scientific reports

OPEN

Hybrid similarity relation based mutual information for feature selection in intuitionistic fuzzy rough framework and its applications

Anoop Kumar Tiwari[1], Rajat Saini[2✉], Abhigyan Nath[3], Phool Singh[4] & Mohd Asif Shah[5,6,7✉]

Fuzzy rough entropy established in the notion of fuzzy rough set theory, which has been effectively and efficiently applied for feature selection to handle the uncertainty in real-valued datasets. Further, Fuzzy rough mutual information has been presented by integrating information entropy with fuzzy rough set to measure the importance of features. However, none of the methods till date can handle noise, uncertainty and vagueness simultaneously due to both judgement and identification, which lead to degrade the overall performances of the learning algorithms with the increment in the number of mixed valued conditional features. In the current study, these issues are tackled by presenting a novel intuitionistic fuzzy (IF) assisted mutual information concept along with IF granular structure. Initially, a hybrid IF similarity relation is introduced. Based on this relation, an IF granular structure is introduced. Then, IF rough conditional and joint entropies are established. Further, mutual information based on these concepts are discussed. Next, mathematical theorems are proved to demonstrate the validity of the given notions. Thereafter, significance of the features subset is computed by using this mutual information, and corresponding feature selection is suggested to delete the irrelevant and redundant features. The current approach effectively handles noise and subsequent uncertainty in both nominal and mixed data (including both nominal and category variables). Moreover, comprehensive experimental performances are evaluated on real-valued benchmark datasets to demonstrate the practical validation and effectiveness of the addressed technique. Finally, an application of the proposed method is exhibited to improve the prediction of phospholipidosis positive molecules. RF(h2o) produces the most effective results till date based on our proposed methodology with sensitivity, accuracy, specificity, MCC, and AUC of 86.7%, 90.1%, 93.0% , 0.808, and 0.922 respectively.

Keywords Rough set, Granular structure, Intuitionisitic fuzzy relation, Intuitionistic Fuzzy Set, Mutual information

The current trend of accumulation of huge amount of data in different databases pertaining to different domains has given rise to the unique opportunity of knowledge discovery/extraction using a plethora of data mining techniques[1]. These techniques[2] can be explored in three ways namely knowledge types, architecture types, and analysis types along with their powerful applications in distinct research and practical domains to solve the interesting real-world problems. Data Mining plays a vital role in establishing smart agriculture application

[1]Department of Computer Science and Information Technology, Central University of Haryana, Mahendergarh 123031, India. [2]Department of Mathematics, School of Basic Sciences, Central University of Haryana, Mahendergarh 123031, India. [3]Department of Biochemistry, Pt. Jawahar Lal Nehru Memorial Medical College, Raipur 492001, India. [4]Department of Mathematics (SoET), Central University of Haryana, Mahendergarh 123031, India. [5]Department of Economics, Kebri Dehar University, 250, Kebri Dehar, Somali, Ethiopia. [6]Centre of Research Impact and Outcome, Chitkara University Institute of Engineering and Technology, Chitkara University, Rajpura 140401, Punjab, India. [7]Division of Research and Development, Lovely Professional University, Phagwara 144001, Punjab, India. ✉email: afcatrajat@gmail.com; drmohdasifshah@kdu.edu.et

1

tools to accomplish real-time data analysis with large volume of data. Data mining tasks[3] offer essential hidden patterns, correlation, and knowledge from the various applications of bioinformatics datasets, viscous dissipation, and activation energy[4,5]. Machine learning methods provide a set of techniques that can be used to create prediction/discriminatory models and subsequent knowledge extraction, which may facilitate in decision making or for better understanding of the concerned domain[6,7]. The "curse of dimensionality" plagues the effectiveness of various machine learning algorithms, but the development of dimensionality reduction methods[8] have considerably impacted in reducing the effects of redundancy present in high dimensional datasets. In the fields of data mining, signal processing, biomedical imaging, agriculture, industrial engineering, and bioinformatics, researchers frequently face obstacles due to "curse of dimensionality" as it leads to enlarge the cost of data storage and extensive computing[9]. Moreover, this issue directly affects both the efficiency and accuracy to cope with different problems[10]. Dimensionality reduction process can easily eliminate redundancy and/or irrelevancy, noise, minimize the complexity of machine learning methods, and enhance the overall accuracy of classification process, and can be identified as an essential and key phase in pattern recognition scheme[11].

Redundant features affects negatively to the various machine learning algorithms mostly resulting in high computation time and less accurate predictive models[12]. It also complicates the model interpretation. Feature reduction methods can be used to mitigate the negative effects of high dimensional data by facilitating the selection of low dimensional non-redundant subset of features. Feature reduction methods have been found to be very effective in a wide variety of research areas, including biological domain[13,14].

Most popular methods of feature reduction algorithms fall under filter and wrapper methods. While wrapper methods are classifier dependent for the evaluation of features[15,16], filter methods use classifier independent feature selection criterion and are generally less computationally intensive[17].

Previously rough set theory[18,19] have been applied very promisingly in feature selection[20]. Although classical rough set theory based feature selection methods[21,22] can only be used on discrete features, which makes it mandatory for discretization of continuous features[23,24]. There is a fair chance of information loss during the process of discretization[25].

The combination of fuzzy[26] and rough sets[27] effectively deals with uncertainty, vague and incomplete data. Rough set theory has been competently employed to produce the most informative features from a dataset consisted of discretized conditional attribute values. This informative feature subset is produced from the original features set with minimum information loss, and termed as reduct. Rough set deals with vagueness, whilst fuzzy set handles uncertainty. Fuzzy set theory ensures that real-valued datasets can be handled without any further discretization. By combining fuzzy set with rough set, information loss due to discretization can be effectively avoided as fuzzy rough set (FRS) can handle real-valued information system (dataset) directly. FRS can be effectively used for mitigating the effects of information loss as a consequence of discretization of features by using fuzzy similarity measures to tackle the continuous feature values[28]. Broadly, FRS aided dimensionality reduction[29] methods can be categorized into two types[30,31] which are based on discernibility matrix and dependency function[32]. Discernibility matrix assisted approaches provide numerous reduct sets[33], whilst dependency function leads to a single feature subset[34].

In FRS aided dimensionality reduction theory, a similarity relation is incorporated between the data points to construct lower and upper approximations. By taking union of the computed lower approximations, we obtain the positive region of decision. Here, the wider is the obtained membership to positive region; greater is the plausibility of instance belonging to an individual category[35]. Based on dependency function, we compute significance of a subset of features. Moreover, the conditional entropy measure is employed in to calculate reduct set for both homogeneous and heterogeneous information system respectively[36–38]. However, it may lead to misclassification of samples when there is a large degree of imbricate between diverse categories of data. Also, it can cope with only with membership of data point to a set, where uncertainty cannot be handled due to both identification and justification. Hence, there is an essential and utmost requirement of distinct kind of mathematical model that can both fit data, and at the same moment it can tackle uncertainty emerging due to identification[39].

Intuitionistic fuzzy (IF) set[40,41] is step ahead that offers two degree of freedom by taking into consideration both membership and non-membership, which can cope with uncertainty that emerges both in judgement and identification[42]. It has been successfully exercised in decision making[43], image segmentation, rule generation, and machine learning[44,45]. In the recent few years, the assemblage of IF[46] and rough sets[47] are employed to establish numerous IF rough set models[48,49] to effectively handle later uncertainty and vagueness in the data[50,51]. Huang et al.[52] proposed a ranking based model for selecting the neighbourhood of objects[53,54] and presented a Dominant IF Decision Table (DIFDT)[55] by using discernibility matrix and assisted discernibility function[23]. They developed IFRS based reduction technique for knowledge extraction from given information system. Huang et al.[56] presented the IF multigranulation rough set (IFMGRS) model and studied different reduction techniques to eliminate redundant granules by introducing reducts for three different types of IFMGRSs in 2014. Tan et al.[57] used the concept of granular structure to introduce an IF rough set model[58] and employed it for feature selection. Tiwari et al.[59] discussed an IF tolerance relation, which was applied to establish IF rough set aided feature selection. Shreevastava et al.[60] addressed different similarity relation assisted technique to deal with both supervised and semi-supervised data. Tiwari et al.[61–63] and Shreevastava et al.[64,65] elaborated different issues related to feature selection technique and presented several lower and upper approximations by using various mathematical ideas. A feature selection to track multiple samples was presented by Li et al.[66] by using IF clustering notion. IF quantifier was introduced by Singh et al.[67] to construct IF rough set model and its application to feature reduction. Jain et al.[39] tried to minimize noise in the data by using the concept of IF granules and incorporated different types of IF relations to introduce feature selection both robust and non-robust. From the recent published articles, it is conspicuous that the use of IF set theory assisted notion for feature selection is still in its incipient stage. Uncertainty is measured in terms of entropy and has its origin in the telecommunications domain[68,69]. Mutual information (MI)[70] aims to measure the relationship between feature and the target.

Further, it can be stated that mutual information (MI)[71] is an interesting quantity that evaluates the dependence between conditional features and has been repeatedly employed to solve an extensive diverse problems. Feature selection techniques can be converted into effective one by incorporating information entropy estimation notion for attribute extraction based on MI[72] and the conventional feature selection approaches on the basis of class seperability. Broadly MI measures the amount of information that can be deduced from a random variable/vector about another random variable/vector[73,74].

Max-relevance-minimum-redundancy method[75,76] is based on the concept of MI and has been relevant in a number of previous studies. It deduces the target MI with minimum redundancy[10,77] among the selected features. A number of MI based feature selection algorithms have been in practice in various domains[72,74]. Fuzzy rough entropy was effectively used to avoid the limitation of rough entropy to handle the real-valued feature data[78,79], but fuzzy rough entropy leads to lessening monotonically with the rise of the dimensions of data, which can promptly reflect the roughness of information systems. This issue was resolved up to certain extent by presenting the extension of fuzzy rough based information entropy with conditional entropy, joint entropy, and mutual information. However, none of the works has handled the noise, vagueness, and uncertainty due to both identification and judgement simultaneously, which is frequently appearing in the current era of high-dimensional datasets due to advancement of internet based technologies. In the current study, a new IFRS based joint entropy, conditional entropy, and mutual information based on a new IF hybrid relation and IF granular structure to handle the different issues such as later uncertainty, vagueness, and imprecision available in the large volume of high dimensional datasets that may degrade the performances of learning algorithms. Firstly, a novel hybrid IF similarity relation is presented. Secondly, joint and conditional entropies are established in IF rough framework. Thirdly, IF rough mutual information is introduced. Then, lower and upper approximations are computed by using presented hybrid IF similarity relation. Thereafter, dependency function is computed by using the defined lower approximation. Next, significance of feature subset is computed by using IF rough mutual information. Further, a heuristic feature selection algorithm is discussed by using both significance and dependency function. IF rough mutual information are employed to measure the later uncertainty and the correlation between features and class. Next, this algorithm is applied on benchmark datasets, and the reduct is computed. The effectiveness of the proposed algorithm is further explained by measuring the performances of seven widely used learning techniques on reduced data produced by our method and four existing approaches. Finally, the proposed method is applied to enhance the overall prediction to discriminate the phsopholipidosis[80] positive (PL+) and phsopholipidosis negative (PL-) molecules. Phospholipidosis is a condition when there is an abnormal buildup of phospholipids in various tissues due to the usage of cationic amphiphilic pharmaceuticals. Phsopholipidosis (PPL) is a reversible condition, and phospholipidosis levels revert to normal once the cationic amphiphilic medications are stopped[81]. Computational prediction of possible inducing characteristics utilizing structure-activity relationship (SAR) can enhance the traditional high throughput screening and drug development pipelines because to its rapidity and cost-effectiveness[82]. The main contributions of the entire study can be highlighted as follows:

Major contributions of the study

- This study establishes a new hybrid IF similarity relation that can deal with both nominal and numerical features.
- An IF granular structure is presented to handle the noise in mixed data.
- IF rough entropy, joint entropy, and conditional entropy is given to handle the later uncertainty with information entropy.
- Further, the idea of an If rough mutual information is discussed.
- Moreover, this If rough mutual information is employed to evaluate both uncertainty and the correlation between conditional feature and decision class.
- Then, a feature selection approach is introduced by using this IF rough mutual information concept.
- Finally, a framework is designed based on our proposed methods to enhance the prediction of phospholipidosis positive molecules.

