Data Preparation by CFS an Essential approach for decision making using C 4.5 for Medical data mining

Ashwinkumar.U.M[1]
[1]Reva Institute of Technology and Management, Bangalore-64
E-mail: ashu12330@rediffmail.com

DR Anandakumar.K.R[2]
[2]S J B Institute of Technology and Management, Bangalore-60

Abstract—**Trauma has become the leading cause of death in day to day life. Every year millions of people die and many more are handicapped due to various types of accidents caused by Trauma and many people become handicapped for the rest of their lives. It is necessary to develop a tool for predicting and preventing trauma. Reducing mortality rate and increasing the Health awareness is the aim. We have used the data mining process, to extract the useful data from large datasets. Feature subset selection is of immense importance in the field of data mining. The increased dimensionality of data makes testing and training of general classification method difficult. Mining on the reduced set of attributes reduces computation time and also helps to make the patterns easier to understand. The CFS approach for feature selection is proposed. As a part of feature selection step we used filter approach algorithm as random search technique for subset generation, wrapped with different classifiers/ induction algorithm namely decision tree C 4.5, Naïve Bayes, as subset evaluating mechanism on standard datasets. It is mandatory to obtain ethical and legal clearance from regional as well as Institutional Ethics Review Board (IERB), before using data mining tools in health care research. We got Ethical clearance from BGS Hospital for using the datasets. These datasets were gathered from the patient files which were recorded in the medical record section of the BGS Hospital Bangalore. Further the relevant attributes identified by proposed filter are validated using classifiers. Experimental results illustrate, employing feature subset selection using proposed filter approach has enhanced classification accuracy. Applying [DM] techniques to the data brings about very interesting and valuable results. It is concluded that in this case, comparing the result of evaluating the models on test set, decision tree works better than NaiveBayes. In this paper, we have also used WEKA Tool for creating the models.**

Keywords-CFS, DataMining, Trauma, GCS,CFS

I. INTRODUCTION

Trauma is an experience that is emotionally painful, distressful or shocking, which often results in lasting mental and physical effects. Every year millions of people die from trauma due to various types of accidents and many more handicapped

- Reducing the mortality rate and increasing Healthcare and Mortality Issues is one of the aims of the planners of development and decision makers
- Trauma involves an event (or repetitive events) and a reaction or response that includes an overwhelming experience of helplessness or powerlessness
- Extracting accident patterns is very important.

In this paper we have used the data mining methods on a real data set to identify the most frequent patterns of accident and know the effective condition in their occurrence. We have used the Decision tree, and Bayes' theorem and compared the results.

1.1.1 Motivation:

Data mining has attracted a great deal of attention in the information industry and in society as a whole in recent years, due to the wide availability of huge amounts of data and imminent need for turning such data into useful information and knowledge. The information and knowledge gained can be used for applications ranging from market analysis, fraud detection, and customer retention, to production control and science exploration

1.2. Introduction to Trauma:

Trauma refers to "a body wound or shock produced by sudden physical injury as from violence or accident." [21]. It can also be described as "a physical wound or injury, such as a fracture or blow. Trauma is the sixth leading cause of death worldwide, accounting for 10% of all mortality, and is a serious public health problem with significant social and economic costs. Trauma has become one of the leading causes of death in present world.

1.21 GCS Definition:

Glasgow Coma Scale or GCS is a neurological scale that aims to give a reliable, objective way of recording the conscious state of a person for initial as well as subsequent assessment. A patient is assessed against the criteria of the scale, and the resulting points give a patient score between 3 (indicating deep unconsciousness) and either 14(original scale) or 15 (the more widely used modified or revised scale).GCS was initially used to assess level of consciousness after head injury, and the scale is now used by first aid and doctors as being applicable to all acute

978-0-7695-4941-5/13 $26.00 © 2013 IEEE
DOI 10.1109/ACCT.2013.14

medical and trauma patients. In hospitals it is also used in monitoring chronic patients in intensive care. The scale was published in 1974 by Graham Teasdale and Bryan J. Jennett, professors of neurosurgery at the University of Glasgow's Institute of Neurological Sciences at the city's Southern General Hospital.

