Feature Selection using Intuitionistic Fuzzy Entropy [1]

R.Parvathi[1] and C.Radhamani[2]

[2,4]*Department of Mathematics*

[2] *Vellalar College for Women (Autonomous) , Erode - 638 107*

[4] *Kongu Arts and Science College (Autonomous), Erode - 638 012*

Tamil Nadu, India.

email: [1]paarvathis@rediffmail.com and [2]palaniradhu@gmail.com

***Abstract* — Feature selection plays an important role in improving the performance of machine learning as it significantly improves the performance of classification by discarding insignificant features from the original set of features. Accuracy in database classification can be achieved through feature selection while at the same time can speedup the classification rate and shorten the function time. The main aim of this work is to choose the most significant features in the feature set to perform a given task. In this paper, intuitionistic fuzzy entropy method is applied in feature extraction and the experimental results show that accuracy in classification is proved with reduced attributes. A comparison is made between the existing methods and the proposed method.**

***Keywords* — Intuitionistic fuzzy sets, temporal intuitionistic fuzzy sets, intuitionistic fuzzy entropy, feature selection, classification**

Mathematics Subject Classification:28E10

I. Introduction

The database accumulation in the environment due to the remarkable advancement in technology resulted in a setback in the effective accuracy of the output. To overcome the difficulty, data mining techniques and feature selection are widely applied in medical field, scientific research and business applications. Data mining is the process of extracting some of the useful information from a database containing a bulk of information. Feature selection reduces the dimensionality of the problem and there by increases the learning accuracy and improves the result. This is a remarkable and proven method to process high dimensional data (data with several features). Feature selection is an optimization algorithm which involves the process of selecting of N features form 2^N subsets. The main purpose of feature selection is to reduce the cost of computation and to improve the performance of the learning algorithm. Feature selection algorithms generally deal with evaluation criteria such as filters and wrapper models. The filter approach deals with the general characteristics of the training data set to select a feature subset and does not involve any learning algorithm where as the wrapper approach involves a pre-determined induction algorithm to analyze the performance of the selected features. The main objective of this paper is to propose an efficient feature selection strategy that has been proven to be an effective and efficient way in data pre-processing to achieve data reduction. The abundance of data in data sets demand development of new algorithms for analyzing the data and there by retaining important information. Entropy is a measure of uncertainity and information and Shannon entropy is a measure of uncertainity associated with a random variable. In many cases, the nature of raw data is imprecise and vague. Hence, it is of great interest to have suitable measures for analyzing the data. Fuzzy sets have the membership value of its members which play a vital role in the data accuracy. Therefore, fuzzy entropy measure based on Shannon entropy is widely employed for extracting the relevant features and classify the data. Fuzzy entropy measure is used in [5] to evaluate the information of pattern distribution by partitioning the data into non-overlapping decision regions. To evaluate and check the consistency of the feature subsets which are hesitant at some point, the design of a new model which can describe the data as it is, is essential and useful. Intuitionistic fuzzy sets give the details of both membership and non-membership along with hesitancy margin of its elements. The measure with these qualities can be used to create new models needed including one with a dynamic property. An intuitionistic fuzzy entropy measure (IFEM) and temporal intuitionistic fuzzy entropy measure (TIFEM) are employed to select relevant features for the classification purpose.

Classification is an essential task to predict some important characteristics of an item based on various parameters. Each parameter is represented by a variable called feature. On the whole, it determines the feature space and the number of features provides the dimension of the feature space. Each classifier maps the feature space into a set of class labels. Support vector machine (SVM) is a linear classification tool used by many researchers which classify items by constructing a hyperplane of dimension $n-1$ for an n dimensional feature space that splits all items into the two classes $+1$ and -1. A set of correctly splitting training examples are used and SVM constructs a maximum margin hyperplane, a hyperplane which maximizes the distance to the closest training examples. k-nearest neighbors algorithm is a non-parametric method used for classification. The input consists of k closest training examples in the feature space and the output is the class membership. An object is classified by a majority vote of its neighbors and the object is assigned to the class most common among its k nearest neighbors. When $k = 1$, the object is assigned to the class of the single nearest neighbor. The neighbors are taken from a set of objects for which the class is known. This is the training set of the algorithm. The training examples are vectors in a multi-dimensional feature space, each with a class label. The training phase store the feature vectors and class labels. In the classification phase, k is a user-defined constant and the test vector is classified by assigning the label which is more frequent among the k training samples nearest to that point. Euclidean, Minkowski, Chebychev and City Block are the commonly used metrics in this classifier.

This paper proposes a new feature selection technique based on IF entropy and TIF entropy measure to handle uncertain and temporal data sets. The proposed method uses the existing k-nearest neighbor and SVM classifiers for classification.

The remainder of the paper is organized as follows. Section 2 is resolved with a brief literature review about Fuzzy Sets(FSs), Intuitionistic Fuzzy Sets(IFSs), temporal intutionistic fuzzy sets(TIFSs) and entropy measures. Section 3 describes the construction of intuitionistic fuzzy sets from a feature set. IFEM methods for feature selection is discussed and implementation of the proposed intuitionistic fuzzy set strategy is done with selected data sets in section 4 and the results are compared with the existing fuzzy entropy based method. Section 5 provides the conclusion of the paper followed by references.

