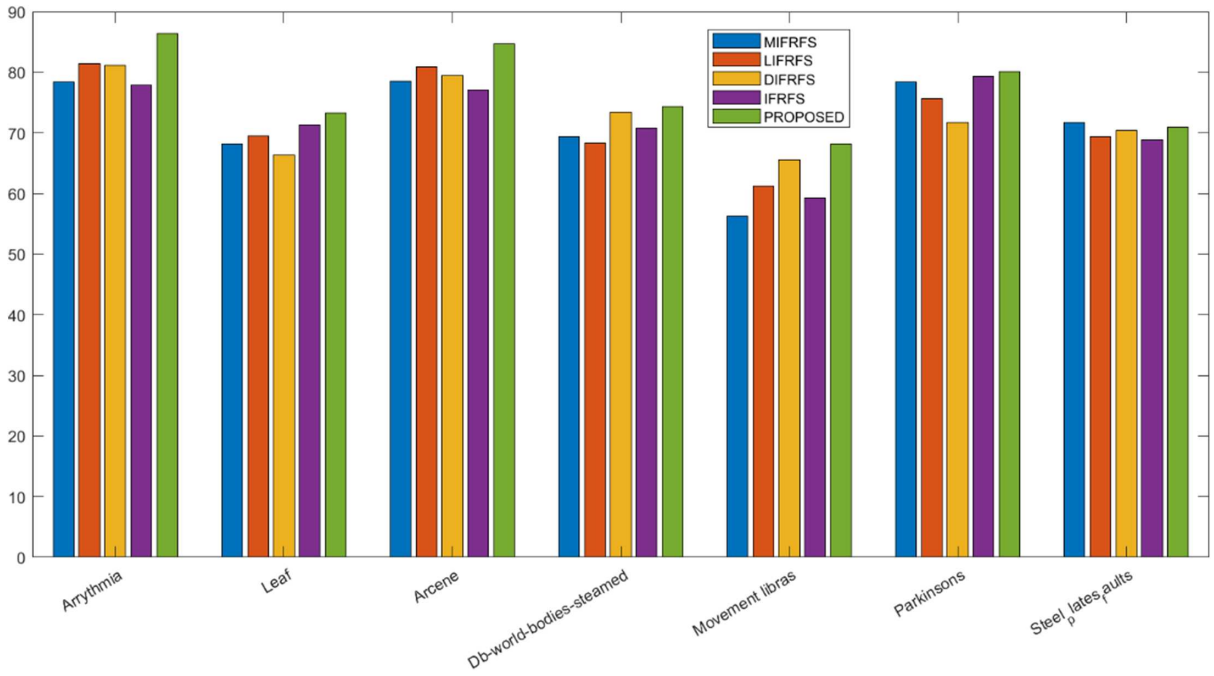


Figure 3
Comparison of mean accuracy using J48 on reduced datasets from existing and proposed methods