TABLE-I GCS scale range

	1	2	3	4	5	6
Eyes	Does not open eyes	Opens eyes in response to painful stimuli	Opens eyes in response to voice	Opens eyes spontaneously	N/A	N/A
Verbal	Makes no sounds	Incomprehensible sounds	Utters inappropriate words	Confused, disoriented	Oriented, converses normally	N/A
Motor	Makes no movements	Extension to painful stimuli (decelerate response)	Abnormal flexion to painful stimuli (decorticate response)	Flexion / Withdrawal to painful stimuli (decelerate response)	Localizes painful stimuli	Obeys commands

The scale comprises three Tests: eye, verbal and motor responses. The three values separately as well as their sum are considered. The lowest possible GCS (the sum) is 3 (deep coma or death), while the highest is 15 (fully awake person). It is as shown in the table 1.0. Individual elements as well as the sum of the score are important. Hence, the score is expressed in the form "GCS 9 = E2 V4 M3 at 07:35".

Generally, brain injury is classified as:

- Severe, with GCS ≤ 8
- Moderate, GCS 9 - 12
- Minor, GCS ≥ 13.

II. Methods

2.1 Feature selection algorithm

Feature selection (also known as subset selection) is a process commonly used in machine learning, wherein a subset of the features available from the data is selected for application of a learning algorithm [8]. The best subset contains the least number of dimensions that most contribute to accuracy; we discard the remaining, unimportant dimensions. This is an important stage of preprocessing and is one of two ways of avoiding the curse of dimensionality (the other is feature extraction).

- There are two approaches: Forward selection Start with no variables and add them one by one, at each step adding the one that decreases the error the most, until any further addition does not significantly decrease the error.
- Backward selection Start with all the variables and remove them one by one, at each step removing the one that decreases the error the most (or increases it only slightly); until any further removal increases the error significantly.

G. Ilczuk and A. Wakulicz-Deja [2] introduced a method of feature selection algorithm based on the Filter category. Additional advantage of feature selection is a reduction of search space, which, as presented in this paper and our entry research [16], reduces a number of decision rules (sometimes by factor 10) without compromising prediction accuracy. This fact is very important in medical domain where achieved results must be explainable and verifiable by experts. There are two main advantages of *Filter* algorithms over *Wrappers* based ones: they require significantly less computational effort and the achieved results do not depend on a specific learning algorithm.

- **QUICKREDUCT** defines a rough set-based attribute reduction family of feature selection algorithms based on concepts developed by Ziarko [17] and Modrzejewski [18]. The algorithm starts with an empty set of variables. Following heuristic is used to add variables to the initial set: the next variable chosen to be added to the candidate reduct is the variable that adds the most to the candidate reducts dependency. The hill climb ends when the dependency reaches one, or when no more variables are left. It must be mentioned, that QUICKREDUCT algorithm does not always generate a reduct. For some cases, the resulting attribute set will be a super -reduct, i.e. it will be possible to reduce it further. QUICKREDUCT in its search does not compromise with reducts offering a near-perfect consistency. It looks for a strict reduct which is not always desirable.
- **Chi-square test.** Chi-square ($\chi 2$) is a non-parametric test of statistical significance for bivariate tabular analysis (cross-breaks) [19]. Essentially the $\chi 2$ test is commonly used for testing independence and/or correlation between two vectors. The test compares observed frequencies with the corresponding expected frequencies. Value 0 of $\chi 2$ means that the corresponding two vectors are statistically independent with each other. At a certain threshold value (e.g., 3.84 at the 95% significance level [19]) an independence assumption between two vectors can be rejected. It can be said, that the higher value $\chi 2$ takes the higher the correlation between the corresponding vectors.