II. Preliminaries And Notations

Definition 2.1: [21] Let X be a non-empty set. A *fuzzy set* (FS) A drawn from X is defined as $A = \{(x, \mu_A(X))/x \in X\}$ where $\mu_A : X \to [0, 1]$ denotes the membership of the element x in the fuzzy set A .

Definition 2.2: [1] Let$X$ be a non-empty set. An *intuitionistic fuzzy set* (IFS) A in X is an object having the form $A = \{\langle x, \mu_A(X), \nu_A(X)\rangle / x \in X\}$ where $\mu_A : X \to [0,1]$ and $\nu_A : X \to [0,1]$ with $0 \le \mu_A(x) + \nu_A(x) \le 1$ for all $x \in X$ represents the degrees of membership and non-membership of the element x to the IFS A. For each IFS, the *intuitionistic index or hesitancy degree* of x in X to the IFS A is $\pi_A(x) = 1 - \mu_A(x) - \nu_A(x)$.

Definition 2.3: [2]

Let E be the universal set and T be a non-empty set of time moments. Then, a *temporal intuitionistic fuzzy set* (TIFS) is an object having the form

$$A(T) = \{< x, \mu_A(x,t), \nu_A(x,t) > /(x,t) \in E \times T\}$$

where

1) $A \subset E$ is a fixed set.
2) $\mu_A(x,t)$ and $\nu_A(x,t)$ denote the degrees of membership and non-membership respectively of the element (x,t) such that $0 \le \mu_A(x,t) + \nu_A(x,t) \le 1$ for every $(x,t) \in E \times T$.

Definition 2.4: [17]

The Shannon entropy $H(X)$ of a discrete random variable X is defined as follows:

$$H(X) = -\sum_{j=1}^{N-1} p(x_j) log_2 p(x_j)$$

or

$$H(X) = -\sum_{j=1}^{N-1} p_j log_2 p_j$$

where p_j denotes $p(x_j)$.

Note that entropy is a function of the distribution of X. It does not depend on the actual values taken by the random variable X, but only on the probabilities. Hence, Shannon entropy is also written as $H(p)$.

Definition 2.5: [5]

The Shannon entropy measures the randomness while the fuzzy entropy measures the fuzziness. Let $X = \{r_1, r_2, ...r_n\}$ be the universal set with n elements and A be the fuzzy set defined with $k < n$ elements whose membership function is defined by $\mu_A(r_i)$. Let the n elements be divided into m classes $\{c_1, c_2, ...c_m\}$. Let the set of elements of the class j is denoted by $S_{c_j}(r_n)$. Then, the match degree D_j of the fuzzy set A for the elements of the class j in an interval is defined as

$$D_j = \frac{\sum_{r \in S_{c_j}} \mu_A(r)}{\sum_{r \in X} \mu_A(r)}$$

The fuzzy entropy $FE_{c_j}(A)$ of the elements in the class j in an interval is defined as follows:

$$FE_{c_j}(A) = -D_j log_2 D_j$$

and the total fuzzy entropy along the universal set is estimated as

$$FE(A) = -\sum_{j=1}^{m} FE_{c_j}(A).$$

Definition 2.6: [9]

A well-defined fuzzy entropy measure must satisfy the following four Luca-Termini axioms [4].

(e1) $E(A) = 0$ iff $A \in 2^N$ where A is a non fuzzy set and 2^N indicates the power set of A.

(e2) $E(A) = 1$ iff $\mu_A(x_i) = 0.5$ for all i, where $\mu_A(x_i)$ indicates the degree of the element x_i to the fuzzy set A.

(e3) $E(A) \leq E(B)$ if A is less fuzzy than B, denoted by $A \prec\prec B$, is defined as
$\mu_A(x_i) \leq \mu_B(x_i)$ for $\mu_B(x_i) \leq 0.5$ and
$\mu_A(x_i) \geq \mu_B(x_i)$ for $\mu_B(x_i) \geq 0.5$ where A and B are fuzzy sets.

(e4) $E(A) = E(\overline{A})$.

Definition 2.7: [14]

A real function $E : \phi(X) \rightarrow [0, 1]$ where $\phi(X)$ is the set of all IFSs defined on X, is named as an *entropy* of IFSs on finite universe X if E satisfies the following four modified properties of Luca-Termini:

(e1) $E(A) = 0$ if A is a crisp set.

(e2) $E(A) = 1$ iff $\mu_A(x_i) = \nu_A(x_i)$, for every $x_i \in X$.

(e3) $E(A) \leq E(B)$ if A is less fuzzy than B, denoted by $A \prec\prec B$, is defined as
$\mu_A(x_i) \leq \mu_B(x_i), \nu_A(x_i) \geq \nu_B(x_i)$ for $\mu_B(x_i) \leq \nu_B(x_i)$ and
$\mu_A(x_i) \geq \mu_B(x_i), \nu_A(x_i) \leq \nu_B(x_i)$ for $\mu_B(x_i) \geq \nu_B(x_i)$.

