Chapter 4

A Fitting Model for Feature Selection

Feature subset selection is an essential machine learning approach aimed at the process of dimensionality reduction of the input space. By removing irrelevant and/or redundant variables, not only it enhances model performance, but also facilitates its improved interpretability. However, the concepts discussed so far can only maintain a maximal dependency function. It cannot preferably illustrate the differences in object classification and does not fit a particular data set well. This problem was handled by using a fitting model for feature selection with fuzzy rough sets. However, intuitionistic fuzzy set theory can deal with uncertainty in a much better way when compared to fuzzy set theory as it considers positive, negative and hesitancy degree of an object simultaneously to belong to a particular set.

4.1 Fitting Model based on Intuitionistic Fuzzy Rough Set

For an information system (U, C, D), the set of samples partitions the decisions into k crisp equivalence classes $U \setminus D = \{D_1, D_2, \ldots, D_k\}$. Then, intuitionistic fuzzy decision of $x \in U$ is defined as follows:

$$\mathbb{D}_i(x) = (\frac{\mid \mu_{[x]_A} \cap D_i \mid}{\mid \mu_{[x]_A} \mid}, \frac{\mid \nu_{[x]_A} \cap D_i \mid}{\mid \nu_{[x]_A} \mid}), i = 1, 2, \ldots k \tag{4.1}$$

where $\mathbb{D}_i(x)$ is an intuitionistic fuzzy set and it indicates the degree of membership and non membership of x to decision class D_i. Obviously, $\{\mathbb{D}_1(x), \mathbb{D}_2(x), \ldots, \mathbb{D}_k(x)\}$ is a intuitionistic fuzzy partition of U.

Let R_a be intuitionistic fuzzy similarity class of samples induced by attribute a, then for any set $A \subseteq C$, intuitionistic fuzzy relation is given by:

$$R_A(x, y) = (\cap_{a \in A} R_a(x, y), \cup_{a \in A} R_a(x, y)), x, y \in U \tag{4.2}$$

Different levels of granularity, acquired from every intuitionistic fuzzy similarity, lead to more classification information. Optimal feature subset is obtained by choosing granularity [22, 62] that leads to optimized accuracy. The $R_A(x, y)$ between sample x and y denotes the similarity between sample based on their membership value and dissimilarity between their non membership value. To remove the impact of noise, low value of R_A can be equated to zero, considering small value being resulted due to noise.

Parameterized intuitionistic fuzzy granule is constructed to achieve this, by introducing $\epsilon \in [0, 1)$ to avoid noise as follows:

$$\mu_{[x]_A^\epsilon}(y) = \begin{cases} 0, & \mu_{R_A}(x, y) < \epsilon \\ \mu_{R_A}(x, y), & \mu_{R_A}(x, y) \geq \epsilon \end{cases}, y \in U$$

$$\nu_{[x]_A^\epsilon}(y) = \begin{cases} 0, & \nu_{R_A}(x, y) < \epsilon \\ \nu_{R_A}(x, y), & \nu_{R_A}(x, y) \geq \epsilon \end{cases}, y \in U$$

Clearly, it can be seen that ϵ impacts the size of intuitionistic fuzzy granule. Therefore, intuitionistic fuzzy similarity is denoted by R_A^ϵ. It is derived from above the following [154]:

Proposition 4.1.1. If $A \subseteq B$, then $R_B^\epsilon \subseteq R_A^\epsilon$

The lower and upper approximation of decision D with respect to attribute A is given by:

$$R^\epsilon \downarrow_A D_i(x) = \begin{cases} (min_{y \in U} & max(1 - \mu_{R_A^\epsilon}(x, y), \mu_{\mathbb{D}_i}(y)), \\ & max_{y \in U} min(1 - \nu_{R_A^\epsilon}(x, y), \nu_{\mathbb{D}_i}(y))), y \in D_i \\ (0, 1), & otherwise \end{cases} \tag{4.3}$$

$$R^\epsilon \uparrow_A D_i(x) = \begin{cases} (max_{y \in U} & min(\mu_{R_A^\epsilon}(x, y), \mu_{\mathbb{D}_i}(y)), \\ & min_{y \in U} max(\nu_{R_A^\epsilon}(x, y), \nu_{\mathbb{D}_i}(y)), y \in D_i \\ (0, 1), & otherwise \end{cases} \tag{4.4}$$