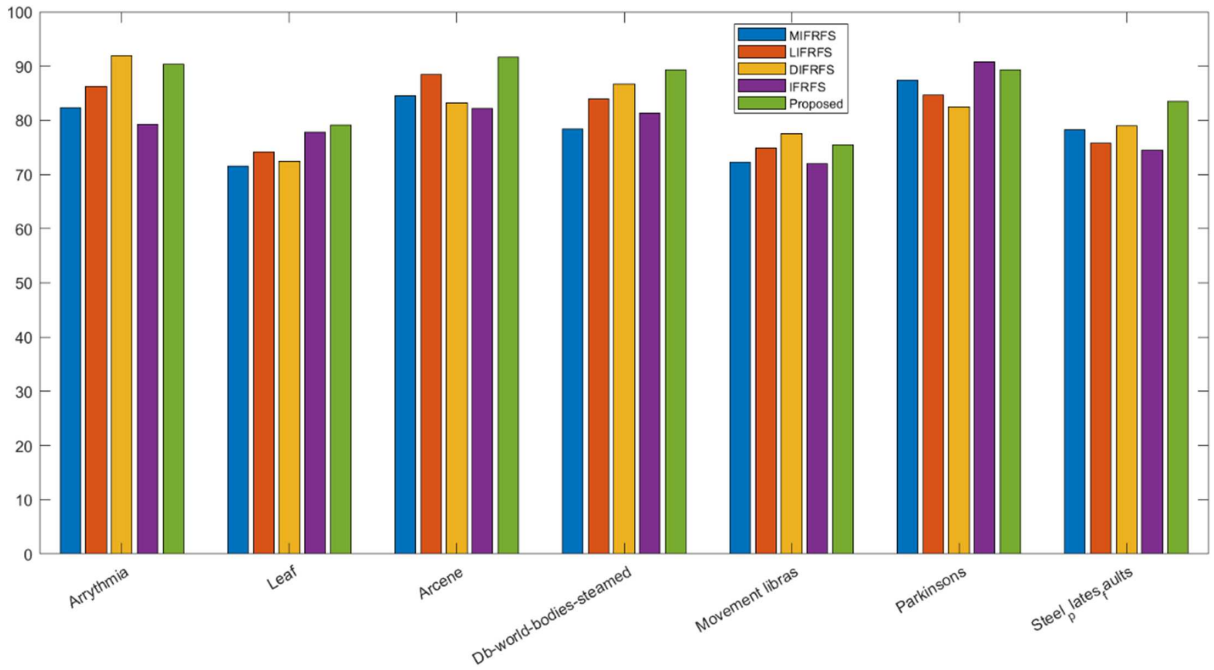
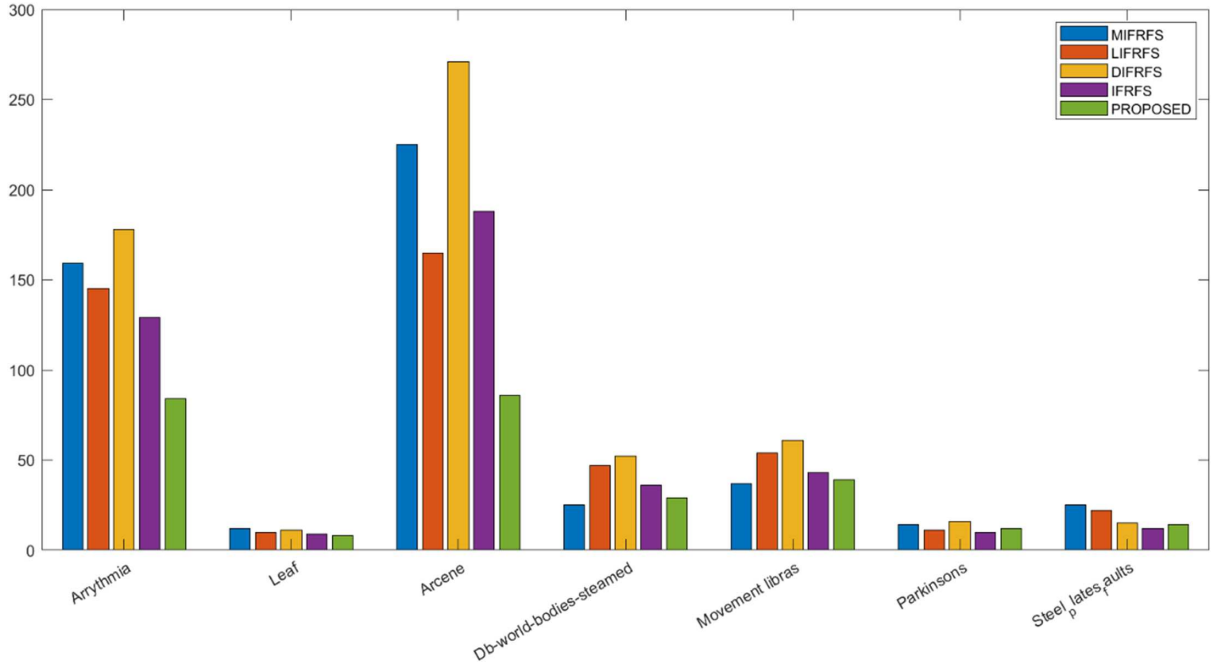


Figure 4
All reduction comparison for different datasets by earlier and proposed approaches



molecules, which are highly accurate for ML predictive models, can lead to the early screening of compounds that induce phospholipidosis during drug discovery workflows, thus minimizing the cost and time related to wet lab experiments.

