CHAPTER-III

Efficient Feature Selection Technique Based on Modified Fuzzy C-means Clustering with Rough Set Theory
CHAPTER-III

EFFICIENT FEATURE SELECTION TECHNIQUE
BASED ON MODIFIED FUZZY C-MEANS
CLUSTERING WITH ROUGH SET THEORY

3.1 Introduction

In the field of Machine Learning Complex Data volume increases every day. Data Mining is becoming the hottest topic in current research in Engineering, Bioinformatics and Applicable Mathematics field. These complex data scientifically acquiring many measurements, and do not know which measure will be relevant to the occurrence of interest. Our goal is to find the significant or relevant variables from a heart, cancer medical data to build the efficient classification model. Diverse data mining methods have been established and are very successful for analyzing these medical data sets. Still, there are many challenging issues linked with data mining and machine learning for mining of high dimensional data. It includes,

1. The curse of dimensionality
2. Huge sparsity
3. Unstructured data
4. Nondeterministic polynomial time hard
5. Speed of modern computers.

Due to the enormous number of features in the medical datasets, feature selection before classification algorithm is an essential part of the process. Along with the feature selection properties embedded in the classification algorithm, it is required to apply some additional feature selection methodology for better classification
accuracy. Only limited number of features has such predictive ability which help classifying the medical datasets with greater accuracy.

Feature selection is a vital preprocessing tool and technique in data mining, especially for high dimensional datasets feature selection plays an imperative role. It has been a dynamic field of research and development for the precedent decades. It entails identifying a subset of the most functional features that produces companionable results as the original entire set of features. In Statistics and machine learning, feature selection, also known as Variable Selection, Attribute Selection or Variable Subset Selection. It is the process of selecting a subset of relevant features (variables, predictors). Feature selection techniques are used for the following reasons.

- Simplification of models to make them easier to interpret by researchers/users.
- Shorter training times.
- To avoid the curse of dimensionality.
- Enhanced generalization by reducing over fitting formally.
- Reduction of variance.
- It enables the machine learning algorithm to train faster.
- It reduces the complexity of a model and makes it easier to interpret.
- It improves the accuracy of a model if the precise subset is chosen.
- It improves the classifier performances.
- It reduces over fitting.
3.2 Related Work

Aspiration of feature selection is to discover a suitable feature subset that produces higher accuracy for classifier in the end user. Hybrid methods for feature selection comprised of combination of filter and wrapper approaches. It has newly been appeared as strong techniques for the problem in this domain. This chapter presents a novel approach for feature selection based on feature clustering using well known Modified Fuzzy C-Means (MFCM) clustering with Rough Set Theory (RST) philosophy for the high dimensional medical data. Changzhong Wang et al., (2016) proposed Feature subset selection based on fuzzy neighborhood rough sets. It discussed the model that is applied to the fields of classification learning and reasoning with uncertainty. High Dimensional datasets are often vague and redundant, creating problem to take decision accurately. The main drawback of the standard FCM for high dimensional data is that the objective function does not take into consideration the spatial information, and the standard FCM algorithm is sensitive to noise and a noisy pixel is always wrongly classified because of its abnormal feature. This chapter aims at applying fuzzy-rough concept to overcome the above limitations.

3.3 Objectives

The main objective of this chapter is to find most relevant features by using efficient feature selection MFCM-RST based feature selection:

- To find the minimum number of most relevant attributes from a given medical dataset.
- To empirically compare the existing feature selection methods accuracy with proposed Modified Fuzzy C-means -RST Feature selection Techniques with standard Datasets.
3.4 Proposed System Architecture

3.4.1 Input Datasets

Due to privacy concern, medical data log is not released by many health care Organizations. Therefore the proposed work uses the benchmark medical dataset compiled for the data mining repository of the University of California, Irvine (UCI). This research work is validated using Cleveland, Hungarian, Switzerland, Breast cancer, Leukemia, Lung cancer dataset. The main advantages of this UCI Machine learning Medical dataset is that the proposed work is capable of providing significant result Napolitano et al., (2016) that are very easily obtained and capturing other researchers in the same domain. Figure 3.1 shows proposed system architecture.
3.4.2 Data Preprocessing

In the Data Mining process Data Preprocessing is an important technique transforming raw data into a human understandable format.