Theoretical background

In this segment, few essential basic notions about IF set, IF relation, IF information system, and mutual information is reviewed. These concepts can be explained/described as follows:

Definition 2.1 IF set An IF set X in $\mathbb{U}$ is well defined collection of samples/objects having the form

$$X = \left\{ <x, \mu_X(x), \nu_X(x)> \,\Big|\, \forall x \in \mathbb{U} \right\} \tag{1}$$

where, $\mathbb{U}$ portrays the set of data points/samples/objects. Moreover, $\mu_X : \mathbb{U} \to [0,1]$ along with $\nu_X : \mathbb{U} \to [0,1]$, which holds the essential condition $0 \leq \mu_X(x) + \nu_X(x) \leq 1, \forall x \in \mathbb{U}$. Here, $\mu_X(x)$ and $\nu_X(x)$ are depicted as the imperative membership and non-membership grades for a given element $x \in \mathbb{U}$. Further, $\pi_X(x) = 1 - \mu_X(x) - \nu_X(x)$ portrays the hesitancy grade of $x \in \mathbb{U}$. Additionally, we have $0 \leq \pi_X(x) \leq 1, \forall x \in \mathbb{U}$. Thus, the obtained ordered pair $<\mu_X, \nu_X>$ is depicted as a requisite IF value.

Definition 2.2 IF information system An IF information system (IFIS) can be exemplified by a quadruple ($\mathbb{U}$,C, V_{IF}, IF), where, we have V_{IF}, which is comprised of all IF values. Further, we have a mapping, which can be portrayed by $IF : \mathbb{U} \times C \to V_{IF}$, in such a way that $IF(x,a) = <\mu_X(x), \nu_X(x)>, \forall x \in \mathbb{U}, \forall a \in C$.

Definition 2.3 IF relation Let $R(x_i, x_j) = (\mu_X(x_i, x_j), \nu_X(x_i, x_j))$ be an IF binary relation induced on the system. $R(x_i, x_j)$ is IF similarity relation if it satisfies :

(1)   Reflexivity: For any given i and j,

$$\mu_R(x_i, x_j) = 1 \text{ and } \nu_R(x_i, x_j) = 0 \tag{2}$$

(2)   Symmetry: For any given i and j,

$$\mu_R(x_i, x_j) = \mu_R(x_j, x_i) \text{ and } \nu_R(x_i, x_j) = \nu_R(x_j, x_i) \tag{3}$$

$$\forall x_i, x_j \in \mathbb{U}$$

Definition 2.4 Mutual information Mutual information (MI) can be expresserd based on broadly depicted entropy and well-known conditional entropy by using the following given equation

$$I(P; D) = H(D) - H(D|P) \tag{4}$$

where, $P \subseteq C$, H(D) and H(D|P) depict information entropy and conditional entropy respectively. Decrease of uncertainty about D gernerated by P is evaluated by mutual information and its inverse is computed in the same way. Mutual information is employed to calculate either volume of information of P enclosed in D or D included in P. H(P) is amount of information contained in P about itself which means I(P;P)=H(P)

Definition 2.5 Significance of conditional feature For a given IFIS and $B \subseteq C$, if we have an arbitrary conditional dimension/feature $b \in (C - B)$, then its significance can be illustrated by the following equation

$$SGF(b, B, D) = I(B \cup b; D) - I(B; D) = H(D|B) - H(D|B \cup b) \tag{5}$$

and $B = \phi, SGF(b, B, D) = H(D) - H(D|b) = I(b; D)$, which is a MI between conditional dimension/feature b and decision feature D. If the calculated value of SGF(b, B, D) is greater, then it insinuates that under the known condition of feature subset B, dimension b is found to be more potential for the available decision feature D.

Proposed work

In the underway segment, we demonstrate a hybrid IF similarity relation, granular structure, and MI. Based on these concepts, a feature selection procedure is introduced to discard irrelevancy and redundancy available in the high-dimensional information systems.

IF Relation: For all $a \in C$, and $x_i, x_j \in \mathbb{U}$, the hybrid similarity $R_a^h(x_i, x_j)$ between x_i and x_j with respect to any given a can be defined by:

$$R_a^h(x_i, x_j) = \begin{cases} 1, & a(x_i) = a(x_j) \text{ and } a \text{ is nominal;} \\ 0, & a(x_i) \neq a(x_j) \text{ and } a \text{ is nominal;} \\ 1 - \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} (|\mu_a(x_i) - \mu_a(x_j)| \\ \qquad \times |\nu_a(x_i) - \nu_a(x_j)|), & a \text{ is numerical and } |\mu_a(x_i) - \mu_a(x_j)| \leq \zeta_a \\ & |\nu_a(x_i) - \nu_a(x_j)| > \zeta_a; \\ 0, & a \text{ is numerical and } |\mu_a(x_i) - \mu_a(x_j)| > \zeta_a \\ & |\nu_a(x_i) - \nu_a(x_j)| \leq \zeta_a; \end{cases} \tag{6}$$

where, $\zeta_a = 1 - R_a^h(x_i, x_j)$ is depicted as an adaptive IF radius. The IF relation and IF relation matrix enticed by $a \in \mathbb{U}$ are R_a^h and $M_{R_a^h} = (r_{ij})_{n \times n}$, where $r_{ij} = R_a^h(x_i, x_j)$.

If we have $C_1 = \{a_1, a_2, \ldots, a_{|C_1|}\} \subseteq \mathbb{C}$, then,

$$R_{C_1}^h(x_i, x_j) = \bigwedge_{l=1}^{|C_1|} R_a^h(x_i, x_j) \tag{7}$$

Proof

(1)   Reflexive: If we take a case when $x_i = x_j$, then, proposed relation follows only two cases, which are first and third. Moreover, other two cases are rejected by default.

Case 1. if $a(x_i) = a(x_j)$ where a is a nominal , then we obtain $R_a^h(x_i, x_j) = R_a^h(x_i, x_i) = 1$

Case 2. If a is numerical and $|\mu_a(x_i) - \mu_a(x_j)| \leq \zeta_a$ and $|\nu_a(x_i) - \nu_a(x_j)| > \zeta_a$, then

$$R_a^h(x_i, x_j) = 1 - \frac{1}{n^2} \sum_{j=1}^{n} \sum_{i=1}^{n} (|\mu_a(x_j) - \mu_a(x_i)| |\nu_a(x_j) - \nu_a(x_i)|)$$

Now, if we put $x_i = x_j$, we get the folllowing results:

$$R_a^h(x_i, x_i) = 1 - \frac{1}{n^2} \sum_{i=1}^{n} (|\mu_a(x_i) - \mu_a(x_i)||\nu_a(x_i) - \nu_a(x_i)|)$$

$R_a^h(x_i, x_i) = 1$, therefore, we get $R_a^h(x_i, x_j)$ as refelxive

(2) Symmetry:

$$R_a^h(x_i, x_j) = \begin{cases} 1, & a(x_i) = a(x_j) \text{ and } a \text{ is nominal;} \\ 0, & a(x_i) \neq a(x_j) \text{ and } a \text{ is nominal;} \\ 1 - \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} (|\mu_a(x_i) - \mu_a(x_j)| \\ \qquad \times |\nu_a(x_i) - \nu_a(x_j)|), & a \text{ is numerical and } |\mu_a(x_i) - \mu_a(x_j)| \leq \zeta_a \\ & |\nu_a(x_i) - \nu_a(x_j)| > \zeta_a; \\ 0, & a \text{ is numerical and } |\mu_a(x_i) - \mu_a(x_j)| > \zeta_a \\ & |\nu_a(x_i) - \nu_a(x_j)| \leq \zeta_a; \end{cases} \qquad (8)$$

$$R_a^h(x_i, x_j) = \begin{cases} 1, & a(x_j) = a(x_i) \text{ and } a \text{ is nominal;} \\ 0, & a(x_j) \neq a(x_i) \text{ and } a \text{ is nominal;} \\ 1 - \frac{1}{n^2} \sum_{j=1}^{n} \sum_{i=1}^{n} (|\mu_a(x_j) - \mu_a(x_i)| \\ \qquad \times |\nu_a(x_j) - \nu_a(x_i)|), & a \text{ is numerical and } |\mu_a(x_j) - \mu_a(x_i)| \leq \zeta_a \\ & |\nu_a(x_j) - \nu_a(x_i)| > \zeta_a; \\ 0, & a \text{ is numerical and } |\mu_a(x_j) - \mu_a(x_i)| > \zeta_a \\ & |\nu_a(x_j) - \nu_a(x_i)| \leq \zeta_a; \end{cases} \qquad (9)$$

Now, it can be identified that

$$R_a^h(x_i, x_j) = R_a^h(x_j, x_i)$$

So , $R_a^h(x_i, x_j)$ is symmetric

Since, $R_a^h(x_i, x_j)$ is both reflexive and symmetric. Hence, we can obviously conclude that $R_a^h(x_i, x_j)$ is an IF similarity relation. □

Granular structure

The IF granule $\forall x_i \in \mathbb{U}$ is elicited by C_1 as follows:

$$\mu_{[X_i]_P^\epsilon}(x_j) = \begin{cases} 0, & \mu_{R_p^h}(x_i, x_j) < \epsilon \\ \mu_{R_p^h}(x_i, x_j), & \mu_{R_p^h}(x_i, x_j) \geq \epsilon \end{cases} \qquad x_j \in \mathbb{U} \qquad (10)$$

,
further,

$$\nu_{[X_i]_P^\epsilon}(x_j) = \begin{cases} 0, & \nu_{R_p^h}(x_i, x_j) < \epsilon \\ \nu_{R_p^h}(x_i, x_j), & \nu_{R_p^h}(x_i, x_j) \geq \epsilon \end{cases} \qquad x_j \in \mathbb{U} \qquad (11)$$

$\forall a \in P$ is subset of C and $\epsilon \in [0, 1]$

By using IF granulation structure, rough entropy can be discussed into IF rough framework, and IF rough entropy of a feature can be described by:

Definition 3.1 The IF rough entropy of C_1 can be given as:

$$ET(C_1) = ET\left(R_{C_1}^h\right) = -\frac{1}{n} \sum_{i=1}^{n} \log_2 \frac{1}{\left|[x_i]_{R_{C_1}^h}\right|} \qquad (12)$$

It is obvious to identify that $0 \leq ET(C_1) \leq \log_2 n$ iff $\forall x_i, x_j \in \mathbb{U}, R_{C_1}^h(x_i, x_j) = 1, \left|[x_i]_{R_{C_1}^h}\right| = n$, so $ET(C_1) = \log_2 n$. In this suit all the sample pairs are found to be identical. Therefore, the obtained granulation space is found to be the largest at this time, on the contrary $\forall x_i \neq x_j \ R_{C_1}^h(x_i, x_j) = 0$, which indicates $\left|[x_i]_{R_{C_1}^h}\right| = 1$. Therefore, $ET(C_1) = \log_2 n = 0$. Now,the granulation space is instituted as the smallest one.

Definition 3.2 The IF joint rough entropy of C_1 and C_2 can be expressed by :

5

$$ET(C_1, C_2) = ET\left(R^h_{C_1 \cup C_2}\right) = -\frac{1}{n}\sum_{i=1}^{n}\log_2 \frac{1}{\left|[x_i]_{R^h_{C_1}} \cap [x_i]_{R^h_{C_2}}\right|}. \tag{13}$$

Definition 3.3 The IF rough conditional entropy of C_2 relative to C_1 can be addressed by the following equation :

$$ET(C_2|C_1) = -\frac{1}{n}\sum_{i=1}^{n}\frac{\left|[x_i]_{R^h_{C_1}}\right|}{\left|[x_i]_{R^h_{C_1}}\right| \cap \left|[x_i]_{R^h_{C_2}}\right|} \tag{14}$$

Definition 3.4 The IF rough mutual information of C_2 and C_1 can be computed as follows;

$$I(C_2; C_1) = -\frac{1}{n}\sum_{i}^{n}\log_2 \frac{\left|[x_i]_{R^h_{C_1}}\right| \cap \left|[x_i]_{R^h_{C_2}}\right|}{\left|[x_i]_{R^h_{C_1}}\right|\left|[x_i]_{R^h_{C_2}}\right|} \tag{15}$$

Definition 3.5 The IF rough mutual information between D and C_1 can be illustrated by the equation:

$$I(D; C_1) = -\frac{1}{n}\sum_{i=1}^{n}\log_2 \frac{\left|[x_i]_{R^h_{C_1}} \cap [x_i]_D\right|}{\left|[x_i]_{R^h_{C_1}}\right||[x_i]_D|} \tag{16}$$

By using this equation, IF rough mutual information $I(d; C_1)$ considers as the correlation between C_1 and decision feature D . If the obtained value of IF rough mutual information between D and C_2 is higher, then, we get more correlated value between C_1 and D.