The current method can open the way for progressive research into the early screening of phospholipidosis-inducing molecules. The dataset comprises 185 compounds in total, which includes 102 classified as PPL+ and 83 as PPL-. All 83 PPL compounds tested negative under electron microscopy. Of the 102 PPL+ compounds, 68 were confirmed positive using electron microscopy, while the remaining 32 were represented as PPL+ on the basis of the presence of foamy macrophages. The proposed methods are implemented on the dataset from Nath and Sahu [69] to create an effectively decreased form by reducing the noise, anomaly, and imprecision in the data and eliminating redundant and irrelevant attributes. Three classifiers of several categories are chosen to measure their

Table 3
Performance metrics of nine classifiers on the reduced dataset using 10-fold cross-validation

Classifiers	Sensitivity	Specificity	Accuracy	AUC	MCC
Naive Bayes	87.23	73.45	81.65	0.901	0.698
SMO	86.8	73.34	79.80	0.871	0.623
IBK	89.38	89.68	88.56	0.906	0.811
RARF	91.50	87.87	92.98	0.965	0.865
PART	75.48	75.96	78.34	0.821	0.598
JRip	79.64	83.68	81.87	0.887	0.686
Random Forest	88.96	89.53	91.52	0.936	0.778
J48	78.34	79.64	78.91	0.789	0.623
(Nath 2019)	86.27	90.20	88.32	0.869	0.765
RARF (h2o)	89.56	92.63	95.98	0.968	0.908

performance on this reduced dataset, with the help of metrics such as AUC, sensitivity, MCC, accuracy, and specificity. Moreover, a comprehensive method that represents the overall performance evaluation of all three classifiers at the optimal decision threshold is discussed using the ROC curve, which provides a visual representation of classifier performance.

The results in Table 3 show the performance evaluation of the nine learning algorithms that are implemented to reduce the dataset generated by our proposed method, which classifies data into positive and negative classes on a 10-fold cross-validation process. The results denote that the RARF(h2o) classifier achieves the overall best performance. Additionally, the high AUC value of RARF(h2o) denotes that it maintains a strong balance between true positive and false positive rates across various thresholds, while the MCC score makes sure that the robustness of the classifier in managing the imbalanced data. Algorithms like Naïve Bayes and Sequential minimal optimization (SMO) performed poorly across the evaluation metrics, with accuracy of 81.65% and 79.80%, respectively, and MCC scores below 0.7. Overall, the results show that the proposed approach increases the performance of the classifiers significantly on the data.

The comparison between the RF(h2o) classifier based on the previous method and the results obtained from the proposed method shows a clear improvement across all key metrics. The existing RF(h2o) methods show 86.7% of sensitivity, 90.1% of accuracy, 93.0% of specificity, an MCC of 0.808, and an AUC of 0.922. The proposed method RF(h2o) yields superior results, with a sensitivity of 89.56%, an accuracy of 95.98%, a specificity of 92.63%, an MCC of 0.908, and an AUC of 0.968. The metrics represent a great sensitivity enhanced by 2.86%, accuracy increased by 5.88%, and MCC gained by 0.100, along with an enhancement in AUC of 0.046. On the other hand, specificity drops from 93.0% to 92.63%; all the improvements in other metrics support the fact that our proposed method obtains better and more consistent results than the current RF(h20) method. Thus, the proposed method is more effective for the classification task.

Furthermore, when compared to the state-of-the-art ensemble model by Nath and Sahu [69], which utilized a stacking ensemble with RF as a meta-classifier, our proposed method continuously demonstrated enhanced capability. Nath and Sahu [69], using their best model with JOElib descriptors and structural alerts, attained an overall accuracy of 88.32%, sensitivity of 86.27%, specificity of 90.20%, MCC of 0.765, and AUC of 0.869. This shows that the proposed RF(h2o) model not only outperforms conventional implementations but also outperforms advanced ensemble methods, setting up its effectiveness and robustness for classification tasks like phospholipidosis prediction.