2.12 Classification

Classification can be used for predicting the class label of data objects. However, in many applications, one may like to predict some missing or unavailable data values rather than class labels. This is usually the case when the predicted values are numerical data, and is often specifically referred to as prediction. Although prediction may refer to both data

value prediction and class label prediction, it is usually confined to data value prediction and thus is distinct from classification. Prediction also encompasses the identification of distribution trends based on the available data

Data classification is a two step process. In the first step, a model is built describing a predetermined set of data classes or concepts. The model is constructed by analyzing database tuples described by attributes. Each tuple is assumed to belong to a predefined class, as determined by one of the attributes, called the class label attribute. In the context of classification, data tuples are also referred to as samples, examples, or objects. The data tuples analyzed to build the model collectively form the training data set. The individual tuples making up the training set are referred to as training samples and are randomly selected from the sample population. Since the class label of each training sample is provided, this step is also known as supervised learning (i.e., the learning of the model is 'supervised' In that it is told to which class each training sample belongs). It contrasts with unsupervised learning (or clustering), in which the class labels of the training samples are not known, and the number or set of classes to be learned may not be known in advance.

The accuracy of a model on a given test set is the percentage of test set samples that are correctly classified by the model. For each test sample, the known class label is compared with the learned model's class prediction for that sample. Note that if the accuracy of the model were estimated based on the training data set, this estimate could be optimistic since the learned model tends to over fit the data (that is, it may have incorporated some particular anomalies of the training data which are not present in the overall sample population). Therefore, a test set is used. If the accuracy of the model is considered acceptable, the model can be used to classify future data tuples or objects for which the class label is not known.

2.13 Prediction

Prediction can be viewed as the construction and usage of a model to assess the class of an unlabeled object, or to assess the value or value ranges of an attribute that a given object is likely to have. In this view, classification and regression are the two major types of prediction problems where classification is used to predict discrete or nominal values, while regression is used to predict continuous or ordered values. In our view, however, we refer to the use of predication to predict class labels as classification and the use of predication to predict continuous values (e.g., using regression techniques) as prediction. This view is commonly accepted in data mining

2.14 Classification by Decision Tree Induction

A decision tree is a flow-chart-like tree structure, where each internal node denotes a test on an attribute, each branch represents an outcome of the test, and leaf nodes represent classes or class distributions. The topmost node in a tree is the root node. A typical decision tree is shown in Figure 1.0. It represents the concept buys computer, that is, it predicts whether or not a customer at AllElectronics is likely to purchase a computer. Internal nodes are denoted by rectangles, and leaf nodes are denoted by ovals... Each internal (non-leaf) node represents a test on an attribute. Each leaf node represents a class (either buys computer = yes or buys computer = no).

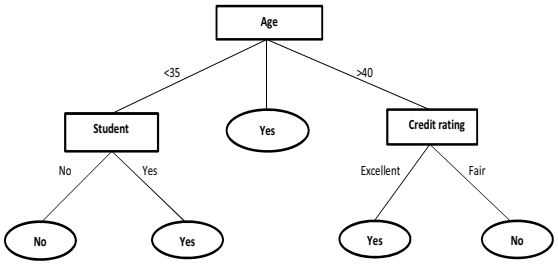

Figure 1.0: A simple decision tree.

In order to classify an unknown sample, the attribute values of the sample are tested against the decision tree. A path is traced from the root to a leaf node which holds the class prediction for that sample. Decision trees can easily be converted to classification rules.

2.15 Attribute Selection Measure.

The information gain measure is used to select the test attribute at each node in the tree. Such a measure is referred to as an attribute selection measure or a measure of the goodness of split. The attribute with the highest information gain (or greatest entropy reduction) is chosen as the test attribute for the current node. This attribute minimizes the information needed to classify the samples in the resulting partitions and reflects the least randomness or \impurity" in these partitions. Such an information-theoretic approach minimizes the expected number of tests needed to classify an object and guarantees that a simple (but not necessarily the simplest) tree is found. Let S be a set consisting of s data samples. Suppose the class label attribute has m distinct values defining m distinct classes, Ci (for i = 1….m). Let si be the number of samples of S in class Ci. The expected information needed to classify a given sample is given by:

$$I(s1, s2 \ldots, sm) = -\sum_{i=1}^{m} pi \log 2(pi) \qquad (1)$$

Where pi is the probability than an arbitrary sample belongs to class Ci and is estimated by si/s. Note that a log function to the base 2 is used since the information is encoded in bits.