Proposition 3.6 If $C_1 \subseteq C_2 \subseteq C$, then $R^h_{C_1} \supseteq R^h_{C_2}$

Proof As discussed by the aforesaid definition 3.1, $R^h_{c_1}(x_i, x_j) = \bigwedge_{l=1}^{|C_1|} R^h_{C_1}(x_i, x_j)$, $R^h_{c_2}(x_i, x_j) = \bigwedge_{l=1}^{|C_2|} R^h_{C_2}(x_i, x_j)$ and

$|C_1| \leq |C_2| \Rightarrow R^h_{c_2}(x_i, x_j) \subseteq R^h_{C_1}(x_i, x_j) \Rightarrow R^h_{C_1} \supseteq R^h_{C_2}$
Now, $R^h_{C_1} \supseteq R^h_{C_2} \iff \forall x_i, x_j \in \mathbb{U}$;
$R^h_{C_1}(x_i, x_j) \geq R^h_{C_2}(x_i, x_j)$ □

Proposition 3.7 If $R^h_{C_1} \subseteq R^h_{C_2}$, then $ET\left(R^h_{C_1}\right) \leq ET\left(R^h_{C_2}\right)$.

Proof For a given $R^h_{C_1} \subseteq R^h_{C_2}$, we have $\forall x_i, x_j \in \mathbb{U}$. Now, we can simply write $R^h_{C_1}(x_i, x_j) \leq R^h_{c_2}(x_i, x_j)$
$\Rightarrow \left|[x_i]_{R^h_{C_1}}\right| \leq \left|[x_i]_{R^h_{C_2}}\right|$
Therefore, we detect the result by using the definition 3.1 as $ET\left(R^h_{C_1}\right) \leq ET\left(R^h_{C_2}\right)$. □

Proposition 3.8 IF $C_1 \subseteq C_2 \subseteq C$ then $ET(C_1) \geq ET(C_2)$

Proof For any given $C_1 \subseteq C_2$, we have the following expression based on the Proposition 3.6,
$R^h_{C_1} \supseteq R^h_{C_2}$. Moreover, by using Proposition 3.7, we can conclude the following result:
$ET(C_1) \geq ET(C_2)$
Proposition 3.8 depcits that IF rough entropy reduces when feature subset accquire larger size, whilst,it grows in case of features subset procures smaller size . It can be easily observed that IF rough entropy definition can evaluate the uncertainty of IF approximation space. □

Proposition 3.9 Suppose $C_1, C_2 \subseteq C$, then $ET(C_1, C_2) \leq \min[ET(C_1), ET(C_2)]$

Proof Since $\forall x_i \in \mathbb{U} [x_i]_{R^h_{C_1}} \cap [x_i]_{R^h_{C_2}} \subseteq [x_i]_{R^h_{C_1}}$ and $[x_i]_{R^h_{C_1}} \cap [x_i]_{R^h_{C_2}} \subseteq [x_i]_{R^h_{C_2}} \Rightarrow \left|[x_i]_{R^h_{C_1}} \cap [x_i]_{R^h_{C_2}}\right| \leq \left|[x_i]_{R^h_{C_1}}\right|$
and $\left|[x_i]_{R^h_{C_1}} \cap [x_i]_{R^h_{C_2}}\right| \leq \left|[x_i]_{R^h_{C_2}}\right|$. By Proposition 3.2, we have $ET(C_1, C_2) \leq ET(C_1)$ and $ET(C_1, C_2) \leq ET(C_2)$.
$\Rightarrow ET(C_1, C_2) \leq \min(ET(C_1), ET(C_2))$. □

Proposition 3.10 IF $C_1 \subseteq C_2 \subseteq C$, then $ET(C_1, C_2) = ET(C_2)$

Proof Since $C_1 \subseteq C_2$, hence, by using the Proposition 3.6, we get
$R^h_{C_1} \supseteq R^h_{C_2} \Rightarrow [x]_{R^h_{C_1}} \supseteq [x]_{R^h_{C_2}} \Rightarrow [x]_{R^h_{C_1}} \cap [x]_{R^h_{C_2}} = [x]_{R^h_{C_2}}$ So, $ET(C_1, C_2) = ET(C_2)$ □

According to the Proposition 3.10, when there are two IF granules produced by two potential feature subsets respectively, then IF joint rough entropy of the calculated two potential feature subsets is equal to the IF rough entropy of the feature subsets corresponding to relatively smaller IF granulation.

Proposition 3.11 $ET(C_2|C_1) = ET(C_2, C_1) - ET(C_1)$.

Proof Based on the Definition 3.3, we have $ET(C_1) + ET(C_2|C_1) = -\frac{1}{n}\sum_{i=1}^{n}\log_2\frac{1}{|[x_i]_{R_{C_1}^h}|} - \frac{1}{n}\sum_{i=1}^{n}$

$\frac{|[x_i]_{R_{C_1}^h}|}{|[x_i]_{R_{C_1}^h}\cap[x_i]_{R_{C_2}^h}|}$ $\Rightarrow ET(C_1) + ET(C_2|C_1) = -\frac{1}{n}\left[\sum_{i=1}^{n}\left[\log_2\frac{1}{|[x_i]_{R_{C_1}^h}|} + \log_2\frac{|[x_i]_{R_{C_1}^h}|}{|[x_i]_{R_{C_1}^h}\cap[x_i]_{R_{C_2}^h}|}\right]\right]$ $\Rightarrow ET(C_1)$

$+ET(C_2|C_1) = -\frac{1}{n}\sum_{i=1}^{n}\left[\log_2\frac{|[x_i]_{R_{C_1}^h}|}{|[x_i]_{R_{C_1}^h}||[x_i]_{R_{C_1}^h}\cap[x_i]_{R_{C_2}^h}|}\right]$ $\Rightarrow ET(C_1) + ET(C_2|C_1) = -\frac{1}{n}\sum_{i=1}^{n}\log_2$

$\Rightarrow ET(C_1) + ET(C_2|C_1) = -\frac{1}{n}\sum_{i=1}^{n}\log_2\frac{1}{|[x_i]_{R_{C_1}^h}\cap[x_i]_{R_{C_2}^h}|}$ $\Rightarrow ET(C_1) + ET(C_2|C_1) = E(C_1, C_2)$

$\Rightarrow ET(C_2|C_1) = ET(C_1, C_2) - ET(C_1)$ $\square$

Proposition 3.12 If $C_1 \subseteq C_2 \subseteq C$, then $ET(C_2|C_1) = 0$

Proof Since, $C_2 \subseteq C_1$, hence, based on the Proposition 3.6, we can conclude that $R_{C_1}^h \subseteq R_{C_2}^h$. Therefore, $\forall x_i$, $[x_i]_{R_{C_1}^h} \subseteq [x_i]_{R_{C_2}^h}$, furthermore, $\forall x_i$, $x_i, [x_i]_{R_{C_1}^h}\cap[x_i]_{R_{C_2}^h} = [x_i]_{R_{C_1}^h}$, now, based on the Definition 3.3, we have

$ET(C_2|C_1) = -\frac{1}{n}\sum_{n=1}^{n}\log_2\frac{\left|[x_i]_{R_{C_1}^h}\right|}{\left|[x_i]_{R_{C_1}^h}\cap[x_i]_{R_{C_2}^h}\right|} \Rightarrow ET(C_2|C_1) = -\frac{1}{n}\sum_{n=1}^{n}\log_2\frac{\left|[x_i]_{R_{C_1}^h}\right|}{\left|[x_i]_{R_{C_1}^h}\right|} = -\frac{1}{n}\sum_{n=1}^{n}\log_2 1 = 0$

$\square$

IF rough mutual information can't only be used to measure the uncertainty of IF approximation space but also can be applied to evaluate the correlation between conditional feature and decision class.

Proposition 3.13 $I(C_1; C_2) = ET(C_2) - ET(C_2|C_1) = ET(C_1) - ET(C_1|C_2)$

Proof Based on the Proposition 3.9, we have $ET(C_2) - ET(C_2|C_1) = -\frac{1}{n}\sum_{i=1}^{n}\log_2\frac{1}{\left|[x_i]_{R_{C_2}^h}\right|} + \frac{1}{n}$

$\sum_{i=1}^{n}\log_2\frac{\left|[x_i]_{R_{C_1}^h}\right|}{\left|[x_i]_{R_{C_1}^h}\cap[x_i]_{R_{C_2}^h}\right|}$ $\Rightarrow ET(C_2) - ET(C_2|C_1) = -\frac{1}{n}\sum_{i=1}^{n}\left[\log_2\frac{1}{\left|[x_i]_{R_{C_2}^h}\right|} - \log_2\frac{\left|[x_i]_{R_{C_1}^h}\right|}{\left|[x_i]_{R_{C_1}^h}\cap[x_i]_{R_{C_2}^h}\right|}\right]$ $\Rightarrow ET$

$(C_2) - ET(C_2|C_1) = -\frac{1}{n}\sum_{i=1}^{n}\log_2\frac{\left|[x_i]_{R_{C_1}^h}\cap[x_i]_{R_{C_2}^h}\right|}{\left|[x_i]_{R_{C_1}^h}\right|\left|[x_i]_{R_{C_2}^h}\right|} = I(C_1; C_2)$ Similarly, we can get $I(C_1; C_2) = ET(C_1)$

$-ET(C_1|C_2)$ $\square$

Proposition 3.14 $I(C_1; C_2) = I(C_2; C_1) = ET(C_1) + ET(C_2) - ET(C_1, C_2)$

Proof Obviously $I(C_1; C_2) = I(C_2; C_1)$ satisfies based on the Definitions 3.1, 3.4, and 3.5. Now, we obtain the following results: $ET(C_1) + ET(C_2) - E(C_1, C_2) = -\frac{1}{n}\sum_{i=1}^{n}\log_2\frac{1}{\left|[x_i]_{R_{C_1}^h}\right|} - \frac{1}{n}\sum_{i=1}^{n}\log_2\frac{1}{\left|[x_i]_{R_{C_2}^h}\right|} + \sum_{i=1}^{n}\log_2$

$\frac{1}{\left|[x_i]_{R_{C_1}^h}\cap[x_i]_{R_{C_2}^h}\right|}$ $\Rightarrow ET(C_1) + ET(C_2) - ET(C_1, C_2) = -\frac{1}{n}\sum_{i=1}^{n}\left[\log_2\frac{1}{\left|[x_i]_{R_{C_1}^h}\right|} + \log_2\frac{1}{\left|[x_i]_{R_{C_2}^h}\right|} - \log_2\right.$

$\left.\frac{1}{\left|[x_i]_{R_{C_1}^h}\cap[x_i]_{R_{C_2}^h}\right|}\right] \Rightarrow ET(C_1) + ET(C_2) - ET(C_1, C_2) = -\frac{1}{n}\sum_{i=1}^{n}\log_2\frac{\left|[x_i]_{R_{C_1}^h}\cap[x_i]_{R_{C_2}^h}\right|}{\left|[x_i]_{R_{C_1}^h}\right|\left|[x_i]_{R_{C_2}^h}\right|} = I(C_1; C_2)$ $\square$

Definition 3.15 For a given IFIS, let P be subset of conditional dimensions/features(C). Thereafter, $\forall Y \in (C - P)$ is found to be the significance as $\Omega(Y, P, D)$, which can be further given by:

$$\Omega(Y, P, D) = I(P \cup Y; D) - I(Y; D) \tag{17}$$

$Y = \phi, \Omega(T, P, D)$, and can be outlined as, $\Omega(Y, D) = ET(D) - ET(D|Y) = I(Y; D)$, which depicts the MI of IF conditional feature T and the decision feature D. If the value of $\Omega(T, P, D)$ increases, then IF conditional dimension/feature T is obtained to be more relevant for a given decision feature D.

Step 1 Input real valued information system $< \mathbb{U}, C \cup D >$ and adaptive radius ζ_a

Step 2 Perform intuitionistic fuzzification by Tan et al. [62] for $< \mathbb{U}, C >$

Step 3 Compute the mutual information I(D; K) for the IFIS based on hybrid similarity relation R^h , where $K \subseteq C$

Step 4 $R \leftarrow \phi$

Step 5 Do

Step 6 set $K \leftarrow R$

Step 7 For every $a \in (C - K)$

Step 8 compute the significance of conditional feature a, i.e. $\Omega(a, K, D)$;

Step 9 Pick out the feature which results in the maximum value of $\Omega(a, K, D)$

Step 10 If

       multiple features produces the maximum value, select the feature whose combination with K results in least value

       then

       $R \leftarrow K \cup a;$

       end if

Step 11 End for

Step 12 Untill I(K;D)=I(C;D);

Step 13 Return R

Algorithm 1. Feature selection alogrithm based on IF mutual information (FSIFMI)

Experimentation

In the current experimental section, the performance of the proposed method is evaluated and compared with the existing fuzzy and IF sets assisted techniques. All the pre-processing concepts are implemented in Matlab 2023[83] and learning algorithms are implemented in WEKA[84]. Firstly, fuzzification and intuitionistic fuzzification of the real valued data is performed by using the methods proposed by Jensen et al.[6] and Tan et al.[57] respectively. Secondly, the reduced datasets are obtained by the previously presented approaches. Thirdly, different threshold parameters values are adjusted for our established method to produce the reduct. Then, reduced datasets are generated by discarding the noise to the maximum level. The reduct is computed by changing the value of ξ from 0.1 to 0.8 in small interval, and the value of ξ providing the maximum performance measures in the experiment is selected as the final one. To perform the entire experimental study, the following setup is exercised to conduct the comprehensive experiments:

Dataset

Ten benchmark datasets are taken from widely discussed University of California from Irvine based Machine Learning Repository[85] to conduct the entire experiments. The required details of these datasets are outlined in Table 1. The dimension and size of these datasets depict that these are small to large datasets as number of data points range from 62 to 4521 and features range from 9 to 10000.