Similar to classical intuitionistic fuzzy rough sets, $R^\epsilon \downarrow_A D_i(x)$ denotes the degree of certainty with which sample x belong to category D_i and $R^\epsilon \uparrow_A D_i(x)$ indicates the possibility of x belonging to category D_i.

Intuitionistic fuzzy positive region is calculated using above defined lower approximation, given by:

$$Pos_A^\epsilon(x) = (max_i \mu_{R^\epsilon \downarrow_A D_i}(x), min_i \nu_{R^\epsilon \downarrow_A D_i}(x)) \tag{4.5}$$

Greater is the size of positive region, the more is the dependency of sample x on feature subset A for its classification. Thereby, dependency degree of attribute A is obtained using formula:

$$\gamma_A = \frac{\sum_{x \in U} \mid Pos_A^{\epsilon}(x) \mid}{\mid U \mid} \tag{4.6}$$

The aim is to find feature subset with maximum dependency degree, as misclassification error is smaller in such case. We have the following result in this context.

Theorem 4.1.1. Given (U, C, D) and $0 < \epsilon < 1$, if $A_1 \subseteq A_2 \subseteq C$, then $Pos_{A_1}^{\epsilon}(D) \subseteq Pos_{A_2}^{\epsilon}(D)$

Proof. From Proposition 4.1.1, $R_{A_2}^{\epsilon} \subseteq R_{A_1}^{\epsilon}$, whenever $A_1 \subseteq A_2 \implies 1 - \mu_{R_{A_2}^{\epsilon}}(x, y) \geq 1 - \mu_{R_{A_1}^{\epsilon}}(x, y)$ and $1 - \nu_{R_{A_2}^{\epsilon}}(x, y) \leq 1 - \nu_{R_{A_1}^{\epsilon}}(x, y), \forall y \in U \implies \mu_{R^{\epsilon}\downarrow_{A_1} D_i}(x) \leq \mu_{R^{\epsilon}\downarrow_{A_2} D_i}(x)$ and $\nu_{R^{\epsilon}\downarrow_{A_1} D_i}(x) \geq \nu_{R^{\epsilon}\downarrow_{A_2} D_i}(x)$, then from definition of lower approximation $\implies R^{\epsilon} \downarrow_{A_1} D(x) \leq R^{\epsilon} \downarrow_{A_2} D(x) \implies Pos_{A_1}^{\epsilon}(D) \subseteq Pos_{A_2}^{\epsilon}(D)$. □

Theorem 4.1.2. Given (U, C, D) and $0 < \epsilon < 1$, if $A_1 \subseteq A_2 \subseteq \cdots \subseteq A_m \subseteq C$, then $\gamma_{A_1}^{\epsilon}(D) \subseteq \gamma_{A_2}^{\epsilon}(D) \subseteq \cdots \subseteq \gamma_{A_m}^{\epsilon}(D) \subseteq C$.

Proof. Obvious from above. □

The above theorem shows that with increase in size of subset, dependency also increases. This guarantees that adding attribute to existing feature set will increase dependency of the new subset obtained. If dependency does not increase on adding an attribute B to feature subset, then that attribute is redundant and can be removed as superfluous attribute, otherwise B is indispensable and cannot be removed. A feature subset *Red* is a reduct set if it has same dependency as a whole set of attributes and removing an attribute decreases its dependency.