(e4) $E(A) = E(\overline{A})$.

The following are two entropy formulae of IFSs which satisfy the above conditions $(e1) - (e4)$.

$$E_1(A) = 1 - \sum_{i=1}^{n} \frac{|\mu_A(x_i) - \nu_A(x_i)|}{n}$$

$$E_2(A) = 1 - \sqrt{\sum_{i=1}^{n} \frac{|\mu_A(x_i) - \nu_A(x_i)|^2}{n}}$$

Definition 2.8: [12] .

Let $\phi(X, T)$ denotes the set of all TIFSs on the finite universe $X = \{x_1, x_2, \ldots, x_n\}$ and $T = \{t_1, t_2, \ldots, t_m\}$. A real function $E : \phi(X, T) \rightarrow [0, 1]$ is called an *entropy* of TIFSs, if E satisfies the following four properties :

(e1) $E(A(T)) = 0$ if A is a crisp set.

(e2) $E(A(T)) = 1$ iff $\mu_A(x, t) = \nu_A(x, t)$, for every $(x, t) \in X \times T$.

(e3) $E(A(T)) \leq E(B(T))$ if $A(T)$ is less fuzzy than $B(T)$ which is defined as
$\mu_A(x, t) \leq \mu_B(x, t), \nu_A(x, t) \geq \nu_B(x, t)$ for $\mu_B(x, t) \leq \nu_B(x, t)$ or
$\mu_A(x, t) \geq \mu_B(x, t), \nu_A(x, t) \leq \nu_B(x, t)$ for $\mu_B(x, t) \geq \nu_B(x, t)$.

(e4) $E(A(T)) = E(\overline{A(T)})$.

.

The *entropy* formulae of a TIFS $A(T)$, which satisfy the entropy properties of TIFSs $(e1) - (e4)$ are given as follows :

$$(i) E_1(A(T)) = 1 - \frac{1}{mn} \sum_{i=1}^{n} \sum_{j=1}^{m} |\mu_A(x_i, t_j) - \nu_A(x_i, t_j)|$$

$$(ii) E_2(A(T)) = 1 - \sqrt{\frac{1}{mn} \sum_{i=1}^{n} \sum_{j=1}^{m} |\mu_A(x_i, t_j) - \nu_A(x_i, t_j)|^2}$$

III. Construction Of Intuitionistic Fuzzy Sets

Fuzzification is the process of modeling crisp sets into fuzzy sets. Membership functions for fuzzy sets are defined through fuzzification. Like wise, construction of IFSs needs membership and non-membership functions along with indeterministic part as hesitation margin. This is done by some standard intuitionistic fuzzification functions [15]. Every uncertain problem existing in day to day life is unique and hence each one needs a unique fuzzification function. But it is not easy to find the functions for all the problems in a short period and as such much work with a plenty of time and cost is required. To meet out this difficulty, there are some standard fuzzification functions available in literature for both FSs and IFSs [15]. Many researches have been focused in this area since the development of FSs and IFSs. To deal with the temporal data in an intuitionistic fuzzy logic controller, where in the intuitionistic fuzzy set concepts and rules are to be applied, intuitionistic fuzzification functions of IFSs are to be generalized. The concept is discussed by the authors in [11] and are modified in this work.

The proposed method of building IFSs is based on entropy measure of IFSs which consists of the following steps:

1. Identifying the number of intervals on each dimension
2. Finding the center and width of each interval
3. Constructing intuitionistic fuzzification functions for each interval i.e., constructing the IFSs.

A. *Identifying the number of intervals on each dimension*

The number of intervals on each dimension has a profound effect on the accuracy of classification as a large number gives a too fine partition and a small one gives a too big partition and both results in poor classification performance. Therefore, concentration should be given in the splitting of intervals.

In this paper, the selection of an optimal number of intervals is done using intuitionistic fuzzy entropy measure. The procedure is described in the following steps.

Step 1 : Normalization of data is carried out.

Step 2 : Set $I = 2$ as the initial number of intervals

Step 3 : Locate the center and width of the intervals using interval average

Step 4 : Determine IFSs by assigning intuitionistic fuzzification functions

Step 5 : Compute the total intuitionistic fuzzy entropy for I and $I - 1$ intervals

Step 6 : If the total entropy of I intervals is greater than that of $I - 1$ intervals, take the number of intervals as $I - 1$, else partition again as $I = I + 1$ and proceed from Step 2.

B. *Finding the center and width of each interval*

The center of each interval is determined and consequently the width and boundaries of the interval can be found. The process is described in the following steps.