12. Conclusion

This study established a new IF similarity relation that can handle the frequently produced both nominal and numerical features and provided a more robust method for feature analysis. To mitigate noise in mixed data, an IF granular structure was proposed to increase the resilience of the real-valued data. Moreover, the research developed IF rough entropy, which included joint entropy and conditional entropy, offering a more comprehensive understanding of uncertainty in information entropy. IF rough MI was also demonstrated to capture both later uncertainty and the correlation between conditional features and decision classes. This foundation presented a novel feature selection approach that gives importance to relevant features by reducing the dimensionality comprised of irrelevant and redundant attributes in the real-valued datasets. The proposed method reduced the size of the data effectively, maintained the classification accuracy, and demonstrated superior performance across various metrics, such as accuracy, sensitivity, specificity, MCC, and AUC when computed on seven benchmark datasets. This research provides key benefits over conventional feature selection methods as it uses both numerical and categorical features, which find and remove the redundant and irrelevant features for optimizing the classification accuracy, and significantly minimizes the size of the data, hence reducing the computational costs. These features give a promising advancement in feature selection, as they can attain dimensionality reduction without sacrificing classification accuracy, which makes it a valuable tool for data scientists who work with high-dimensional datasets. Consequently, this approach represented a meaningful contribution to the field of feature selection, which optimizes the performance of the classification and increases the efficiency of the data analysis process. Finally, this approach was employed to improve the discrimination performance of positive and negative samples of phospholipidosis molecules. In the future, this method can be upgraded with more precise tolerance parameters. Moreover, a probabilistic IFS-assisted hybrid similarity relation can be presented to enhance the prediction performance.

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

The data that support the findings of this study are openly available at https://archive.ics.uci.edu/.

Author Contribution Statement

Aneesh Kumar Mishra: Conceptualization, Methodology, Validation, Data curation, Writing – original draft, Writing – review & editing. **Neelesh Kumar Jain:** Software, Formal analysis, Supervision, Project administration. **Ravindra Kumar Singh:** Investigation, Resources, Visualization.

References

[1] Cui, Z., Wang, L., Su, C., Kang, Y., & Zhang, Q. (2025). Dimensional consistency-based modeling of disc cutter wear in hard rock tunneling utilizing integrated multivariate analysis. *Bulletin of Engineering Geology and the Environment*, *84*(4), 202. https://doi.org/10.1007/s10064-025-04221-2

[2] Mitić, N. S., Malkov, S. N., Maljković Ružičić, M. M., Veljković, A. N., Čukić, I. L., Lin, X., Brusić, V., ..., & Brusić, V. (2025). Correlation-based feature selection of single cell transcriptomics data from multiple sources. *Journal of Big Data*, *12*(1), 4. https://doi.org/10.1186/s40537-024-01051-z

[3] 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

[4] Rabiee, M., Mirhashemi, M., Pangburn, M. S., Piri, S., & Delen, D. (2024). Towards explainable artificial intelligence through expert-augmented supervised feature selection. *Decision Support Systems*, *181*, 114214. https://doi.org/10.1016/j.dss.2024.114214

[5] Pawlak, Z., Grzymala-Busse, J., Slowinski, R., & Ziarko, W. (1995). Rough sets. *Communications of the ACM*, *38*(11), 88–95. https://doi.org/10.1145/219717.219791

[6] Hamidzadeh, J., Mehravaran, Z., & Harati, A. (2025). Feature selection by utilizing kernel-based fuzzy rough set and entropy-based non-dominated sorting genetic algorithm in multi-label data. *Knowledge and Information Systems*, *67*(4), 3789–3819. https://doi.org/10.1007/s10115-025-02341-5

[7] Skowron, A., & Ślęzak, D. (2022). Rough sets turn 40: From information systems to intelligent systems. *Annals of Computer Science and Information Systems*, *30*, 23–34. https://doi.org/10.15439/2022F310

[8] Kumbure, M. M., Luukka, P., & Collan, M. (2020). A new fuzzy k-nearest neighbor classifier based on the Bonferroni mean. *Pattern Recognition Letters*, *140*, 172–178. https://doi.org/10.1016/j.patrec.2020.10.005

[9] Qu, K., Gao, P., Dai, Q., Sun, Y., & Hua, X. (2025). Attribute reduction using self-information uncertainty measures in optimistic neighborhood extreme-granulation rough set. *Information Sciences*, *686*, 121340. https://doi.org/10.1016/j.ins.2024.121340

[10] Abd Elaziz, M., Abu-Donia, H. M., Hosny, R. A., Hazae, S. L., & Ibrahim, R. A. (2022). Improved evolutionary-based feature selection technique using extension of knowledge based on the rough approximations. *Information Sciences*, *594*, 76–94. https://doi.org/10.1016/j.ins.2022.01.026

[11] Chen, Y., & Wei, X. (2024). Axiomatic approaches to three types of L-valued rough sets. *International Journal of Machine Learning and Cybernetics*, *15*(12), 5469–5493. https://doi.org/10.1007/s13042-024-02252-x