4.2 Feature Selection based on Fitting Model

In this section, a greedy forward algorithm for feature selection is proposed. The algorithm begins with empty set and iteratively adds attribute to the set with maximum dependency until dependency increases further. It has been abbreviated as FMIFRFS.

Algorithm 4.2 Heuristic algorithm based on FMIFRFS

Input intuitionistic fuzzy information system

Find the intuitionistic fuzzy decision classes $U \setminus D = \{\mathbb{D}_1, \mathbb{D}_2, \ldots, \mathbb{D}_k\}$

Initialize $C = \{a_1, a_2, \ldots, a_m\}, Red = \{\}$

repeat

Set $T \leftarrow Red$

for every $a \in C - Red$ **do**

Compute intuitionistic fuzzy similarity $R^{\epsilon}_{Red \cup \{a\}}$

Compute lower approximation $R^{\epsilon} \downarrow_{Red \cup \{a\}} D(x)$, for each $x \in U$

Calculate degree of dependency $\gamma_{Red \cup \{a\}}$

end for

Find attribute $a \in C - Red$ with greatest $\gamma_{Red \cup \{a\}}(D)$ and set $Red \leftarrow T \cup \{a\}$

until $\gamma_{Red} = 1$ or $\gamma_T(D) = \gamma_{Red}(D)$

return *Red*

The proposed algorithm is illustrated using example dataset given in 4.1.

Firstly, dataset is normalized into interval $[0, 1]$, then the normalized values are converted into intuitionistic fuzzy values. Finally, intuitioinistic fuzzy similarity r_{ij} between x and y is obtained using formula in equation (4.7).

$$r_{ij}(x, y) = (1 - \frac{1}{num}\sqrt{\sum_{i=1}^{num}(\mu(x) - \mu(y))^2}, \frac{1}{num}\sqrt{\sum_{i=1}^{num}(\nu(x) - \nu(y))^2}) \tag{4.7}$$

TABLE 4.1: Example dataset

Features / **Instances**	a_1	a_2	a_3	a_4	a_5	**D**
x_1	0.08	0.08	0.1	0.24	0.9	4
x_2	0.06	0.06	0.05	0.25	0.33	2
x_3	0.1	0.1	0.15	0.65	0.3	3
x_4	0.08	0.08	0.08	0.98	0.24	2
x_5	0.09	0.15	0.4	0.1	0.66	3
x_6	0.15	0.02	0.34	0.4	0.01	1
x_7	0.24	0.75	0.32	0.18	0.86	4
x_8	0.276	0.225	0.81	0.27	0.33	2

TABLE 4.2: Similarity relation obtained from example dataset for attribute a_1

	x_1	x_2	x_3	x_4	x_5	x_6	x_5	x_6
x_1	(1, 0)	(0.85, 0.15)	(0.88, 0.19)	(1, 0)	(0.93, 0.09)	(0.81, 0.41)	(0.63, 0.30)	(0.42, 0.37)
x_2	(0.85, 0.15)	(1, 0)	(0.74, 0.34)	(0.85, 0.15)	(0.79, 0.24)	(0.67, 0.56)	(0.77, 0.14)	(0.56, 0.22)
x_3	(0.88, 0.19)	(0.74, 0.34)	(1, 0)	(0.88, 0.19)	(0.94, 0.10)	(0.93, 0.21)	(0.51, 0.49)	(0.30, 0.56)
x_4	(1, 0)	(0.85, 0.15)	(0.88, 0.19)	(1, 0)	(0.93, 0.09)	(0.81, 0.41)	(0.63, 0.30)	(0.42, 0.37)
x_5	(0.93, 0.09)	(0.79, 0.24)	(0.94, 0.10)	(0.93, 0.09)	(1, 0)	(0.87, 0.32)	(0.56, 0.39)	(0.35, 0.46)
x_6	(0.81, 0.41)	(0.67, 0.56)	(0.93, 0.21)	(0.81, 0.41)	(0.87, 0.32)	(1, 0)	(0.44, 0.71)	(0.23, 0.78)
x_7	(0.63, 0.30)	(0.77, 0.14)	(0.51, 0.49)	(0.63, 0.30)	(0.56, 0.39)	(0.44, 0.71)	(1, 0)	(0.78, 0.07)
x_6	(0.42, 0.37)	(0.56, 0.22)	(0.30, 0.56)	(0.42, 0.37)	(0.35, 0.46)	(0.23, 0.78)	(0.78, 0.07)	(1, 0)

