



OPEN **Covering assisted intuitionistic fuzzy bi-selection technique for data reduction and its applications**

Rajat Saini¹, Anoop Kumar Tiwari²✉, Abhigyan Nath³, Phool Singh⁴, S. P. Maurya⁵ & Mohd Asif Shah^{6,7,8}✉

The dimension and size of data is growing rapidly with the extensive applications of computer science and lab based engineering in daily life. Due to availability of vagueness, later uncertainty, redundancy, irrelevancy, and noise, which imposes concerns in building effective learning models. Fuzzy rough set and its extensions have been applied to deal with these issues by various data reduction approaches. However, construction of a model that can cope with all these issues simultaneously is always a challenging task. None of the studies till date has addressed all these issues simultaneously. This paper investigates a method based on the notions of intuitionistic fuzzy (IF) and rough sets to avoid these obstacles simultaneously by putting forward an interesting data reduction technique. To accomplish this task, firstly, a novel IF similarity relation is addressed. Secondly, we establish an IF rough set model on the basis of this similarity relation. Thirdly, an IF granular structure is presented by using the established similarity relation and the lower approximation. Next, the mathematical theorems are used to validate the proposed notions. Then, the importance-degree of the IF granules is employed for redundant size elimination. Further, significance-degree-preserved dimensionality reduction is discussed. Hence, simultaneous instance and feature selection for large volume of high-dimensional datasets can be performed to eliminate redundancy and irrelevancy in both dimension and size, where vagueness and later uncertainty are handled with rough and IF sets respectively, whilst noise is tackled with IF granular structure. Thereafter, a comprehensive experiment is carried out over the benchmark datasets to demonstrate the effectiveness of simultaneous feature and data point selection methods. Finally, our proposed methodology aided framework is discussed to enhance the regression performance for IC50 of Antiviral Peptides.

Keywords Rough set, Granular structure, Dimensionality reduction, IF set, Instance selection

The proliferation of technological breakthroughs, along with the decreasing cost of equipment, has led to a significant rise in the collection of data, both in terms of volume and variety. Consequently, this has resulted in the emergence of multiple data resources. The concept of big data encompasses a vast collection of data that is characterized by its immense volume and intricate nature. The significant increase in data has led to its widespread access across various platforms, including both free and corporate resources. The current information-centric landscape necessitates the procurement, amalgamation, and examination of large-scale data sets in order to unravel intricate medical and scientific challenges.

Reducing data is a crucial step in the preprocessing of data. The primary aim of data reduction is to decrease the size or complexity of the original dataset by retaining just the most essential and representative information.

¹Department of Mathematics, School of Basic Sciences, Central University of Haryana, Mahendergarh 123031, India. ²Department of Computer Science and Information Technology, Central University of Haryana, Mahendergarh 123031, India. ³Department of Biochemistry, Pt. Jawahar Lal Nehru Memorial Medical College, Raipur 492001, India. ⁴Department of Mathematics (SoET), Central University of Haryana, Mahendergarh 123031, India. ⁵Department of Geophysics, Institute of Science, Banaras Hindu University, Varanasi 221005, India. ⁶Department of Economics, Kebri Dehar University, 250 Kebri Dehar, Somali, Ethiopia. ⁷Division of Research and Development, Lovely Professional University, Phagwara, Punjab 144001, India. ⁸Department of Economics, Kardan University, Parwan e Du, Kabul 1001, Afghanistan. ✉email: dranoop@cuh.ac.in; ohaasif@gmail.com

The clear benefits of data minimization revolve around mitigating time complexity and preventing excessive storage requirements. Conversely, the utilization of simplified models can be facilitated through the utilization of decreased data, hence potentially enhancing the quality of the analysis outcomes. Commonly used data reduction techniques in the field include instance selection and feature selection¹.

These techniques can be categorized into three parts namely filter method, wrapper method, and embedded method. In filter techniques, the entire process of data reduction remains independent from different machine learning algorithms as it does not incorporate any feedback from these algorithms. In wrapper techniques, the response from the machine algorithms is incorporated to evaluate the quality of the reduced data, and hence the overall performances of these algorithms can be improved for the specific domains. Embedded techniques are used to avoid the limitations of both filter and wrapper methods. Data reduction is incorporated as the part of the machine learning algorithms in embedded techniques, hence these techniques can observe advantages of both wrapper and filter methods.

Instance selection² is a crucial strategy in the field of data reduction. It involves the selection of a subset of data that effectively fulfills the original objectives of a data mining application³, as if the entire dataset were utilized. Through the process of instance selection⁴, it is often possible to obtain representative instances by eliminating redundant instances⁵, irrelevant instances, or errors from the original data. This approach facilitates the extraction of crucial information and enables the acquisition of high-quality results with reduced computational time, particularly when dealing with large-scale datasets. The topic of instance selection⁵ has been extensively examined across various application domains. Feature selection⁶ is a significant method for reducing data, which has been employed to enhance learning performance and reduce dimensionality⁷ by eliminating repetitive or irrelevant features.

The solution to this type of problem has been proposed by many earlier studies, although the majority of them simply conducted instance selection or feature selection⁸. Few of them simultaneously thought about choosing both instances and characteristics⁹. Additionally, the majority of earlier studies did not take into account massive data and noisy data. In order to simultaneously select features and instances in big data¹⁰, numerous prior researches have put forth potential solutions to this particular challenge, but with a predominant focus on instance selection or feature selection methodologies¹¹. A subset of individuals contemplated the selection of both examples and qualities in tandem. Furthermore, a significant proportion of previous research failed to consider the implications of extensive and unreliable data to effectively address the challenge of selecting both features and instances in massive data¹². In following studies, researchers have employed feature selection techniques¹³ based on genetic algorithms (GA) to address a range of difficulties. These include enhancing the accuracy of molecular illness identification¹⁴, providing valuable assistance to neurosurgeons during surgical procedures, among others. Particle swarm optimization (PSO) has also been employed in the feature selection process.

The work conducted by the authors in reference¹⁵ involved the modification of the continuous Particle Swarm Optimization (PSO) algorithm to a binary PSO algorithm for the purpose of feature selection¹³. In order to enhance the accuracy of classification, the research conducted in reference¹⁶ employed a combination of Particle Swarm Optimization (PSO) and Linear Discriminant Analysis (LDA)¹⁶. In order to mitigate the problem of premature convergence in feature selection, a Particle Swarm Optimization (PSO) algorithm incorporating three distinct methods was introduced in a previous study.

Moreover, the utilization of genetic algorithms (GA)¹⁷ has been widely observed in the context of instance selection¹⁸. Ahmad et al.¹⁹ employed the cooperative Particle Swarm Optimization (PSO) algorithm to choose instances in their study. Here, the enhanced immune binary particle swarm optimization (PSO) algorithm as a method for classifying time-series data was presented.

While there have been previous studies that focused on either feature selection²⁰ or instance selection, only a limited number of publications, such as GA²¹, have concurrently examined tackling both feature selection²² and instance selection. Furthermore, a significant proportion of previous algorithms did not incorporate multiple neighborhood search methods concomitantly in order to enhance the efficiency of local solution discovery and increase the likelihood of algorithm convergence.

However, these conventional approaches usually faces the loss of information due to use of expert suggestions. Further, rough set theory²³ was incorporated to perform simultaneous feature and instance selections^{24–26} that minimize this information loss²⁷. In the rough set theory^{28,29}, entire bireduct process is performed based on the information from the dataset itself³⁰. Ma et al.³¹ presented a min-max feature-sample bireduct method (RBR), where unification of reduct model in rough set theory³² was discussed. Bi-clustering was mainly employed notion for demonstrating bireducts in rough set framework. This concept was consisted of two facets: feature subsets that are consistent with decision class and sample subset where summarisation is found to be consistent. But, discretization of data in rough set theory aided techniques³³ again leads to information loss. This issue was further handled by fuzzy rough set theory (FRS)^{34,35} by the amalgamation of fuzzy set³⁵ and rough set. FRS model is defined as an effective generalization of classical rough set model. This model is comprised of the advantages of both rough as well as fuzzy sets, which has been broadly and mainly applied to cope with the data reduction of real-valued datasets^{36,37}. Jensen et al.³⁸ initially employed this model to propose the idea of fuzzy rough aided dependency functions and presented an effective way to compute the reduct³⁹. Chen et al. established the concept of fuzzy assisted discernibility matrix based on lower approximations and used it to compute the feature subset of a decision system⁴⁰. Bhatt et al. proposed a squeezed computing domain to perform dimensionality reduction, which enhanced the computing efficiency in an effective way⁴¹. Mieszkowicz-rolka discussed a variable precision model in fuzzy rough framework to eradicate noise in the high-dimensional datasets⁴². Dai et al. introduced a dimensionality reduction technique with FRS by using a maximal discernibility pair concept⁴³. Misclassification and disruption due to noise in data was handled with variable precision model by Zhao et al.⁴⁴. In the recent years, dependency function idea in FRS framework was improved by formulating proportion of membership cardinality related to fuzzy positive region to the complete dataset with different distance functions and granular

structures^{34,45}. FRS model was further extended based on information from hypergraph idea to improve feature selection process in terms of time and space complexity⁴⁶. Multi-label feature selection was presented in the recent days based on multi-correlation, robust multi-neighborhood, parameterized covering space, and mutual information^{37,43}. Moreover, data reduction was also discussed by eliminating redundant instances. Jensen et al.⁴⁷ initiated this process by improving the fuzzy rough positive region concept with threshold parameters. Optimization strategies were integrated with fuzzy rough set theory⁴⁸ to develop simultaneous size and feature selection approaches^{49,50}. These optimization concepts included sophisticated search strategies⁵¹, where population were considered on the basis of natural occurrences and patterns via iterations⁵². An additive dimensionality^{53,54} and example reductions⁵⁵ was discussed by Anarki et al.⁵⁶ through fuzzy rough set theory⁵⁷ with a shuffled frog jumping technique. Further, frequency assisted heuristic method was employed to offer the simultaneous reduction⁵⁸ of attribute and instances⁵⁹ in large volume of high-dimensional data. Additionally, a harmony-search based technique by considering the threshold parameters was presented to determine the bireduct (FBR) by using the fuzzy rough sets as addressed by Mac Parthalain et al.⁶⁰. Furthermore, Zhang et al.⁶¹ outlined a bi-selection model (CFBR), where both irrelevant and/or redundant samples and dimension were eliminated based on representative instances^{47,62} along with fuzzy granular structure⁴⁵. But, fuzzy set theory⁶³ was incapable in dealing with later uncertainty i.e. uncertainty due to identification. This obstacle can be easily and effectively handled by introducing the bireduct method with set theory⁶⁴. A heuristic search aided IF rough bireduct idea was discussed by Jain et al.⁶⁵. This method was the extension of Parthalain et al.⁶⁰ fuzzy rough bireduct concept, where features were selected based on the notion of dependency function and instances were eliminated by employing positive region. However, Simultaneous instance and feature elimination methods by amalgamation of IF and rough sets^{24,66} is in its nascent stage due to scarcity of works till date. Moreover, the data point and feature selection by using IF rough sets^{67–69} has not been taken care from the perspective of the significance degrees of the IF granules⁷⁰. Further, a mining algorithm⁷¹ appears inefficient while coping with errors available in the entire datasets when compared to reduced dataset⁷². Additionally, noise elimination has not been considered efficaciously in the IF instance selection with IF granular structure⁷³. In this paper, we consider all the above discussed limitations in the earlier models. Initially, a novel IF similarity relation is established. Then, IF lower and upper approximations are described with the help of this similarity relation. Next, a novel IF granular structure is defined on the basis of this similarity relation⁷⁴. Moreover, relevant prepositions are proved to justify the proposed concepts. Further, significance degree of IF granule is employed to discard redundant size by removing the noise. Furthermore, significance-degree-preserved feature selection is elaborated. Thereafter, simultaneous sample and feature selection⁷⁵ techniques are investigated over the benchmark datasets⁷⁶ to perform a comprehensive experimental study to show the superiority of the entire method. Finally, a framework is presented based on our proposed technique to improve the regression performance for IC50 of Antiviral Peptides (AVPs). The schematic depiction of entire methodology is given by Figure 5. Now, we can outline the way of handling the stated issues as below:

- (1) This study presented IF aided ideas for simultaneous sample and dimensionality reduction methods. IF highlights both membership and non-membership grades for a particular sample, which leads to cope with uncertainty due to both judgement and identification i.e. later uncertainty.
- (2) Here, IF and rough sets were combined to deal with both later uncertainty and vagueness.
- (3) IF granular structure, covering, and proximity concepts effectively avoided noise in the information systems.
- (4) Representative instances were identified during elimination of redundancy in size and importance-degree-preserved reduct computation ensured the removal of redundancy and irrelevancy in conditional features.

Finally, we can highlight the major contributions of the entire work as follows:

Major contributions of the current study

- (1) This study has been initiated by establishing a novel IF similarity relation for a pair of IF data points.
- (2) Further, an IF rough set model is presented based on this relation by assembling IF and rough sets notions to handle uncertainty due to both judgement and identification.
- (3) Next, a parameterized IF granular structure is discussed based on the lower approximation to deal with noise.
- (4) Thereafter, covering of all the available instances is defined by using the feedback from IF granular structure.
- (5) Moreover, proximity of a particular instance relative to different samples available in the dataset is computed.
- (6) Furthermore, covering and importance degree notions are incorporated to identify representative instances.
- (7) This covering idea is further employed to perform the importance degree based instance elimination to remove redundant size.
- (8) Finally, importance-degree-preserved feature subset selection is presented to discard redundant and irrelevant dimensions.

Theoretical background

In this segment, we present subsequent discussions of basic knowledge related to IF set, IF decision system and IF relation⁷⁷.

Definition 2.1 (IF set) An IF set X in U^d , where U^d is well defined collection of data points/objects with

$$X = \left\langle x, \rho_X(x), \eta_X(x) \right\rangle, \forall x \in U^d \quad (1)$$

Moreover, ρ_X and η_X are depicted by mappings

$$\rho_X : U^d \rightarrow [0, 1] \quad (2)$$

and

$$\eta_X : U^d \rightarrow [0, 1] \quad (3)$$

Here, interesting condition holds, which is specified by:

$0 \leq \rho_X(x) + \eta_X(x) \leq 1, \forall x \in U^d$. Here, the two values $\rho_X(x)$ and $\eta_X(x)$ are narrated as membership and non-membership grades respectively, $\forall x \in U^d$. Further, hesitancy degree can be narrated by $\pi_X \forall x \in U^d$, which is persuasively computed by the widely employed expression $\pi_X(x) = 1 - \rho_X(x) - \eta_X(x)$. It can be noted here that $0 \leq \pi_X(x) \leq 1, \forall x \in U^d$. The arrayed pair $\{\rho_X, \eta_X\}$ is identified as an essential IF value.

Definition 2.2 (IF information system) An IF information system (IFIS) is portrayed by a quadruple (U^d, A, V_L, L) . Here, V_L is comprised of all IF values, and is exemplified by $L : U^d \times A \rightarrow V_L$, in such a way that $IF(x, a) = \langle \rho_X(x), \eta_X(x) \rangle, \forall x \in U^d$ and $\forall a \in A$

Definition 2.3 (IF decision system) An IFIS (U^d, A, V_L, L) is delineated as an IF decision system iff $A = C \cup D$, where C is an available non-empty finite accumulation of conditional dimensions/attributes and D is an available non-empty accumulation of decision dimension.