Step 1 : Set the initial number of intervals

Step 2 : Calculate the interval centers as the average value of the elements in each interval

Step 3 : Calculate the Euclidean distance between the i^{th} element x_i and the j^{th} interval center C_j as $d_j = |C_j - x_i|$

Step 4 : Find the distances of each element using Euclidean distance method from the center of each interval

and select the least distance center
Step 5 : Reassign the elements in the intervals by taking the closest elements in each interval
Step 6 : Recompute the centers of each interval
Step 7 : Fix the center of each interval and proceed to the next step
Step 8 : Calculate the width of the intervals as the difference between the end values of the elements.

C. Constructing intuitionistic fuzzification functions for each interval

Intuitionistic fuzzy sets are constructed for the feature sets by assigning intuitionistic fuzzification functions to each interval. Various standard intuitionistic fuzzification (IF) functions are existing in literature [15]. The membership and non-membership values of each element of the feature in different intervals are calculated using IF functions. Fixing the hesitation margin as a constant value ϵ(between 0 and 1) depending on the nature of the problem, triangular IF functions are defined for each feature as follows:
Case 1 : Left most interval
Let c_1 be the center of the left most interval and c_2 be the center of the consecutive interval. Taking c_1 as the center and c_2 as the right boundary, the membership and non-membership functions of the left most interval for each element x in the interval is defined by $\mu(x)$ and $\nu(x)$ as follows:

$$\mu(x) = \begin{cases} \frac{c_1+x}{2c_1} - \frac{\epsilon}{2}, & for \quad x \leq c_1 \\ max\{1 - \frac{|x-c_1|}{|c_2-c_1|}, 0\} - \frac{\epsilon}{2}, & for \quad x > c_1 \end{cases}$$

$$\nu(x) = \begin{cases} 1 - \frac{c_1+x}{2c_1} - \frac{\epsilon}{2}, & for \quad x \leq c_1 \\ 1 - max\{1 - \frac{|x-c_1|}{|c_2-c_1|}, 0\} - \frac{\epsilon}{2}, & for \quad x > c_1 \end{cases}$$

Case 2 : Right most interval
In this case, the center of the just previous interval is chosen as the left boundary to the right most interval. Then, the membership and non-membership functions of the right most interval for each element x in the interval is defined by $\mu(x)$ and $\nu(x)$ as follows:

$$\mu(x) = \begin{cases} max\{1 - \frac{|x-c_4|}{|c_4-c_3|}, 0\} - \frac{\epsilon}{2}, & for \quad x \leq c_4 \\ \frac{2-x-c_4}{2(1-c_4)} - \frac{\epsilon}{2}, & for \quad x > c_4 \end{cases}$$

$$\nu(x) = \begin{cases} 1 - max\{1 - \frac{|x-c_4|}{|c_4-c_3|}, 0\} - \frac{\epsilon}{2}, & for \quad x \leq c_4 \\ 1 - \frac{2-x-c_4}{2(1-c_4)} - \frac{\epsilon}{2}, & for \quad x > c_4 \end{cases}$$

Case 3 : Middle intervals
For the middle intervals, the left and right interval centers to the considered middle interval are considered as the left and right boundaries respectively. Then, the membership and non-membership functions of the middle intervals for each element x in the interval is defined by $\mu(x)$ and $\nu(x)$ as follows:

$$\mu(x) = \begin{cases} max\{1 - \frac{|c_3-x|}{|c_3-c_2|}, 0\} - \frac{\epsilon}{2}, & for \quad x \leq c_3 \\ max\{1 - \frac{|c_3-x|}{|c_4-c_3|}, 0\} - \frac{\epsilon}{2}, & for \quad x > c_3 \end{cases}$$

$$\nu(x)=\begin{cases}1-max\{1-\frac{|x-c_3|}{|c_3-c_2|},0\}-\frac{\epsilon}{2}, & for \quad x\le c_3\\ 1-max\{1-\frac{|x-c_3|}{|c_4-c_3|},0\}-\frac{\epsilon}{2}, & for \quad x> c_3\end{cases}$$

D. Constructing temporal intuitionistic fuzzification functions for each interval

Temporal Intuitionistic fuzzy sets with time domain T are constructed by assigning intuitionistic fuzzification functions to each interval as in the case of IFSs. The membership and non-membership values of each element of the feature in different intervals are calculated using IF functions for TIFSs. Fixing the hesitation margin as a constant value ϵ(between 0 and 1), triangular IF functions are defined as follows:
Consider the number of intervals as four for the feature set and two for the time domain T. Let c_1, c_4 be the centers of the left and right most intervals and c_2, c_4 be the centers of the consecutive middle intervals. Let c' and c'' be the centers of time domain intervals.
Case 1 : Left most interval
Taking c_1 as the center and c_2 as the right boundary, the membership and non-membership functions of the left most interval for each element x at time t in the left most interval is defined by $\mu(x,t)$ and $\nu(x,t)$ as follows:

$$\mu(x,t)=\begin{cases}\frac{((c_1+c')+(x+t)}{2(c_1+c')}-\frac{\epsilon}{2}, & for \quad x\le c_1, t\le c'\\ \frac{((c_1+t)+(x+t)}{2(c_1+t)}-\frac{\epsilon}{2}, & for \quad x\le c_1, t> c'\\ max\{1-\frac{|(x+t)-(c_1+c')|}{|(c_2+c'')-(c_1+c')|},0\}-\frac{\epsilon}{2}, & for \quad x> c_1, t> c'\\ max\{1-\frac{|(x+c')-(c_1+t)|}{|(c_2+c'')-(c_1+c')|},0\}-\frac{\epsilon}{2}, & for \quad x> c_1, t\le c'\end{cases}$$

$$\nu(x,t)=\begin{cases}1-\frac{((c_1+c')+(x+t)}{2(c_1+c')}-\frac{\epsilon}{2}, & for \quad x\le c_1, t\le c'\\ 1-\frac{((c_1+t)+(x+t)}{2(c_1+t)}-\frac{\epsilon}{2}, & for \quad x\le c_1, t> c'\\ 1-max\{1-\frac{|(x+t)-(c_1+c')|}{|(c_2+c'')-(c_1+c')|},0\}-\frac{\epsilon}{2}, & for \quad x> c_1, t> c'\\ 1-max\{1-\frac{|(x+c')-(c_1+t)|}{|(c_2+c'')-(c_1+c')|},0\}-\frac{\epsilon}{2}, & for \quad x> c_1, t\le c'\end{cases}$$

Case 2 : Right most interval
In this case, the center of the just previous interval is chosen as the left boundary to the right most interval. Then, the membership and non-membership functions of the right most interval for each element x at time t in the right most interval is defined by $\mu(x,t)$ and $\nu(x,t)$ as follows:

$$\mu(x,t)=\begin{cases}max\{1-\frac{|(x+t)-(c_4+c'')|}{|(c_4+c'')-(c_3+c')|},0\}-\frac{\epsilon}{2}, & for \quad x\le c_4, t\le c''\\ max\{1-\frac{|(x+t)-(c_4+t)|}{|(c_4+t)-(c_3+c')|},0\}-\frac{\epsilon}{2}, & for \quad x\le c_4, t> c''\\ \frac{2-(x+t)-(c_4+c'')}{2(1-(c_4+c''))}-\frac{\epsilon}{2}, & for \quad x> c_4, t> c''\\ \frac{2-(x+t)-(c_4+t)}{2(1-(c_4+t))}-\frac{\epsilon}{2}, & for \quad x> c_4, t\le c''\end{cases}$$

$$\nu(x,t)=\begin{cases}1-max\{1-\frac{|(x+t)-(c_4+c'')|}{|(c_4+c'')-(c_3+c')|},0\}-\frac{\epsilon}{2}, & for \quad x\le c_4, t\le c''\\ 1-max\{1-\frac{|(x+t)-(c_4+t)|}{|(c_4+t)-(c_3+c')|},0\}-\frac{\epsilon}{2}, & for \quad x\le c_4, t> c''\\ 1-\frac{2-(x+t)-(c_4+c'')}{2(1-(c_4+c''))}-\frac{\epsilon}{2}, & for \quad x> c_4, t> c''\\ 1-\frac{2-(x+t)-(c_4+t)}{2(1-(c_4+t))}-\frac{\epsilon}{2}, & for \quad x> c_4, t\le c''\end{cases}$$

Case 3 : Middle intervals
For the middle intervals, the left and right interval centers to the selected interval are considered as the left and

right boundaries respectively. Then, the membership and non-membership functions of the middle intervals for each element x in the interval at time t is defined by $\mu(x,t)$ and $\nu(x,t)$ as follows:

$$\mu(x,t) = \begin{cases} max\{1 - \frac{|(c_3+c')-(x+t)|}{|(c_3+c')-(c_2+t)|}, 0\} - \frac{\epsilon}{2}, & for \quad x \le c_3 \\ max\{1 - \frac{|(c_3+t)-(x+c')|}{|(c_3+t)-(c_2+c')|}, 0\} - \frac{\epsilon}{2}, & for \quad x \le c_3 \\ max\{1 - \frac{|(c_3+c')-(x+t)|}{|(c_4+c'')-(c_3+c')|}, 0\} - \frac{\epsilon}{2}, & for \quad x > c_3 \\ max\{1 - \frac{|(c_3+t)-(x+c')|}{|(c_4+c'')-(c_3+t)|}, 0\} - \frac{\epsilon}{2}, & for \quad x > c_3 \end{cases}$$

$$\nu(x,t) = \begin{cases} 1 - max\{1 - \frac{|(c_3+c')-(x+t)|}{|(c_3+c')-(c_2+t)|}, 0\} - \frac{\epsilon}{2}, & for \quad x \le c_3 \\ 1 - max\{1 - \frac{|(c_3+t)-(x+c')|}{|(c_3+t)-(c_2+c')|}, 0\} - \frac{\epsilon}{2}, & for \quad x \le c_3 \\ 1 - max\{1 - \frac{|(c_3+c')-(x+t)|}{|(c_4+c'')-(c_3+c')|}, 0\} - \frac{\epsilon}{2}, & for \quad x > c_3 \\ 1 - max\{1 - \frac{|(c_3+t)-(x+c')|}{|(c_4+c'')-(c_3+t)|}, 0\} - \frac{\epsilon}{2}, & for \quad x > c_3 \end{cases}$$

IV. Feature Selection Using Intuitionistic Fuzzy Entropy And Other Existing Methods

The remarkable progress in the science of information technology lead computers to store huge volume of data. High dimensional data can be a great challenge to deal with and affects the performance of the learning algorithm. Hence, feature selection is needed as a preprocessing step to deal with this type of data. Feature selection has been studied by several researchers using different methods to improve the rate of classification by removing the redundant features from the feature set. Fuzzy entropy is applied as a feature selection strategy and is focused by many researchers [5], [6] and [18].