1. In data preprocessing thousands of medical records from the dataset are integrated into one table.
2. Deletes the record whichever fail to match with the standard format from the record set.
3. If any data type value fall out of specified range of the average data value, then those records are also deleted.
4. Missing data, Incorrect Records are also be deleted.
5. Remove Multivalve attributes.
6. Remove missing attribute instances.

3.5 Feature Selection using Modified Fuzzy C-Means (MFCM) Clustering Algorithm with Rough Set Theory

In this chapter, MFCM-RST is used to identify the most relevant features from the heart disease, cancer data sets, which are used in the disease classification pattern. Initially medical data set records involves data preprocessing, Feature selection, classification and optimization has been proposed and implemented. This kind of data mining techniques are necessary since it is acquiring high data volume and complex to process individual patient record in real time to classify heart disease, cancer disease pattern. Here, input dataset will be as high dimensional or high features which are a great barrier for classification. Therefore, feature dimension reduction method will be applied to reduce the features space without losing the accuracy of classification. Here, clustering with rough set theory based feature selection will be
developed and used to reduce the feature dimension. Figure 3.2 shows Modified Fuzzy C-Means (MFCM-RST) clustering algorithm with rough set theory will be combined and used to feature selection process.

![MFCM-RST Feature Selection Work Flow](image)

**Figure 3.2:** MFCM-RST Feature Selection Work Flow

### 3.5.1 Steps for MFCM-RST Feature Selection

#### 3.5.1.1 Subset Formation

It is useful when the required number of clusters is pre-determined. Thus, the algorithm tries to put each of the data points to one of the clusters. What makes FCM different is that it does not decide the absolute membership of a data point to a given cluster; instead, it calculates the likelihood (the degree of membership) that a data point will belong to that cluster. Hence, depending on the accuracy of the clustering which is required in practice, appropriate tolerance measures can be put in place.
Since the absolute membership is not calculated, FCM can be extremely fast because the number of iterations required to achieve a specific clustering exercise corresponds to the required accuracy. Clustering techniques are mostly unsupervised methods that can be used to organize data into groups based on similarities among the individual data items. Most clustering algorithms do not rely on assumptions common to conventional statistical methods, such as the underlying statistical distribution of data, and therefore they are useful in situations where little prior knowledge exists. The potential of clustering algorithms to reveal the underlying structures in data can be exploited in a wide variety of applications, including classification, image processing, pattern recognition, modeling and identification. This chapter presents an overview of fuzzy clustering algorithms based on the \( c \) means functional.

The FCM algorithm assigns cluster centre to each category by using fuzzy memberships.

\[
J_m = \sum_{i=1}^{J} \sum_{j=1}^{F} (\mu_{ij})^m \|x_i - z_j\|^2 
\]  
... (3.1)

In equation (3.1), \( x_i \) represents the features \( w(s_{ni}) \), \( c(s_{nj}) \) extracted from the input database, \( Z_j \) is the \( j \)th cluster centre and \( m \) is the constant value. The membership function represents the probability that a cluster center belongs to a specific cluster. In the FCM algorithm, the probability is dependent on the distance between the pixel and each individual cluster center in the feature domain. The membership functions and cluster centers are updated by the equations (3.2) and (3.3).

\[
\mu_{ij} = \frac{1}{\sum_{k=1}^{J} \left( \|x_i - z_j\|^2 \right)^{m-1}} \]  
... (3.2)
Repeat the algorithm until the coefficients’ change between two iterations is no more than $\xi$, for the given sensitivity threshold.

$$\max_j \left\| \mathcal{U}_j^{(k)} - \mathcal{U}_j^{(k+1)} \right\| < \xi$$ ... (3.3)