Dataset	Instances	Features	Reduct size				
			FSFrMI	GIFRFS	TIFRFS	FRFS	IFRFSMI
Bank marketing	4521	16	10	12	15	15	14
Breast cancer	699	9	8	9	9	8	8
Dbworld-bodies	64	4702	97	128	187	88	8
Arcene	200	10000	453	287	303	268	169
Colon	62	32	24	27	21	18	8
Gsar-biodegradation	1055	41	31	36	29	33	25
Fertility diagnosis	100	9	8	6	8	7	7
Thyroid- hypothyroid	3163	25	11	17	19	15	12
Heart disease	294	13	11	10	10	12	9
Wdbc	569	21	17	14	18	10	8

Table 1. Dataset characteristics and reduct size.

Classifiers

Seven different learning methods[86] are applied to demonstrate the performance measures on the reduced datasets obtained from different feature selection techniques. RealAdaBoost with random forest as base classifier (RARF) and IBK are employed for the objective of evaluating overall classification accuracies with standard deviation by using diverse validation techniques for ten benchmark reduced datasets. Moreover, we applied naive bayes, SMO, IBK, RARF, PART, JRip,J48, and random forest (RF) to evaluate the performances based on various evaluation metrics for the reduced Nath et al.[87] dataset for evaluating the effectiveness of the proposed technique when compared to existing method for discriminating PL+ and PL- molecules.

Dataset split: Feature selection process is carried out over complete information system. After production of reduced datasets, individual learning algorithm is evaluated based on percentage split of 66:34 and kd-fold cross validation. In percentage split technique, dataset is randomly divided into two parts, where training is done on 66% of the entire dataset, while 34% of the dataset is employed to perform testing. In kd-fold cross validation, whole dataset is randomly separated into kd subsets, where kd-1 parts form training set, whilst one is employed to form testing set. After kd such repetitions, average value of different evaluation metrics is considered as final performance. In the current study, the value of kd is taken as 10.

Performance evaluation metrics

The prediction performance measures of the seven learning algorithms from different categories are evaluated using both broadly elaborated threshold-dependent and threshold-independent assessment parameters. These assessment parameters are ascertained based on the calculated values of true positive (TRP), true negative (TRN), false positive (FLP), and false negative (FLN). TRP is computed number of correctly predicted positive data points; TRN is calculated number of correctly predicted negative data points. FLN is representation for the number of incorrectly predicted positive samples, while FLP is depiction for the number of incorrectly predicted negative samples. We employ different parameters namely: Sensitivity (Sn), Specificity (Sp), Accuracy (Ac), AUC, and MCC to measure the overall performances of the individual learning algorithms. Now, these evaluation parameters can be mathematically discussed as follows:

Sn: This calculates the overall percentage of correctly classified PPL+, which is specified by:

$$Sn = \frac{TRP}{(TRP + FLN)} \times 100 \tag{18}$$

Sp: This includes the efficacious percentage of correctly classified PPL−, which is produced by:

$$Sp = \frac{TRN}{(TRN + FLP)} \times 100 \tag{19}$$

Ac: The percentage of required overall correctly classified PPL+ and PPL−, which can be stated as:

$$Ac = \frac{TRP + TRN}{(TRP + FLN + TRN + FLP)} \times 100 \tag{20}$$

AUC: It is applied to observe the important and required area under the receiver operating characteristic curve (ROC), the more tends its count towards 1, the better will be the obtained predictor.

MCC: Mathew's correlation coefficient is a very much potential and the most awaited parameters, which is computed with the help of following equation:

$$MCC = \frac{TRP \times TRN - FLN \times FLP}{\sqrt{((TRP + FLP)(TRP + FLN)(TRN + FLN)(TRN + FLP)}} \times 100 \tag{21}$$

This parameter is applied not only to clarify the effectiveness of the binary classifications but also to justify its efficiency. An MCC value tends towards 1 to specify that the predictor is the promising one.

Results and discussion

The details of the ten benchmark datasets along with the reduct as produced by four existing as well as presented methods is depicted in Table 1. Real-valued datasets are converted into fuzzy and IF values by using widely discussed Jensen et al.[6] and Tan et al.[57] concepts. Entire reduction process is accomplished over complete data by using both fuzzy and IF aided techniques. FSFrMI[72], GIFRFS[57], TIFRFS[59], and FRFS[6] are the earlier efficacious and effective techniques, which are incorporated to perform the comparative results (Table 2). Our proposed method produced reduct set range from 7 to 169, where reduct size is smaller when compared to reduct size by earlier approaches, except bank marketing and thyroid-hypothyroid datasets. For bank marketing dataset, FSFrM and GIFRFS resulted in relatively less size data, whilst smaller size is produced by FSFrMI and FRFS for thyroid-hypothyroid and fertility diagnosis datasets respectively in contrast with IFRFSMI. Moreover, for breast cancer, FSFrM and FRFS provide the similar size, whilst, for fertility diagnosis dataset FRFS produce similar size of the data when compared to the results presented by proposed method. From the recorded reduct in Table 1, it can be observed that our proposed technique is generating more reduced dimensions for most of the cases related to all the ten datasets rather than recently established powerful methods. We have presented the visualization of reduction process based on different methods in Fig. 1, which clearly indicates that our proposed method produces high percentage of overall feature elimination with the increment of total conditional features. Then, IBK and RARF are chosen to show the learning performances in terms of standard deviation with overall accuracies for the reduced datasets generated by four existing and our proposed techniques, where 10-fold cross validation
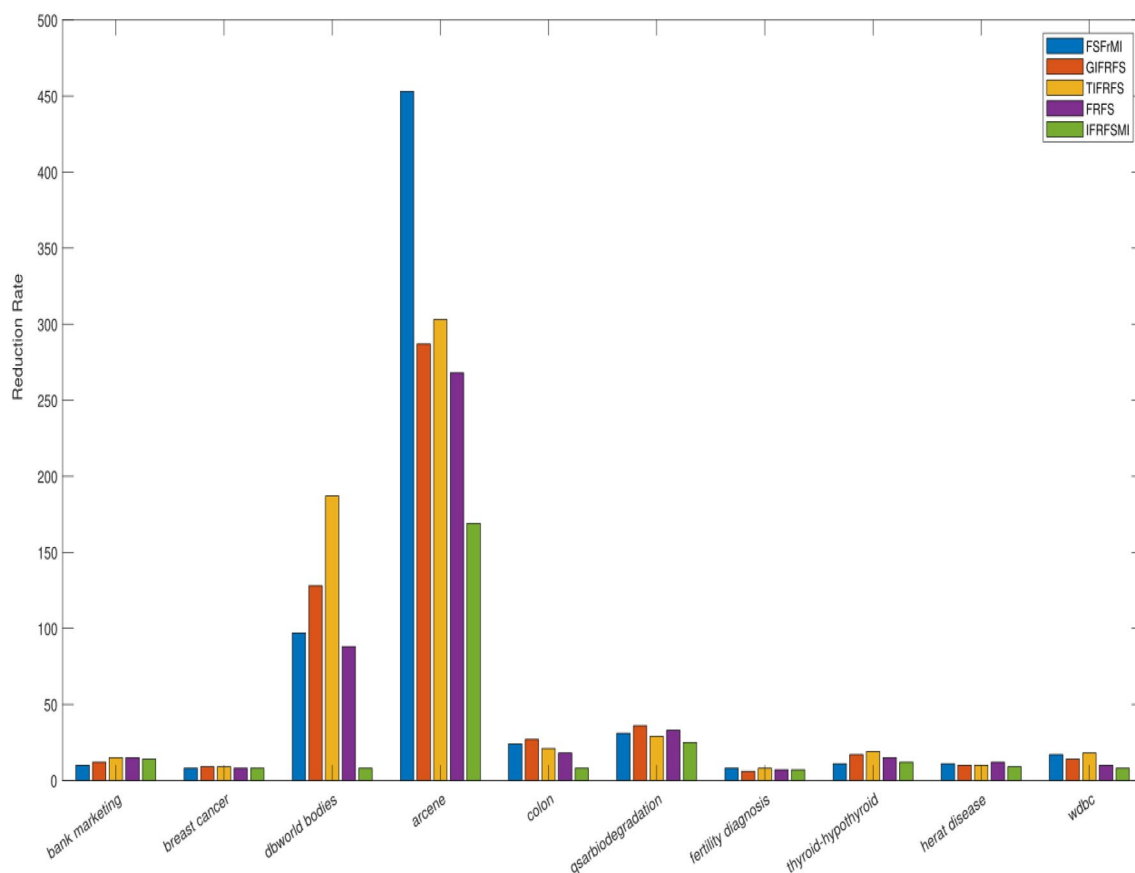
Figure 1. Comparison of overall reduction for different daasets by previous and proposed methods.

is employed to avoid the overfitting. These results are reported in Table 2, where the ranks are outlined in the superscript of all the individual results. From the results available in Table 2, it is obvious that our proposed method is dispensing the better results in contrast with the results of other previous approaches regardless of reduced data produced by previous approaches, except the outcome for breast cancer and heart disease datasets. For breast cancer dataset, TIFRFS presents better outcome when compared to IFRFSMI by using both IBK and RARF, while, for heart disease dataset TIFRFS gave the best result with RARF. For colon and heart disease datasets, GIFRFS and TIFRFS leads to identical results as compared to IFRFSMI based results by IBK. Similar results are presented by RARF for fertility diagnosis and wdbc datasets based on the reduced datasets produced by FSFrMI and GIFRFS respectively in contrast with proposed method based reduced datasets. Entire results can be visualized by Figs. 2 and 3. These figures depict that proposed concept are very much effective for both low and high-dimensional datasets as the reduced datasets produced by this method always leads to increment of overall accuracies of the different learning algorithms regardless of their dimensionality size.

Our assumptions to verify the significance of our proposed method are as follows:

Null Hypothesis: All the employed methods are equivalent.

Alternate Hypothesis: There is significant difference among the employed methods.

Two widely accepted testing approaches namely Freidman test[88] and Bonferoni Dunn test[89] are applied to validate the significance of the presented method. Freidman test is used to perform comparative study of multiple models. Further, Bonferoni Dunn is employed to obtain which method is significantly different from proposed technique. The null hypothesis can be rejected at α% level of significance if the values between their average ranks is higher rather than critical distance value. In the current study, average ranks by both IBK and RARF based on our proposed method are recorded as the minimum value (Table II). These values are clearly depicting the superiority of our established models. Moreover, F-statistics computed values based on IFRFSMI are obtained larger for both IBK and RARF when compared to F-tabular value. F-statistics computed values for IBK and RARF are 23.09 and 32.38 (Table II), whilst F-tabular value is 2.634 ($F(4,36) = 2.634$ at 5% level of significance). Therefore, based on Dunn Test our proposed method is found as significantly different.

Case study: an application to discriminate PL+ and PL- molecules
One of the prime applications of machine learning based methods in cheminformatics is the reduction of enormous chemical space with respect to some property of interest. The reduced chemical space can then be validated using wet lab based experiments, thus making the fidelity of machine learning methods of outmost importance.