Definition 2.4 (IF relation) Let $R(x, y)$ be an required IF similarity relation, where it holds/satisfies following essential conditions Reflexivity:

$$\rho_R(x, x) = 1 \text{ and } \eta_R(x, x) = 0 \quad (4)$$

Symmetry:

$$\rho_R(x, y) = \rho_R(y, x) \text{ and } \eta_R(x, y) = \eta_R(y, x), \forall x, y \in U^d \quad (5)$$

Proposed work

In this segment, we incorporate the essential conditions to establish the new extensions of IF relation and IF granular structure. Further, an effective bireduct notion is discussed by presenting the instance/data point and feature/dimension selection approaches.

Definition 3.1 (IF relation) Let U^d be the assemblage of data points/tuples i.e. universe of discourse. Then, an extension IF similarity relation as R^c over U^d can be defined by the following prominent equation:

$$R^c(x_i, x_j) = \sum_{i=1}^n \sum_{j=1}^n \frac{(1 - \frac{1}{2}(|\rho_a(x_i) - \rho_a(x_j)| + |\eta_a(x_i) - \eta_a(x_j)|))}{n} \quad (6)$$

$$\forall x_i, x_j \in U^d \text{ and } \forall a \in B$$

Proof P(1). we know that value of $\rho_a(x_i)$, $\rho_a(x_j)$, $\eta_a(x_i)$ and $\eta_a(x_j)$ always lies between 0 and 1.

$$\Rightarrow 0 \leq |\rho_a(x_i) - \rho_a(x_j)| \leq 1 \text{ and } 0 \leq |\eta_a(x_i) - \eta_a(x_j)| \leq 1$$

$$\Rightarrow 0 \leq \frac{|\rho_a(x_i) - \rho_a(x_j)| + |\eta_a(x_i) - \eta_a(x_j)|}{2} \leq 1$$

$$\Rightarrow 0 \leq \frac{|\rho_a(x_i) - \rho_a(x_j)| + |\eta_a(x_i) - \eta_a(x_j)|}{2} \leq 1$$

$$\Rightarrow 0 \leq \frac{|\rho_a(x_i) - \rho_a(x_j)| + |\eta_a(x_i) - \eta_a(x_j)|}{2} \leq 1$$

$$\Rightarrow \sum_{i=1}^n \sum_{j=1}^n 0 \leq \sum_{i=1}^n \sum_{j=1}^n \frac{1 - \frac{|\rho_a(x_i) - \rho_a(x_j)| + |\eta_a(x_i) - \eta_a(x_j)|}{2}}{n} \leq \sum_{i=1}^n \sum_{j=1}^n \frac{1}{n}$$

$$\Rightarrow 0 \leq R^c(x_i, x_j) \leq 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i, x_i) = 1$$

$$\Rightarrow R^c(x_i,$$

$$\Rightarrow R^c(x_i, x_j) = \sum_{j=1}^n \sum_{n=1}^n \frac{(1-\frac{1}{2}(|-\rho_a(x_j)-\rho_a(x_i)|+|-\eta_a(x_j)-\eta_a(x_i)|))}{n}$$
$$\Rightarrow R^c(x_i, x_j) = \sum_{j=1}^n \sum_{n=1}^n \frac{(1-\frac{1}{2}(|\rho_a(x_j)-\rho_a(x_i)|+|\eta_a(x_j)-\eta_a(x_i)|))}{n}$$
$$\Rightarrow R^c(x_i, x_j) = R^c(x_j, x_i) \Rightarrow R^c(x_i, x_j) \text{ is symmetric}$$

By P(1), P(2), P(3) $R^c(x_i, x_j)$ is an established similarity relation. □

Example 3.2 An IF decision system is chosen, which represents judgment problem. In this dataset, the object set $X=\{x_1, x_2, x_3, x_4, x_5\}$ comprises of 5 objects. Five conditional attributes $A= \{a_1, a_2, a_3, a_4, a_5\}$ exist in conditional attribute set and decision attribute d with two decision class $\{D_1, D_2\}$. Instances in decision class $D_1 = \{x_1, x_3, x_4\}$ and $D_2 = \{x_2, x_5\}$ relation R^c lies between 0 and 1 for dataset (Table 1) as shown in Table 2.

Definition 3.3 (IF rough set) Let U^d be the assemblage of samples/instances and R^c be a presented IF similarity relation (6). Now, for any given $X \subseteq U^d$, we can provide the lower and upper approximations of X relative to R^c , and for any $B \subseteq A$ are described respectively by the below equation:

$$\underline{R}_B^c(X) = \left\{ \frac{\left(\rho_{\underline{R}_B^c(X)}^c(x_i), \eta_{\underline{R}_B^c(X)}^c(x_i) \right)}{x_i} \middle| x_i \in U^d \right\}$$
(7)

$$\overline{R}_B^c(X) = \left\{ \frac{\left(\rho_{\overline{R}_B^c(X)}^c(x_i), \eta_{\overline{R}_B^c(X)}^c(x_i) \right)}{x_i} \middle| x_i \in U^d \right\}$$
(8)

, here, we have

$$\rho_{\underline{R}_B^c(X)}^c(x_i) = \inf_{x_j \in U^d} \max \left(\eta_{R^c}(x_i, x_j), \rho_x(x_j) \right)$$
(9)

Objects	Attributes					
	a_1	a_2	a_3	a_4	a_5	d
x_1	$\langle 0.2, 0.4 \rangle$	$\langle 0.1, 0.7 \rangle$	$\langle 0.2, 0.6 \rangle$	$\langle 0.6, 0.4 \rangle$	$\langle 0.2, 0.8 \rangle$	1
x_2	$\langle 0.1, 0.7 \rangle$	$\langle 0.1, 0.8 \rangle$	$\langle 0.3, 0.6 \rangle$	$\langle 0.5, 0.2 \rangle$	$\langle 0.2, 0.7 \rangle$	2
x_3	$\langle 0.1, 0.8 \rangle$	$\langle 0.1, 0.8 \rangle$	$\langle 0.2, 0.8 \rangle$	$\langle 0.5, 0.4 \rangle$	$\langle 0.6, 0.4 \rangle$	1
x_4	$\langle 0.1, 0.9 \rangle$	$\langle 0.6, 0.3 \rangle$	$\langle 0.2, 0.7 \rangle$	$\langle 0.2, 0.8 \rangle$	$\langle 0.6, 0.4 \rangle$	1
x_5	$\langle 0.4, 0.6 \rangle$	$\langle 0.2, 0.6 \rangle$	$\langle 0.2, 0.8 \rangle$	$\langle 0.2, 0.8 \rangle$	$\langle 0.2, 0.8 \rangle$	2

Table 1. An information system decision table.

Relation	Value
$R^c(x_1, x_1)$	1
$R^c(x_1, x_2) = R^c(x_2, x_1)$	0.88
$R^c(x_1, x_3) = R^c(x_3, x_1)$	0.83
$R^c(x_1, x_4) = R^c(x_4, x_1)$	0.68
$R^c(x_1, x_5) = R^c(x_5, x_1)$	0.84
$R^c(x_2, x_2)$	1
$R^c(x_2, x_3) = R^c(x_3, x_2)$	0.87
$R^c(x_2, x_4) = R^c(x_4, x_2)$	0.70
$R^c(x_2, x_5) = R^c(x_5, x_2)$	0.80
$R^c(x_3, x_3)$	1
$R^c(x_3, x_4) = R^c(x_4, x_3)$	0.81
$R^c(x_4, x_5) = R^c(x_5, x_3)$	0.77
$R^c(x_4, x_4)$	1
$R^c(x_4, x_5) = R^c(x_5, x_4)$	0.78
$R^c(x_5, x_5)$	1

Table 2. Similarity relation between objects.

$$\eta_{\underline{R}_B^c}(x_i) = \sup_{x_j \in U^d} \min(\rho_{R^c}(x_i, x_j), \eta_x(x_j)) \quad (10)$$

$$\rho_{\overline{R}_B^c}(x_i) = \sup_{x_j \in U^d} \min(\rho_{R^c}(x_i, x_j), \rho_x(x_j)) \quad (11)$$

$$\eta_{\overline{R}_B^c}(x_i) = \inf_{x_j \in U^d} \max(\eta_{R^c}(x_i, x_j), \eta_x(x_j)) \quad (12)$$

Now, the available pair $(\rho_{\underline{R}_B^c}(x_i), \eta_{\underline{R}_B^c}(x_i))$ along with $\rho_{\overline{R}_B^c}(x_i), \eta_{\overline{R}_B^c}(x_i)$ are portrayed as an IF numbers of data point/object x_i to the well depicted lower and upper approximations. If, we have $\underline{R}^c(x) = \overline{R}^c(x)$, then, we obtain $(\underline{R}^c(x), \overline{R}^c(x))$, which is stated as an essential IF-definable set.

Definition 3.4 (IF granular structure) Let $(U^d, A \cup D)$ be an aforesaid IF decision system with U^d and a subset of conditional dimensions $B \subseteq A$. Now, the proposed IF similarity relation is employed to achieve different levels of granularity, which results in better classification performance. Here, we compute optimal feature subset based on different levels of granularity to acquire the optimized overall accuracy. The $R_c^B(x_i, x_j)$ between the instances x_i and x_j can be used to indicate the similarity and dissimilarity between the instances by using the information from membership and non-membership values respectively. The overall impact of noise can be eliminated or minimize by putting the value of R_c^B equal to zero by considering that small value is achieved due to availability of noise. This goal can be effectively achieved by constructing parameterized IF granule with $\lambda, \delta \in (0, 1]$ to discard noise as follows:

(i) (i) For membership grade:

$$[x_i]_{\rho_{R_B^c}}^{\lambda_i} = \begin{cases} \lambda_i, & \rho_{R_B^c}(x_i, x_j) \geq \lambda_i \\ 0, & \text{otherwise.} \end{cases} \quad (13)$$

(ii) (ii) For non-membership grade

$$[x_i]_{\eta_{R_B^c}}^{\delta_i} = \begin{cases} \delta_i, & \eta_{R_B^c}(x_i, x_j) \geq \delta_i \\ 0, & \text{otherwise.} \end{cases} \quad (14)$$

From the above expressions, it can be easily identified that the two parameters λ and δ successfully impact the different size of IF granules. Here, we can observe that the IF granule is based on the diversity function of IF similarity relation. Both the parameters influence the different size of the IF granule: the larger is the values of these parameters results in the larger IF granule. Hence, we can easily compute optimized feature subset and size of the large volume of high-dimensional datasets consisted of uncertainty and noise. Here, we have $\lambda_i = \rho_{R_A^c}[x_i]_D(x_i) > 0$. Further, $[x_i]_{\rho_{R_B^c}}^{\lambda_i}$ can be exemplified as the IF granule (13) persuaded by x_i with respect to $B \subseteq A$.

$\delta_i = \eta_{R_A^c}[x_i]_D(x_i) > 0$. Further, $[x_i]_{\eta_{R_B^c}}^{\delta_i}$ can be exemplified as the IF granule (14) persuaded by x_i with respect to $B \subseteq A$.

Now, the collection of the IF granules $[x_i]_{\rho_{R_B^c}}^{\lambda_i}$ stimulated by all the data points/samples in U^d relative to the given subset of features/attributes B , and is depicted by the expression $\rho GrS(U^d, B)$, where,

$$\rho GrS(U^d, B) = \left\{ [x_i]_{\rho_{R_B^c}}^{\lambda_i} : x_i \in U^d, i = 1, 2, \dots, n, B \subseteq A, \lambda_i = \rho_{R_A^c}[x_i]_D(x_i) \right\} \quad (15)$$

Now, we have $\delta_i = \eta_{R_A^c}[x_i]_D(x_i) > 0$. Moreover, $[x_i]_{\eta_{R_B^c}}^{\delta_i}$ can be illustrated by the IF granule stimulated by x_i relative to the subset of dimensions B . Then, the set of the intuitionistic fuzzy granules $[x_i]_{\eta_{R_B^c}}^{\delta_i}$ stimulated by all the given samples/objects in U^d relative to available subset of attributes/dimensions B , and is depicted by the expression $\eta GrS(U^d, B)$, where,

$$\eta GrS(U^d, B) = \left\{ [x_i]_{\eta_{R_B^c}}^{\delta_i} : x_i \in U^d, i = 1, 2, \dots, n, B \subseteq A, \delta_i = \eta_{R_A^c}[x_i]_D(x_i) \right\} \quad (16)$$

Example 3.5 Here, we use the example data as given in Table 1 to select the instance for different values of λ_i and δ_i . As in Tables 3, 4, and 5, various objects/instances and attributes are chosen as values of λ_i and δ_i change.

To compute the importance degree of IF granule, we propose the essential definition, which can be elucidated as follows:

Objects	Attributes					
	a_1	a_2	a_3	a_4	a_5	d
x_1	$\langle 0.2, 0.4 \rangle$		$\langle 0.2, 0.6 \rangle$	$\langle 0.6, 0.4 \rangle$	$\langle 0.2, 0.8 \rangle$	1
x_2			$\langle 0.3, 0.6 \rangle$		$\langle 0.2, 0.7 \rangle$	2
x_3			$\langle 0.2, 0.8 \rangle$	$\langle 0.5, 0.4 \rangle$	$\langle 0.6, 0.4 \rangle$	1
x_4		$\langle 0.6, 0.3 \rangle$	$\langle 0.2, 0.7 \rangle$	$\langle 0.2, 0.8 \rangle$	$\langle 0.6, 0.4 \rangle$	1
x_5	$\langle 0.4, 0.6 \rangle$	$\langle 0.2, 0.6 \rangle$	$\langle 0.2, 0.8 \rangle$	$\langle 0.2, 0.8 \rangle$	$\langle 0.2, 0.8 \rangle$	2

Table 3. Case 1: $\lambda_i > 0.1$ and $\delta_i > 0.3$.

Objects	Attributes					
	a_1	a_2	a_3	a_4	a_5	d
x_1				$\langle 0.6, 0.4 \rangle$		1
x_2						2
x_3				$\langle 0.5, 0.4 \rangle$	$\langle 0.6, 0.4 \rangle$	1
x_4		$\langle 0.6, 0.3 \rangle$			$\langle 0.6, 0.4 \rangle$	1
x_5	$\langle 0.4, 0.6 \rangle$					2

Table 4. Case 2: $\lambda_i > 0.3$ and $\delta_i > 0.2$.