TABLE 4.3: Granularity $[x]^{\epsilon}_{a_1}$ obtained from example dataset for attribute a_1

	x_1	x_2	x_3	x_4	x_5	x_6	x_5	x_6
x_1	(1, 0)	(0.85, 0)	(0.88, 0)	(1, 0)	(0.93, 0)	(0.81, 0)	(0, 0)	(0, 0)
x_2	(0.85, 0)	(1, 0)	(0.74, 0)	(0.85, 0)	(0.79, 0)	(0, 0)	(0.77, 0)	(0, 0)
x_3	(0.88, 0)	(0.74, 0)	(1, 0)	(0.88, 0)	(0.94, 0)	(0.93, 0)	(0, 0)	(0, 0)
x_4	(1, 0)	(0.85, 0)	(0.88, 0)	(1, 0)	(0.93, 0)	(0.81, 0)	(0, 0)	(0, 0)
x_5	(0.93, 0)	(0.79, 0)	(0.94, 0)	(0.93, 0)	(1, 0)	(0.87, 0)	(0, 0)	(0, 0)
x_6	(0.81, 0)	(0, 0)	(0.93, 0)	(0.81, 0)	(0.87, 0)	(1, 0)	(0, 0.71)	(0, 0.78)
x_7	(0, 0)	(0.77, 0)	(0, O)	(0, 0)	(0, 0)	(0, 0.71)	(1, 0)	(0.78, 0)
x_6	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0.78)	(0.78, 0)	(1, 0)

where $\mu(x)$ and $\nu(x)$ are membership and non membership degree, respectively of an instance x to the attribute set A and num is the number of attributes in set A. Hence, the similarity relation for attribute a_1 is given in table 4.2.

Thereby, the granularity is obtained as in table 4.3 using $\epsilon = 0.7$.

The decision attribute partitions decision class into four sets as:

TABLE 4.4: Intuitionistic fuzzy decision matrix

	$\mathbb{D}_1$	$\mathbb{D}_2$	$\mathbb{D}_3$	$\mathbb{D}_4$
x_1	(0.14, na)	(0.33, na)	(0.33, na)	(0.18, na)
x_2	(0.37, na)	(0.36, na)	(0.30, na)	(0.32, na)
x_3	(0.17, na)	(0.30, na)	(0.36, na)	(0.16, na)
x_4	(0.14, na)	(0.33, na)	(0.33, na)	(0.18, na)
x_5	(0.15, na)	(0.31, na)	(0.35, na)	(0.17, na)
x_6	(0.22, 0)	(0.18, 52)	(0.40, 0)	(0.18, 0.47)
x_7	(0, 1)	(0.60, 0)	(0, 0)	(0.39, 0)
x_6	(0, 1)	(0.55, 0)	(0, 0)	(0.44, 0)

TABLE 4.5: Lower approximation obtained from example dataset for attribute a_1

	$R^\epsilon \downarrow_{a_1} D_1$	$R^\epsilon \downarrow_{a_1} D_2$	$R^\epsilon \downarrow_{a_1} D_3$	$R^\epsilon \downarrow_{a_1} D_4$
x_1	(0, 1)	(0, 1)	(0, 1)	(0.16, 1)
x_2	(0, 1)	(0.30, 1)	(0, 1)	(0, 1)
x_3	(0, 1)	(0, 1)	(0.30, 1)	(0, 1)
x_4	(0, 1)	(0.18, 1)	(0, 1)	(0, 1)
x_5	(0, 1)	(0, 1)	(0.30, 1)	(0, 1)
x_6	(0.15, 1)	(0, 1)	(0, 1)	(0, 1)
x_7	(0, 1)	(0, 1)	(0, 1)	(0.32, 1)
x_6	(0, 1)	(0.55, 1)	(0, 1)	(0, 1)