An efficient fuzzy classifier with feature selection based on fuzzy entropy is stated in [5]. Input feature sets are fuzzified using suitable membership function and corresponding fuzzy entropy is estimated. Minimal feature subsets are found by selecting the top k features with minimum fuzzy entropy. Fuzzy entropy is applied to find information of pattern distribution in pattern space. With this information, pattern space is partitioned into non-overlapping decision regions for pattern classification and hence a good classification performance is achieved.

In this section, a new feature selection approach based on intuitionistic fuzzy entropy is presented using intuitionistic fuzzy entropy for non-temporal and temporal data sets. Information content of all the features are available by the computation of entropies. Since a feature with less fuzzy entropy possess less fuzziness, that feature is chosen to obtain a better classification. A threshold is set to remove the features. The aim is to retain the features with less entropies as they could give a higher contribution to the classification results. The steps involved in the process of the proposed algorithm based on intuitionistic fuzzy entropy is described as follows:

Step 1 : Select a feature from the feature set

Step 2 : Construct the IFSs / TIFSs for each interval as given in section 3

Step 3 : Determine the IF entropy for all the IFSs / TIFSs constructed in step 2 by using the formulae given in Definition 2.7 / Definition 2.8 respectively

Step 4 : Determine the IF entropy for all intervals in the selected feature

Step 5 : Find the total entropy of the selected feature

Step 6 : Repeat the procedure for all the features

Step 7 : Out of all features in the feature space, select the features with minimum total entropy by setting a threshold.

A. Feature selection methods on different data sets

The solution to the chosen problem lies in the perfect prediction of results. Some popular classifiers are used to make predictions after the feature selection process is over. In this section, the increasingly popular tool in classification, the k-nearest neighbors algorithm for dimensionality reduction is used as a classifier to validate the accuracy of the proposed method. The bench-mark data sets like Iris data set, Wisconsin breast cancer data sets and the Texas weather data are considered to investigate the performance of the proposed feature selection strategy. The experimental results with feature selection are compared with regular classification results (without feature selection) in each case.

1) Iris data base: The Iris database created by R A Fisher is a popular database used for classification of Iris flower. It includes three classes namely, Verginica, Setosa and Versicolor and has four continuous features namely sepal length, sepal width, petal length and petal width. The data consists of 150 instances, 50 for each class. The Setosa class is linearly separable from the other two while the other two classes overlap each other. Two out of four features are selected by the proposed intuitionistic fuzzy entropy based feature selection

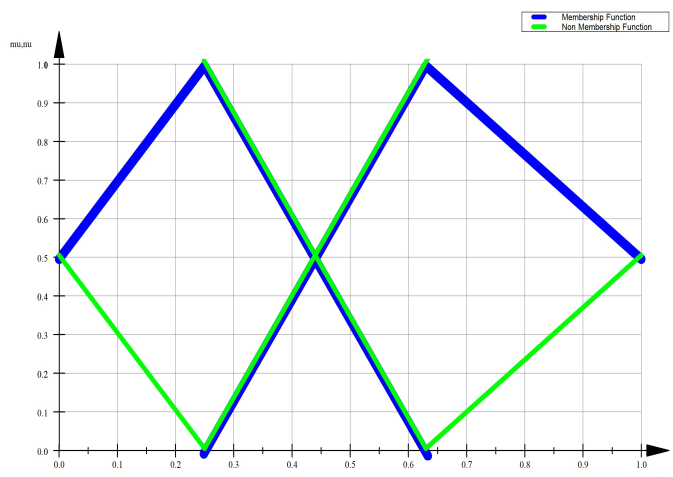

Fig. 4.1: Fuzzification Functions of the Feature 'Sepal Length' for Iris Data

method. The experimental results of the proposed algorithm are tested under two circumstances, one with all the four features (Sepal length, Sepal width, Petal length, Petal width) and the other with only the two selected features. 3 fold and 5 fold classification is done using k-nearest neighbors algorithm. The accuracy rate of classification with four features and two features are compared. Table 1 shows the experimental results of the k-nearest neighbors classifier using MATLAB. From the results, one can also conclude that if the width of either the petal or sepal is known for a particular class of an Iris flower, the other can be predicted with maximum accuracy.