Objects	Attributes					
	a_1	a_2	a_3	a_4	a_5	d
x_1		$\langle 0.1, 0.7 \rangle$	$\langle 0.2, 0.6 \rangle$		$\langle 0.2, 0.8 \rangle$	1
x_2	$\langle 0.1, 0.7 \rangle$	$\langle 0.1, 0.8 \rangle$	$\langle 0.3, 0.6 \rangle$		$\langle 0.2, 0.7 \rangle$	2
x_3	$\langle 0.1, 0.8 \rangle$	$\langle 0.1, 0.8 \rangle$	$\langle 0.2, 0.8 \rangle$			1
x_4	$\langle 0.1, 0.9 \rangle$		$\langle 0.2, 0.7 \rangle$	$\langle 0.2, 0.8 \rangle$		1
x_5	$\langle 0.4, 0.6 \rangle$	$\langle 0.2, 0.6 \rangle$	$\langle 0.2, 0.8 \rangle$	$\langle 0.2, 0.8 \rangle$	$\langle 0.2, 0.8 \rangle$	2

Table 5. Case 3: $\lambda_i > 0.05$ and $\delta_i > 0.5$.

Definition 3.6 Let $(U^d, A \cup D)$ be an aforementioned given IF decision system, where, $U^d, B \subseteq A$ and $[x_i]_{\rho_{R_B}^{\lambda_i}} \in \rho GrS(U^d, B)$, for all given $x_j \in U^d$. Now, if $[x_i]_{\rho_{R_B}^{\lambda_i}} > 0$, it can be elucidated that the available data points/instances x_j can be covered by the discussed IF granule $[x_i]_{\rho_{R_B}^{\lambda_i}}$. Here, the collection of all available instances/objects can be covered by $[x_i]_{\rho_{R_B}^{\lambda_i}}$, and is depicted with

$$CR([x_i]_{\rho_{R_B}^{\lambda_i}}) = \{x_j : [x_i]_{\rho_{R_B}^{\lambda_i}}(x_j) > 0, x_j \in U^d\} \quad (17)$$

Definition 3.7 Let $(U^d, A \cup D)$ be any given IF decision system, where, $U^d, B \subseteq A$ along with $[x_i]_{\eta_{R_B}^{\delta_i}} \in \eta GrS(U^d, B)$ for any specified tuple/object $x_j \in U^d$. Now, If we have $[x_i]_{\eta_{R_B}^{\delta_i}} > 0$, then, we identify that the instances/examples x_j can be covered by the IF granule $[x_i]_{\eta_{R_B}^{\delta_i}}$. Here, The collection of samples/instances can be covered by the expression $[x_i]_{\eta_{R_B}^{\delta_i}}$, and can be efficiently presented by the following interesting equation:

$$CR([x_i]_{\eta_{R_B}^{\delta_i}}) = \{x_j : [x_i]_{\eta_{R_B}^{\delta_i}}(x_j) > 0, x_j \in U^d\} \quad (18)$$

Based on the Definitions 3.6 and 3.7, it can be observed that the tuple/object available in $CR([x_i]_{\rho_{R_B}^{\lambda_i}})$, additionally $CR([x_i]_{\eta_{R_B}^{\delta_i}})$ can easily be differentiated by the $[x_i]_{\rho_{R_B}^{\lambda_i}}$ and $[x_i]_{\eta_{R_B}^{\delta_i}}$ from the set of tuples/instances U^d . Moreover, The higher values of $|CR([x_i]_{\rho_{R_B}^{\lambda_i}})|$ and $|CR([x_i]_{\eta_{R_B}^{\delta_i}})|$ depicts that the more instances/samples can be covered by $[x_i]_{\rho_{R_B}^{\lambda_i}}$ and $[x_i]_{\eta_{R_B}^{\delta_i}}$, this clearly outlines that $[x_i]_{\rho_{R_B}^{\lambda_i}}$ and $[x_i]_{\eta_{R_B}^{\delta_i}}$ can efficiently hold more powerful discerning ability.

Proposition 3.8 Let $(U^d, A \cup D)$ be a given well-described IF decision system, where U^d . If we have an arbitrary instance/object $x_i \in U^d$, then both $[x_i]_{\rho_{R_B}^{\lambda_i}}^{\lambda_i} \subseteq [x_i]_{\rho_{R_H}^{\lambda_i}}^{\lambda_i}$ and $|CR([x_i]_{\rho_{R_B}^{\lambda_i}}^{\lambda_i})| < |CR([x_i]_{\rho_{R_H}^{\lambda_i}}^{\lambda_i})|$ satisfies for an available $H \subseteq B \subseteq A$.

Proof If, for any given tuple/example x_j , which holds the condition $[x_i]_{\rho_{R_B}^{\lambda_i}}^{\lambda_i} > 0$, then we have $\rho_{R_B}^{\lambda_i}[x_i] > \lambda_i$. Now, since, $R_B^c \subseteq R_H^c$ satisfies for an available $H \subseteq B$, and we have the criteria $\rho_{R_H}^{\lambda_i}[x_i] > \rho_{R_B}^{\lambda_i}[x_i] > \lambda_i$ and

$\Rightarrow \rho_{R_H}^{\lambda_i} > \lambda_i$, then we get $[x_i]_{\rho_{R_H}^{\lambda_i}}^{\lambda_i}(x_j) = \lambda_i > 0$. Hence, we obtain $[x_i]_{\rho_{R_B}^{\lambda_i}}^{\lambda_i}(x_j) = [x_i]_{\rho_{R_H}^{\lambda_i}}^{\lambda_i}(x_j)$ Now, on the contrary, for any given sample/example x_j , which holds $[x_i]_{\rho_{R_B}^{\lambda_i}}^{\lambda_i}(x_j) = 0$, therefore, we identify that $\rho_{R_B}^{\lambda_i}[x_i](x_j) \leq \lambda_i$. So, either the expression $\rho_{R_B}^{\lambda_i}[x_i](x_j) > \lambda_i \Rightarrow [x_i]_{\rho_{R_H}^{\lambda_i}}^{\lambda_i}(x_j) = \lambda_i > 0$ or the expression $\rho_{R_H}^{\lambda_i}[x_i](x_j) \leq \lambda_i \Rightarrow [x_i]_{\rho_{R_H}^{\lambda_i}}^{\lambda_i}(x_j) = 0$.

Therefore, we get $[x_i]_{\rho_{R_B}^{\lambda_i}}^{\lambda_i}(x_j) \leq [x_i]_{\rho_{R_H}^{\lambda_i}}^{\lambda_i}(x_j)$.

In outline, $[x_i]_{\rho_{R_H}^{\lambda_i}}^{\lambda_i}(x_j) \leq [x_i]_{\rho_{R_B}^{\lambda_i}}^{\lambda_i}(x_j)$ satisfies for any given $x_j \in U$, which results in $[x_i]_{\rho_{R_B}^{\lambda_i}}^{\lambda_i} \subseteq [x_i]_{\rho_{R_H}^{\lambda_i}}^{\lambda_i}$, thus obvious result can be obtained as $|CR[x_i]_{\rho_{R_B}^{\lambda_i}}^{\lambda_i}| \leq |CR([x_i]_{\rho_{R_H}^{\lambda_i}}^{\lambda_i})|$.

Further, It can be handily detected from the aforementioned Proposition 3.8 that the above discussed IF granule $[x_i]_{\rho_{R_B}^{\lambda_i}}^{\lambda_i}$ holds the interesting monotonicity condition. In particular, $[x_i]_{\rho_{R_{B \cup \{a\}}}^{\lambda_i}}^{\lambda_i} \subseteq [x_i]_{\rho_{R_B}^{\lambda_i}}^{\lambda_i}$ satisfies for a given dimension $a \in A \setminus B$, which efficiently results in

$$|CR([x_i]_{\rho_{R_{B \cup \{a\}}}^{\lambda_i}}^{\lambda_i})| \leq |CR([x_i]_{\rho_{R_B}^{\lambda_i}}^{\lambda_i})|$$

Proposition 3.9 Let $(U^d, A \cup D)$ be an available IF decision system, where U^d . Now, for an existing arbitrary data point $x_i \in U^d$, we have both $[x_i]_{\eta_{R_B}^{\delta_i}}^{\delta_i} \subseteq [x_i]_{\eta_{R_F}^{\delta_i}}^{\delta_i}$ as well as $|CR([x_i]_{\eta_{R_B}^{\delta_i}}^{\delta_i})| \leq |CR([x_i]_{\eta_{R_F}^{\delta_i}}^{\delta_i})|$, which satisfies for any $H \subseteq B \subseteq A$

Proof It can be identified as an obvious proof similar to proposition 3.8 □

Definition 3.10 Let $(U^d, A \cup D)$ be an IF decision system, where U^d is the collection of examples/instances. Now, we can illustrate the importance degree of IF granule $[x_i]_{\rho_{R_A}^{\lambda_i}}^{\lambda_i} \in \rho_{GrS}(U^d, A)$ by the following expression:

$$H([x_i]_{\rho_{R_A}^{\lambda_i}}^{\lambda_i}) = \lambda_i |CR([x_i]_{\rho_{R_A}^{\lambda_i}}^{\lambda_i})| \quad (19)$$

Definition 3.11 Let $(U^d, A \cup D)$ be an above defined IF decision system, where U^d represents the given instances/samples. Then, we can depict the importance degree of IF granule $[x_i]_{\eta_{R_A}^{\delta_i}}^{\delta_i} \in \eta_{GrS}(U^d, A)$ by the below expression:

$$H([x_i]_{\eta_{R_A}^{\delta_i}}^{\delta_i}) = \delta_i |CR([x_i]_{\eta_{R_A}^{\delta_i}}^{\delta_i})| \quad (20)$$

Based on the above demonstrated definition 3.10 and definition 3.11, we have the importance degree of the IF granules $[x_i]_{\rho_{R_A}^{\lambda_i}}^{\lambda_i}$ and $[x_i]_{\eta_{R_A}^{\delta_i}}^{\delta_i}$, which is illustrated by two important factors; the IF lower approximation assess λ_i (for membership) and δ_i (for non-membership), and the essential cardinality of coverage data points/samples set $|CR([x_i]_{\rho_{R_A}^{\lambda_i}}^{\lambda_i})|$ and $|CR([x_i]_{\eta_{R_A}^{\delta_i}}^{\delta_i})|$ can be described as the more influential discriminating proficiency of the IF granule holds (17) (18). Moreover, λ_i and δ_i is employed to contemplate the determinacy of the discerning proficiency of $[x_i]_{\rho_{R_A}^{\lambda_i}}^{\lambda_i}$ and $[x_i]_{\eta_{R_A}^{\delta_i}}^{\delta_i}$. Further, if the IF granule is elicited by sample/instance x_i , then, the importance degree $H([x_i]_{\rho_{R_A}^{\lambda_i}}^{\lambda_i})$ along with $H([x_i]_{\eta_{R_A}^{\delta_i}}^{\delta_i})$ is again illustrated as the importance degree of example/sample x_i (19) (20).

Let $(U^d, A \cup D)$ be an aforementioned IF decision system and decision class $U^d/D_t = \{D_t1, D_t2, D_t3, \dots, D_tl\}$ be the important decision split. Now, we present the forward object/instance selection idea, which can be classified into two categories as follows:

- (i) Initially, we contemplate to determine the noise in each available decision class D_t (1, 2, 3, ..., l), thereafter, the noise is eliminated from D_t . Now, we calculate $CR([x_i]_{\rho_{R_A}^{\lambda_i}}^{\lambda_i})$ along with $CR([x_i]_{\eta_{R_A}^{\delta_i}}^{\delta_i})$ for any given $x_i \in D_t$. Further, the coverage potential of the noise is found to be lower. Moreover, we consider the sample x_{i0} along with $|CR([x_{i0}]_{\rho_{R_A}^{\lambda_{i0}}}^{\lambda_{i0}})| = 1$ as well as $|CR([x_{i0}]_{\eta_{R_A}^{\delta_{i0}}}^{\delta_{i0}})| = 1$, which is obtained as the crucial potential possible noise provided in the IF decision system $(U^d, A \cup D)$. Subsequently, we have to

further identify whether x_{i_0} is computed as noise or not. In particular, we have to detect for the k -nearest neighbors of computed value x_{i_0} , which is delineated as the k objects that are nearest to the value x_{i_0} . Hereafter, the computed proximity of x_{i_0} relative to the different instances given in U^d , which is expressed by the following equation:

$Pro_A(x_{i_0}, x_i) = 1 - R_A^c(x_{i_0}, x_i) (x_i \in U^d, i \neq i_0)$. Moreover, we describe the cardinality of any given $x_i \in U$ by the below demonstration :

$$K_{x_{i_0}} = \frac{1 + \rho_A(x_{i_0}) - \eta_A(x_{i_0})}{2} \quad (21)$$

Here, R_A^c is demonstrated as the IF instance selection relative to the given conditional dimensions/features set A . Now, the collection of the k -nearest neighbors with respect to x_{i_0} , which can illustrated as $k_n(x_{i_0})$ equal to the number of samples in $k_n(x_{i_0})$ consisted of distinct labels from the available label of x_{i_0} is found as higher value with respect to the $k_{x_{i_0}}$, then, the sample/object x_{i_0} is considered as noise. Thereafter, it should be highlighted that the existing label of an object x_i is obtained as factual value of the decision dimension dp , i.e., $dp(x_i)$. Moreover, the selection of the parameter k is implied in the section of the comprehensive numerical experimental study.

- (ii) Now, we identify the representative samples on the basis of definition 3.4 and definition 3.6. It can be noted that $D_t (t = 1, 2, \dots, l)$ portray the decision class, where the noise has been effectively removed. Here, we consider that the essential importance degree of a specified data point can also be pondered by the Definition 3.1 with respect to each data point $x_j \in D_t$. Now, we determine $H([x_j]_{\rho_{R_A^c}}^{\lambda_j})$, and detect the instance x_{j_0} , which holds $\max_{j \in \{1, 2, \dots, |D_t|\}} H([x_j]_{\rho_{R_A^c}}^{\lambda_j})$. Consequently, the objects/data points covered by the $[x_{j_0}]_{\rho_{R_A^c}}^{\lambda_{j_0}}$, merely, x_{j_0} is discarded from the D_t .

Now, for any given $x_{j_1} \in CR([x_{j_0}]_{\rho_{R_A^c}}^{\lambda_{j_0}})$, we assume that $CR([x_{j_1}]_{\rho_{R_A^c}}^{\lambda_{j_1}}) = \phi$, which proceeds to discard x_{j_1} from $CR([x_j]_{\rho_{R_A^c}}^{\lambda_j}) (x_j \in CR([x_{j_0}]_{\rho_{R_A^c}}^{\lambda_{j_0}}), j = 1, 2, \dots, |D_t|)$.

Here, the entire depicted process is iterated till $CR([x_j]_{\rho_{R_A^c}}^{\lambda_j}) = \phi$ for any given $x_j \in D_t (j = 1, 2, \dots, |D_t|)$, we get the equivalent value as $\max_{j \in \{1, 2, \dots, |D_t|\}} H([x_j]_{\rho_{R_A^c}}^{\lambda_j}) = 0$. Likewise, it can be inferred as $H([x_j]_{\rho_{R_A^c}}^{\delta_j}) = 0$.

From the above discussion, we can conclude that a given IF decision system $(U^d, A \cup D)$ converted into an IF decision subsystem $(U^{d*}, A \cap D)$, where U^{d*} can be explained as a representative data points set.