$U \setminus D = \{D_1, D_2, D_3, D_4\} = \{\{x_6\}, \{x_2, x_4, x_8\}, \{x_3, x_5\}, \{x_1, x_7\}\}$. Intuitionistic fuzzy decision matrix is obtained as given in table 4.4.

Some of non-membership values of decision matrix are 'na' as the corresponding non membership value is 0 in $[x]^\epsilon_{a_1}$. Thereby, lower approximation is obtained as in table 4.5.

Now, degree of dependency of decision attribute over a_1 is calculated by proposed concept as $\gamma_{a_1} = 0.1441$. Similarly, degrees of dependencies of decision attribute

over other conditional attributes are:

$$\gamma_{a_2} = 0.1928$$

$$\gamma_{a_3} = 0.1676$$

$$\gamma_{a_4} = 0.1751$$

$$\gamma_{a_5} = 0.2621$$

Therefore a_4 is selected as the potential reduct set. Combining with other attributes, this process iterates and after termination of algorithm, the reduct set is $\{a_1, a_4, a_5\}$.

4.3 Experimentation

In the current study, the performance of the proposed model is evaluated and compared with existing fitting model based on fuzzy-rough feature selection [154] (FM-FRFS). Firstly, the dataset is fuzzified using simple algorithm. Then, fuzzified data is converted to intitionistic fuzzy dataset. These algorithms employed forward search to obtain optimal feature subset. The intuitionistic fuzzy similarity r_{ij} between instances x and y is computed by:

$$r_{ij} = (1- \mid \mu(x) - \mu(y) \mid, \mid \nu(x) - \nu(y) \mid) \tag{4.8}$$

where $\mu(x)$ and $\nu(x)$ are membership and non membership degree, respectively of an instance x to a set. Twelve benchmark datasets from the UCI Repository [108] is used to represent the performance of the proposed approach. The details of these datasets are mentioned in table 4.6. Further, the choice of ϵ depends on the amount of noise present in the dataset. The value of ϵ is varied from 0.1 to 0.9 in a small interval, and the value of ϵ giving highest classification accuracy is selected.

Three different machine learning algorithms namely PART [49], JRip [26] and J48 [120] are used for the purpose of evaluating classification accuracy using full dataset. While kNN (k = 3) [131] and SVM [117] were employed to test performance on

dataset using 10-fold cross validation.

In table 4.6, the size of the reduct set produced by FMFRFS as well as FMIFRFS using full training set and 10-fold cross validation technique respectively are also recorded. Overall classification accuracies along with standard deviation are evaluated by using PART, JRip, J48 for both original datasets and reduced datasets as produced by FMFRFS and FMIFRFS on full training sets, as recorded in table 4.7. Moreover, average classification accuracies along with standard deviation are again evaluated by using kNN (k=3), SVM for both original datasets and reduced datasets as produced by FMFRFS and FMIFRFS on 10-fold cross validation, as depicted in table 4.8. From the experimental results, it can be observed that the current technique usually provides smaller subset of features than existing method. For some of the datasets, FMIFRFS produces larger subsets when compared with FMFRFS but these reduct sets are more accurate as the performance of different learning algorithms for these sets are better when compared with FMFRFS based reduct sets. From the experiments, it can be observed that the average accuracies of different classifiers for the reduced datasets produced by FMIFRFS is always more than those of reduced datasets produced by FMFRFS and the values of standard deviation are vice-versa. Wang et. al. [154] has revealed that FMFRFS is better performing approach than other existing feature selection techniques. Therefore, the proposed approach outperforms all the existing approach till date.