In equation (3), $\xi$ is a termination criterion between 0 and 1, whereas $k$ are the iteration steps. The clusters centroid values are computed by using the equation (3.4).

$$z_j = \frac{\sum_{i=1}^{J} u_{ij} w x_j}{\sum_{i=1}^{J} u_{ij} w}$$ ... (3.4)

To enhance the performance of the fuzzy-C-means clustering method, adaptiveness is invoked by measuring the Clustering effectiveness ($\alpha$) and Absolute density ($\beta$). On the basis of these two, set two thresholds to ensure the clustering being good. After the FCM process, obtain the number of cluster set such as $I_1$, $I_2$, $I_3$, $\ldots$, $I_n$. This clustered dataset is used for the further processing. The function FCM takes a data set and a desired number of clusters and returns optimal cluster centers and membership grades for each data point. It can use this information to build a fuzzy inference system by creating membership functions that represent the fuzzy qualities of each cluster.

- Find 2 clusters using Fuzzy C-means clustering.
- Classify each data point into the cluster with the largest membership value.
- Plot the clustered data and cluster centers

**MFCM settings:**

1. Choose a Sample Dataset
2. Initialize Number of clusters
Tune the 3 Optional Parameters

- exponent,
- maximum number of iterations
- minimum amount of improvement

The function FCM takes a high dimensional medical data set with a desired number of clusters that returns optimal cluster centers and membership grades for each data point. Figure 3.3 shows Kernel Specification.

<table>
<thead>
<tr>
<th>Kernel parameter</th>
<th>Initial cluster centers</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma$</td>
<td>C1</td>
</tr>
</tbody>
</table>

**Figure 3.3: Kernel Specification**

```matlab
%% Split the data using MFCM
k = 2;
T = data(:,end);
iter = 0;
[center] = MFCMClust(T,k);
[a1,b1] = find(T <= center(1));
[a2,b2] = find(T >= center(1));
D1 = data(a1,1:end-1);
D2 = data(a2,1:end-1);
function [center, U, obj_fcn] = MFCMClust(data, cluster_n, kernel_b,options)
KernelMat = gaussKernel(center,data,kernel_b);
um = mf.*KernelMat * data;
den = sum(mf.*KernelMat,2);
```
center_new = num./((den*ones(1,feature_n)));
kdist = distkfcm(center_new, data, kernel_b);
obj_fcn = sum(sum((kdist.^2).*mf));
tmp = kdist.^(-1/(expo-1));
U_new = tmp./(ones(cluster_n, 1)*sum(tmp));

function out = distkfcm(center, data, kernel_b)
cluster_n = size(center, 1);
data_n = size(data, 1);
out = zeros(cluster_n, data_n);
for i = 1:cluster_n
    vi = center(i,:);
    out(i,:) = 2-2*gaussKernel(vi,data,kernel_b);
end

function out = gaussKernel(center,data,kernel_b)
if nargin == 2
    kernel_b = 150;
end
dist = zeros(size(center, 1), size(data, 1));
for k = 1:size(center, 1),
    dist(k,:) = sqrt(sum(((data-ones(size(data,1),1)*center(k,:)).^2)',1));
end
out = exp(-dist.^2/kernel_b^2);
3.5.1.2 Attribute Selection

After the clustering process using FCM, have to do the attribute selection process. Here, calculate the minimum $A_{\text{min}}$ and maximum $A_{\text{max}}$ value of each column (or each attribute) in the dataset $D_1$ and $D_2$. If the $A_{\text{min}}$ and $A_{\text{max}}$ value is similar to corresponding dataset $I_1$ and $I_2$, shall be neglect that column.

```matlab
% attribute selection.....
attribute = [];
for i = 1 : size(minmaxval2,1)
    if
        (minmaxval1(i,1)~=minmaxval2(i,1)||minmaxval1(i,2)~=minmaxval2(i,2))
            attribute = [attribute i];
        end
    end
    end
Traindata = [Traindata1;Traindata2];
[range] = Find_range(Traindata,attribute);
```

3.5.1.3 Discretization

Discretization is one of the most important prominent data preprocessing tasks in data mining and Knowledge Discovery. It regards a data reduction mechanism because it diminishes data from a huge domain of numeric values to a subset of categorical values. Many data mining algorithms use discretized data, which can only deal with discrete attributes. This method causes that the learning methods show remarkable improvements in learning speed and accuracy. It generally leads to a loss of information. This minimization of such information loss is the main goal of a
discretizer. The Discretization consists of transforming a continuous variable into a discrete attribute, by defining a set of intervals.