Input: An IF decision system $(U^d, A \cup D)$,
Output: Sample subset U^{d*}
initialize $U^{d*} = \phi$
Compute the IF granule
 $[x_i]_{\eta_{R_B}}^{\delta_i}$ with respect to $A \subseteq B$
for $i=1$ to n do
compute covering set
 $CR([x_i]_{\rho_{R_A^c}}^{\lambda_i})$ and importance degree $H([x_i]_{\eta_{R_B}}^{\delta_i})$
end for
while
 $\max_{j=1, 2, 3, \dots, |D_t|} H([x_j]_{\rho_{R_A^c}}^{\lambda_j}) > 0$ do
add instance x_{j_0} to U^{d*} satisfying $H([x_{j_0}]_{\rho_{R_A^c}}^{\lambda_{j_0}}) = \max_{j=1, 2, 3, \dots, |D_t|} H([x_j]_{\rho_{R_A^c}}^{\lambda_j})$
Update $U^d = U^d - (x_{j_0})_{\rho_{R_A^c}}^{\lambda_{j_0}}$
for each $x_j \in (x_{j_0})_{\rho_{R_A^c}}^{\lambda_{j_0}}$ do
remove instance x_{j_0} from D_t
end for
for each $x_{j_1} \in CR([x_{j_0}]_{\rho_{R_A^c}}^{\lambda_{j_0}})$
if $CR([x_j]_{\rho_{R_A^c}}^{\lambda_j}) = \phi$
remove x_{j_1} from $CR([x_j]_{\rho_{R_A^c}}^{\lambda_j})$
end for
for $i=1$ to $|U^d|$ do
recompute $CR([x_i]_{\rho_{R_B}}^{\lambda_i})$ and importance degree $H([x_i]_{\rho_{R_A^c}}^{\delta_i})$
end for
 $CR([x_{j_1}]_{\rho_{R_A^c}}^{\lambda_{j_1}}) = \phi$
end while
return U^{d*}

Algorithm 1. An instance elimination procedure to remove noise by identifying the representative data points

Importance-degree-preserved IF feature selection method

In the high-dimensional datasets, redundancy and irrelevancy in dimensions can be observed frequently. Now, for above computed IF decision system $(U^{d*}, A \cup D)$, if we select the critical attribute/dimension set A^* by preserving the essential importance degrees of the aforementioned IF granules, then we can obtain a more compact IF decision system $(U^{d*}, A^* \cup D)$, where the size and dimension of the datasets can be found effectively smaller. Thus, we can accomplish the objective of the interesting data reduction task. Therefore, in the current segment, importance-degree-preserved dimensionality/feature elimination for the IF decision system $(U^{d*}, A \cup D)$ is presented. Thereafter, we present the corresponding attribute/dimension selection algorithm.

Now, we will exercise all the related work based on the IF decision subsystem $(U^{d*}, A \cup D)$ to determine a new dimensionality reduction notion. Here, the collection of representative examples U^{d*} can be partitioned by using the decision equivalence relation R_D^c towards a category of disjoint subsets $U^{d*}/D = \{[x_i]_D^*, x_i \in U^{d*}\}$, such that $[x_i]_D^*$ is described as the decision class of $(U^{d*}, A \cup D)$, where any given x_i stands in U^{d*} . It can be outlined that, for an obtained IF decision system $(U^{d*}, A \cup D)$, we have IF lower approximation estimate of x_i with respect to $[x_i]_D^*$, which is expressed by:

$$\lambda_i^* = \rho_{R_A^c} [x_i]_D^* (x_i) = \inf_{x_j \in U^d} \max \{ \eta_{R_A^c} (x_i, x_j), \rho_X (x_j) [x_i]_D^* (x_j) \} \quad (22)$$

Additionally,

$$\delta_i^* = \eta_{R_A^c} [x_i]_D^* (x_i) = \sup_{x_j \in U^d} \min \{ \rho_{R_A^c} (x_i, x_j), \eta_X (x_i) [x_i]_D^* (x_i) \} \quad (23)$$

this computation appears to be different from the λ_i, δ_i for any given $(U^d, A \cup D)$ alongside $(U^{d*}, A \cup D)$, which represents the collection of all the IF granules inserted by the examples available in U^{d*} related to B , which can be given by the following expression:

$$\rho_{GrS}(U^{d*}, B) = \{ [x_i]_{\rho_{R_B^c}}^{\lambda_i^*}, x_i \in U^{d*}, B \subseteq A, \lambda_i^* = \rho_{R_A^c} [x_i]_D^* (x_i) \} \quad (24)$$

$$\eta_{GrS}(U^{d*}, B) = \{ [x_i]_{\eta_{R_B^c}}^{\delta_i^*}, x_i \in U^{d*}, B \subseteq A, \delta_i^* = \eta_{R_A^c} [x_i]_D^* (x_i) \} \quad (25)$$

Definition 3.12 Let $(U^{d*}, A \cup D)$ be an aforesaid IF decision System. Now, for any $B \subseteq A$, we can compute the importance degree of above demonstrated IF granule as follows:

$$[x_i]_{\rho_{R_B^c}}^{\lambda_i^*} \in \rho_{GrS}(U, B) \quad (26)$$

is

$$F \left[[x_i]_{\rho_{R_B^c}}^{\lambda_i^*} \right] = \frac{\left| CR \left([x_i]_{\rho_{R_A^c}}^{\lambda_i^*} \right) \right|^2}{\left| CR \left([x_i]_{\rho_{R_B^c}}^{\lambda_i^*} \right) \right|} \rho_{R_B^c} [x_i]_D^* (x_i) \quad (27)$$

Here, we have

$$\rho_{\overline{R_B^c}(X)}(x_i) = \sup_{x_j \in U} \min \left(\rho_{R^c}(x_i, x_j), \rho_X(x_j) \right) [x_i]_D^*, \quad (28)$$

In particular, if we have $U^{d*} = U$, then the definition 3.1 clearly indicates that the required importance degree of the aforementioned IF granule

$$[x_i]_{\rho_{R_B^c}}^{\lambda_i} \in \rho_{GrS}(U, B) \quad (29)$$

can be given as:

$$F \left[[x_i]_{\rho_{R_B^c}}^{\lambda_i} \right] = \frac{\left| CR \left([x_i]_{\rho_{R_A^c}}^{\lambda_i} \right) \right|^2}{\left| CR \left([x_i]_{\rho_{R_B^c}}^{\lambda_i} \right) \right|} \rho_{R_B^c} [x_i]_D^* (x_i) \quad (30)$$

. Moreover, the calculated importance degree of IF granule

$$[x_i]_{\rho_{R_B^c}}^{\lambda_i} \in \rho_{GrS}(U^{d*}, A \cup B) \quad (31)$$

, can be expressed by:

$$F\left[x_i\right]_{\rho_{R_A^c}^{\lambda_i}} = \frac{\left|CR\left([x_i]_{\rho_{R_A^c}^{\lambda_i}}\right)\right|^2}{\left|CR\left([x_i]_{\rho_{R_A^c}^{\lambda_i}}\right)\right|} \rho_{R_A^c}^{\lambda_i} [x_i]_D^*(x_i) \quad (32)$$

$$= \left|CR\left([x_i]_{\rho_{R_A^c}^{\lambda_i}}\right)\right| \lambda_i \quad (33)$$

Definition 3.13 Let $(U^{d*}, A \cup D)$ be an obtained IF decision system. Now, for any given $B \subseteq A$, we can get the importance degree of the IF granule as follows:

$$[x_i]_{\rho_{R_B^c}^{\delta_i}^*} \in \rho_{G_r S}(U^{d*}, A \cup B) \quad (34)$$

can be further depicted by:

$$F\left[x_i\right]_{\eta_{R_B^c}^{\delta_i}^*} = \frac{\left|CR\left([x_i]_{\eta_{R_B^c}^{\delta_i}^*}\right)\right|^2}{\left|CR\left([x_i]_{\eta_{R_B^c}^{\delta_i}^*}\right)\right|} \eta_{R_B^c}^{\delta_i} [x_i]_D^*(x_i) \quad (35)$$

$$\eta_{\overline{R^c} B(X)}(x_i) = \inf_{x_j \in U} \max(\eta_{R^c}(x_i, x_j), \eta_x(x_j)), [x_i]_D^* \quad (36)$$

In particular, if we have $U^{d*} = U$, then, the definition 3.7 clarifies that the importance degree of IF granule can be illustrated as:

$$[x_i]_{\eta_{R_B^c}^{\delta_i}^*} \in \eta_{Gr S}(U^{d*}, A \cup B) \quad (37)$$

can be given by:

$$F\left[x_i\right]_{\eta_{R_B^c}^{\delta_i}^*} = \frac{\left|CR\left([x_i]_{\eta_{R_B^c}^{\delta_i}^*}\right)\right|^2}{\left|CR\left([x_i]_{\eta_{R_B^c}^{\delta_i}^*}\right)\right|} \eta_{R_B^c}^{\delta_i} [x_i]_D^*(x_i) \quad (38)$$

. Thus, the required importance degree of IF granule can be demonstrated as:

$$[x_i]_{\rho_{R_A^c}^{\lambda_i}^*} \in \rho_{c_r S}(U^{d*}, A \cup B) \quad (39)$$

can be further depicted by the expression:

$$F\left[x_i\right]_{\eta_{R_A^c}^{\delta_i}^*} = \frac{\left|CR\left([x_i]_{\eta_{R_A^c}^{\delta_i}^*}\right)\right|^2}{\left|CR\left([x_i]_{\eta_{R_A^c}^{\delta_i}^*}\right)\right|} \eta_{R_A^c}^{\delta_i} [x_i]_D^*(x_i) \quad (40)$$

$$= \left|CR\left([x_i]_{\eta_{R_A^c}^{\delta_i}^*}\right)\right| \delta_i \quad (41)$$

Property 3.14 Let $(U^{d*}, A \cup D)$ be an aforesaid IF decision system. Now, we have $F\left([x_i]_{\rho_{R_E^c}^{\delta_i}^*}\right) \leq F\left([x_i]_{\rho_{R_B^c}^{\delta_i}^*}\right)$, which satisfies for any given $x_i \in U$ along with $E \subseteq B \subseteq A$.

Proof For an available IF decision system $(U^{d*}, A \cup D)$, we have an arbitrary example $x_i \in U^{d*}$, based on proposition 1, it can be obtained that $\left|CR\left([x_i]_{\rho_{R_E^c}^{\lambda_i}^*}\right)\right| \leq \left|CR\left([x_i]_{\rho_{R_B^c}^{\lambda_i}^*}\right)\right|$, which satisfies for any provided $E \subseteq B$. Moreover, we have

$$\rho_{R_E^c}^{\lambda_i} [x_i]_D^*(x_i) = \inf_{x_j \in U^{d*}} \max\left\{(\eta_{R_E^c}(x_i, x_j), \rho_x(x_i)), [x_i]_D^*(x_j)\right\} \leq$$

$$\rho_{R_B^c}^{\lambda_i} [x_i]_D^*(x_i) = \inf_{x_j \in U^{d*}} \max\left\{(\eta_{R_B^c}(x_i, x_j), \rho_x(x_i)), [x_i]_D^*(x_j)\right\}$$

, satisfies for any provided $x_i \in U^{d*}$. Further, we get

$$F\left([x_i]_{\rho_{R_E^c}^{\lambda_i}^*}\right) = \frac{\left|CR\left([x_i]_{\rho_{R_E^c}^{\lambda_i}^*}\right)\right|^2}{\left|CR\left([x_i]_{\rho_{R_E^c}^{\lambda_i}^*}\right)\right|} \rho_{R_E^c}^{\lambda_i} [x_i]_D^*(x_i) \leq \frac{\left|CR\left([x_i]_{\rho_{R_A^c}^{\lambda_i}^*}\right)\right|^2}{\left|CR\left([x_i]_{\rho_{R_B^c}^{\lambda_i}^*}\right)\right|} \rho_{R_B^c}^{\lambda_i} [x_i]_D^*(x_i) = F\left([x_i]_{\rho_{R_B^c}^{\lambda_i}^*}\right) \quad \square$$

we are aware of importance degree of IF granular structure $[x_i]_{\rho_{R_E^c}^{\delta_i}^*}$ established by two determining factor: the IF lower approximation λ_i and the cardinality of the coverage object set $|CR([x_i]_{\rho_{R_B^c}^{\lambda_i}^*})|$ and in summary, importance degree of IF satisfies monotonicity.

Property 3.15 Let $(U^{d*}, A \cup D)$ be an aforesaid IF decision system. Now, $F([x_i]_{\eta_{R_E}^{d*}}) \leq F([x_i]_{\eta_{R_B}^{d*}})$, satisfies for any given $x_i \in U$ along with the available $E \subseteq B \subseteq A$.

Proof The proof can be given in the same way as property 3.14 $\square$

Definition 3.16 Let $(U^{d*}, A \cup D)$ be an aforementioned IF decision system. Now, a computed $B \subseteq A$ can be described as importance degree preserved aided consistent collection in the computed IF decision system $(U^{d*}, A \cup D)$. Now, if $F([x_i]_{\rho_{R_B}^{d*}}) = F([x_i]_{\rho_{R_A}^{d*}})$, satisfies for an available $x_i \in U^{d*}$, B can be depicted as importance degree preserved feature subset of $(U^{d*}, A \cup D)$. Further, if B is calculated as importance degree preserved assisted consistent collection in the calculated IF decision system $(U^{d*}, A \cup D)$. Thereafter, there holds $x_{i_0} \in U^{d*}$ in such a way that $F([x_{i_0}]_{\rho_{R_B-\{a\}}^{d*}}) < F([x_{i_0}]_{\rho_{R_B}^{d*}})$, which satisfies for a given $a \in B$

Definition 3.17 Let $(U^{d*}, A \cup D)$ be an aforesaid IF decision system. Now, $B \subseteq A$ can be explained as an importance degree preserved assisted consistent collection of the calculated IF decision system $(U^{d*}, A \cup D)$. Further, if $F([x_i]_{\eta_{R_B}^{d*}}) = F([x_i]_{\eta_{R_A}^{d*}})$, satisfies for a provided $x_{i_0} \in U$. Moreover, B can be illustrated as importance degree preserved feature subset of a computed IF decision system $(U^{d*}, A \cup D)$. Furthermore, if B is obtained as an importance degree preserved aided consistent collection of $(U^{d*}, A \cup D)$, where $x_{i_0} \in U^{d*}$ holds in such a way that $F([x_{i_0}]_{\eta_{R_B-\{a\}}^{d*}}) < F([x_{i_0}]_{\eta_{R_B}^{d*}})$, which satisfies for a given $a \in B$.