Variation of classification accuracy and reduct size with noise parameter ϵ is depicted in figure 4.1, which is obtained by using 10-fold cross validation by conducting series of experiments.

TABLE 4.6: Dataset characteristics and reduct size

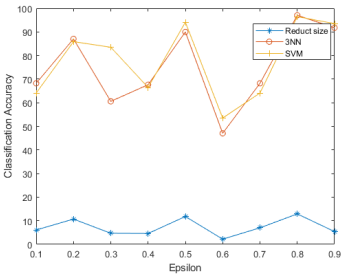
Dataset	Instances	Attributes	Reduct size			
			FMFRFS		Proposed algorithm	
			Full training set	10 fold cross validation	Full training set	10 fold cross validation
Wine	178	13	2	11.9	6	12.9
Heart	267	13	2	6.6	7	11.4
Ionosphere	351	34	2	11.3	9	8.7
Balloon Scale+Stretch	20	4	2	3.3	2	2
Balloon Scale-Stretch	20	4	2	2.9	2	2
Dbworld-bodies-stemmed	64	3721	4	3.7	6	5.2
Dbworld-bodies	64	4702	4	3.7	6	17
Cardiotocography-3class	2126	35	2	2	6	12
Lung Cancer	32	56	9	9.5	5	5.3
Soyabean small	47	21	4	4	2	2
Trains	10	26	6	3	3	3
Zoo	10	16	2	8.2	8	6.8

TABLE 4.7: Comparison of classification accuracies for original datasets and reduced datasets by proposed model, and FMFRFS using full training

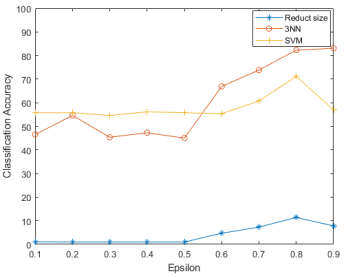
Dataset	Original			FMFRFS			Proposed model		
	JRip	PART	J48	JRip	PART	J48	JRip	PART	J48
Wine	92.82±5.79	92.16±6.36	93.90±6.00	78.52±10.10	75.32±9.91	79.08±9.63	94.66±4.82	93.82±5.17	95.86±4.73
Heart	77.63±7.55	79.27±7.45	79.73±6.40	64.32±8.05	66.19±8.48	64.66±7.87	80.38±7.82	81.73±7.21	82.03±7.21
Ionosphere	90.83±4.66	89.55±4.82	89.74±4.38	74.93±4.91	74.93±4.91	74.93±4.91	90.89±4.89	91.34±4.85	91.97±4.26
Balloon Scale + Stretch	100.0±0.00	100.0±0.00	100.0±0.00	56.00±21.65	60.00±20.10	60.00±20.1	100.0±0.00	100.0±0.00	100.0±0.00
Balloon Scale - Stretch	100.0±0.00	100.0±0.00	100.0±0.00	55.00±21.00	60.00±20.10	60.00±20.1	100.0±0.00	100.0±0.00	100.0±0.00
Dbworld-bodies-stemmed	80.00±15.7	79.88±15.3	80.93±15.7	54.52±5.29	54.52±5.29	54.52±5.29	89.86±11.2‘	90.98±10.83	89.86±11.2
Dbworld-bodies	77.81±14.3	80.57±14.1	76.71±15.1	53.60±6.20	53.14±6.66	53.60±6.20	87.40±10.7	88.98±10.58	87.40±10.7
Cardiotocography-3class	98.50±0.91	98.63±0.83	98.67±0.85	78.89±2.09	78.48±1.39	79.54±1.99	98.40±0.90	98.40±0.90	98.40±0.90
Lung Cancer	47.08±27.7	50.58±23.7	44.75±23.9	44.58±23.85	45.50±20.43	37.42±21.2	66.23±23.4	61.17±23.75	67.08±24.9
Soyabean small	100.0±0.00	97.65±7.12	97.65±7.12	83.60±15.09	81.85±12.82	84.05±14.3	100.0±0.00	100.0±0.00	100.0±0.00
Trains	90.0±30.15	51.0±50.24	90.0±30.15	60.49±49.24	57.00±49.76	60.0±49.24	90.0±30.15	82.00±38.61	90.0±30.15
Zoo	93.41±7.28	89.81±8.37	92.61±7.33	71.32±5.83	60.43±3.06	71.32±5.83	97.05±4.76	92.59±7.39	97.05±4.76