Let attribute A have v distinct values, $\{a1, a2 \ldots, av\}$ avg. Attribute A can be used to partition S into v subsets, $\{S1, S2 \ldots, Sv\}$ Sv, where Sj contains those samples in S that have value aj of A. If A were selected as the test attribute (i.e., best attribute for splitting), then these subsets

would correspond to the branches grown from the node containing the set S. Let sij be the number of samples of class Ci in a subset Sj . The entropy, or expected information based on the partitioning into subsets by A is given by:

$$E(A) = \sum_{j=1}^{v} \frac{s1j + s2j + \dots + smj}{s} I(s1j, \dots smj) \quad (2)$$

The term $\sum_{j=1}^{v} \frac{s1j + s2j + \dots + smj}{s}$ acts as the weight of the jth subset and is the number of samples in the subset (i.e., having value aj of A) divided by the total number of samples in S. The smaller the entropy value is, the greater the purity of the subset partitions. The encoding information that would be gained by branching on A is

Gain(A)=I(s1,s2……..sm)-E(A) (3)

2.16 Tree Purning

When a decision tree is built, many of the branches will reflect anomalies in the training data due to noise or outliers. Tree pruning methods address this problem of over fitting the data. Such methods typically use statistical measures to remove the least reliable branches, generally resulting in faster classification and an improvement in the ability of the tree to correctly classify independent test data. How does tree pruning work?" There are two common approaches to tree pruning. In the prepruning approach, a tree is pruned" by halting its construction early (e.g., by deciding not to further split or partition the subset of training samples at a given node). Upon halting, the node becomes a leaf. The leaf may hold the most frequent class among the subset samples, or the probability distribution of those samples. The postpruning approach removes branches from a fully grown" tree. A tree node is pruned by removing its branches. The cost complexity pruning algorithm is an example of the postpruning approach. The pruned node becomes a leaf and is labeled by the most frequent class among its former branches. For each non-leaf node in the tree, the algorithm calculates the expected error rate that would occur if the subtree at that node were pruned. Next, the expected error rate occurring if the node were not pruned is calculated using the error rates for each branch, combined by weighting according to the proportion of observations along each branch. If pruning the node leads to a greater expected error rate, then the subtree is kept. Otherwise, it is pruned. After generating a set of progressively pruned trees, an independent test set is used to estimate the accuracy of each tree. The decision tree that minimizes the expected error rate is preferred.

2.17 Extracting Classification Rules from Decision Trees

IF-THEN rules. One rule is created for each path from the root to a leaf node. Each attribute-value pair along a given path forms a conjunction in the rule antecedent (IF" part). The leaf node holds the class prediction, forming the rule consequent (THEN" part). The IF-THEN rules may be easier for humans to understand, particularly if the given tree is very large [1].

The decision tree of Figure 1.0 can be converted to classification IF-THEN rules by racing the path from the root node to each leaf node in the tree. The rules extracted from Figure1.2 are:

IF age = \<30" AND student = no THEN buys computer = no.

IF age = \<30" AND student = yes THEN buys computer = yes.

IF age = \30-40" THEN buys computer = yes.

IF age = \>40" AND credit rating = excellent THEN buys computer = yes.

IF age = \>40" AND credit rating = fair THEN buys computer = no.

C4.5, a later version of the ID3 algorithm, uses the training samples to estimate the accuracy of each rule. Since this would result in an optimistic estimate of rule accuracy, C4.5 employs a pessimistic estimate to compensate for the bias. Alternatively, a set of test samples independent from the training set can be used to estimate rule accuracy.

2.17 Feature Selection Algorithm

Feature selection is often an essential data processing step prior to applying a learning algorithm [2]. If processed information contains irrelevant, unreliable or redundant data then a process of knowledge discovery is more difficult and achieved results are complicate to analyze. One way to remove the unneeded information is a selection of a subset of attributes from an original dataset for further processing. Depending on purposes of data mining this selection can be focused on:

- Finding the minimally sized feature subset that is necessary and sufficient to the target concept.
- Improving classification accuracy or decreasing a number of selected features without significantly decreasing the prediction accuracy of a selected classifier.