Definition 3.18 Let $(U^{d*}, A \cup D)$ be a computed IF decision system. Now, if $B \subseteq A$, then the sum of computed importance degrees for all the calculated objects/samples in U^{d*} relative to B , which can be indicated by the following appealing equation:

$$\rho\Psi_{U^{d*}}(B) = \sum_{x_i \in U^{d*}} F([x_i]_{\rho_{R_B}^{d*}}) \quad (42)$$

Definition 3.19 Let $(U^{d*}, A \cup D)$ be a calculated IF decision system. Now, if $B \subseteq A$, then the addition of the aforesaid importance degree for all the examples in representative collection of samples U^{d*} relative to feature subset B , which is given by the expression

$$\eta\Psi_{U^{d*}}(B) = \sum_{x_i \in U^{d*}} F([x_i]_{\eta_{R_B}^{d*}}) \quad (43)$$

Property 3.20 Let $(U^{d*}, A \cup D)$ be aforementioned IF decision system. Then, we have $\rho\Psi_{U^{d*}}(M) \leq \rho\Psi_{U^{d*}}(B)$, which satisfies for any given $M \subseteq B \subseteq A$

Proof On the basis of Definition 3.16, it can be easily implied with the help of Property 3.14. $\square$

Property 3.21 Let $(U^{d*}, A \cup D)$ be aforesaid IF decision system. Now, $\eta\Psi_{U^{d*}}(M) < \eta\Psi_{U^{d*}}(B)$, satisfies for any provided $M \subseteq B \subseteq A$

Proof On the basis of Definition 3.17, it can be easily implied with the help of Property 3.15. $\square$

Theorem 3.22 Let $(U^{d*}, A \cup D)$ be a computed IF decision system. Now, if $B \subseteq A$ is a calculated importance degree preserved feature subset with respect to $(U^{d*}, A \cup D)$ if and only if $\rho\Psi_{U^{d*}}(B) = \rho\Psi_{U^{d*}}(A)$ along with $\rho\Psi_{U^{d*}}(B - \{a\}) < \rho\Psi_{U^{d*}}(B)$ based on a given $a \in B$.

Proof Let $B \subseteq A$ is a calculated importance degree preserved aided consistent collection of the $(U^{d*}, A \cup D) \iff F([x_i]_{\rho_{R_B}^{d*}}) = F([x_i]_{\rho_{R_A}^{d*}})$ for each $x_i \in U^{d*} \iff \sum_{x_i \in U^{d*}} F([x_i]_{\rho_{R_B}^{d*}}) = \sum_{x_i \in U^{d*}} F([x_i]_{\rho_{R_A}^{d*}}) \iff \rho\Psi_{U^{d*}}(B) = \rho\Psi_{U^{d*}}(A)$ by using the Definition 3.6.

Moreover, if we have $x_{i_0} \in U^{d*}$ in such a way that $F([x_{i_0}]_{\rho_{R_B-\{a\}}^{d*}}) < F([x_{i_0}]_{\rho_{R_B}^{d*}})$, which effectively satisfies for any provided $a \in B$, $\iff \sum_{x_i \in U^{d*}} H([x_i]_{\rho_{R_B-\{a\}}^{d*}}) < \sum_{x_i \in U^{d*}} F([x_i]_{\rho_{R_B}^{d*}})$, which efficiently satisfies for an available $a \in B$, $\iff \rho\Psi_{U^{d*}}(B - \{a\}) < \rho\Psi_{U^{d*}}(B)$ for any $a \in B$.

Further, it can be observed based on the Definition 3.16 that B can be identified that an importance degree preserved feature subset iff $\rho\Psi_{U^{d*}}(B) = \rho\Psi_{U^{d*}}(A)$ and $\rho\Psi_{U^{d*}}(B - \{a\}) < \rho\Psi_{U^{d*}}(B)$, for all $a \in B$. $\square$

Let $(U^*, X \cup D)$ be an IF decision system with $X = \{x_1, x_2, \dots, x_j\}$, and U^* be a representative instance set. Assume that the attributes $x_{i_1}, x_{i_2}, \dots$ are added into the null set one by one according to the magnitude of their respective obtained $\rho\Psi_{U^*}$. The process continues until there exists some $t \in \{1, 2, \dots, m\}$ such that $\rho\Psi_{U^*}(\{x_{i_1}, x_{i_2}, \dots, x_{i_t}\}) = \rho\Psi_{U^*}(X)$. It is obtained from above Property that $\rho\Psi_{U^*}(\{x_{i_1}\}) \leq \rho\Psi_{U^*}(\{x_{i_1}, x_{i_2}\}) \leq \dots \leq \rho\Psi_{U^*}(\{x_{i_1}, x_{i_2}, \dots, x_{i_m}\}) = \rho\Psi_{U^*}(A)$.

Theorem 3.23 Let $(U^{d*}, A \cup D)$ be a calculated IF decision system. Now, $B \subseteq A$ can be depicted as an importance degree preserved feature subset of an available $(U^{d*}, A \cup D)$ iff $\eta\Psi_{U^{d*}}(B) = \eta\Psi_{U^{d*}}(A)$ and $\eta\Psi_{U^{d*}}(B - \{a\}) < \eta\Psi_{U^{d*}}(B)$ for any $a \in B$.

Proof The proof can be given similar to Theorem 3.22 □

Input: An IF decision system $(U^{d*}, A \cup D)$
Output: Reduct B
Initialize: $B = \phi$ and $\rho\Psi_{U^{d*}}(B) = 0$
Now, compute $H([x_i]_{\rho R_B^c}^{\lambda_i^*})$ and $\rho GrS(U^{d*}, B)$
while
 $\eta\Psi_{U^{d*}}(B - \{a\}) < \eta\Psi_{U^{d*}}(B)$
for each $a_t \in (A - B)$ do
compute $\eta\Psi_{U^{d*}}(B \cup \{a_t\})$
end for
select $a_{t_0} \in (A - B)$ satisfying $\sum_{x_i \in U^{d*}} H([x_{i_0}]_{\rho R_{B-\{a_t\}}^c}^{\lambda_{i_0}^*}) < \sum_{x_i \in U^{d*}} H([x_{i_0}]_{\rho R_B^c}^{\lambda_{i_0}^*})$
update $\rho\Psi_{U^{d*}}(B) = \rho\Psi_{U^{d*}}(B \cup \{a_t\})$
 $B = B \cup \{a_t\}$
check $\eta\Psi_{U^{d*}}(B) = \eta\Psi_{U^{d*}}(A)$
end while
return B

Algorithm 2. Feature subset selection procedure based on importance degree preserved notion

Experimentation

In this segment, we have conducted thorough experiments, which demonstrates the overview and experimental verification of our illustrated methods. Here, we perform the classification performance for the effectiveness of the proposed models. Moreover, data reduction has been employed to improve the regression performance for the values of AVPs. These experiments are conducted by using Matlab 2023 and WEKA 3.8⁷⁸.

Performance evaluation parameters

Machine learning techniques have been evaluated with both threshold dependent and threshold independent evaluation metrics. These metrics can be calculated from the widely described quadrants of the confusion matrix, which can be outlined as True Positive (TRP), True Negative (TRN), False Positive (FLP), and False Negative (FLN). Based on these parameters, following performance evaluation metrics have been considered for addressing the overall results after data reduction: Sensitivity: This parameter depicts the average percentage of correctly predicted positive class value, which is determined by:

$$\text{Sensitivity} = \frac{\text{TRP}}{((\text{TRP} + \text{FLN}))} \times 100 \quad (44)$$

Specificity: This parameter offers the overall percentage of correctly predicted negative class values, which is estimated by:

$$\text{Specificity} = \frac{\text{TRN}}{((\text{TRN} + \text{FLP}))} \times 100 \quad (45)$$

Accuracy: This metric explains the average percentage of correctly classified positive and negative class values, which can be assessed as follows:

$$\text{Accuracy} = \frac{(\text{TRP} + \text{TRN})}{(\text{TRP} + \text{FLP} + \text{TRN} + \text{FLN})} \times 100 \quad (46)$$

AUC: It validates the calculated value of area under the receiver operating characteristic curve (ROC). If the obtained value of AUC is nearer to 1, then, the predictor is found to be as the better performing one. MCC: Mathew's correlation coefficient is widely and expansively used performance metric for determining better binary classification task. A MCC value calculated as 1 depicts that the classifier is the best performer when compared to others, which can be defined by using the well-known following equation:

$$\text{MCC} = \frac{(\text{TRP} \times \text{TRN} - \text{FLP} \times \text{FLN})}{\sqrt{((\text{TRP} + \text{FLP})(\text{TRP} + \text{FLN})(\text{TRN} + \text{FLP})(\text{TRN} + \text{FLN}))}} \quad (47)$$

Moreover, the regression analysis is evaluated by using the broadly discussed correlation coefficient (CC) and mean absolute errors (MAE). If the values of correlation coefficient is higher and mean absolute error is less, then the regression performance is outlined as the better one.

Precision: This evaluation parameter can be calculated by the following equation:

Dataset	Instances	Attributes	Reduction							
			RBR		IFBR		CFBR		Proposed method	
			Instances	Attributes	Instances	Attributes	Instances	Attributes	Instances	Attributes
Bank marketing	4521	16	3987	14	4121	12	4278	14	4023	11
Hepatitis	155	19	151	15	147	14	137	12	147	6
Dburorld-bodies	64	4702	59	168	46	109	41	144	61	7
Arcene	200	10000	182	243	165	154	178	203	190	60
Mushroom	8124	21	7456	12	7365	16	7711	15	7149	7
qsar-biodegradation	1055	41	987	23	898	27	902	31	1012	15
Fertility diagnosis	100	9	91	7	95	6	89	7	86	5
Ionosphere	351	33	323	21	319	7	334	18	319	7
Thyroid-hypothyroid	3163	25	2977	17	3124	14	3107	19	3099	8
Wdbc	569	21	543	18	533	15	509	12	529	6

Table 6. Benchmark datasets characteristics and their reduction.

$$Precision = \frac{TRP}{TRP + FLP} \tag{48}$$

Recall: This evaluation parameter can be calculated as follows:

$$Recall = \frac{TRP}{TRP + FLN} \tag{49}$$

F-measure: This evaluation metric can be computed by the following expression:

$$F - measure = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{50}$$

Result and discussion

We have evaluated previous as well as proposed method based on ten benchmark datasets, which are taken from UCI repository and their characteristics are depicted in Table 6. In this study, different experiments are carried out on the basis of both 10-fold cross validation and 70:30 percentage split techniques. Initially, the datasets are converted into IF information systems by employing Tan et al.⁷⁰ method. Lower approximation is computed based on the proposed relation. Importance degree of IF granule is computed based on lower approximation. Then, redundant instances are eliminated by using this importance degree. Further, feature selection is applied based on importance degree preserved concept as discussed in “Proposed work”. The reduction rate by earlier and current approaches are reported in Table 6. Instances removed by RBR is higher in number rather than our technique for bank-marketing, dbworld-bodies, Arcene, qsar-biodegradation, and thyroid-hypothyroid, but features are discarded relatively less in number for all the datasets. RBR reduces the instances from 59 to 7456, and features from 7 to 243 as mentioned in Table 6. IFBR obliterates more samples for dbworld-bodies, Arcene, and qsar-biodegradation than proposed approach, but removal of features is found less for all the datasets. For IFBR, instances and attributes varies from 46 to 7365 and from 6 to 154 respectively. Next, CFBR eradicates more number of data points for hepatitis, dbworld-bodies, Arcene, qsar-biodegradation, and Wdbc, but dimensionality is obtained higher in count for all the datasets yet again. For CFBR, samples and dimensions vary from 41 to 7711 and from 7 to 203 respectively. Our proposed method produced the reduced datasets, where size and dimension changes from 61 to 7149 and from 5 to 60. Now, we can conclude that overall dimension in terms of both samples and features are highly reduced by our method when compared to previous techniques, as displayed by Fig. 1. Now, two broadly illustrated classifiers namely JRip and Random Forest (RF) are applied to evaluate the performances over the reduced datasets produced by RBR, IFBR, CFBR, and our technique based on overall accuracies and standard deviation. The proposed method produces better results except in the few cases for the datasets namely hepatitis, qsar-biodegradation, fertility diagnosis, ionosphere, and Wdbc. IFBR generates more effective results for hepatitis, qsar-biodegradation, and fertility diagnosis by JRip with accuracies and standard deviation of 89.24%, 85.54%, 81.41% and 8.31, 2.54, 10.18 respectively. Further, this method gives the equal results with respect to our method by both JRip and RF for ionosphere dataset with average accuracies and standard deviation of 95.55%, 98.12% and 3.78, 2.67 respectively. RBR yields in better performance for hepatitis dataset by RF 92.58% and 5.98 and equivalent measure for qsar-biodegradation by JRip with accuracies and standard deviation of 87.80% and 3.87. CFBR presents more efficacious evaluation measure by JRip for fertility diagnosis with average accuracies and standard deviation of 83.17 % and 9.99, whilst equivalent metrics values for Wdbc by both JRip and RF with average accuracies and standard deviation of 96.20 %, 97.86 % and 2.47, 2.67 respectively. From Table 7, it is obvious that our method is performing much better when compared to the existing approaches, and the best results are achieved for mushroom dataset by RF with average accuracy 99.99% and standard deviation 0.04 %. The increase in overall accuracies and decrease in corresponding standard deviation clearly depicts that the proposed methodology improve the prediction performance of the learning models irrespective of

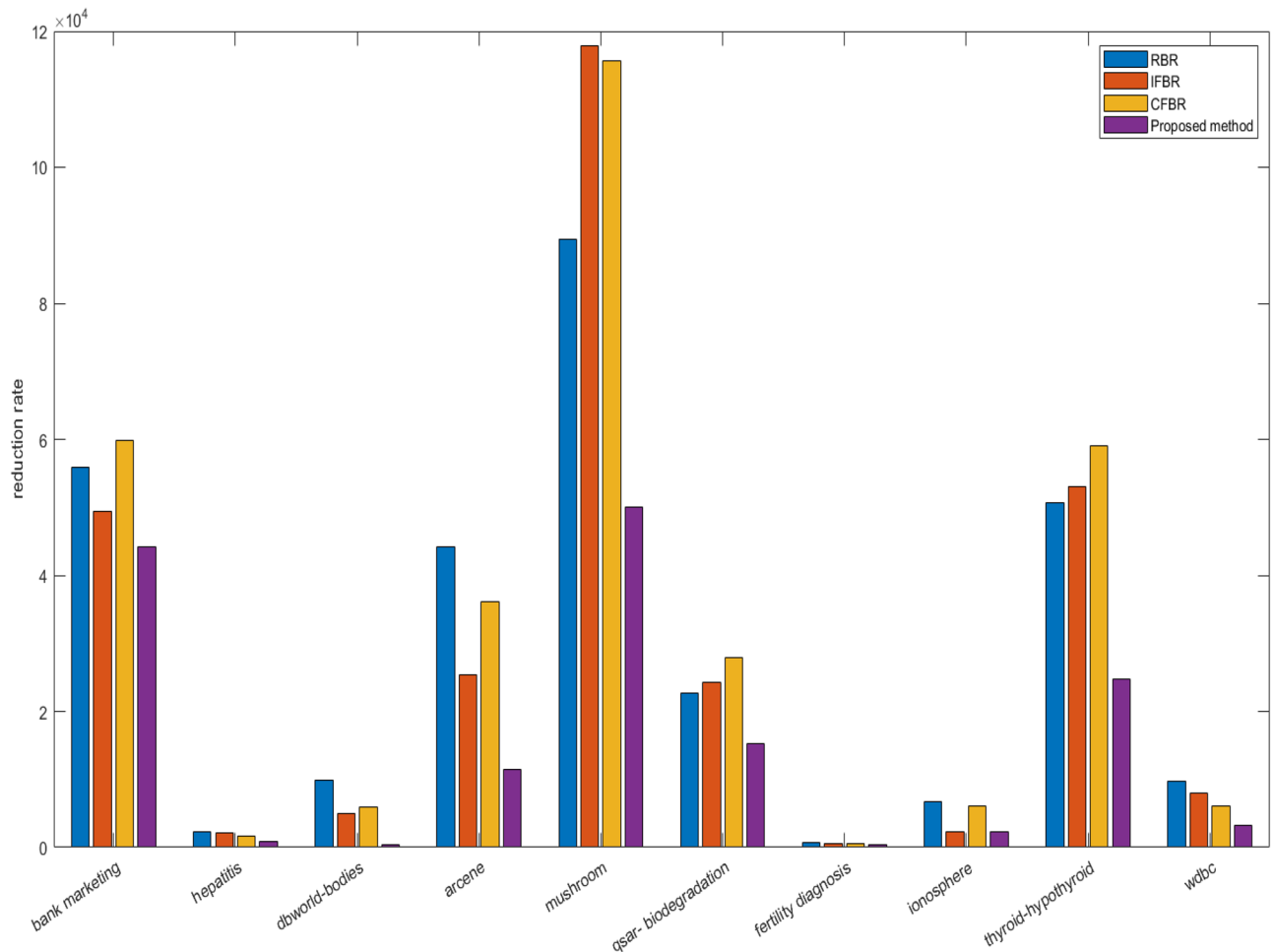


Figure 1. Comparison of overall reduction for different datasets by previous and proposed methods.