TABLE 4.8: Comparison of classification accuracies for original datasets and reduced datasets by proposed model, and FMFRFS using 10 fold cross validation

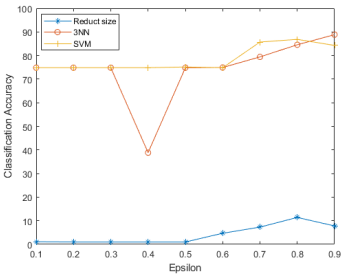
Dataset	Original		FMFRFS		Proposed model	
	KNN	SVM	KNN	SVM	KNN	SVM
Wine	95.29±5.40	95.88±4.84	94.11±5.54	94.11±6.20	97.05±4.99	96.47±7.44
Heart	79.61±8.31	72.69±13.74	73.07±8.10	56.53±7.48	82.30±8.91	71.15±7.53
Ionosphere	84.57±3.35	88.00±8.05	89.71±8.21	87.42±8.43	88.85±5.62	84.28±6.49
Balloon Scale + Stretch	100.0±0.00	100.0±0.00	70.00±48.30	70.00±48.30	100.00±0.00	100.00±0.00
Balloon Scale-Stretch	100.0±0.00	100.0±0.00	70.00±42.16	70.00±42.16	100.00±0.00	100.00±0.00
Dbworld-bodies-stemmed	58.33±18.00	86.67±13.14	60.00±11.66	53.33±10.54	70.00±15.31	78.33±11.24
Dbworld-bodies	53.33±23.30	86.67±13.14	53.33±20.48	53.33±20.48	70.00±23.03	80.00±23.30
Cardiotocography-3class	98.86±0.71	92.07±10.37	75.33±3.35	41.32±8.97	98.34±1.13	98.39±0.77
Lung Cancer	43.33±22.49	46.67±32.20	53.33±32.30	53.33±32.30	63.33±10.54	53.33±32.20
Soyabean small	100.0±0.00	100.0±0.00	85.00±17.48	92.50±12.07	100.00±0.00	100.00±0.00
Trains	50.00±52.70	70.00±48.30	70.00±48.30	50.00±52.70	70.00±48.30	90.00±31.62
Zoo	93.00±6.74	95.00±9.71	85.00±13.54	84.00±14.29	94.00±8.43	92.00±9.18