Feature Selection is a process that attempts to select a subset of features, satisfying a combination of application and methodology-dependent criteria: minimizing the cardinality of the feature subset; ensuring classification accuracy does not significantly decrease; and approximating the original class distribution with the class distribution given the selected features. Attribute selection techniques can be categorized using different criteria

One approach called *wrapper* uses a statistical re-sampling technique (e.g. cross validation) together with a target learning algorithm to estimate an accuracy of feature subsets [8]. As it is discussed in literature *wrapper* approach should result in a better prediction accuracy on new, unseen data. The main limitation of the approach is an additional computational cost resulting from frequently repeated cross-

validation. This process which is used for a validation of each feature subset makes this method inapplicable for processing of large datasets.

The second approach called *filter* as shown in Figure 1.1 uses a general characteristics of data to filter out undesirable features independently of a learning algorithm and without the knowledge of the classifying properties of the data [3]. Since *filters* are not coped with classifier systems or other learner they use quality metric to evaluate features. Four main types of such metric are commonly used:

- Distance metrics. During feature selection a distance between samples is kept maximal to improve separability.
- Information measures. They measure information gain for each feature (or subsets of features) and attempt to maximize it.
- Dependence/Correlation metrics. These metrics identify redundant features by calculating the correlation between them and other features. Afterwards redundant features are removed.
- Consistency metrics. This new class of metrics employs the training data to assess their consistency, given the subset of features currently evaluated.

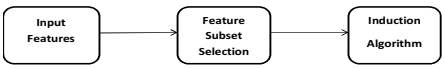

Figure 1.1: Filter based Feature Selection algorithm

Filter based feature selection algorithm are faster than wrapper method, hence they are best suited for medical domain datasets containing information about several thousand of patients described with hundreds of attributes are common we concentrate in our research on appliance of *filter* methods for feature selection

Filter class has 3 types:

- QUICKREDUCT
- CFS (Correlation based Feature selection)
- CHI-SQUARE SET
  - We have implemented feature selection algorithm is implanted CFS (correlation based feature selection).

2.18 CFS Method:

Described by Hall Correlation based Feature Selection (CFS) uses a heuristic which measures the usefulness of individual features for predicting the class label along with the level of intercorrelation among them [4]. This defined by equation 4 heuristic should filter out: irrelevant features because they are poor predictors of the class and redundant features which should be ignored because of their high correlation with each other [5].

$$H(s) = \frac{krci}{\sqrt{k + k(k-)}} \qquad (1)$$

H(s) is the heuristic function and *k* is the number of features in the subset, *rci* is the mean feature correlation with the class, and *rii* _ is the average feature intercorrelation. For computing the correlations necessary for equation (1) a number of information based measures of association were proposed such as: the uncertainty coefficient, the gain ratio or the minimum description length principle [6,4,5]. The below equation are used to calculate the Symmetric Uncertainty.

$$H(y) = \sum_{y=0}^{n} p(y) \log_2(p(y)) \qquad (2)$$

$$\text{Gain} = H(Y) - H(Y \mid X) = H(X) - H(X \mid Y)$$

$$\text{gain ratio} = \frac{\text{gain}}{2H(Y)}$$

$$\text{Symmetric Uncertainty} = 2x\left[\frac{\text{gain}}{H(y) + H(x)}\right] \qquad (5)$$

In our research we use CFS heuristic with Best First search strategy [7]. This strategy starts with an empty set of attributes and generates all single feature expansions which are possible. Then the subset of attributes with the highest value of the evaluation function is chosen and the procedure repeats. A stop criterion is defined as a number of subsets that results in no improvement.

2.19 C 4.5 Algorithm

We have used the C4.5 algorithm as the classification algorithm in our project for generating the decision trees. The C4.5 algorithm is having the best accuracy in the mentioned classification algorithm. It works even better when we use it after applying the feature selection algorithm.

C4.5 has additional features such as handling missing values, categorization of continuous attributes, and pruning of decision trees, rule derivation and others. C4.5 constructs a very big tree by considering all attribute values and finalizes the decision rule by pruning. It uses a heuristic approach for pruning based on the statistical significance of splits

2.20 Features of C4.5 Algorithm

There are several features of C4.5. Two features of C4.5 algorithm are discussed below.