reduction rate. Results from Table 6 and Table 7 highlights that our proposed methodology produces the most informative size and dimensions when compared to existing approaches in the literature. These results can be effectively visualized by the given Figs. 2 and 3. The average accuracies produced by different machine learning algorithms have been usually obtained as effective values based on our proposed when compared to the previous approaches are obtained higher irrespective of reduced number of dimensions and size. Now, we apply the widely and effectively explained statistical test namely Freidman test⁷⁹ and Bonferoni Dunn test⁸⁰ to evince the significance of the current approach. Freidman test is exercised to demonstrate multiple hypotheses testing by using F-statistics, which can be illustrated by the following equations:

$$F = \frac{(m-1)Z^2}{m(l-1) - Z^2} \quad (51)$$

and

$$Z^2 = \frac{12m}{l(l+1)} \left(\sum_{i=1}^l r_i^2 - \frac{l(l+1)^2}{4} \right) \quad (52)$$

Here, m and l indicate the count of employed datasets and algorithms, and r_i depict the calculated value of average rank of i th algorithm, which is computed from all the information systems. Further, if the F-statistics with respect to the distribution of Freidman is obtained greater than existing critical value (F_{crit}) at $(l-1)$ and $(l-1)(m-1)$ degrees of freedom, then the null hypothesis which explains that all algorithms are equivalent in the context of average classification accuracy is declared as rejected. In this scenario, Bonferroni Dunn is applied to identify that a particular algorithm is significantly divergent from the presented method. Two methods can be said as significantly divergent at $\alpha\%$ significance level if and only if the distance available between their r_i is larger than a critical distance (Cd_α), which can be outlined by the following expression:

$$Cd_\alpha = q_\alpha \sqrt{\frac{l(l+1)}{6m}} \quad (53)$$

Dataset	Classifier	RBR	IFBR	CFBR	Proposed method
Bank	JRip	86.44±3.22 ⁴	89.65±2.31 ²	88.12±2.32 ³	90.27±1.15 ¹
Marketing	RF	89.23±3.25 ⁴	90.22±3.01 ³	91.59±2.65 ²	94.37±1.01 ¹
Hepatitis	JRip	86.19±6.52 ⁴	89.24±8.31 ¹	87.23±8.11 ³	89.14±7.67 ²
	RF	92.34±5.98 ¹	90.34±7.22 ³	89.44±7.97 ⁴	91.86±7.22 ²
dbworld	JRip	86.41±10.24 ⁴	87.12±11.21 ³	89.21±12.56 ²	91.74±10.08 ¹
Bodies	RF	89.32±9.65 ⁴	90.54±9.98 ³	91.32±10.87 ²	93.93±8.32 ¹
Arcene	JRip	81.27±11.12 ³	80.31±8.78 ⁴	82.21±9.95 ²	84.32±9.07 ¹
	RF	84.27±9.32 ³	84.23±7.98 ⁴	87.65±7.43 ²	90.58±6.99 ¹
Mushroom	JRip	92.68±2.71 ³	91.12±4.37 ⁴	93.89±2.21 ²	99.95±0.13 ¹
	RF	95.44±1.51 ³	95.21±1.86 ⁴	97.28±1.09 ²	99.99±0.04 ¹
qsarbio-degradation	JRip	84.80±3.87 ^{2.5}	85.54±2.54 ¹	82.21±2.89 ⁴	84.80±3.87 ^{2.5}
	RF	89.45±3.23 ³	90.62±3.87 ²	86.29±2.98 ⁴	92.63±2.43 ¹
Fertility diagonsis	JRip	75.59±10.01 ⁴	81.41±10.18 ²	83.17±9.99 ¹	78.44±9.68 ³
	RF	81.43±9.73 ⁴	86.23±11.21 ^{2.5}	86.23±11.21 ^{2.5}	89.04±10.28 ¹
Ionosphere	JRip	91.21±3.95 ⁴	95.55±3.78 ^{1.5}	92.45±4.11 ³	95.55±3.78 ^{1.5}
	RF	95.31±3.01 ⁴	98.12±2.67 ^{1.5}	96.29±3.19 ³	98.12±2.67 ^{1.5}
Thyroid-hypothyroid	JRip	92.41±2.19 ³	91.46±1.65 ⁴	98.22±1.05 ²	99.32±0.47 ¹
	RF	95.41±2.01 ³	93.59±2.86 ⁴	99.11±0.98 ²	99.59±0.36 ¹
wdbc	JRip	95.21±2.98 ^{3.5}	95.21±2.98 ^{3.5}	96.20±2.47 ^{1.5}	96.20±2.47 ^{1.5}
	RF	96.79±3.02 ³	96.28±1.86 ⁴	97.86±2.21 ^{1.5}	97.86±2.21 ^{1.5}
Average	JRip	3.50	2.60	2.35	1.55
Rank	RF	3.20	3.10	2.50	1.20
F statistics	JRip	5.68			
	RF	9.29			

Table 7. Classification accuracy with standard deviation comparison with other state of art bireduct algorithms.

In this expression, q_{α} is a tabular value at $\alpha\%$ significance level.

The test results are recorded in Table 7, where individual rank is given in the superscript of respective results. In this study, the average rank values by our proposed method for all the ten reduced datasets based on both machine learning algorithms namely JRip and RF are computed as 1.55 and 1.20 as reported in Table 7. These values are lower when compared to the average rank values based on other existing techniques, which clearly indicates the effectiveness of the proposed method. Moreover, F-statistics values by using both the learning algorithms based on our proposed technique are calculated as 5.68 and 9.29, whilst F-tabular value at 5% level of significance is $F(3,27) = 2.960$, which exhibits that the null hypothesis is rejected on the basis of Freidman test. Hence, all the four algorithms are obtained as statistically different. Further, by using Bonferoni Dunn test, it can be concluded that our proposed technique is statistically superior to earlier algorithms for both the learning algorithms at 5% significance level.

Here, we have employed the three most prominent state-of-the-art techniques namely RBR, IFBR, and CFBR to perform the comparative performances. RBR struggles with information loss due to discretization process prior to apply the size and dimension reduction. This leads to degrade in the performance of the learning methods. Next, CFBR is an effective bireduct approach to handle the real-valued datasets, but it fails to handle the uncertainty due to both judgement and identification. This technique produces the reduced dataset, which usually guides the overfitting issue for complex data. IFBR technique can effectively handles the issues with RBR and CFBR, but it faces obstacles while handling overlapping available in the real-valued datasets. Moreover, IFBR usually takes more time to accomplish the reduction process. Our proposed method addresses all these obstacles by producing the reduced dataset for the real-valued datasets without performing discretization with the proposed similarity relation. IF set assisted idea helps in coping with uncertainty due to both judgement and identification. Covering concept avoids the overlapping issues during reduction process. The experimental results clearly declare the justification of theoretical proposal of the method. Tabular and pictorial results indicate that our proposed notion is usually providing effective results both in terms of data reduction and performance measures of the machine learning algorithms. Parameterized concepts in the proposed technique allows to discard the issues such redundancy, irrelevancy, and noise available in the large volume of high-dimensional datasets by adjusting the values of the parameters for the datasets from different domains.

Case study: enhanced regression prediction model analysis for IC50 of antiviral peptides

Bioactive peptides known as antiviral peptides (AVPs) have virucidal action and show promise as antiviral medicines. They are peptides of shorter length with the capacity to stop the spread of viral infections. The research

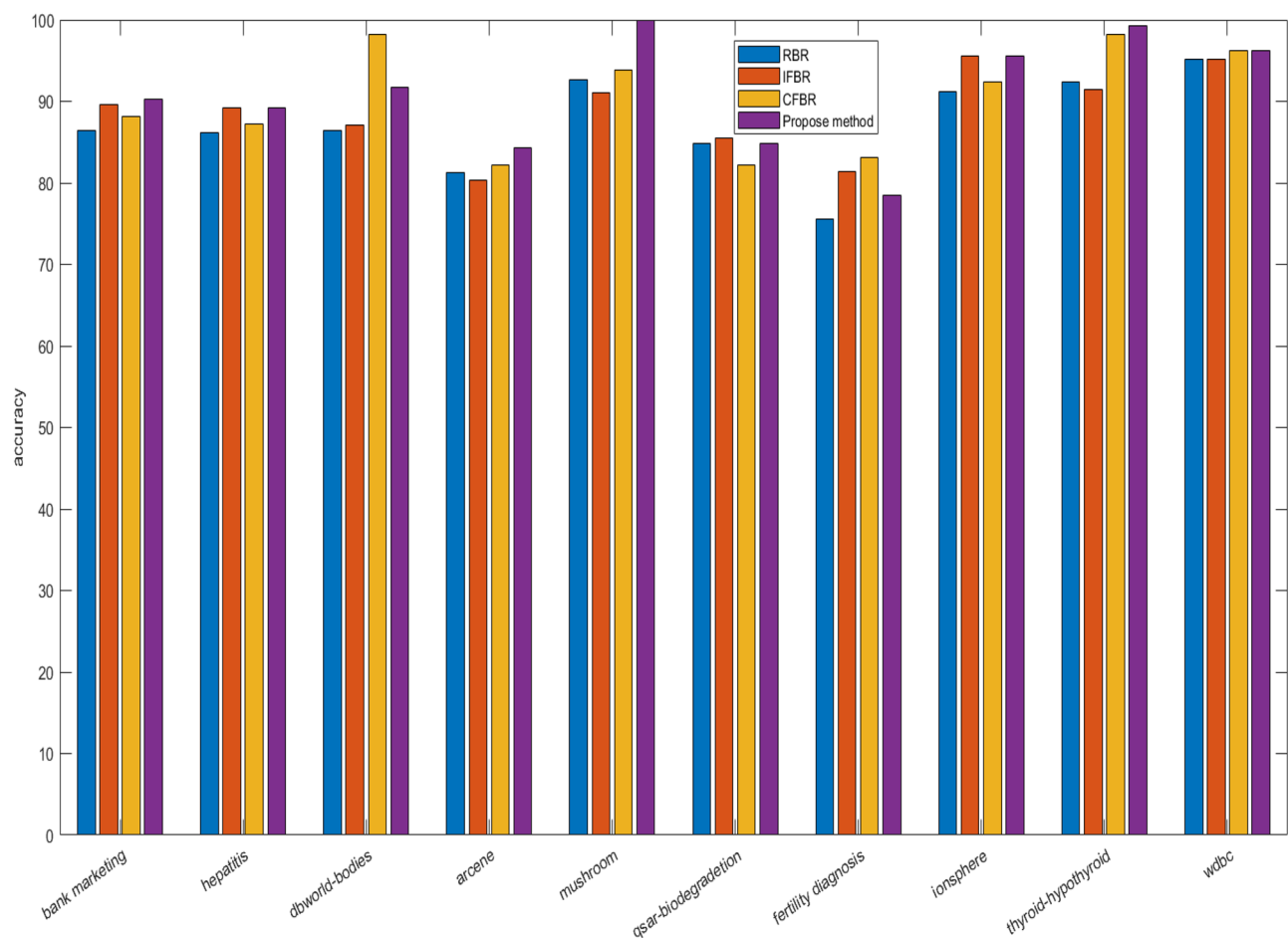


Figure 2. Comparison of average accuracies by JRip for different reduced datasets as produced by existing and proposed methods.

community has recently become interested in the medicinal application of antiviral peptides. For the development of new treatments for viral infections, AVP development and identification are essential. The prediction potency of AVPs measured in IC₅₀ can provide valuable insights into the physicochemical space of high potency AVPs and facilitate in their rationale design. The proposed method is developed for selection of non-redundant features and informative samples for classification problems, but we further experimented with regression based modes IC₅₀ value prediction. We first converted the AVP dataset consisting of IC₅₀ values into a binary class dataset, followed by selection of informative instances and non-redundant features, then conversion of class labels with original IC₅₀ values and further development of regression models on the bireduct dataset.