Dataset	Classifier	FSFrMI	GIFRFS	TIFRFS	FRFS	IFRFSMI
Bank	IBK	84.75±2.88[2]	83.79±3.22[3]	83.01±2.19[4]	81.21±2.22[5]	86.28±1.29[1]
Marketing	RARF	87.23±3.11[3]	86.18±2.33[4]	87.59±1.21[2]	83.18±1.99[5]	89.37±0.86[1]
Breast	IBK	81.11±0.76[5]	86.24±3.83[3]	96.11±2.11[1]	84.29±2.89[4]	95.67±2.43[2]
Cancer	RARF	89.34±4.12[4]	93.34±3.02[3]	97.12±1.95[1]	88.66±3.22[5]	96.04±2.36[2]
Dbworld	IBK	89.16±7.27[4]	90.86±9.25[3]	91.89±7.23[2]	88.89±8.23[5]	94.74±8.28[1]
Bodies	RARF	90.25±6.88[5]	92.19±7.23[3]	93.55±7.89[2]	90.55±7.69[4]	97.21±6.00[1]
Arcene	IBK	71.47±10.25[3]	70.72±9.01[4]	72.09±10.12[2]	71.09±10.44[4]	74.00±9.53[1]
	RARF	75.69±7.55[3]	74.69±8.65[4]	77.55±9.28[2]	72.35±9.68[5]	83.45±9.09[1]
Colon	IBK	75.88±6.18[4]	78.12±5.84[2,5]	79.06±5.19[1]	73.06±7.88[5]	78.12±5.84[2,5]
	RARF	79.21±3.29[4]	80.41±2.99[3]	81.17±3.33[2]	77.17±3.33[5]	82.81±12.55[1]
Qsarbio-degradation	IBK	78.27±4.33[4]	77.69±3.87[3]	79.51±5.11[2]	75.87±4.45[5]	82.09±12.55[1]
	RARF	80.28±5.19[4]	81.33±4.66[3]	82.06±3.77[2]	79.16±4.78[5]	86.74±3.04[1]
Fertility diagnosis	IBK	83.21±9.88[2]	81.41±10.18[4]	83.17±9.99[3]	80.17±9.87[5]	84.30±9.98[1]
Thyroid- hypothyroid	RARF	87.20±6.68[1,5]	83.69±7.65[4]	85.23±5.77[3]	82.87±6.45[5]	87.20±6.68[1,5]
	IBK	92.33±3.22[3]	91.23±2.66[4]	95.16±2.77[2]	88.33±2.34[5]	97.87±0.69[1]
	RARF	95.21±2.88[3]	93.41±1.18[4]	97.17±2.55[2]	92.17±1.87[5]	99.11±0.46[1]
Heart disease	IBK	79.26±1.03[3]	78.46±2.28[4]	81.16±1.99[1,5]	76.25±2.99[5]	81.16±1.99[1,5]
	RARF	81.27±1.79[3]	80.38±1.23[4]	83.69±1.18[1]	78.98±1.55[5]	82.74±1.50[2]
Wdbc	IBK	95.68±0.28[2]	93.46±1.28[4]	95.16±1.87[3]	89.33±2.65[5]	96.06±0.11[1]
	RARF	96.41±2.28[4]	97.73±2.99[1,5]	97.69±3.19[3]	91.26±3.59[5]	97.73±2.99[1,5]
Average	IBK	3.20	3.55	2.15	4.80	1.30
Rank	RARF	3.45	3.35	2.00	4.90	1.30
F statistics	IBK	23.09				
	RARF	32.38				

Table 2. Comparison of overall accuracies with standard deviation for the datasets produced by FSFrMI GIFRFS, TIFRFS, FRFS, and IFRFSMI by using 10-fold cross validation.

One of the hallmarks of phospholipidosis is the accumulation of phospholipids in the various types of tissues for eg. kidneys, eyes etc. mostly caused by cationic amphiphilic molecules. Highly accurate machine learning prediction models can facilitate in screening of phospholipidosis inducing compounds in early stages of drug discovery workflows, thereby reducing the cost and time associated with wet lab based experiments (Fig. 4).

The present methodology can open new possibilities for further research in early screening of phospholipidosis inducing molecules.

Now, our proposed approach is applied to Nath et al.[87] dataset to produce the effective reduced form by minimizing noise, uncertainty, imprecision available in the data along with removal of redundant, and irrelevant attributes. Thereafter, seven classifiers from different categories are investigated to evaluate their performances over this reduced dataset based on sensitivity, AUC, Specificity, MCC, and accuracy, which have reported in Tables 3, 4, 5 and 6. Moreover, for original and reduced data, a commodious approach to represent theoverall performance measures of all the seven classifiers at the best decision threshold can be given by Receiver Operating Characteristic (ROC) curve, which furnishes a visual explanation of the classifiers performance. Figures 5 and 6 depict ROC curves for original and reduced dataset based on 10-fold cross validation. These figures indicate that RARF algorithm achieved the best AUC in comparison to all the other algorithms(> 0.89).

To compare with the performance evaluation metrics for the phospholipidosis dataset, we used the same package in R (https://https://cran.r-project.org/web/packages/h2o/index.html)as used in the original work (Nath et al.[87]). We used a grid search strategy to obtain the best hyperparameters for the random forest algorithm Hyperpaprametersntrees = c(20,50,100,500),max depth = c(20,40,60,80),sample rate = c(0.2,1,0.01). Further, we used the same of features (JOELib+Structural alerts), which are calculated using the ChemMine tools webserver (https://chemminetools.ucr.edu/). The dataset consisted of 102 phospholipidosis inducing compounds (positive samples) and 83 phospholipidosis non-inducing compounds (negative samples), thus constituting a total of 185 molecules. Schematic representation for entire process is given by Fig. 7. In the current methodology, we start the process with a dataset consisted of phospholipidosis positive molecules and phospholipidosis negative molecules. Then, descriptor generator converts the initial data into target data. Further, SMOTE is applied to obtain the balanced dataset. Next, this dataset is converted into intuitionistic fuzzy information system by using Tan et al.[57] approach. Thereafter, our proposed feature subset selection method is applied to remove noise, vagueness,
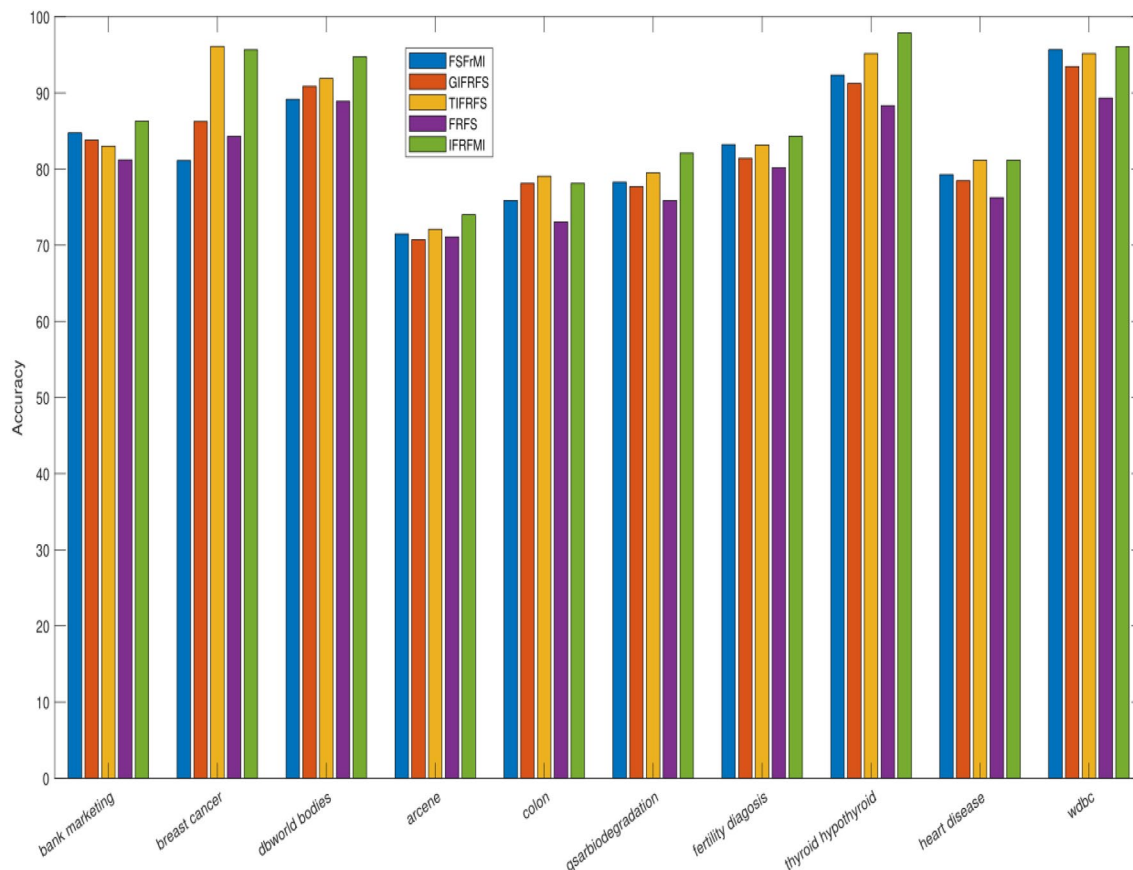
Figure 2. Comparison of average accuracies by IBK for different reduced datasets as produced by existing and proposed methods.

irrelevancy, redundancy, and uncertainty to obtain reduced dataset. Moreover, several classifiers are used to discriminate positive and negative classes. Finally, RARF is identified as the best performer.

The performance evaluation metrics for the current method and the previous ensemble based method are presented in Table 7. The dataset preprocessing introduced in the current work resulted in enhanced performance evaluation metrics for the RF algorithm in comparison to the previously published results. Notably a 2 percent rise on overall accuracy is observed. As the dataset is slightly imbalanced, a rise in MCC for the current method proves the usefulness of the dataset preprocessing step. The ROC plot for the RF(h2o) model is presented in Fig. 4. An AUC value of 0.922 indicates an acceptable prediction model for phospholipidosis inducing molecules. In the end of the entire study, the list of abrreviations, signs, and symbols are presented in Table 8.

Ethical approval
This article does not contain any studies with human participants or animals performed by any of the authors.

Conclusion
Dimensionality reduction broadly aims to obtain a feature subset from existing original feature set by using certain powerful evaluation criterion. Since dimensionality reduction can produce efficient feature subset, where feature selection has found as an interesting central technique for data pre-processing in various beneficial and interesting data mining tasks. Conventional fuzzy rough set frequently incorporates dependency function as an evaluation criterion of feature subset selection. However, this method only maintained the maximum membership grade of a data point to one decision class and found to be unable in discarding later uncertainty and noise up to certain extent, which cannot characterize the classification error. To avoid these issues, we presented a novel intuitionistic fuzzy aided technique, where feature selection method is established by integrating information entropy with IF rough set concept.
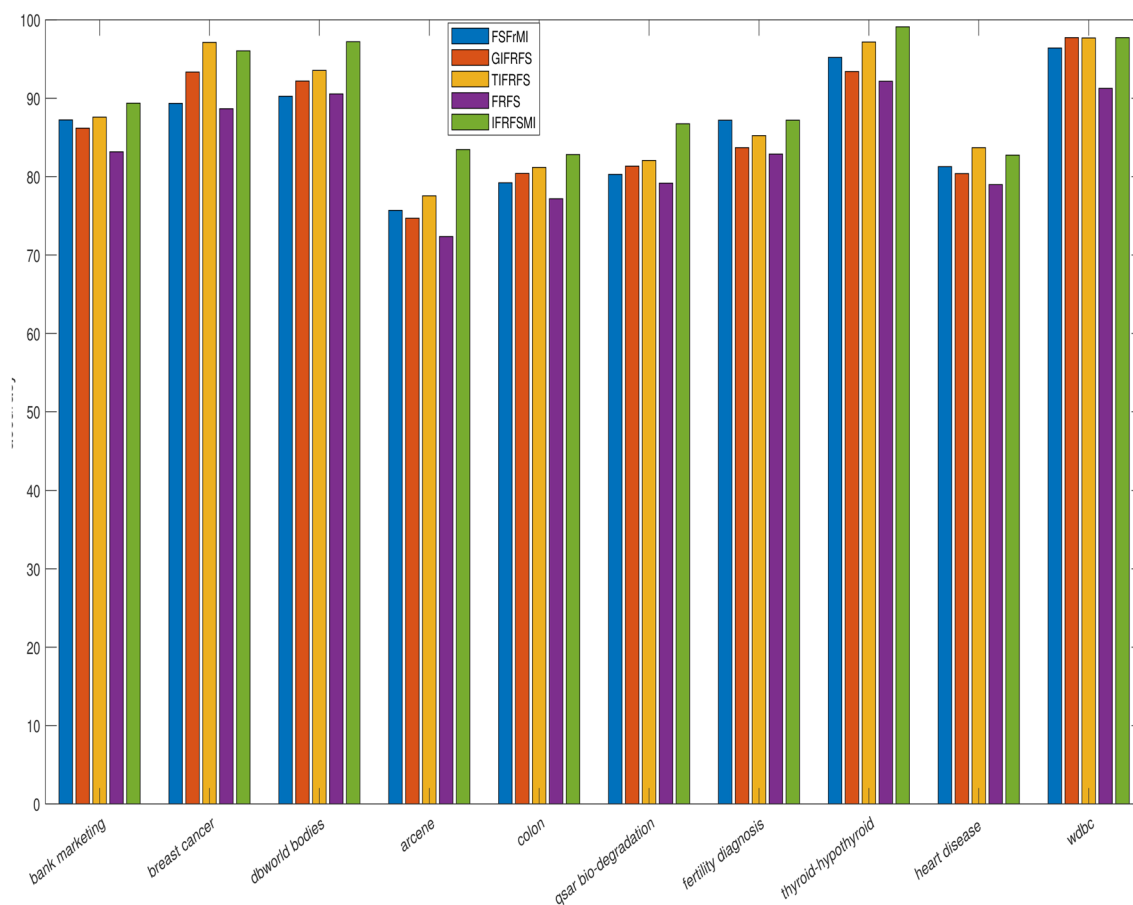
Figure 3. Comparison of average accuracies by RARF for different reduced datasets as produced by existing and proposed methods.