The two main issues are:

- How to determine the number of intervals;
- How to determine the cut points.

The best discretization is the one performed by a domain expert. Indeed, it takes into account other information than those only provided by the available dataset. But this knowledge is often unavailable. The data-driven methods are distinguished by their context (supervised or unsupervised); the kind of information they handle; and the strategy used (top down vs. bottom-up in the most of the cases). The discretization is part of the learning process shown in Figure 3.4.

**Figure 3.4: Discretization Process**

Discretization reduces the number of values for a continuous attribute. Some methods can only use nominal data $K$, the number of interval is a parameter Divide the dataset into $K$ ranges of equal size. Calculate data from the width of the intervals (a) derive the $(K-1)$ cut points ($b_1, b_2, \text{etc.}$).
a = \frac{\text{max} - \text{min}}{k} \quad \ldots (3.5)

b_1 = \text{min} + a

b_2 = b_1 + a = \text{min} + 2a

Presuppose a dataset \( S \) consisting of \( N \) examples attributes and \( C \) class labels a
discretization scheme. \( D_A \) exist on the continuous attributes \( A \in M \), which separation
this attribute into \( k \) discrete and disjoint intervals. Represented as \( \{(d_0, d_1), (d_1, d_2)\ldots (d_{kA-1}, d_{kA})\} \)

\[ d_0 \quad - \quad \text{minimum value} \]

\[ d_{kA} \quad - \quad \text{maximum value} \]

The set of cut points \( A \) represents in ascending order

\[ P_A = \{d_1, d_2, \ldots d_{kA-1}\} \]

Discretization is a significant step in data processing to convert the data into
specific interval, means that the range of values is confined into a specific interval.
Here, we have used one discretization function based on the predictable way. We
perform the discretization process at first, we identify the maximum and minimum
values of every attribute, and the \( K \) interval is tracked by taking the ratio between the
deviated value and the \( K \) value.

\[
\text{for each "} j\text{" } D^j = \left[ \frac{\max(A_j) - \min(A_j)}{3} \right] \quad \ldots (3.6)
\]

\[
D^i = \min(A_j) \leq \left[ \min(A_j) + D^j \right] \quad \ldots (3.7)
\]

\[
D^{ii} = \left[ \min(A_j) + D^j \right] \leq \left[ \min(A_j) + 2 \bullet D^j \right] \quad \ldots (3.8)
\]

\[
D^{iii} = \left[ \min(A_j) + 2 \bullet D^j \right] \leq \max(A_j) j \quad \ldots (3.9)
\]
Using equation (3.6) adjust all the feature values in the specific interval. Now obtain the new feature values, which feature values, vary from specific interval. Then every value that comes under within the range is replaced with the interval value so that the input data is transformed to the discretized data. Consequently, the training dataset $D^{TR}$ is concerted to the discretized format $D^D$ where, the entire data element $D^D$ contains only the L, M, and H if $k=3$.

Machine learning algorithms such as Support Vector Machines have been used for classification in high dimensional data due to their robustness to the dimensionality of the data. It shows that discretization can facilitate significantly the classification performance of these algorithms that are sensitive to the dimensionality of the data.

3.5.1.4 Reduct and Core Analysis

![Figure 3.5: Rough Set Feature Reduction](image)

Rough set provides a method to decide the importance and necessity of features. It is an extension of set theory proposed for knowledge discovery in data sets. Given a Dataset with discretized features, it is possible to find a reduct of original features that are most predictive in terms of classification accuracy. Clustering is an unsupervised classification that partitions an input data set into a desired number of subgroups or clusters. It is a process of dividing the data points into their natural groups so that the data in the same group or cluster are similar to one
another and different from the data points in other groups. In Rough Set theory, Reduct and core are the two most important concept.