[12] Zhao, J., Wan, R., & Miao, D. (2024). Conflict analysis triggered by three-way decision and Pythagorean fuzzy rough set. *International Journal of Computational Intelligence Systems*, *17*(1), 17. https://doi.org/10.1007/s44196-023-00378-4

[13] Zhang, P., Li, T., Yuan, Z., Luo, C., Liu, K., & Yang, X. (2024). Heterogeneous feature selection based on neighborhood combination entropy. *IEEE Transactions on Neural Networks and Learning Systems*, *35*(3), 3514–3527. https://doi.org/10.1109/TNNLS.2022.3193929

[14] Yang, X., Chen, H., Li, T., Zhang, P., & Luo, C. (2022). Student-*t* kernelized fuzzy rough set model with fuzzy divergence for feature selection. *Information Sciences*, *610*, 52–72. https://doi.org/10.1016/j.ins.2022.07.139

[15] Jensen, R., & Shen, Q. (2009). New approaches to fuzzy-rough feature selection. *IEEE Transactions on Fuzzy Systems*, *17*(4), 824–838. https://doi.org/10.1109/TFUZZ.2008.924209

[16] Kumar, A., & Prasad, P. S. V. S. S. (2020). Scalable fuzzy rough set reduct computation using fuzzy min–max neural network preprocessing. *IEEE Transactions on Fuzzy Systems*, *28*(5), 953–964. https://doi.org/10.1109/TFUZZ.2020.2965899

[17] Chen, J., Mi, J., & Lin, Y. (2020). A graph approach for fuzzy-rough feature selection. *Fuzzy Sets and Systems*, *391*, 96–116. https://doi.org/10.1016/j.fss.2019.07.014

[18] Mondal, S., & Singh, R. (2025). *Optimizing cybersecurity budgets in financial networks: A comparative study of genetic algorithms and trust-region methods*. Research Square. https://doi.org/10.21203/rs.3.rs-7323238/v1

[19] An, S., Zhao, E., Wang, C., Guo, G., Zhao, S., & Li, P. (2023). Relative fuzzy rough approximations for feature selection and classification. *IEEE Transactions on Cybernetics*, *53*(4), 2200–2210. https://doi.org/10.1109/TCYB.2021.3112674

[20] Wang, C., Qian, Y., Ding, W., & Fan, X. (2022). Feature selection with fuzzy-rough minimum classification error criterion. *IEEE Transactions on Fuzzy Systems*, *30*(8), 2930–2942. https://doi.org/10.1109/TFUZZ.2021.3097811

[21] Dai, J., Liu, Y., & Chen, J. (2021). Feature selection via max-independent ratio and min-redundant ratio based on adaptive weighted kernel density estimation. *Information Sciences*, *568*, 86–112. https://doi.org/10.1016/j.ins.2021.03.049

[22] Awais, M., & Salahuddin, T. (2024). Radiative magnetodydrodynamic cross fluid thermophysical model passing on parabola surface with activation energy. *Ain Shams Engineering Journal*, *15*(1), 102282. https://doi.org/10.1016/j.asej.2023.102282

[23] Priya, M. P., Roopa, C. K., & Harish, B. S. (2024). Modified Intuitionistic Fuzzy Clustering Method (MIFCM) for microarray image spot segmentation. *Procedia Computer Science*, *235*, 878–888. https://doi.org/10.1016/j.procs.2024.04.083

[24] Atanassov, K. T. (1999). Intuitionistic fuzzy sets. In J. Kacprzyk (Ed.), *Intuitionistic fuzzy sets: Theory and applications* (pp. 1–137). Physica-Verlag Heidelberg. https://doi.org/10.1007/978-3-7908-1870-3_1

[25] Jensen, R., & Shen, Q. (2004). Fuzzy–rough attribute reduction with application to web categorization. *Fuzzy Sets*

and Systems, *141*(3), 469–485. https://doi.org/10.1016/S0165-0114(03)00021-6

[26] Chakrabarty, K., Gedeon, T., & Koczy, L. (1998). Intuitionistic fuzzy rough set. In *Proceedings of 4th Joint Conference on Information Sciences*, 211–214.