Classifiers	Sensitivity	Specificity	Accuracy	AUC	MCC
Davie Bayes	75.5	81.4	78.4	0.828	0.570
SMO	81.4	85.3	83.3	0.833	0.667
IBK	82.4	80.4	81.4	0.806	0.628
RARF	81.4	85.3	83.3	0.908	0.667
PART	75.5	73.5	74.5	0.718	0.490
JRip	66.7	69.6	68.1	0.723	0.363
RandomForest	83.3	82.4	82.8	0.893	0.657
J48	74.5	74.5	74.5	0.769	0.510

Table 3. Performance evaluation metrics of eight classifiers for original dataset consisting of PL+ and PL-molecules based on 10-fold cross validation.

- Initially, we established a hybrid IF similarity relation, which is further employed to present a novel IF rough joint and conditional entropies.
- Then, IF granular structure was introduced based on the proposed hybrid similarity relation.
- Thereafter, IF rough set model was described by using the aforesaid relation.
- Based on these entropies and granular structure, we suggested a mutual information idea to compute the significance of the feature subset for a decision class.
- Next, mathematical theorems are validated to justify the correctness of the proposed ideas.

Classifiers	Sensitivity	Specificity	Accuracy	AUC	MCC
Navie Bayes	85.3	70.6	77.9	0.846	0.565
SMO	81.4	68.6	75.0	0.750	0.504
IBK	87.3	87.3	87.3	0.811	0.745
RARF	88.2	84.3	86.3	0.925	0.726
PART	71.6	72.5	72.1	0.778	0.441
JRip	74.5	80.4	77.5	0.811	0.550
RandomForest	84.3	84.3	84.3	0.915	0.686
J48	74.5	75.5	75.0	0.752	0.500

Table 4. Performance evaluation metrics of eight classifiers for reduced dataset generated by proposed approach consisting of PL+ and PL- molecules based on 10-fold cross validation.

Classifiers	Sensitivity	Specificity	Accuracy	AUC	MCC
Navie Bayes	70.3	84.4	76.8	0.831	0.548
SMO	70.3	87.5	78.3	0.789	0.581
IBK	75.7	84.4	79.7	0.789	0.599
RARF	78.4	81.3	79.7	0.893	0.595
PART	56.8	81.3	68.1	0.700	0.388
JRip	78.4	65.6	72.5	0.733	0.445
RandomForest	75.7	81.3	78.3	0.868	0.568
J48	70.3	71.9	71.0	0.735	0.420

Table 5. Performance evaluation metrics of eight classifiers for original dataset consisting of PL+ and PL- molecules based on percentage split of 66:34.

Classifiers	Sensitivity	Specificity	Accuracy	AUC	MCC
Navie Bayes	86.5	71.9	79.7	0.851	0.593
SMO	73.0	81.3	76.8	0.771	0.541
IBK	81.1	84.4	82.6	0.834	0.653
RARF	91.9	87.5	89.9	0.903	0.796
PART	78.4	84.4	81.2	0.890	0.626
LRip	54.1	93.8	72.5	0.735	0.512
RandomForest	81.1	87.5	84.1	0.904	0.684
J48	83.8	87.5	85.5	0.842	0.711

Table 6. Performance evaluation metrics of eight classifiers for reduced dataset generated by proposed approach consisting of PL+ and PL- molecules based on percentage split of 66:34.

Classifiers	Sensitivity	Specificity	Accuracy	AUC	MCC
RF(h2o)	86.7	93.0	90.1	0.922	0.808
Nath et. al[87]	86.2	90.1	88.2	0.896	0.725

Table 7. Perfomance evaluation metrics for the RF algorithm with previous method.
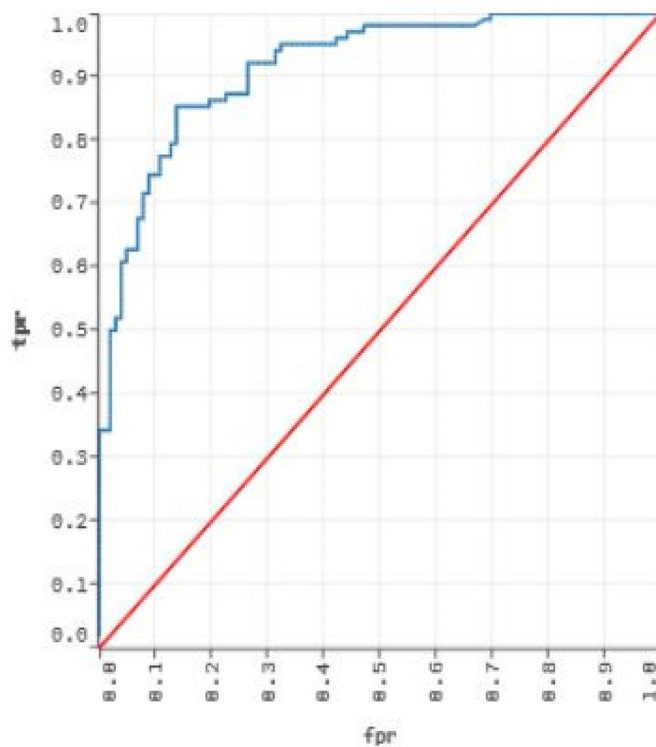
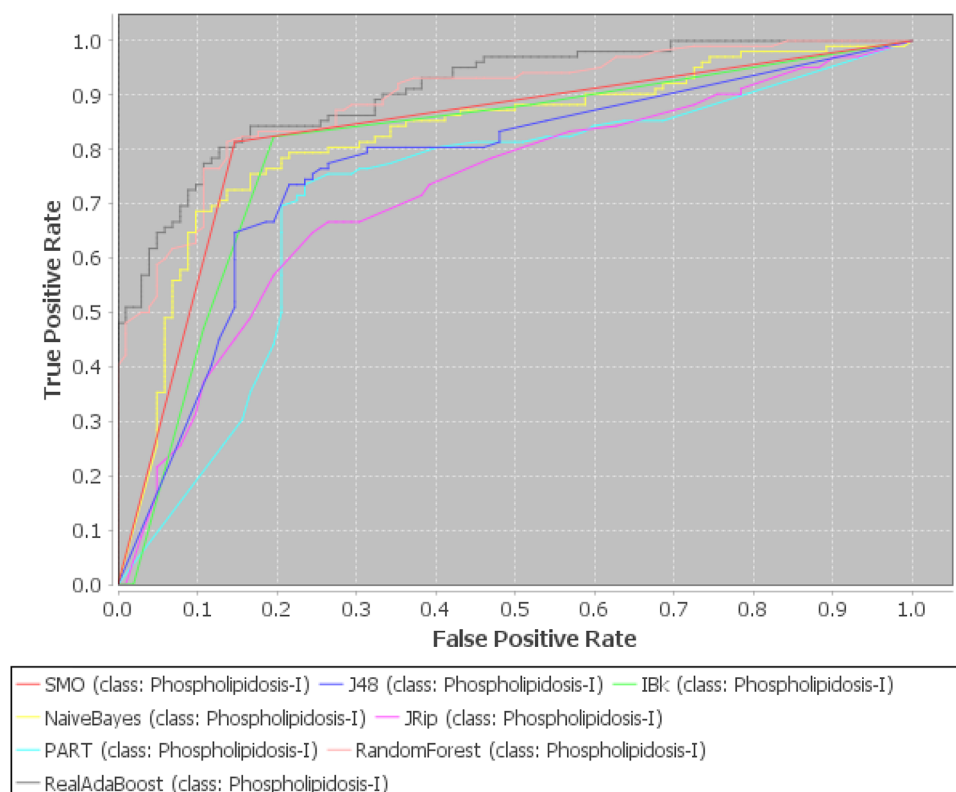**Figure 4.** ROC for the RF algorithm on phospholipidosis dataset.

Figure 5. ROC curve for orginal dataset for various machine learing algorithms.
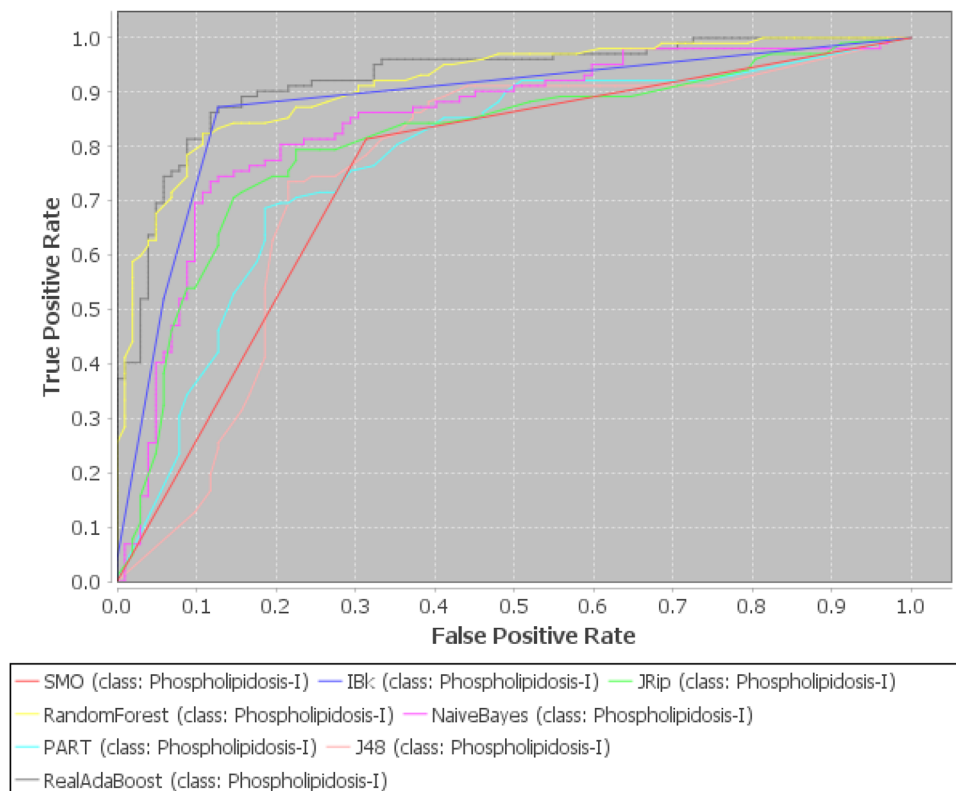
Figure 6. ROC curve for reduced dataset by various machine learing algorithms.

- By using the significance notion a heuristic IF rough feature selection algorithm is represented. Then, we apply this heuristic algorithm on ten benchmark datasets to illustrate extensive experiments.
- Finally, proposed method is successfully employed to enhance the prediction performance for identifying PL+ and PL- molecules.

For dbworld-bodies dataset, our method has eliminated 99.83% features. Moreover, performance measures of learning algorithms were evaluated based on the reduced data produced by four existing and our proposed methods, where results clearly indicate superiority of the proposed technique. For thyroid- hypothyroid dataset, RARF has reported an accuracy of 99.11% and standard deviation of 0.46% for IFRFSMI based reduced dataset. For the discrimination of PL+ and PL- molecules, the best sensitivity is achieved based on 66:34 validation technique with 91.9%. The best overall result was obtained by RF(h2o) with sensitivity, specificity, accuracy, AUC, and MCC of 86.7%, 93.0%, 90.1%, 0.922, and 0.808 respectively.