Reduct is a reduced subset of original set which retains the accuracy of original set shown in Figure 3.5. There are usually several such subsets of attributes and those which are minimal are called reducts. Reduct is often used in the attribute selection process to reduce unnecessary attributes towards decision making applications.

1. Reduct of a decision table is a set of condition attributes that is sufficient to define the decision attribute.
2. Any reduct enables us to reduce condition attribute.
3. A reduct does not contain redundant attribute toward a classification task.
4. It reduces computation cost for rule generation.
5. Finding all the reduct is an NP-hard problem.

In a decision table may find multiple reduct and some rule would appear more frequently in some reduct than others. There are so many methods of finding reduct of a decision table. The reducts can be obtained by using the reduct generation algorithms. Using the discernibility matrix, the reduct of a decision table can be found. The core can be found as the set of all singleton entries in the discernibility matrix. The reduct is the minimal element in the discernibility matrix, which intersects all the element of the discernibility matrix.

In this chapter a novel feature selection algorithm is presented for finding the most important attribute of the dataset using reduct (reduced set of necessary attributes) and core. Reduct and core study is a significant step of rough set theory.
categorize the significance of the attributes so that the relevant attributes can be able to recognize without compromising the accuracy of classification. Now the dimensionality reduction can simply be considered as removal of some attributes from the decision table (actually some features from the feature vector) preserving its basic classification capability. In this case we would face two conditions;

i) Whether a decision table contains some redundant or superfluous data, if yes then,

ii) Collect those redundant data and remove them.

To perform such reduction of attributes in rough set theory select indispensable attributes. Indiscernibility Relation is the relation between two objects or more, where all the values are identical in relation to a subset of considered attributes. It can be observed that the set is composed of attributes that are directly related to patients. A reduct is a set of necessary minimum data, since the original proprieties of the system or information table are maintained. Therefore, the reduct must have the capacity to classify objects, without altering the form of representing the knowledge. The process of reduction of information is presented below; it can be observed that the data is of a discreet type. An information system or information table can be viewed as a table, consisting of objects (rows) and attributes (columns). It is used in the representation of data that will be utilized by Rough Set, where each object has a given amount of attributes. These objects are described in accordance with the format of the data table, in which rows are considered objects for analysis and columns as attributes. IR are selected using two fundamental concepts: reducts and core: they are defined as follows:

Let Q be a subset of P and let a belong to Q.
• A is dispensable in Q if I(Q-{a}); otherwise a is indispensable in Q.

• Set Q is independent if all its attributes are indispensible.

• Subset Q’ of Q is a reduct of Q if Q’ is independent and I(Q’)=I(Q).

Thus a reduct is the minimal subset of attributes that enables the same classification of elements of the universe as the whole set of attributes i.e., attributes that do not belong to a reduct are redundant or superfluous.

\[ Core = \bigcap \text{reduct} \quad \ldots (3.10) \]

• Rough set theory provides efficient algorithms, techniques and tools for finding hidden patterns in data.

• It consents to reduce original data, and to determine minimal sets of data with the same knowledge as in the original dataset.

• It allows in automatic way to the sets of decision rules from original data set.

• It is simple and easy to understand.

Reduct is computed by taking the relative discernability function and eliminating unnecessary attributes. The equivalence class is signifying the equivalence class is denoted as ‘E’.