[27] Zhan, J., & Sun, B. (2020). Covering-based intuitionistic fuzzy rough sets and applications in multi-attribute decision-making. *Artificial Intelligence Review*, *53*(1), 671–701. https://doi.org/10.1007/s10462-018-9674-7

[28] Pandey, K., Mishra, A., Rani, P., Ali, J., & Chakrabortty, R. (2023). Selecting features by utilizing intuitionistic fuzzy entropy method. *Decision Making: Applications in Management and Engineering*, *6*(1), 111–133. https://doi.org/10.31181/dmame07012023p

[29] de, S. K., Biswas, R., & Roy, A. R. (1998). Intuitionistic fuzzy database. In *Second International Conference on IFS*, *4*(2), 34–41.

[30] Zhang, X., Li, M., Shao, S., & Wang, J. (2024). (I, O)-fuzzy rough sets based on overlap functions with their applications to feature selection and image edge extraction. *IEEE Transactions on Fuzzy Systems*, *32*(4), 1796–1809. https://doi.org/10.1109/TFUZZ.2023.3335108

[31] Mahmood, T., Ahmmad, J., Ali, Z., & Yang, M.-S. (2023). Confidence level aggregation operators based on intuitionistic fuzzy rough sets with application in medical diagnosis. *IEEE Access*, *11*, 8674–8688. https://doi.org/10.1109/ACCESS.2023.3236410

[32] Mazarbhuiya, F. A., & Shenify, M. (2023). An intuitionistic fuzzy-rough set-based classification for anomaly detection. *Applied Sciences*, *13*(9), 5578. https://doi.org/10.3390/app13095578

[33] Xin, X. W., Shi, C. I., Sun, J. B., Xue, Z. A., Song, J. H., & Peng, W. M. (2023). A novel attribute reduction method based on intuitionistic fuzzy three-way cognitive clustering. *Applied Intelligence*, *53*(2), 1744–1758. https://doi.org/10.1007/s10489-022-03496-2

[34] Huang, G., Xiao, L., & Zhang, G. (2021). Risk evaluation model for failure mode and effect analysis using intuitionistic fuzzy rough number approach. *Soft Computing*, *25*(6), 4875–4897. https://doi.org/10.1007/s00500-020-05497-0

[35] Tiwari, S. P., Singh, A. P., & Pandey, S. (2021). Note on 'intuitionistic L-fuzzy rough sets, intuitionistic L-fuzzy preorders and intuitionistic L-fuzzy topologies'. *Fuzzy Information and Engineering*, *13*(2), 196–198. https://doi.org/10.1080/16168658.2021.1921376

[36] Hu, M., Guo, Y., Wang, R., & Wang, X. (2024). Attribute reduction with fuzzy kernel-induced relations. *Information Sciences*, *669*, 120589. https://doi.org/10.1016/j.ins.2024.120589

[37] Esmail, H., Maryam, J., & Habibolla, L. (2013). Rough set theory for the intuitionistic fuzzy information. *Systems International Journal of Modern Mathematical Sciences*, *6*(3), 132–143.

[38] Huang, B., Li, H. X., & Wei, D. K. (2012). Dominance-based rough set model in intuitionistic fuzzy information systems. *Knowledge-Based Systems*, *28*, 115–123. https://doi.org/10.1016/j.knosys.2011.12.008

[39] He, J., Qu, L., Wang, Z., Chen, Y., Luo, D., & Wen, C.-F. (2022). Attribute reduction in an incomplete categorical decision information system based on fuzzy rough sets. *Artificial Intelligence Review*, *55*(7), 5313–5348. https://doi.org/10.1007/s10462-021-10117-w

[40] Shreevastava, S., Tiwari, A. K., & Som, T. (2018). Intuitionistic fuzzy neighborhood rough set model for feature selection. *International Journal of Fuzzy System Applications*, *7*(2), 75–84. https://doi.org/10.4018/IJFSA.2018040104

[41] Tiwari, A. K., Shreevastava, S., Shukla, K. K., & Subbiah, K. (2018). New approaches to intuitionistic fuzzy-rough attribute reduction. *Journal of Intelligent & Fuzzy Systems*, *34*(5), 3385–3394. https://doi.org/10.3233/JIFS-169519

[42] Zhang, X., Zhou, B., & Li, P. (2012). A general frame for intuitionistic fuzzy rough sets. *Information Sciences*, *216*, 34–49. https://doi.org/10.1016/j.ins.2012.04.018