The advantages of our proposed methodology can be outlined as bellow:

- This study presents a new hybrid similarity relation that can handle mixed data in intuitionistic fuzzy framework.
- Adaptive radius is computed in the recursive way from relation itself, which ensures the information loss.
- IF granular structure is implemented to deal with noise in mixed data as it is based on our proposed hybrid relation.
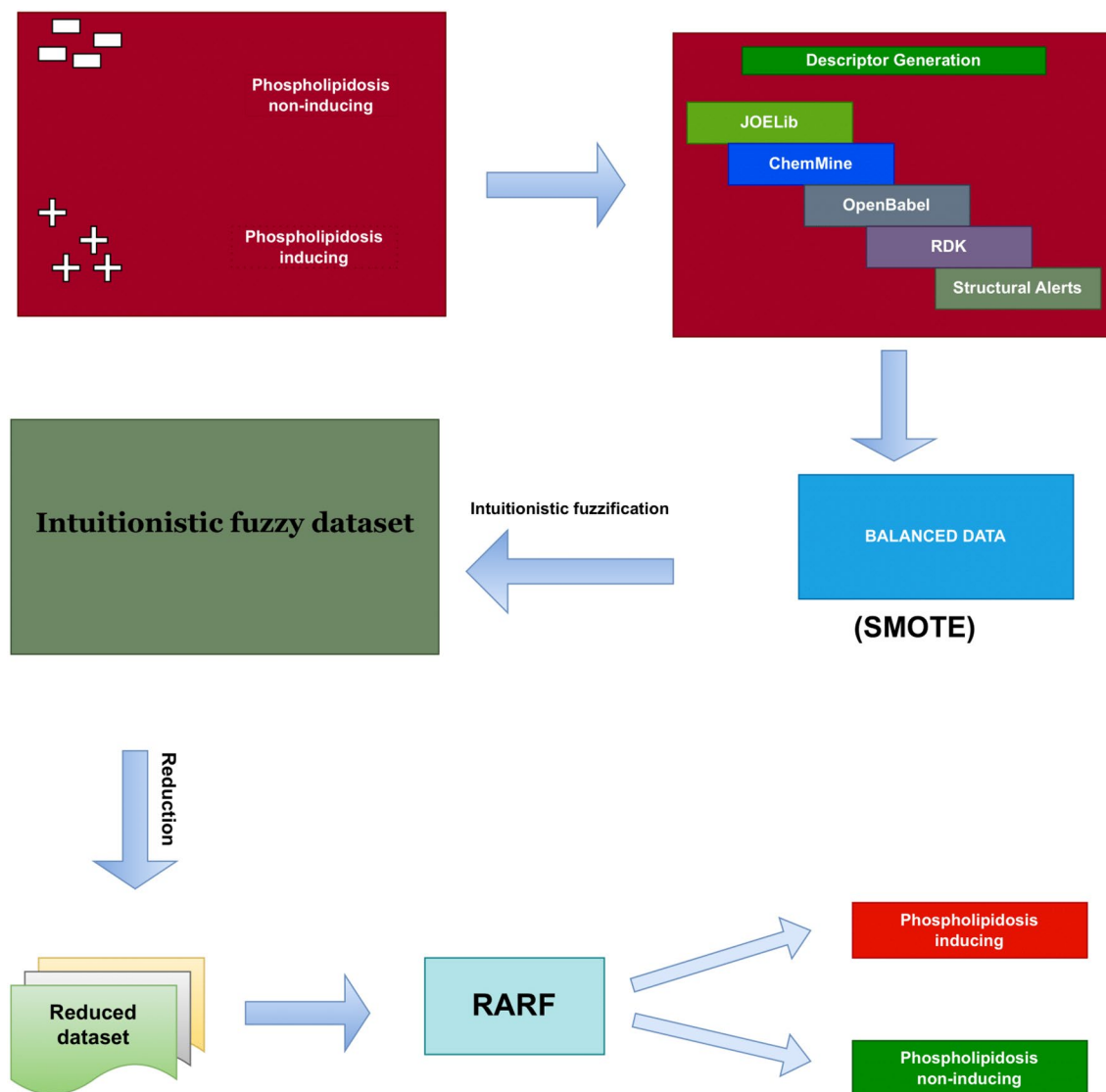
Figure 7. Schematic representation for generating classifier for phospholipidosis.

- IF rough mutual information is implemented to cope with noise and later uncertainty based on the proposed IF granular structure.
- This study presents a new methodology to discriminate PL+ and PL- molecules in an efficient and efficacious way.

In future, the proposed hybrid similarity relation can be improved by providing a more effective definition of adaptive radius. Further, inner and outer significance can be computed by assembling mutual information in robust IF rough framework to establish efficient approach to calculate the correlation between feature subset and class.

Abbreviations/symbols/signs	Explanation
IFS	Intuitionistic Fuzzy Set
FRS	Fuzzy rough set
DIFDT	Dominant intuitionistic fuzzy decision table
IFRS	Intuitionistic fuzzy rough set
IFMGRS	Intuitionistic fuzzy multigranulation rough set
MI	Mutual information
PL+	Phsopholipidosis positive
PL-	Phsopholipidosis negative
IFIS	Intuitionistic fuzzy information system
RARF	RealAdaBoost random forest
TRP	True positive
TRN	True negative
FLP	False positive
FLN	False negative
Sn	Sensitivity
Sp	Specificity
Ac	Accuracy
AUC	Area under curve
MCC	Mathews correlation coefficient
ROC	Receiver operating characteristic
SMO	Sequential minimal optimization
IBK	Instance based learner
FSFrMI	Feature selection based on fuzzy rough mutual information
GIFRFS	Granular structure based intuitionistic fuzzy rough feature selection
TIFRFS	Tolrence based fuzzy rough feature selection
FRFS	Fuzzy rough feature selection
IFRFSMI	Intuitionistic fuzzy rough feature selection based on mutual information
μ	Membership grade
ν	Non-membership grade
ϕ	Hesitancy grade
R_a^h	Hybrid similarity relation
ζ_a	Adaptive intuitionistic fuzzy radius
ET	Entropy
I	Mutual information
$\sum$	Summation
$\cup$	Union
$\cap$	Intersection
ϵ	Epsilon
$\forall$	Forall
$\in$	Belong
Ω	Significance

Table 8. The list of Abbreviations,Symbols, and Signs.

Data availability
The data supporting this study's findings are available from the corresponding author (Mohd Asif Shah) upon reasonable request.

References
1. Issad, H. A., Aoudjit, R. & Rodrigues, J. J. A comprehensive review of data mining techniques in smart agriculture. *Eng. Agric. Environ. Food* **12**(4), 511–525 (2019).
2. Li, J. *et al.* Feature selection: A data perspective. *ACM Comput. Surv. (CSUR)* **50**(6), 1–45 (2017).
3. Papakyriakou, D. & Barbounakis, I. S. Data mining methods: A review. *Int. J. Comput. Appl.* **183**(48), 5–19 (2022).
4. Awais, M. & Salahuddin, T. Radiative magnetodydrodynamic cross fluid thermophysical model passing on parabola surface with activation energy. *Ain Shams Eng. J.* **15**(1), 102282 (2024).

5. Awais, M. & Salahuddin, T. Variable thermophysical properties of magnetohydrodynamic cross fluid model with effect of energy dissipation and chemical reaction. *Int. J. Mod. Phys. B*, 2450197 (2023).
6. Jensen, R. & Shen, Q. Semantics-preserving dimensionality reduction: Rough and fuzzy-rough-based approaches. *IEEE Trans. Knowl. Data Eng.* **16**(12), 1457–1471 (2004).
7. Awais, M., Salahuddin, T. & Muhammad, S. Effects of viscous dissipation and activation energy for the MHD Eyring-powell fluid flow with Darcy-Forchheimer and variable fluid properties. *Ain Shams Eng. J.* **15**(2), 102422 (2024).
8. Chauhan, D. & Mathews, R. Review on dimensionality reduction techniques. In *Proceeding of the International Conference on Computer Networks, Big Data and IoT (ICCBI-2019)* 356–362 (Springer International Publishing, 2020).
9. Hu, J. *et al.* Orthogonal learning covariance matrix for defects of grey wolf optimizer: Insights, balance, diversity, and feature selection. *Knowl.-Based Syst.* **213**, 106684 (2021).
10. Jia, W., Sun, M., Lian, J. & Hou, S. Feature dimensionality reduction: A review. *Complex Intell. Syst.* **8**(3), 2663–2693 (2022).
11. Tubishat, M., Idris, N., Shuib, L., Abushariah, M. A. & Mirjalili, S. Improved Salp Swarm Algorithm based on opposition based learning and novel local search algorithm for feature selection. *Expert Syst. Appl.* **145**, 113122 (2020).
12. Chandrashekar, G. & Sahin, F. A survey on feature selection methods. *Comput. Electr. Eng.* **40**(1), 16–28 (2014).
13. Remeseiro, B. & Bolon-Canedo, V. A review of feature selection methods in medical applications. *Comput. Biol. Med.* **112**, 103375 (2019).
14. Saeys, Y., Inza, I. & Larranaga, P. A review of feature selection techniques in bioinformatics. *Bioinformatics* **23**(19), 2507–2517 (2007).
15. Bommert, A., Sun, X., Bischl, B., Rahnenführer, J. & Lang, M. Benchmark for filter methods for feature selection in high-dimensional classification data. *Comput. Stat. Data Anal.* **143**, 106839 (2020).
16. Cai, J., Luo, J., Wang, S. & Yang, S. Feature selection in machine learning: A new perspective. *Neurocomputing* **300**, 70–79 (2018).
17. Dash, M. & Liu, H. Feature selection for classification. *Intell. Data Anal.* **1**(1–4), 131–156 (1997).
18. Pawlak, Z. Rough sets. *Int. J. Comput. Inf. Sci.* **11**, 341–356 (1982).
19. Pawlak, Z., Grzymala-Busse, J., Slowinski, R. & Ziarko, W. Rough sets. *Commun. ACM* **38**(11), 88–95 (1995).
20. Sivasankar, E., Selvi, C. & Mahalakshmi, S. Rough set-based feature selection for credit risk prediction using weight-adjusted boosting ensemble method. *Soft. Comput.* **24**(6), 3975–3988 (2020).
21. Bania, R. K. & Halder, A. R-HEFS: Rough set based heterogeneous ensemble feature selection method for medical data classification. *Artif. Intell. Med.* **114**, 102049 (2021).
22. Thangavel, K. & Pethalakshmi, A. Dimensionality reduction based on rough set theory: A review. *Appl. Soft Comput.* **9**(1), 1–12 (2009).
23. Campagner, A., Ciucci, D. & Hüllermeier, E. Rough set-based feature selection for weakly labeled data. *Int. J. Approx. Reason.* **136**, 150–167 (2021).
24. Jensen, R. Rough set-based feature selection: A review. In *Rough Computing: Theories, Technologies and Applications* 70–107 (2008).
25. Raza, M. S. & Qamar, U. *Understanding and Using Rough Set Based Feature Selection: Concepts, Techniques and Applications* (Springer, 2017).
26. Zadeh, L. A. Fuzzy sets. *Inf. Control* **8**(3), 338–353 (1965).
27. Dubois, D. & Prade, H. Putting rough sets and fuzzy sets together. In *Intelligent Decision Support: Handbook of Applications and Advances of the Rough Sets Theory* (ed. Slowinski, R.) 203–232 (Springer, 1992).
28. Chen, J., Mi, J. & Lin, Y. A graph approach for fuzzy-rough feature selection. *Fuzzy Sets Syst.* **391**, 96–116 (2020).
29. Qiu, Z. & Zhao, H. A fuzzy rough set approach to hierarchical feature selection based on Hausdorff distance. *Appl. Intell.* **52**(10), 11089–11102 (2022).
30. Sang, B., Yang, L., Chen, H., Xu, W. & Zhang, X. Fuzzy rough feature selection using a robust non-linear vague quantifier for ordinal classification. *Expert Syst. Appl.* **230**, 120480 (2023).
31. Yin, T., Chen, H., Li, T., Yuan, Z. & Luo, C. Robust feature selection using label enhancement and β-precision fuzzy rough sets for multilabel fuzzy decision system. *Fuzzy Sets Syst.* **461**, 108462 (2023).
32. Wang, C., Huang, Y., Ding, W. & Cao, Z. Attribute reduction with fuzzy rough self-information measures. *Inf. Sci.* **549**, 68–86 (2021).
33. Zhang, X., Mei, C., Chen, D. & Yang, Y. A fuzzy rough set-based feature selection method using representative instances. *Knowl.-Based Syst.* **151**, 216–229 (2018).
34. Wang, C., Huang, Y., Shao, M. & Fan, X. Fuzzy rough set-based attribute reduction using distance measures. *Knowl.-Based Syst.* **164**, 205–212 (2019).
35. Wang, C., Wang, Y., Shao, M., Qian, Y. & Chen, D. Fuzzy rough attribute reduction for categorical data. *IEEE Trans. Fuzzy Syst.* **28**(5), 818–830 (2019).
36. Yang, X., Chen, H., Li, T. & Luo, C. A noise-aware fuzzy rough set approach for feature selection. *Knowl.-Based Syst.* **250**, 109092 (2022).
37. Yang, X., Chen, H., Li, T., Zhang, P. & Luo, C. Student-t kernelized fuzzy rough set model with fuzzy divergence for feature selection. *Inf. Sci.* **610**, 52–72 (2022).
38. Yuan, Z. *et al.* Attribute reduction methods in fuzzy rough set theory: An overview, comparative experiments, and new directions. *Appl. Soft Comput.* **107**, 107353 (2021).
39. Jain, P., Tiwari, A. K. & Som, T. A fitting model based intuitionistic fuzzy rough feature selection. *Eng. Appl. Artif. Intell.* **89**, 103421 (2020).
40. Annamalai, C. Intuitionistic fuzzy sets: New approach and applications (2022).
41. Dan, S. *et al.* Intuitionistic type-2 fuzzy set and its properties. *Symmetry* **11**(6), 808 (2019).
42. Atanassov, K. T. & Stoeva, S. Intuitionistic fuzzy sets. *Fuzzy Sets Syst.* **20**(1), 87–96 (1986).
43. Cornelis, C., De Cock, M. & Kerre, E. E. Intuitionistic fuzzy rough sets: At the crossroads of imperfect knowledge. *Expert Syst.* **20**(5), 260–270 (2003).
44. Zhan, J., Masood Malik, H. & Akram, M. Novel decision-making algorithms based on intuitionistic fuzzy rough environment. *Int. J. Mach. Learn. Cybern.* **10**, 1459–1485 (2019).
45. Zhang, Z. Attributes reduction based on intuitionistic fuzzy rough sets. *J. Intell. Fuzzy Syst.* **30**(2), 1127–1137 (2016).
46. Atanassov, K. T. & Atanassov, K. T. *Intuitionistic Fuzzy Sets* (Springer, 1999).
47. Tseng, T.-L.B. & Huang, C.-C. Rough set-based approach to feature selection in customer relationship management. *Omega* **35**(4), 365–383 (2007).
48. Zhang, X., Zhou, B. & Li, P. A general frame for intuitionistic fuzzy rough sets. *Inf. Sci.* **216**, 34–49 (2012).
49. Zhou, L. & Wu, W.-Z. On generalized intuitionistic fuzzy rough approximation operators. *Inf. Sci.* **178**(11), 2448–2465 (2008).
50. Jain, P. & Som, T. Multigranular rough set model based on robust intuitionistic fuzzy covering with application to feature selection. *Int. J. Approx. Reason.* **156**, 16–37 (2023).
51. Liu, Y. & Lin, Y. Intuitionistic fuzzy rough set model based on conflict distance and applications. *Appl. Soft Comput.* **31**, 266–273 (2015).
52. Huang, B., Zhuang, Y.-L., Li, H.-X. & Wei, D.-K. A dominance intuitionistic fuzzy-rough set approach and its applications. *Appl. Math. Model.* **37**(12–13), 7128–7141 (2013).