Finding hidden patterns in data;

• Finds minimal sets of data (data reduction);

• Evaluates significance of data;

• Generates sets of decision rules from data;

• Facilitates the interpretation of obtaining result

• Rough set theory handles discrete attributes.
3.6 Implementation steps for Proposed Algorithm

1. Read the Data
2. Split the data using MFCM
3. Apply Rough set theory
4. Separate the splited data by rough set theory

Main program
clc
clear all
close all
warning off
dbstop if error
addpath('Sub code')

%% Read the Dataset directory
F = dir('.\Dataset\*.xls');
for i1 = 1:length(F)
    %% Read the Data
    data = xlsread(['."Dataset\" F(i1).name']);
    Feat = [];
    [nanx,nany] = find(isnan(data)==1);
    for i = 1:length(nanx)
        data(nanx(i),nany(i)) = 0;
    end
    %% Split the data using MFCM
    k = 2;
    T = data(:,end);
iter = 0;

[center] = MFCMClust(T,k);

[a1,b1] = find(T <= center(1));

[a2,b2] = find(T >= center(1));

D1 = data(a1,1:end-1);

D2 = data(a2,1:end-1);

%% Apply Rough set theory

[Rule,V] = Rough_set(D1,D2);

%% Seperate the splited data by rough set theory

inc = 1;

D = data(:,V);

for i = 1:size(D,2)

    R1 = Rule(:,i);

    [a,b] = find(R1 ~= 0);

    if ~isempty(a)

        Feat(:,inc) = D(:,i);

        inc = inc+1;

    end

end

Tar = data(:,end);

Tar(Tar > 0) = 2;

Tar(Tar == 0) = 1;

%% Save the Results

% save(['MFCM RS' F(i1).name(1:end-4)],'Feat','Tar')
3.7 Computational Results

The proposed methodology is experimented with the dataset namely Cleveland, Hungarian, Switzerland, leukemia, lung, Breast. These datasets are taken from the UCI machine learning repository. It was recorded using Intel® Atom™ CPU N455@ 1.67GHZ processor and 2GB RAM. The proposed methodology implemented by Mat lab. Table-3.1 shows the Characteristics of UCI Datasets which are used in experiments.

The computational Analysis result shows cluster center value for each cluster c1, c2 of the high dimensional datasets in Table-3.2.

Table-3.1: Characteristics of UCI Datasets which are used in Experiments

<table>
<thead>
<tr>
<th>Datasets</th>
<th>Number of instances</th>
<th>Number of attributes</th>
<th>Datasets characteristics</th>
<th>Attribute characteristics</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cleveland</td>
<td>303</td>
<td>14</td>
<td>Multivariate</td>
<td>Categorical, Integer, Real</td>
<td>2</td>
</tr>
<tr>
<td>Hungarian</td>
<td>294</td>
<td>14</td>
<td>Multivariate</td>
<td>Categorical, Integer, Real</td>
<td>2</td>
</tr>
<tr>
<td>Switzerland</td>
<td>123</td>
<td>14</td>
<td>Multivariate</td>
<td>Categorical, Integer, Real</td>
<td>2</td>
</tr>
<tr>
<td>Breast Cancer</td>
<td>60</td>
<td>2001</td>
<td>Multivariate</td>
<td>Categorical, Integer, Real</td>
<td>2</td>
</tr>
<tr>
<td>Leukemia</td>
<td>72</td>
<td>7129</td>
<td>Multivariate</td>
<td>Categorical, Integer, Real</td>
<td>2</td>
</tr>
<tr>
<td>Lung</td>
<td>62</td>
<td>2002</td>
<td>Multivariate</td>
<td>Categorical, Integer, Real</td>
<td>2</td>
</tr>
</tbody>
</table>

Table-3.3 shows selected numbers of Data Points in Each Cluster for Datasets, Table-3.4 presents selected number of features, Table-3.5 illustrates comparative analysis of selected numbers of features with various feature selection techniques.
Table-3.2: Cluster Centroid

<table>
<thead>
<tr>
<th>Sl. No.</th>
<th>Datasets</th>
<th>C1 center</th>
<th>C2 center</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Clevand</td>
<td>1.206147</td>
<td>0.557318</td>
</tr>
<tr>
<td>2.</td>
<td>Hungarian</td>
<td>0.311423</td>
<td>0.414436</td>
</tr>
<tr>
<td>3.</td>
<td>Switzerland</td>
<td>1.821604</td>
<td>1.787654</td>
</tr>
<tr>
<td>4.</td>
<td>Breast cancer</td>
<td>0.682458</td>
<td>0.142903</td>
</tr>
<tr>
<td>5.</td>
<td>Leukemia</td>
<td>1.55441</td>
<td>1.722316</td>
</tr>
<tr>
<td>6.</td>
<td>Lung</td>
<td>1.619517</td>
<td>2.249127</td>
</tr>
</tbody>
</table>