A. Continuous Attributes Categorization: Earlier versions of decision tree algorithms were unable to deal with continuous attributes. 'An attribute must be categorical value' was one of the preconditions

for decision trees [34]. Another condition is 'decision nodes of the tree must be categorical' as well. Decision tree of C4.5 algorithm illuminates this problem by partitioning the continuous attribute value into discrete set of intervals which is widely known as discretization. For instance, if a continuous attribute *C* needs to be processed by C4.5 algorithm, then this algorithm creates a new Boolean attributes *Cb* so that it is true if $C<b$ and false otherwise [38]. Then it picks values by choosing a best suitable threshold.

B. Handling Missing Values: Dealing with Missing values of attribute is another feature of C4.5 algorithm. Some of these are Case Substitution, Mean Substitution, Hot Deck Imputation, Cold Deck Imputation, Nearest Neighbor Imputation [38]. However C4.5 uses probability values for missing value rather assigning existing most common values of that attribute. This probability values are calculated from the observed frequencies in that instance.

2.21 Limitations of C4.5

- **Empty branches:** Constructing tree with meaningful value is one of the crucial steps for rule generation by C4.5 algorithm. In our experiment, we have found many nodes with zero values or close to zero values. These values neither contribute to generate rules nor help to construct any class for classification task. Rather it makes the tree bigger and more complex.
- **Insignificant branches:** Numbers of selected discrete attributes create equal number of potential branches to build a decision tree. But all of them are not significant for classification task. These insignificant branches not only reduce the usability of decision trees but also bring on the problem of over fitting.

Over fitting happens when algorithm model picks up data with uncommon characteristics. This cause many fragmentations is the process distribution. Statistically insignificant nodes with very few samples are known as fragmentations [1]. Generally C4.5 algorithm constructs trees and grows it branches 'just deep enough to perfectly classify the training examples'. This strategy performs well with noise free data. But most of the time this approach over fits the training examples with noisy data. Currently there are two approaches are widely using to bypass this over-fitting in decision tree learning [39]. Those are:

- If tree grows very large, stop it before it reaches maximal point of perfect classification of the training data
- Allow the tree to over-fit the training data then postprune tree.

Thus we have used the C4.5 algorithm after applying the Feature selection algorithm with filter class method. The main objective of this project is to boost up the classification accuracy and simultaneously roll back timing to build a classification model. We have emphasized reducing reduce input space using entropy and several correlation coefficients formulas [40]. The proposed method shows better performance for each data file. The results demonstrate that the predictive power of the best models obtained with the three evaluated algorithms after feature selection was found to be in the range of 63.38±2.18–74.68±1.43%. The highest disease classification accuracy was reached by C4.5, which also provides the most informative model in the form of a decision tree [41].

To remove irrelevant and/or redundant features and to improve classification performance, feature selection was applied as a pre- processing step.

3.0 Data set Preparation

This system is used as an approach for using data mining in classifying mortality rate related to accidents. These data were gathered from the patient files which were recorded in the medical record section of the BGS Hospital in Bangalore.

Feature selection Pre-Processing is used in reducing the size of database, creating a more manageable set of attributes for modeling and time reduction in generating scores. By using this Pre-Processing the predictive model is only based upon a subset of predictors [35, 36].

Feature selection has three steps**:** Screening, Ranking and Selecting.

A. Screening step removes variables and cases that do not provide useful information for prediction. The following variables are removed:

- Variables that have missing values in more than 70% records
- Variables that represent case ID
- Categorical variables that have a single category counting for more than 70% cases

B. Ranking step considers one predictor at a time to see how well each predictor alone predicts the target variable. In the last step, between suggested fields, user can select some of them for participate in creating model. After execute screening steps on 166 Primary fields in a set of 100 patient records, 32 fields are suggested by Weka tool that we select 10 of them for creating model, these fields shown in table II.