[43] Choudhary, A., de, A., Ahmed, K., & Shankar, R. (2025). An integrated fuzzy intuitionistic sustainability assessment framework for manufacturing supply chain: A study of UK based firms. *Annals of Operations Research*, *349*(2), 687–730. https://doi.org/10.1007/s10479-019-03452-3

[44] Li, K., Chen, C.-Y., & Zhang, Z.-L. (2023). Mining online reviews for ranking products: A novel method based on multiple classifiers and interval-valued intuitionistic fuzzy TOPSIS. *Applied Soft Computing*, *139*, 110237. https://doi.org/10.1016/j.asoc.2023.110237

[45] Singh, S., Shreevastava, S., Som, T., & Jain, P. (2019). Intuitionistic fuzzy quantifier and its application in feature selection. *International Journal of Fuzzy Systems*, *21*(2), 441–453. https://doi.org/10.1007/s40815-018-00603-9

[46] Isaee, M., Fatemi, A., & Nematbakhsh, M. (2025). Addressing the challenges of open n-ary relation extraction with a deep learning-driven approach. *Information Sciences*, *692*, 121643. https://doi.org/10.1016/j.ins.2024.121643

[47] Cornelis, C., de Cock, M., & Kerre, E. E. (2003). Intuitionistic fuzzy rough sets: At the crossroads of imperfect knowledge. *Expert Systems*, *20*(5), 260–270. https://doi.org/10.1111/1468-0394.00250

[48] Cornelis, C., Deschrijver, G., & Kerre, E. E. (2004). Implication in intuitionistic fuzzy and interval-valued fuzzy set theory: Construction, classification, application. *International Journal of Approximate Reasoning*, *35*(1), 55–95. https://doi.org/10.1016/S0888-613X(03)00072-0

[49] Zhou, L., Wu, W.-Z., & Zhang, W.-X. (2009). On characterization of intuitionistic fuzzy rough sets based on intuitionistic fuzzy implicators. *Information Sciences*, *179*(7), 883–898. https://doi.org/10.1016/j.ins.2008.11.015

[50] Zhou, L., & Wu, W.-Z. (2008). On generalized intuitionistic fuzzy rough approximation operators. *Information Sciences*, *178*(11), 2448–2465. https://doi.org/10.1016/j.ins.2008.01.012

[51] Tiwari, A. K., Nath, A., Pandey, R. K., & Maratha, P. (2024). A novel intuitionistic fuzzy rough instance selection and attribute reduction with kernelized intuitionistic fuzzy C-means clustering to handle imbalanced datasets. *Expert Systems with Applications*, *251*, 124087. https://doi.org/10.1016/j.eswa.2024.124087

[52] Ghosh, S. K., Ghosh, A., & Bhattacharyya, S. (2022). Recognition of cancer mediating biomarkers using rough approximations enabled intuitionistic fuzzy soft sets based similarity measure. *Applied Soft Computing*, *124*, 109052. https://doi.org/10.1016/j.asoc.2022.109052

[53] Hesameddini, E., Jafarzadeh, M., & Latifizadeh, H. (2013). Rough set theory for the intuitionistic fuzzy information systems. *International Journal of Modern Mathematical Sciences*, *6*(3), 132–143.

[54] Yu, P., & Zhao, X. (2025). A three-way decision model for multi-granular support intuitionistic fuzzy rough sets based on

overlap functions. *Artificial Intelligence Review*, *58*(5), 143. https://doi.org/10.1007/s10462-025-11139-4

[55] Li, X., & Huang, X. (2020). A novel three-way investment decisions based on decision-theoretic rough sets with hesitant fuzzy information. *International Journal of Fuzzy Systems*, *22*(8), 2708–2719. https://doi.org/10.1007/s40815-020-00836-7

[56] Jain, P., Tiwari, A., & Som, T. (2024). Intuitionistic fuzzy rough set model based on k-means and its application to enhance prediction of aptamer–protein interacting pairs. *Journal of Ambient Intelligence and Humanized Computing*, *15*(9), 3575–3586. https://doi.org/10.1007/s12652-024-04837-4

[57] Xue, Z. A., Jing, M. M., Li, Y. X., & Zheng, Y. (2023). Variable precision multi-granulation covering rough intuitionistic fuzzy sets. *Granular Computing*, *8*(3), 577–596. https://doi.org/10.1007/s41066-022-00342-1