Table-3.3: Selected Numbers of Data Points in Each Cluster for Datasets

<table>
<thead>
<tr>
<th>Sl. No.</th>
<th>Datasets</th>
<th>Selected No. of features D1</th>
<th>Selected No. of features D2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Cleveland</td>
<td>219,13</td>
<td>13,84</td>
</tr>
<tr>
<td>2.</td>
<td>Hungarian</td>
<td>188,13</td>
<td>106,13</td>
</tr>
<tr>
<td>3.</td>
<td>Switzerland</td>
<td>56,13</td>
<td>67,13</td>
</tr>
<tr>
<td>5.</td>
<td>Leukemia</td>
<td>397,129</td>
<td>217,129</td>
</tr>
</tbody>
</table>
Table-3.4: Selected Number of Features

<table>
<thead>
<tr>
<th>Sl. No.</th>
<th>MFCM - RST</th>
<th>Selected No. of Features</th>
<th>Selected No. of Instances</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Cleveland</td>
<td>6</td>
<td>64</td>
</tr>
<tr>
<td>2.</td>
<td>Hungarian</td>
<td>6</td>
<td>294</td>
</tr>
<tr>
<td>3.</td>
<td>Switzerland</td>
<td>5</td>
<td>123</td>
</tr>
<tr>
<td>4.</td>
<td>Breast cancer</td>
<td>20</td>
<td>62</td>
</tr>
<tr>
<td>5.</td>
<td>Leukemia</td>
<td>71</td>
<td>60</td>
</tr>
<tr>
<td>6.</td>
<td>Lung</td>
<td>492</td>
<td>180</td>
</tr>
</tbody>
</table>

Table-3.5: Comparative Analysis of Selected Numbers of Features

<table>
<thead>
<tr>
<th>Sl. No.</th>
<th>Data set</th>
<th>FRS</th>
<th>FCM-RST</th>
<th>MFCM-RST</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>#O F's</td>
<td>#S F's</td>
<td>#O F's</td>
</tr>
<tr>
<td>1.</td>
<td>Cleveland</td>
<td>13</td>
<td>7</td>
<td>13</td>
</tr>
<tr>
<td>2.</td>
<td>Hungarian</td>
<td>13</td>
<td>6</td>
<td>13</td>
</tr>
</tbody>
</table>

(#OF’s-Original Features, #SF’s-Selected Features)

3.8 Conclusion

In this chapter a Modified Fuzzy C-Means Algorithm (MFCMS) with RST have shown how this algorithm can be used to select features for large data sets. MFCM-RST is an objective function-based clustering algorithm, which partitions the data set minimizing the sum of distances. Proposed a new modify spatial FCM that incorporates the spatial information into the membership function to improve the segmentation results. The new method was tested on high dimensional data and evaluated by using various cluster validity functions. Preliminary results showed that the effect of noise was considerably less with the new algorithm than with the
conventional FCM. Overall applying feature selection will always provide benefits such as providing insight into the data, better classifier model, enhance generalization and identification of irrelevant variables. For the results in this work, use the SVM classifier accuracy and the number of reduced features to compare the feature selection techniques. It has also successfully used Feature selection for improving predictor performance and for fault prediction analysis. Also, the performance achieved by the combination MFCMS-RST was better than some filter and wrapper approaches to feature selection proposed in the literature and applied to the same data sets. As shown in our experiments, proposed feature selection method can largely reduces the dimensions of medical expression Datasets and enhances the classification performance. Compared with previously published classifier methods, our proposed feature selection method achieved higher classification accuracy with much fewer features.