In Table 9, we have tabulated the classification results after converting the regression problem into binary classification problem by using median IC₅₀ value as the cutoff for generating the binary classes for selection of non-redundant features and informative samples. The rationale for selecting median as the cutoff, is to divide the whole AVP dataset based on potency (measured in terms of IC₅₀) and obtain a balanced dataset. The performance evaluation metrics for all the algorithms are similar. High sensitivity and accuracies values indicate that positive class prediction can be highly improved with the proposed methodology. AUC and MCC values are outlining that the proposed methodology is generating the relatively better predictor. In Table 8, the performance evaluation metrics for the different learning algorithms based on reduced datasets produced by our bi-selection method for the regression problem of predicting pIC₅₀ values are tabulated. GP-PuK, SMO-PuK, RF, and Vote methods surpassed the previously reported best CC of 0.797. The best CC value is obtained by the vote method (CC = 0.8652) as given in Table 8. The dataset consisted of 683 AVPs with IC₅₀ values, in line with the previous method⁸¹, where IC₅₀ values are converted to pIC₅₀ (negative log of half maximal inhibitory concentration) values, which facilitates in training as it compactly compresses the range of IC₅₀ values. The features for the representation of AVPs consisted of SWScore (local alignment scores) which are effective for identifying patches of identical or similar amino acid residues that are conserved locally and regions of significant local similarity and evolutionary information in the form of PSSM (position specific scoring matrices) obtained from PSI-BLAST algorithm (parameters—e-value = 2000, word-size = 2, scoring matrix = PAM30, threshold = 16, gap open = 9, gap extend = 1, and word-size = 15). For better understanding of the different results, ROC curve is given by Fig. 4 to display the performances of machine learning methods. The original dataset for IC₅₀/pIC₅₀ prediction consisted of 449 length feature vector (each Antiviral protein is represented by 449 features, consisting of sequence and evolutionary information). In domains like AVPs IC₅₀ value prediction, the use of lower prediction

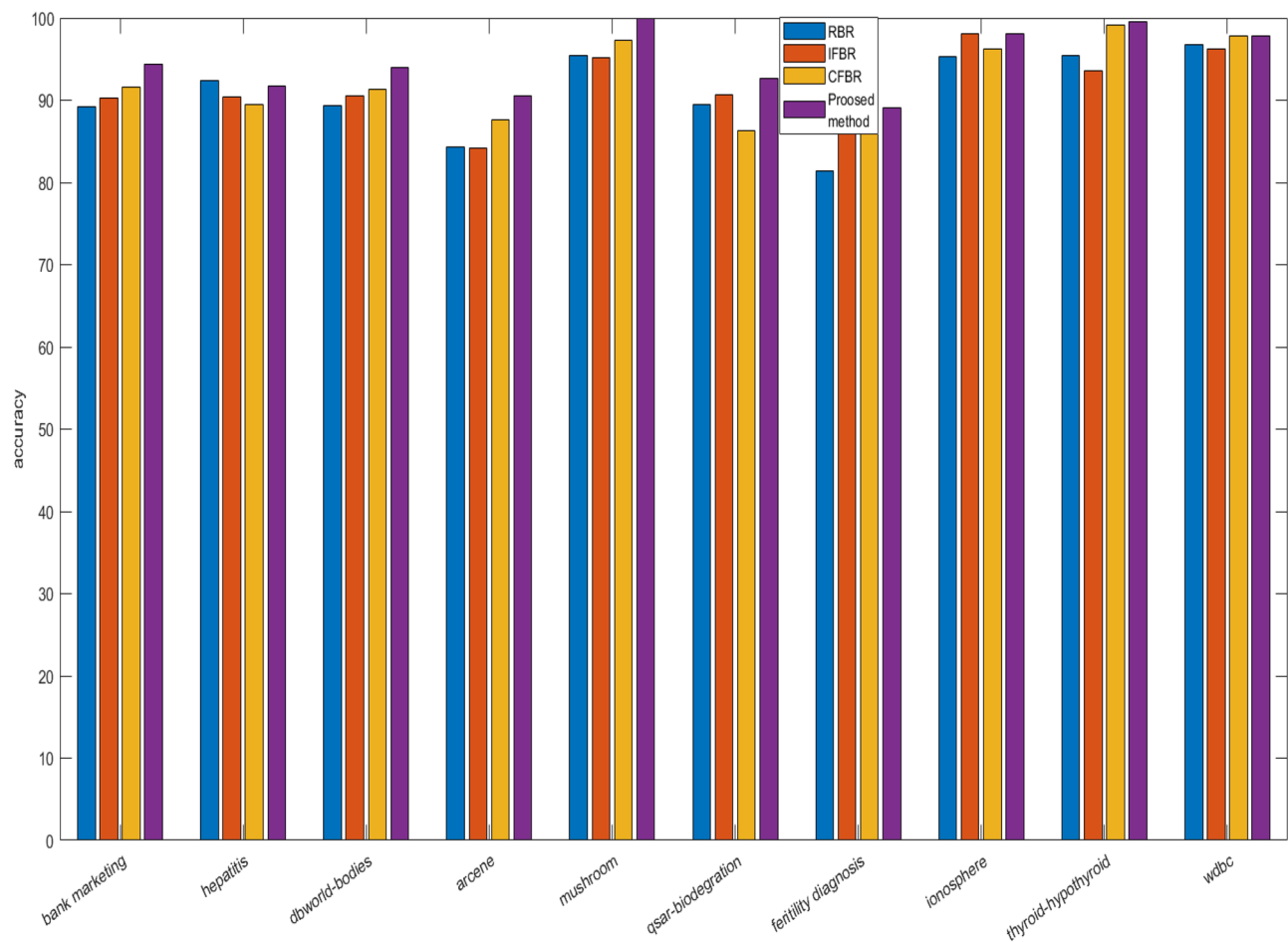


Figure 3. Comparison of average accuracies by RF for different reduced datasets as produced by existing and proposed methods.

Algorithm	CC	MAE
GP-poly kernel	0.7101	0.7284
GP-PUK	0.8318	0.5539
GP-RBF	0.6452	0.8312
SMO-RBF	0.6863	0.7523
SMO-poly kernel	0.7289	0.6709
SMO-PUK	0.8599	0.3632
Decision table	0.7839	0.4977
Linear regression	0.7368	0.6837
IBK	0.7943	0.4116
RF	0.8577	0.4208
Vote (SMO-PUK+RF)	0.8652	0.3835
⁸¹	0.797	NA

Table 8. Performance evaluation metrics for different learning algorithms for pIC50.

accuracy. The vote method consisting of support vector machine and random forest, achieved a value of 0.865 for correlation coefficient in comparison to original dataset where the vote with the same combination achieved 0.797. A rise of 7 points can be observed for the correlation coefficient metric (Table 8). This supports the utility of the proposed approach. The values of CC and MAE depict that our proposed methodology produces the reduced datasets by avoiding redundancy, uncertainty, and noise to enhance the regression performance better than the existing techniques.

Now, we conduct the experimental study in the field of signal processing and bioinformatics to validate the wider perspective of the proposed method as follows:

Learning methods	Sensitivity	Specificity	Accuracy	MCC	AUC
SMO	71.6	90.0	80.6	0.625	0.808
MLP	78.1	83.1	80.6	0.612	0.863
ROF	91.5	79.2	84.7	0.704	0.925
PART	84.4	80.8	82.6	0.652	0.848
J48	84.4	82.3	83.3	0.666	0.843
RF	89.7	83.8	86.7	0.736	0.950
RARF	90.9	85.6	88.2	0.766	0.956

Table 9. Performance evaluation metrics/parameters of different machine learning methods based on tenfold cross validation.

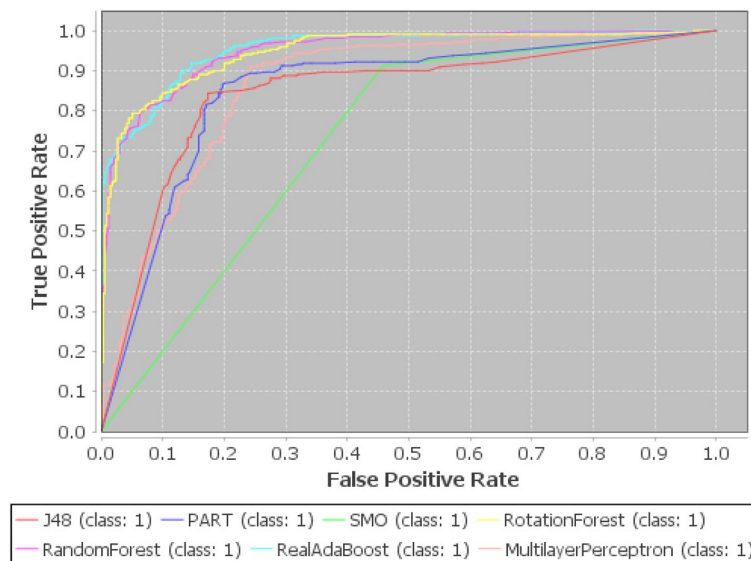


Figure 4. ROC curve for the visualization of performances of different learning algorithms for the binary dataset obtained from AVP dataset.

In the signal processing dataset⁸², the objective is to identify the various forms and manifestations of cardiac arrhythmia and assign it to one of the sixteen groups. There are currently 279 feature values that characterize 452 patient records. Class 01 pertains to normal ECG readings. Class 09 to left bundle branch block, class 10 to right bundle block, class 06 to sinus bradycardia, class 07 to ventricular premature contraction (PVC). Class 02 denotes ischemic alterations (coronary artery disease), class 03 denotes an old anterior myocardial infarction, class 04 denotes an old inferior myocardial infarction, and class 05 denotes sinus tachycardia. Class 11 corresponds to the first degree of AV block, Class 12 to the second degree, Class 13 to the third degree, Class 14 to left ventricle hypertrophy, Class 15 to atrial fibrillation or flutter, and Class 16 to the remaining cases.

In Table 10, the performance evaluation metrics for the various machine learning algorithms are tabulated for the reduced Arrhythmia dataset as produced by our presented models. A rise in all the performance evaluation

Learning methods	Precision	Recall	F-measure	Accuracy	AUC
SMO	84.2	79.5	81.8	79.5	0.826
Rotation forest	86.8	88.3	87.5	88.3	0.915
PART	81.0	81.3	81.2	81.3	0.858
J48	86.1	85.2	85.7	85.2	0.919
Vote	89.9	88.6	89.3	88.6	0.939
Random forest	87.2	86.0	86.6	86.0	0.940

Table 10. Performance evaluation metrics of different machine learning methods for reduced Arrhythmia dataset using 10-fold cross validation.

Learning methods	Precision	Recall	F-measure	Accuracy	AUC
SMO	74.3	72.4	73.3	72.4	0.741
Rotation forest	92.7	92.7	92.7	92.7	0.934
PART	88.3	88.3	88.3	88.3	0.899
J48	88.3	91.2	89.7	91.2	0.916
Vote	95.2	96.3	95.8	96.3	0.989
Random forest	93.7	92.5	93.1	92.5	0.962

Table 11. Performance evaluation metrics of different machine learning methods for reduced dataset consisted of antioxidant and non-antioxidant peptides using tenfold cross validation.

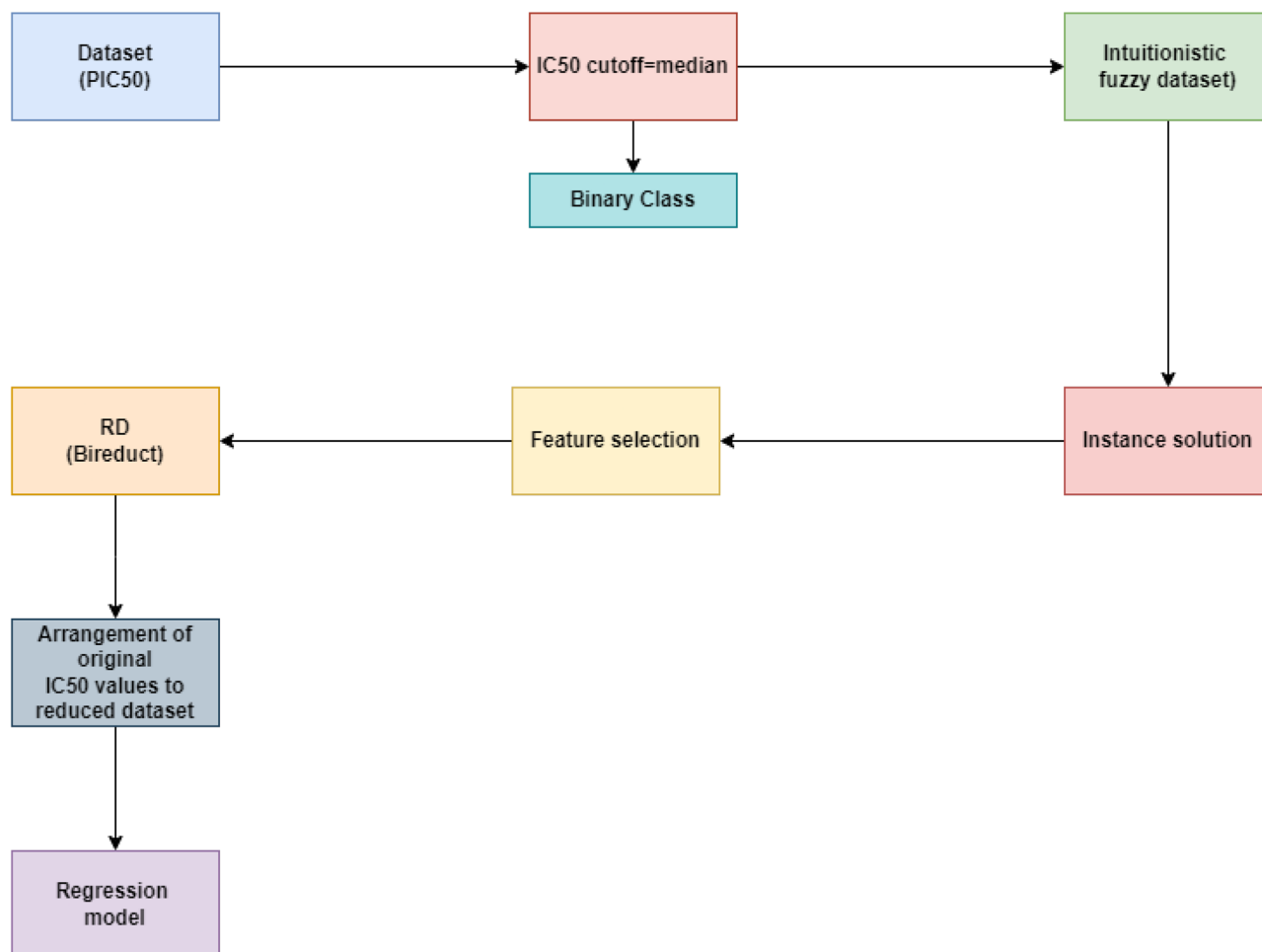


Figure 5. Schematic representation of the entire proposed methodology for regression analysis.

metrics can be observed for all the machine learning algorithms using the reduced dataset. The best performance is achieved by Vote method reaching an accuracy of 88.6% on the reduced dataset (a rise of 18 points in comparison to original dataset). The best values are achieved as precision, recall, F-measure, and accuracy of 89.9%, 88.6%, 89.3%, and 88.6% respectively.

The UniProt database was used to carefully select proteins with antioxidant activity for the dataset. Uncertain proteins, such as those with unusual letters like “B,” “X,” and “Z,” were excluded, and only proteins with antioxidant activity that had been experimentally verified were included. After this comprehensive screening, it was found that 710 protein sequences were the initial positive samples. For the negative samples, 1567 PDB proteins with identification values less than 20 % were chosen using the PISCES-culled approach. To cut down on redundancy and avoid homology bias, peptides that shared more than 60% of similar sequences were removed from the benchmark dataset using the CD-HIT program. Ultimately, 1805 protein sequences were added to the new dataset as negative samples⁸³.

In Table 11, the performance evaluation metrics for the various machine learning algorithms are tabulated for the reduced dataset produced by our proposed models for discriminating the antioxidant and non-antioxidant

peptides. A rise in all the performance evaluation metrics can be observed for all the machine learning algorithms using the reduced dataset. The best performance is achieved by vote method reaching an overall accuracy of 96.3 % on the reduced dataset (a rise of 10 points in comparison to original dataset). An effective improvement in precision, recall, F-measure, and AUC is also reported based on our proposed method. The evaluation metrics values are guiding the conclusion that proposed methodology is producing the effective predictor for biological, signal processing, and bioinformatics data.