[58] Zhang, T., Gao, J., Liu, H., Tan, Q., Wang, Y., & Huang, N. (2025). An applied study of multi-stage multi-attribute group decision-making based on intuitionistic fuzzy sets—Selection of hydrogen energy projects in the context of "dual-carbon". *Journal of the Operational Research Society*, *76*(10), 2020–2038. https://doi.org/10.1080/01605682.2025.2450289

[59] Shreevastava, S., Singh, S., Tiwari, A. K., & Som, T. (2021). Different classes ratio and Laplace summation operator based intuitionistic fuzzy rough attribute selection. *Iranian Journal of Fuzzy Systems*, *18*(6), 67–82. https://doi.org/10.22111/ijfs.2021.6334

[60] Salahuddin, T. (2023). *Numerical techniques in MATLAB: Fundamental to advanced concepts. USA: Chapman and Hall/CRC*, https://doi.org/10.1201/9781003385288

[61] Tiwari, A. K., Saini, R., Nath, A., Singh, P., & Shah, M. A. (2024). Hybrid similarity relation based mutual information for feature selection in intuitionistic fuzzy rough framework and its applications. *Scientific Reports*, *14*(1), 5958. https://doi.org/10.1038/s41598-024-55902-z

[62] Shreevastava, S., Maratha, P., Som, T., & Tiwari, A. K. (2025). A novel (alpha, beta)-indiscernibility-assisted intuitionistic fuzzy-rough set model and its application to dimensionality reduction. *Optimization*, *74*(9), 2091–2111. https://doi.org/10.1080/02331934.2023.2270683

[63] Atanassov, K. (1984). Intuitionistic fuzzy relations. In *Third International Symposium on Automation and Scientific Instrumentation* 56–57.

[64] Burillo, P., & Bustince, H. (1996). Entropy on intuitionistic fuzzy sets and on interval-valued fuzzy sets. *Fuzzy Sets and Systems*, *78*(3), 305–316. https://doi.org/10.1016/0165-0114(96)84611-2

[65] Frank, E., Hall, M., Trigg, L., Holmes, G., & Witten, I. H. (2004). Data mining in bioinformatics using Weka. *Bioinformatics*, *20*(15), 2479–2481. https://doi.org/10.1093/bioinformatics/bth261

[66] Tan, A., Wu, W.-Z., Qian, Y., Liang, J., Chen, J., & Li, J. (2019). Intuitionistic fuzzy rough set-based granular structures and attribute subset selection. *IEEE Transactions on Fuzzy Systems*, *27*(3), 527–539. https://doi.org/10.1109/TFUZZ.2018.2862870

[67] Kelly, M., Longjohn, R., & Nottingham, K. (2023). *The UCI Machine Learning Repository.* https://archive.ics.uci.edu

[68] Singh, P., & Singh, N. (2024). Role of data mining techniques in bioinformatics. In I. Management Association (Ed.), *Research anthology on bioinformatics, genomics, and computational biology* (pp. 1406–1417). IGI Global Scientific Publishing. https://doi.org/10.4018/979-8-3693-3026-5.ch063

[69] Nath, A., & Sahu, G. K. (2019). Exploiting ensemble learning to improve prediction of phospholipidosis inducing potential. *Journal of Theoretical Biology*, *479*, 37–47. https://doi.org/10.1016/j.jtbi.2019.07.009

[70] Wang, Z., Chen, H., Yuan, Z., Yang, X., Zhang, P., & Li, T. (2022). Exploiting fuzzy rough mutual information for feature selection. *Applied Soft Computing*, *131*, 109769. https://doi.org/10.1016/j.asoc.2022.109769

[71] Rank, M., Chaturvedi, N., Vyas, K. D., & Singh, A. (2025). Biodesalination of RO reject water by salinity-tolerant bacteria and nutrient fortification for agricultural applications-A sustainable energy-efficient solution to valorize wastewater. *Desalination*, *600*, 118484. https://doi.org/10.1016/j.desal.2024.118484

[72] Cavallo, B., & Ishizaka, A. (2025). A comparative study on precision of direct evaluations, the pairwise comparisons method and the best-worst method. *Omega*, *130*, 103175. https://doi.org/10.1016/j.omega.2024.103175

How to Cite: Mishra, A. K., Jain, N. K., & Singh, R. K. (2026). Intuitionistic Fuzzy Rough Mutual Information Aided Feature Subset Selection and Its Applications. *Journal of Computational and Cognitive Engineering*. https://doi.org/10.47852/bonviewJCCE62026152