2) Wisconsin breast cancer data base: The data set is based on 569 instances characterized by thirty real valued input features created by Dr.William H.Wolberg, W.Nick Street and Olvi L. Mangasarian in 1995. It was obtained from the UCI Machine Learning Repository [22]. Features are computed from a digitized image of a Fine Needle Aspiration (FNA) of a breast mass and describe the characteristics of the cell nuclei present in the image. The ten real valued features computed for each cell nucleus are Radius (mean of distances from

TABLE I: Testing Recognition Rates of Iris Database

Distance	Chebychev	City Block	Euclidean	Minkowski
3 fold cross validation with 4 features	96.69	94.68	95.38	96.03
3 fold cross validation with 2 features	93.38	92.68	93.38	93.8

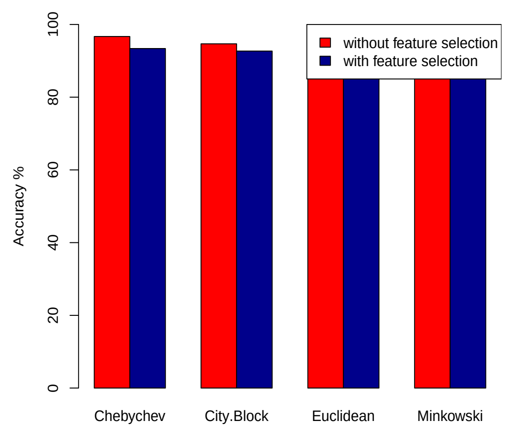

Fig. 4.2: Accuracy Results of Iris Database using IFEM Method of Feature Selection

center to point on the perimeter), Texture (standard deviation of grey-scale values), Perimeter(perimeter of the cell nucleus), Area (area of the cell nucleus), Smoothness (local variance in radius lengths), Compactness ($perimeter^2/area - 1.0$), Concavity (severity of concave portions of the contour), Concave points (Number of concave portions of the contour), Symmetry (symmetry of the cell nuclei) and Fractal dimension (coastline approximation-1). The mean, standard error and worst or largest (mean of the three largest values) of these features were computed for each image, resulting in 30 features. The features are used to predict benign and malignant cancer growths. The class distribution of the database is 357 benign and 212 malignant tumors.

The aim of the proposed algorithm is to achieve a feature subset with minimum number of features which could provide an efficient classification. The classification performance of the proposed feature selection method is compared with the one having all original features. The classification accuracy is obtained using SVM and k-nearest neighbors classification algorithm and the comparative results of the proposed method with and without feature selection is reported in Table 2 given below.

TABLE II: Classification Results on Wisconsin Database

Distance	Chebychev	City Block	Euclidean	Minkowski	SVM
3 fold cross validation with 30 features	92.98	94.73	95.61	95.44	97.19
3 fold cross validation with 20 features	92.45	95.25	95.26	93.68	96.49
5 fold cross validation with 30 features	93.16	95.26	95.27	95.28	97.53
5 fold cross validation with 20 features	93.87	95.96	95.80	95.11	96.32

3) Texas weather data base: The Texas Commission on Environmental Quality (TCEQ) data is a recorded meteorological data for 7 years between 1998 and 2004 which is available in http://archive.ics.uci.edu/ml. TCEQ ozone data consists of 72 features and two classes, 0 for no ozone day and 1 for ozone day. i.e., the class label is divided into 0 or 1. Three fold, five fold and ten fold classification is done with and without feature selection using k-means clustering algorithm.

The classification performance of the proposed feature selection method is compared with the one having all original features. The classification accuracy results are presented in Table 3.

TABLE III: Testing Recognition Rates of Texas Weather Database

Distance	Chebychev	City Block	Euclidean	Minkowski
3 fold classification with 72 features	100	98.26	99.94	100
3 fold classification with 42 features	89.4	91.1	90.6	90.8
5 fold classification with 72 features	100	98.27	99.95	100
5 fold classification with 42 features	88.9	90.53	90	89.64
10 fold classification with 72 features	100	98.32	99.95	100
10 fold classification with 42 features	89.11	90.48	89.89	90.16

B. Comparison of proposed feature selection with the existing methods

In section 4.1, feature selection is done for different data sets to reduce the number of attributes using IFEM method. K- nearest neighbour and SVM classification algorithms are applied for performing classification. In this section, the experimental results found on Iris and Wisconsin data sets are used to make comparison with the existing fuzzy entropy method [5]. Out of four features, two are selected in Iris data set and twenty out of thirty features are selected in Wisconsin data sets. A comparative study on the two feature selection methods are shown in table 4.