Conclusion

The current study combined the notion of IF and rough sets for data reduction both in terms of instances and features. These reductions demonstrate that the reduced data can be produced whilst still maintaining the informative predictive facets, and hence avoid the complexity available in the large volume of real-valued datasets. An IF similarity relation was presented to establish IF similarity relation. Then, we investigated an IF granule based on this relation and thresholds. Relevant propositions were put forward to clarify the effectiveness of the proposed concepts theoretically. Next, instance selection algorithm was addressed by using importance degree of this granular structure, which considers representative instances by removing noise in the data. Thereafter, significance-degree-preserved feature selection idea was introduced, which was employed to eliminate redundant and irrelevant attributes. Further, the entire bireduct was applied over benchmark datasets to perform extensive experiments. From the comparative study with earlier discussed models clearly indicated the dominance of the presented bireduct method. Lastly, a methodology was highlighted as claimed by our proposed method to improve the regression analysis for IC50.

In the future, we intend to establish a new intuitionistic fuzzification of real-valued data to discard the possible information loss due to existing parameterized concept. Moreover, this method can be further improved by providing the generalized computation of the threshold parameters.

Data availability

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

Received: 7 December 2023; Accepted: 13 May 2024

Published online: 12 June 2024

References

- Li, J. *et al.* Feature selection: A data perspective. *ACM Comput. Surv. (CSUR)* **50**(6), 1–45 (2017).
- Olvera-López, J. A., Carrasco-Ochoa, J. A., Martínez-Trinidad, J. F. & Kittler, J. A review of instance selection methods. *Artif. Intell. Rev.* **34**, 133–143 (2010).
- Frank, E., Hall, M., Holmes, G., Kirkby, R., Pfahringer, B., Witten, I. H., & Trigg, L. Weka—a machine learning workbench for data mining. In *Data Mining and Knowledge Discovery Handbook*. 1269–1277 (2010).
- Brighton, H. & Mellish, C. Advances in instance selection for instance-based learning algorithms. *Data Min. Knowl. Discov.* **6**, 153–172 (2002).
- Fu, Y., Zhu, X. & Li, B. A survey on instance selection for active learning. *Knowl. Inf. Syst.* **35**, 249–283 (2013).
- Chandrashekar, G. & Sahin, F. A survey on feature selection methods. *Comput. Electric. Eng.* **40**(1), 16–28 (2014).
- 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).
- Benitez Caballero, M. J., Medina, J., Ramirez Poussa, E. & Ilezak, D. Bireducts with tolerance relations. *Inf. Sci.* **435**, 26–39 (2018).
- Stawicki, S. & Slezak, D. *Recent Advances in Decision Bireducts: Complexity, Heuristics and Streams* (2013).
- De Souza, J. T., Do Carmo, R. A. F. & De Campos, G. A. L. A novel approach for integrating feature and instance selection. In *2008 International Conference on Machine Learning and Cybernetics* (2008).
- Tsai, C.-F., Eberle, W. & Chu, C.-Y. Genetic algorithms in feature and instance selection. *Knowl.-Based Syst.* **39**, 240–247 (2013).
- Benítez-Caballero, M. J., Medina, J. & Ramírez-Poussa, E. *Towards a Classification of Rough Set Bireducts* (2020).
- Rostami, M., Berahmand, K., Nasiri, E. & Forouzandeh, S. Review of swarm intelligence-based feature selection methods. *Eng. Appl. Artif. Intell.* **100**, 104210 (2021).
- Gangavarapu, T. & Patil, N. A novel filter-wrapper hybrid greedy ensemble approach optimized using the genetic algorithm to reduce the dimensionality of high-dimensional biomedical datasets. *Appl. Soft Comput.* **81**, 105538 (2019).
- Xue, B., Zhang, M. & Browne, W. N. New fitness functions in binary particle swarm optimisation for feature selection. In *2012 IEEE Congress on Evolutionary Computation* (2012).
- Lin, S.-W. & Chen, S.-C. PSOLDA: A particle swarm optimization approach for enhancing classification accuracy rate of linear discriminant analysis. *Appl. Soft Comput.* **9**(3), 1008–1015 (2009).
- Cao, L., Li, J., Zhou, Y., Liu, Y. & Liu, H. Automatic feature group combination selection method based on GA for the functional regions clustering in DBS. *Comput. Methods Prog. Biomed.* **183**, 105091 (2020).
- Garci, S., Triguero, I., Carmona, C. J. & Herrera, F. Evolutionary-based selection of generalized instances for imbalanced classification. *Knowl.-Based Syst.* **25**(1), 3–12 (2012).
- Ahmad, S. S. & Pedrycz, W. Feature and instance selection via cooperative PSO. In *2011 IEEE International Conference on Systems, Man, and Cybernetics* (2011).
- Khaire, U. M. & Dhanalakshmi, R. Stability of feature selection algorithm: A review. *J. King Saud Univ.-Comput. Inf. Sci.* **34**(4), 1060–1073 (2022).
- Kuncheva, L. I. & Jain, L. C. Nearest neighbor classifier: Simultaneous editing and feature selection. *Pattern Recognit. Lett.* **20**(11–13), 1149–1156 (1999).
- Jensen, R. & Shen, Q. *Computational Intelligence and Feature Selection: Rough and Fuzzy Approaches* (2008).
- Pawlak, Z. Rough sets. *Int. J. Comput. Inf. Sci.* **11**, 341–356 (1982).
- 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).
- Slezak, D. & Stawicki, S. The problem of finding the simplest classifier ensemble is NP-hard—a rough-set-inspired formulation based on decision bireducts. In *International Joint Conference on Rough Sets* (2020).

26. Verbiest, N., Cornelis, C. & Herrera, F. FRPS: A fuzzy rough prototype selection method. *Pattern Recognit.* **46**(10), 2770–2782 (2013).
27. Zhang, X. & Yao, Y. Tri-level attribute reduction in rough set theory. *Expert Syst. Appl.* **190**, 116187 (2022).
28. 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*. 203–232 (Springer, 1992).
29. Thangavel, K. & Pethalakshmi, A. Dimensionality reduction based on rough set theory: A review. *Appl. Soft Comput.* **9**(1), 1–12 (2009).
30. Sammany, M. & Medhat, T. Dimensionality reduction using rough set approach for two neural networks-based applications. In *Rough Sets and Intelligent Systems Paradigms: International Conference, RSEISP 2007, Warsaw, Poland, June 28–30, 2007. Proceedings* (2007).
31. Ma, X.-A. & Yao, Y. Min-max attribute-object bireducts: On unifying models of reducts in rough set theory. *Inf. Sci.* **501**, 68–83 (2019).
32. Pawlak, Z. Rough set theory and its applications to data analysis. *Cybern. Syst.* **29**(7), 661–688 (1998).
33. 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).
34. Sheeja, T. & Kuriakose, A. S. A novel feature selection method using fuzzy rough sets. *Comput. Indus.* **97**, 111–116 (2018).
35. Zadeh, L. A. Fuzzy sets. *Inf. Control* **8**(3), 338–353 (1965).
36. Tsang, E. C., Song, J., Chen, D. & Yang, X. Order based hierarchies on hesitant fuzzy approximation space. *Int. J. Mach. Learn. Cybern.* **10**, 1407–1422 (2019).
37. Yin, T. *et al.* Exploiting feature multi-correlations for multilabel feature selection in robust multi-neighborhood fuzzy β covering space. *Inf. Fusion* **104**, 102150 (2024).
38. Jensen, R. & Shen, Q. Fuzzy-rough sets assisted attribute selection. *IEEE Trans. Fuzzy Syst.* **15**(1), 73–89 (2007).
39. Jensen, R. & Shen, Q. Fuzzy-rough attribute reduction with application to web categorization. *Fuzzy Sets Syst.* **141**(3), 469–485 (2004).
40. Chen, D., Zhang, L., Zhao, S., Hu, Q. & Zhu, P. A novel algorithm for finding reducts with fuzzy rough sets. *IEEE Trans. Fuzzy Syst.* **20**(2), 385–389 (2011).
41. Bhatt, R. B. & Gopal, M. On fuzzy-rough sets approach to feature selection. *Pattern Recognit. Lett.* **26**(7), 965–975 (2005).
42. Mieszkowicz-Rolka, A. & Rolka, L. Variable precision fuzzy rough sets. In *Transactions on Rough Sets I* (Peters-Andrzej Skowron, J.F. Ed.). 144–160. (Springer, 2004).
43. Dai, J., Huang, W., Zhang, C. & Liu, J. Multi-label feature selection by strongly relevant label gain and label mutual aid. *Pattern Recognit.* **145**, 109945 (2024).
44. Zhao, S., Tsang, E. C. & Chen, D. The model of fuzzy variable precision rough sets. *IEEE Trans. Fuzzy Syst.* **17**(2), 451–467 (2009).
45. Wang, C. *et al.* A fitting model for feature selection with fuzzy rough sets. *IEEE Trans. Fuzzy Syst.* **25**(4), 741–753 (2016).
46. Chen, J., Mi, J. & Lin, Y. A graph approach for fuzzy-rough feature selection. *Fuzzy Sets Syst.* **391**, 96–116 (2020).
47. Jensen, R. & Cornelis, C. Fuzzy-rough instance selection. In *International Conference on Fuzzy Systems*. 1–7. (IEEE, 2010).
48. Lin, Y., Li, Y., Wang, C. & Chen, J. Attribute reduction for multi-label learning with fuzzy rough set. *Knowl.-Based Syst.* **152**, 51–61 (2018).
49. Jain, P., Tiwari, A. & Som, T. Fuzzy rough assisted missing value imputation and feature selection. *Neural Comput. Appl.* **35**(3), 2773–2793 (2023).
50. Som, T., Shreevastava, S., Tiwari, A. K. & Singh, S. Fuzzy rough set theory-based feature selection: A review. In *Mathematical Methods in Interdisciplinary Sciences*. 145–166 (2020).
51. Wang, C., Qian, Y., Ding, W. & Fan, X. Feature selection with fuzzy-rough minimum classification error criterion. *IEEE Trans. Fuzzy Syst.* **30**(8), 2930–2942 (2021).
52. 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 Recognit.* **56**, 1–15 (2016).
53. Jensen, R. & Shen, Q. New approaches to fuzzy-rough feature selection. *IEEE Trans. Fuzzy Syst.* **17**(4), 824–838 (2008).
54. Ji, W. *et al.* Fuzzy rough sets and fuzzy rough neural networks for feature selection: A review. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **11**(3), e1402 (2021).
55. Mansouri, P., Asady, B. & Gupta, N. The bisection-artificial bee colony algorithm to solve fixed point problems. *Appl. Soft Comput.* **26**, 143–148 (2015).
56. Anaraki, J. R., Samet, S., Eftekhari, M. & Ahn, C. W. A fuzzy-rough based binary shuffled frog leaping algorithm for feature selection. arXiv preprint [arXiv:1808.00068](https://arxiv.org/abs/1808.00068) (2018).
57. Huang, B., Li, H.-X. & Wei, D.-K. Dominance-based rough set model in intuitionistic fuzzy information systems. *Knowl.-Based Syst.* **28**, 115–123 (2012).
58. Wang, Z., Chen, H., Yuan, Z. & Li, T. Fuzzy-rough hybrid dimensionality reduction. *Fuzzy Sets Syst.* **459**, 95–117 (2023).
59. Liu, Z. & Pan, S. Fuzzy-rough instance selection combined with effective classifiers in credit scoring. *Neural Process. Lett.* **47**, 193–202 (2018).
60. Mac Parthaláin, N., Jensen, R. & Diao, R. Fuzzy-rough set bireducts for data reduction. *IEEE Trans. Fuzzy Syst.* **28**(8), 1840–1850 (2019).
61. Zhang, X., Mei, C., Li, J., Yang, Y. & Qian, T. Instance and feature selection using fuzzy rough sets: A bi-selection approach for data reduction. *IEEE Trans. Fuzzy Syst.* (2022).
62. 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).
63. Mizumoto, M. & Tanaka, K. Fuzzy sets and their operations. *Inf. Control* **48**(1), 30–48 (1981).
64. Atanassov, K. T. & Stoeva, S. Intuitionistic fuzzy sets. *Fuzzy Sets Syst.* **20**(1), 87–96 (1986).
65. Jain, P., Tiwari, A. K. & Som, T. An intuitionistic fuzzy bireduct model and its application to cancer treatment. *Comput. Indus. Eng.* **168**, 108124 (2022).
66. 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).
67. Xue, Z. & a., Zhao, L., Sun, L., Zhang, M., & Xue, T. Three-way decision models based on multigranulation support intuitionistic fuzzy rough sets. *Int. J. Approx. Reas.* **124**, 147–172 (2020).
68. Zhang, X., Zhou, B. & Li, P. A general frame for intuitionistic fuzzy rough sets. *Inf. Sci.* **216**, 34–49 (2012).
69. Zhou, L., Wu, W.-Z. & Zhang, W.-X. On intuitionistic fuzzy rough sets and their topological structures. *Int. J. Gen. Syst.* **38**(6), 589–616 (2009).
70. Tan, A. *et al.* Intuitionistic fuzzy rough set-based granular structures and attribute subset selection. *IEEE Trans. Fuzzy Syst.* **27**(3), 527–539 (2018).
71. 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).
72. 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).

73. Tan, A., Shi, S., Wu, W.-Z., Li, J. & Pedrycz, W. Granularity and entropy of intuitionistic fuzzy information and their applications. *IEEE Trans. Cybern.* **52**(1), 192–204 (2020).
74. 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).
75. Jain, P., Tiwari, A. K. & Som, T. A fitting model based intuitionistic fuzzy rough feature selection. *Eng. Appl. Artif. Intell.* **89**, 103421 (2020).
76. Asuncion, A. & Newman, D. *UCI Machine Learning Repository* (2007).
77. 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).
78. Hall, M. *et al.* The WEKA data mining software: An update. *ACM SIGKDD Explor. Newslett.* **11**(1), 10–18 (2009).
79. Friedman, M. A comparison of alternative tests of significance for the problem of m rankings. *Ann. Math. Stat.* **11**(1), 86–92 (1940).
80. Dunn, O. J. Multiple comparisons among means. *J. Am. Stat. Assoc.* **56**(293), 52–64 (1961).
81. Nath, A. Prediction for understanding the effectiveness of antiviral peptides. *Comput. Biol. Chem.* **95**, 107588 (2021).
82. Guvenir, H. A., Acar, B., Demiroz, G. & Cekin, A. A supervised machine learning algorithm for arrhythmia analysis. In *Computers in Cardiology*. 433–436. (IEEE, 1997).
83. Shao, L. *et al.* Identification of antioxidant proteins with deep learning from sequence information. *Front. Pharmacol.* **9**, 1036 (2018).

Author contributions

Rajat Saini: Conceptualization, Problem formulation, Methodology, Original draft preparation, Reviewing and Editing, and Final drafting. Anoop Kumar Tiwari: Data curation, Programming, Simulation, Validation, Numerical analysis, Visualization, and System set-up. Abhigyan Nath: Programming, Writing, Validation, Simulation, Editing, Numerical analysis. Phool Singh: Mathematical Modelling, Visualization, and Investigation. S.P. Maurya: Revision, Programming, Reviewing, and Editing, Numerical analysis. Mohd Asif Shah: Supervision, Programming, Validation, Writing, Reviewing, and Editing.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to A.K.T. 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.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024