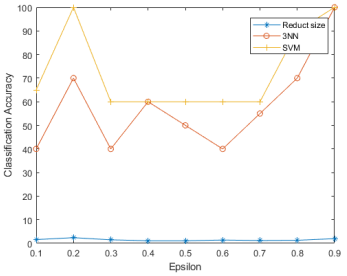
Wine



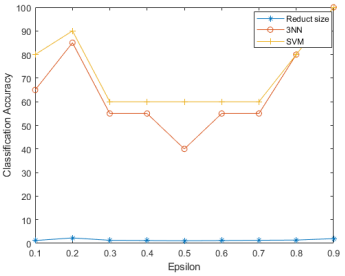
Heart



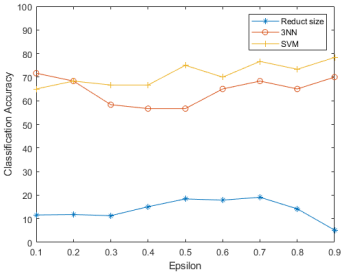
Ionosphere



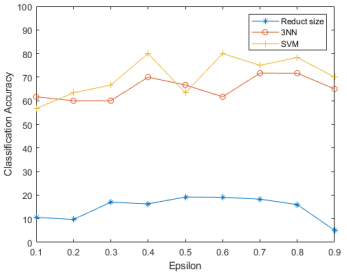
Balloon Scale+Stretch



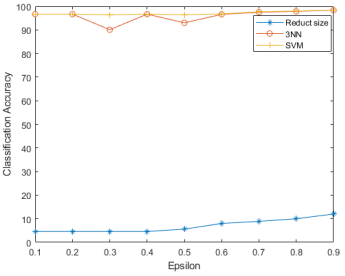
Balloon Scale-Stretch



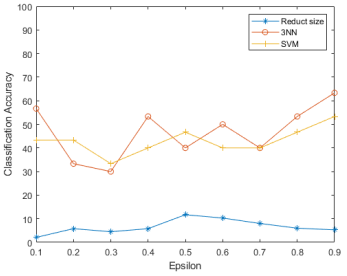
Dbworld-bodies-stemmed



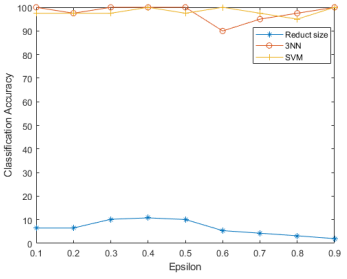
Dbworld-bodies



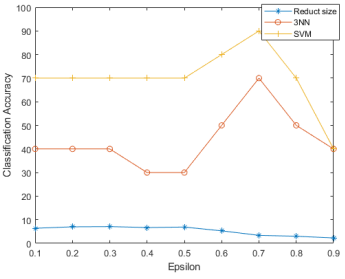
Cardiotocography-3class



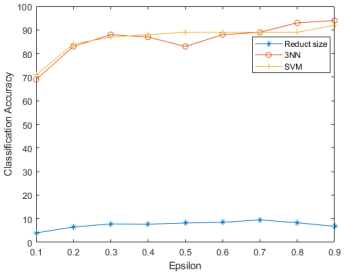
Lung Cancer



Soyabean small



Trains



Zoo

FIGURE 4.1: Variation of classification accuracy and reduct size with epsilon by proposed method

4.4 Summary

The traditional fuzzy-rough dependency cannot reveal better the learning ability of a subset of features as it only tries to keep the fuzzy positive region maximal and it cannot suitably fit data. Wang et. al. [154] handled this problem by introducing a fitting model for feature selection with fuzzy rough sets. However, fuzzy set theory has certain limitations and it cannot handle the uncertainty in the case where it is not found only in judgment but also in the identification. It is anticipated that the human decision-making process and activities require human expertise and knowledge which are inevitability imprecise and that could be simulated by using intuitionistic fuzzy set concept as it considers membership, non-membership and hesitancy functions simultaneously. This chapter introduces a fitting intuitionistic fuzzy rough set model to cope with above mentioned problems. This model fitted data well and avoided misclassification properly. Firstly, Intuitionistic fuzzy decision of an object was established using neighborhood concept. Then, intuitionistic fuzzy approximations are introduced using intuitionistic fuzzy decision along with parameterized intuitionistic fuzzy granule. Furthermore, an intuitionistic fuzzy dependency function is presented. Moreover, a heuristic greedy forward algorithm is applied to compute the reduct set. Finally, our proposed technique is applied on the benchmark datasets and a comparative study is presented. From the experimental results, we observe that presented algorithm provides more accurate reduct set than existing algorithm especially for those information systems in which various categories have a great degree of overlap. However, the model can only be applied for complete datasets. Noise, human error, improper communication, etc usually leads to missing values in the dataset gathered. Feature seletion for incomplete datasets is discussed in the next chapter.