TABLE-II Field guide

Field guide	Name of the field	Value set	Value description
1	Patient_Gender	[1, 2]	boy[1] , girl[2]
2	Age	[1,…,15]	
3	Trauma_Type	[1,…,9]	car accident[1], falling[2], bicycle[3], burning[4], drown[5], cutting and explosion[6], motor[7], botulism and poisoning[8], foreign object[9]
4.	Trauma_Location	[1,…,4]	Home , school and Club[1], avenue and street[2], pathway[3], nature[4]
5.	Trauma_Anatomic _ Location	[1,…,6]	Head[1], ribcage[2], tummy[3], backbone[4], dynamic organs[5] and multiple trauma[6]
6.	GCS	[1,…15]	
7.	Time_spacing	At hour	
8.	Bed_Ridden	At day	
9.	Dead_Type	[0,…,5]	Alive[0], shock[1], coma[2], arrest[3], cardiac[4],multiple organ failure[5]
10.	Reffrence_Type	[1,2,3]	Common people[1], Emergency ambulance[2], reference from other hospital[3]
11.	IS_Dead	[1, 2]	Dead[1], alive[2]

III RESULTS

A. Input File Formats: The inputs for the tool include the data files which have extensions like .rc and .DATA for Feature selection algorithm. The .rc file contains the description of the data file format, it gives details about the number of attributes, attribute values and the class labels and class values. The .DATA file contains the patient records, as per the description mentioned in the .rc file. The file formats of .rc and .DATA

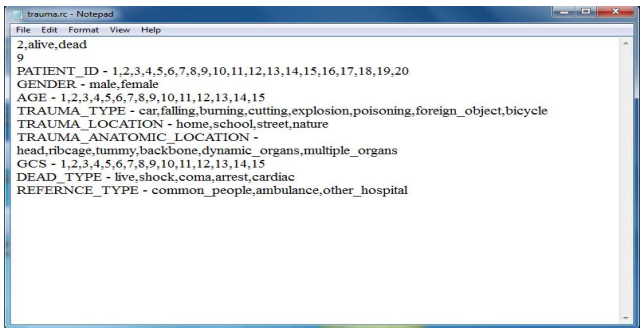

Figure II.I trauma ".rc" file

3.1Features Selection Process

The feature selection method is done as described. First the data file with extensions .rc and .DATA are taken as input to the Feature Selection algorithm, which are chosen from the file selection menu by clicking the file choose button from the tool as shown in the figure III. The selected file is displayed in the text area of the tool. Then the algorithm is sun by clicking RUN button in the tool. After running algorithm the output is displayed in the tool

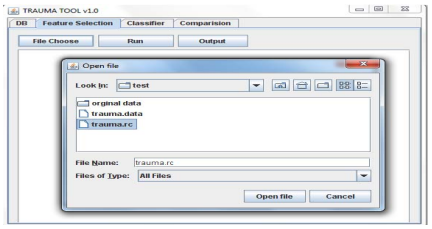

Figure III Feature selection process with File selection menu

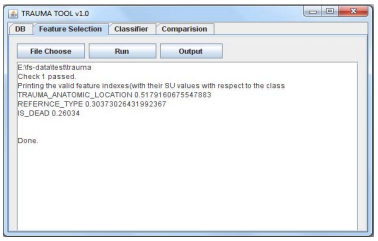

Figure IV Feature selection process output

The output consists of the subset of the attributes from the original data file, which are selected from the algorithm by correlation factor with the class. This is shown in the figure IV

3.2 Classification Process

The classification process is as described, the classification is done based on the attribute subset given out from the feature selection method. We have chosen the attribute either of IS_DEAD and Trauma _type attributes for root node for classification method. The output of the classification process is based on the root node, which gives the decision tree with prediction for the trauma in the patients and even the death mortality of the patients due to trauma is also predicted from the decision tree. The extractions of rules from the decision tree are used for the prevention of the trauma. The output of the classification process is as shown in the following figure V

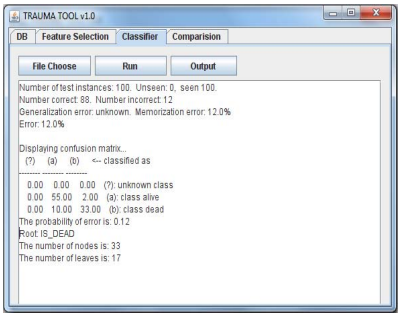

Figure V Classification process with output.