TABLE IV: Performance Analysis of Feature Selection Methods on Iris Data Sets using K-nn Classifier

Feature Selection method	Chebychev	City Block	Euclidean	Minkowski
Fuzzy entropy method	91.33	90.66	91.33	91.33
Intuitionistic fuzzy entropy method	93.38	92.68	93.38	93.8

TABLE V: Performance Analysis of Feature Selection Methods on Wisconsin Database using K-nn Classifier

Feature Selection methods	Chebychev	City Block	Euclidean	Minkowski	SVM
3 fold cross validation using fuzzy entropy method	91.75	93.67	93.15	92.63	94.90
3 fold cross validation using intuitionistic fuzzy entropy method	92.45	95.25	95.26	93.68	96.49
5 fold cross validation using fuzzy entropy method	93.01	94.21	93.69	93.87	95.25
5 fold cross validation using intuitionistic fuzzy entropy method	93.87	95.96	95.80	95.11	96.32

It can be noted that maximum accuracy is achieved using the proposed intuitionistic fuzzy entropy based feature selection method. Hence the proposed method shows better performance based on classification accuracy.

V. Conclusion

Feature selection has been done in a new method to retrieve the most relevant features from the feature space. This in turn reduces the unnecessary and irrelevant features as well as improves the classification accuracy with minimum computational effort. The main intention of this paper is to select the features for classification of a vague and huge data set with many features. Entropy measures of IFSs and TIFSs are used in feature selection. The experimental results are compared with the existing fuzzy entropy based method and prove that the proposed method is perfectly fit to locate the hidden information.

References

[1] K.T.Atanassov, "Intuitionistic fuzzy sets", VII ITKR's Session, Sofia, 1983.

[2] K.T.Atanassov, "Intuitionistic fuzzy sets: Theory and Applications", Physica-Verlag, Heidelberg, 1999.

[3] P.Burillo and H.Bustince, *Entropy for intuitionistic fuzzy sets and on interval-valued fuzzy sets*, Fuzzy Sets and Systems, 78 (1996) 305-316.

[4] A.De Luca and S.Termini, *A definition of non probabilistic entropy in the setting of fuzzy set theory*, Information and Control, 20 (1972)301-312.

[5] Hahn-Ming Lee, Chih-Ming Chen, Jyh-Ming Chen and Yu-Lu Jou, *An efficient fuzzy classifier with feature selection based on fuzzy entropy*, IEEE Transactions on Systems, Man and Cybernetics-Part B: Cybernetics, Vol.31, 3 (2001) 425-432.

[6] P.Jaganathan and R.Kuppuchamy, *A threshold fuzzy entropy based feature selection for medical database classification*, Computers in Biology and Medicines, 43 (2013) 2222-2229.

[7] Jiawei Han, Micheline Kamber and Jian Pei, "Data Mining: Concepts and Techniques", Morgan Kaufmann Publishers, An imprint of Elsevier, 2012.

[8] C.R.Kavitha and T.Mahalakshmi, *Feature selection methods for classification- A comparison*, International Journal of Research in Engineering and Technology, Vol.6, 2017.

[9] B.Kosko, *Fuzzy entropy and conditioning*, Information Sciences, Vol.40, 1986, 165-174.

[10] D.Nauck and R.Kruse, *A neuro-fuzzy method to learn fuzzy classification rules from data*, Fuzzy Sets Syst., Vol. 89, 3(1997) 277-288.

[11] R.Parvathi and C. Radhamani, *Multi-parameter temporal intuitionistic fuzzy sets*, Notes on Intuitionistic Fuzzy Sets, Vol.22, 1(2016) 35 - 47.

[12] C. Radhamani, *Entropy measures of temporal intuitionistic fuzzy sets*, International Journal of Fuzzy Mathematical Archive, Vol.15 1(2018) 91-103.

[13] Peter Revesz and Thomas Triplet, *Temporal data classification using linear classifiers*, Information Systems, 36(2011) 30-41.

[14] Qiansheng zhang, Hongyan xing and Lihua Wu, *Entropy, similarity measure, Inclusion measure of Intuitionistic fuzzy sets and their relationships*, International Journal of Computational Intelligence Systems, Vol.5, 3(2012) 519-527.

[15] C.Radhika and R.Parvathi, *Intuitionistic fuzzification functions*, Global Journal of Pure and Applied Mathematics, Vol.12, 2(2016) 1211 - 1227.

[16] E.Szmidt, *Entropy for intuitionistic fuzzy sets*, Fuzzy Sets and Systems, 118 (2001) 467-677.

[17] C.E.Shannon, *A mathematical theory of communication*, Bell syst. Tech. Journal, Vol.27,1948, 379-423.

[18] D.SheelaJeyarani and A.Pethalakshmi, *An efficient fuzzy entropy based feature selection algorithm for high dimensional data*, International Journal of Advanced Research in Computer Science, Vol.6, 6 (2015).

[19] Z.S.Xu, *Some similarity measures of intuitionistic fuzzy sets and their application to multiple attribute decision making*, Fuzzy Optimization and Decision Making, 6 (2007) 109-121.

[20] L.A.Zadeh, *Fuzzy Sets and systems*, In Proceedings of Symposium on System Theory, Polytechnic Institute of Brooklyn, Newyork, 1965, 29 - 37.

[21] L.A.Zadeh, *Fuzzy Sets*, Information and Control, 8 (1965) 338 - 353.

[22] http://archive.ics.uci.edu/ml/machine-learning-databases/breast-cancer-wisconsin/breast-cancer.