53. Wang, C., Huang, Y., Shao, M., Hu, Q. & Chen, D. Feature selection based on neighborhood self-information. *IEEE Trans. Cybern.* **50**(9), 4031–4042 (2019).

54. Xu, J., Shen, K. & Sun, L. Multi-label feature selection based on fuzzy neighborhood rough sets. *Complex Intell. Syst.* **8**(3), 2105–2129 (2022).

55. Huang, B., Li, H., Feng, G. & Zhou, X. Dominance-based rough sets in multi-scale intuitionistic fuzzy decision tables. *Appl. Math. Comput.* **348**, 487–512 (2019).

56. Huang, B., Guo, C.-X., Zhuang, Y.-L., Li, H.-X. & Zhou, X.-Z. Intuitionistic fuzzy multigranulation rough sets. *Inf. Sci.* **277**, 299–320 (2014).

57. Tan, A. *et al.* Intuitionistic fuzzy rough set-based granular structures and attribute subset selection. *IEEE Trans. Fuzzy Syst.* **27**(3), 527–539 (2018).

58. Zhou, L., Wu, W.-Z. & Zhang, W.-X. On characterization of intuitionistic fuzzy rough sets based on intuitionistic fuzzy implicators. *Inf. Sci.* **179**(7), 883–898 (2009).

59. Tiwari, A. K., Shreevastava, S., Som, T. & Shukla, K. K. Tolerance-based intuitionistic fuzzy-rough set approach for attribute reduction. *Expert Syst. Appl.* **101**, 205–212 (2018).

60. Shreevastava, S., Tiwari, A. & Som, T. Feature subset selection of semi-supervised data: An intuitionistic fuzzy-rough set-based concept. In *Proceedings of International Ethical Hacking Conference 2018: eHaCON 2018, Kolkata, India* (2019).

61. Tiwari, A. K., Shreevastava, S., Subbiah, K. & Som, T. An intuitionistic fuzzy-rough set model and its application to feature selection. *J. Intell. Fuzzy Syst.* **36**(5), 4969–4979 (2019).

62. Tiwari, A. K., Shreevastava, S., Shukla, K. K. & Subbiah, K. New approaches to intuitionistic fuzzy-rough attribute reduction. *J. Intell. Fuzzy Syst.* **34**(5), 3385–3394 (2018).

63. Tiwari, A. K., Shreevastava, S., Subbiah, K. & Som, T. An intuitionistic fuzzy-rough set model and its application to feature selection. *J. Intell. Fuzzy Syst.* **36**(5), 4969–4979 (2019).

64. Shreevastava, S., Singh, S., Tiwari, A. & Som, T. Different classes ratio and Laplace summation operator based intuitionistic fuzzy rough attribute selection. *Iran. J. Fuzzy Syst.* **18**(6), 67–82 (2021).

65. Shreevastava, S., Tiwari, A. K. & Som, T. Intuitionistic fuzzy neighborhood rough set model for feature selection. *Int. J. Fuzzy Syst. Appl. (IJFSA)* **7**(2), 75–84 (2018).

66. Li, L. Q., Wang, X. L., Liu, Z. X. & Xie, W. X. A novel intuitionistic fuzzy clustering algorithm based on feature selection for multiple object tracking. *Int. J. Fuzzy Syst.* **21**, 1613–1628 (2019).

67. Singh, S., Shreevastava, S., Som, T. & Jain, P. Intuitionistic fuzzy quantifier and its application in feature selection. *Int. J. Fuzzy Syst.* **21**, 441–453 (2019).

68. Sun, L., Wang, L., Ding, W., Qian, Y. & Xu, J. Feature selection using fuzzy neighborhood entropy-based uncertainty measures for fuzzy neighborhood multigranulation rough sets. *IEEE Trans. Fuzzy Syst.* **29**(1), 19–33 (2020).

69. Sun, L., Zhang, X., Qian, Y., Xu, J. & Zhang, S. Feature selection using neighborhood entropy-based uncertainty measures for gene expression data classification. *Inf. Sci.* **502**, 18–41 (2019).

70. Fang, L. *et al.* Feature selection method based on mutual information and class separability for dimension reduction in multidimensional time series for clinical data. *Biomed. Signal Process. Control* **21**, 82–89 (2015).

71. Fernandes, A. D. & Gloor, G. B. Mutual information is critically dependent on prior assumptions: Would the correct estimate of mutual information please identify itself?. *Bioinformatics* **26**(9), 1135–1139 (2010).

72. Wang, Z. *et al.* Exploiting fuzzy rough mutual information for feature selection. *Appl. Soft Comput.* **131**, 109769 (2022).

73. Xie, L., Lin, G., Li, J. & Lin, Y. A novel fuzzy-rough attribute reduction approach via local information entropy. *Fuzzy Sets Syst.* **473**, 108733 (2023).

74. Xu, F., Miao, D. & Wei, L. Fuzzy-rough attribute reduction via mutual information with an application to cancer classification. *Comput. Math. Appl.* **57**(6), 1010–1017 (2009).

75. Fang, H., Tang, P. & Si, H. Feature selections using minimal redundancy maximal relevance algorithm for human activity recognition in smart home environments. *J. Healthc. Eng.* **2020**, 1–13 (2020).

76. Xie, S. *et al.* A new improved maximal relevance and minimal redundancy method based on feature subset. *J. Supercomput.* **79**(3), 3157–3180 (2023).

77. Maji, P. & Garai, P. On fuzzy-rough attribute selection: Criteria of max-dependency, max-relevance, min-redundancy, and max-significance. *Appl. Soft Comput.* **13**(9), 3968–3980 (2013).

78. Zhang, X., Mei, C., Chen, D. & Li, J. Feature selection in mixed data: A method using a novel fuzzy rough set-based information entropy. *Pattern Recogn.* **56**, 1–15 (2016).

79. Zhang, X., Mei, C., Chen, D., Yang, Y. & Li, J. Active incremental feature selection using a fuzzy-rough-set-based information entropy. *IEEE Trans. Fuzzy Syst.* **28**(5), 901–915 (2019).

80. Anderson, N. & Borlak, J. Drug-induced phospholipidosis. *FEBS Lett.* **580**(23), 5533–5540 (2006).

81. Breiden, B. & Sandhoff, K. Emerging mechanisms of drug-induced phospholipidosis. *Biol. Chem.* **401**(1), 31–46 (2020).

82. Shayman, J. A. & Abe, A. Drug induced phospholipidosis: An acquired lysosomal storage disorder. *Biochim. Biophys. Acta (BBA)-Mol. Cell Biol. Lipids* **1831**(3), 602–611 (2013).

83. Salahuddin, T. *Numerical Techniques in MATLAB: Fundamental to Advanced Concepts* (CRC Press, 2023).

84. Frank, E., Hall, M., Trigg, L., Holmes, G. & Witten, I. H. Data mining in bioinformatics using Weka. *Bioinformatics* **20**(15), 2479–2481 (2004).

85. Asuncion, A. & Newman, D. UCI machine learning repository. In: Irvine, CA, USA (2007).

86. Hall, M. *et al.* The WEKA data mining software: An update. *ACM SIGKDD Explor. Newsl.* **11**(1), 10–18 (2009).

87. Nath, A. & Sahu, G. K. Exploiting ensemble learning to improve prediction of phospholipidosis inducing potential. *J. Theor. Biol.* **479**, 37–47 (2019).

88. Friedman, M. A comparison of alternative tests of significance for the problem of m rankings. *Ann. Math. Stat.* **11**(1), 86–92 (1940).

89. Dunn, O. J. Multiple comparisons among means. *J. Am. Stat. Assoc.* **56**(293), 52–64 (1961).

Author contributions

Funding

nature portfolio

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to R.S. or M.A.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Terms and Conditions

Springer Nature journal content, brought to you courtesy of Springer Nature Customer Service Center GmbH ("Springer Nature").

Springer Nature supports a reasonable amount of sharing of research papers by authors, subscribers and authorised users ("Users"), for small-scale personal, non-commercial use provided that all copyright, trade and service marks and other proprietary notices are maintained. By accessing, sharing, receiving or otherwise using the Springer Nature journal content you agree to these terms of use ("Terms"). For these purposes, Springer Nature considers academic use (by researchers and students) to be non-commercial.

These Terms are supplementary and will apply in addition to any applicable website terms and conditions, a relevant site licence or a personal subscription. These Terms will prevail over any conflict or ambiguity with regards to the relevant terms, a site licence or a personal subscription (to the extent of the conflict or ambiguity only). For Creative Commons-licensed articles, the terms of the Creative Commons license used will apply.

We collect and use personal data to provide access to the Springer Nature journal content. We may also use these personal data internally within ResearchGate and Springer Nature and as agreed share it, in an anonymised way, for purposes of tracking, analysis and reporting. We will not otherwise disclose your personal data outside the ResearchGate or the Springer Nature group of companies unless we have your permission as detailed in the Privacy Policy.

While Users may use the Springer Nature journal content for small scale, personal non-commercial use, it is important to note that Users may not:

1. use such content for the purpose of providing other users with access on a regular or large scale basis or as a means to circumvent access control;

2. use such content where to do so would be considered a criminal or statutory offence in any jurisdiction, or gives rise to civil liability, or is otherwise unlawful;

3. falsely or misleadingly imply or suggest endorsement, approval , sponsorship, or association unless explicitly agreed to by Springer Nature in writing;

4. use bots or other automated methods to access the content or redirect messages

5. override any security feature or exclusionary protocol; or

6. share the content in order to create substitute for Springer Nature products or services or a systematic database of Springer Nature journal content.

In line with the restriction against commercial use, Springer Nature does not permit the creation of a product or service that creates revenue, royalties, rent or income from our content or its inclusion as part of a paid for service or for other commercial gain. Springer Nature journal content cannot be used for inter-library loans and librarians may not upload Springer Nature journal content on a large scale into their, or any other, institutional repository.

These terms of use are reviewed regularly and may be amended at any time. Springer Nature is not obligated to publish any information or content on this website and may remove it or features or functionality at our sole discretion, at any time with or without notice. Springer Nature may revoke this licence to you at any time and remove access to any copies of the Springer Nature journal content which have been saved.

To the fullest extent permitted by law, Springer Nature makes no warranties, representations or guarantees to Users, either express or implied with respect to the Springer nature journal content and all parties disclaim and waive any implied warranties or warranties imposed by law, including merchantability or fitness for any particular purpose.

Please note that these rights do not automatically extend to content, data or other material published by Springer Nature that may be licensed from third parties.

If you would like to use or distribute our Springer Nature journal content to a wider audience or on a regular basis or in any other manner not expressly permitted by these Terms, please contact Springer Nature at

onlineservice@springernature.com