The output includes the number of attributes and the number of correctly classified attributes and incorrectly classified attributes. The confusion matrix is also displayed with the correct and incorrect predictions. The class value gives the death mortality rate of the hospital due to trauma. The decision tree nodes and leaves are also given it is shown as in the following figure V.

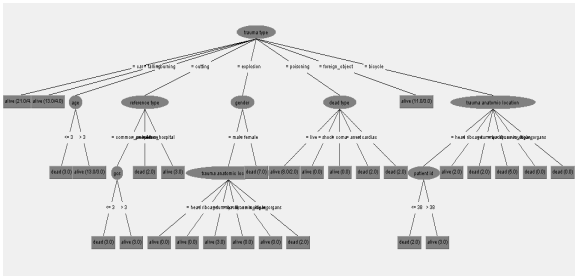

Figure VI Decision tree of trauma.

The decision tree is shown in either graphical representation or else in the textual levels format, which gives the tree in different levels. Each traverse from root node to leaf will generate a rule.

Finally the project's prediction methods, which is includes applying of the feature selection algorithm before C4.5 algorithm is compared with the WEKA's J48 algorithm and Navie Bayes algorithms. The comparison is shown in the graph plotted against each other in the. Following figure 7VII.9. The efficiency of all algorithms is shown in the graph.

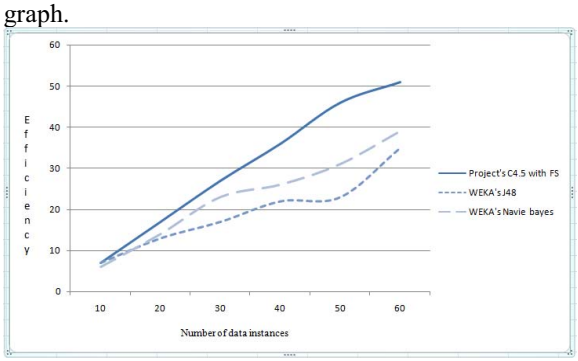

Figure VII Graph comparison

Finally the project's prediction methods, which is includes applying of the feature selection algorithm before C4.5 algorithm is compared with the WEKA's J48 algorithm and Navie Bayes algorithms. The comparison is shown in the graph plotted against each other in the following figure 7.9. The efficiency of all algorithms is shown in the graph

IV CONCLUSION AND FUTURE WORK

In this paper we have proposed the CFS approach for feature selection as a part of feature selection step, we used filter approach algorithm as random search technique for subset generation, wrapped with different classifiers/ induction algorithm namely decision tree C4.5.our results show better performance compared to J48 in weka tool and Naïve bayes, our future work is to use signal data and test the results.

ACKNOWLEDGMENT

We are thankful for DRDO for supporting this project.

REFERENCES

[1] Jiawei Han and Micheline Kamber Simon Fraser University "Data mining concepts and techniques " 2000.

[2] Grzegorz Ilczuk1 and Alicja Wakulicz-Deja2, "Selection of Important Attributes for Medical Diagnosis Systems" 2007.

[3] John, G. and Kohavi, R. and Pfleger, K.: Irrelevant Features and the Subset Selection Problem. In: International Conference on Machine Learning, New Jersey (1994), 121–129.

[4] Hall, M.: Correlation-based Feature Selection for Machine Learning. Ph.D diss.
Hamilton, NZ: Waikato University, Department of Computer Science, 1998.

[5] Ghiselli, E.: Theory of Psychological Measurement. McGraw-Hill Book, New York (1964)

[6] Quinlan, J. R: Induction of Decision Trees. In: Mach. Learn., (2003) 81–106

[7] Rich, E. and Knight, K.: Artificial Intelligence. McGraw-Hill Science, New York (1990)

[8] Feature Selection Martin Sewell 2007.

[9] KIRA, Kenji, and Larry A. RENDELL, 1992. A practical approach to feature selection. In: Derek H. SLEEMAN and Peter EDWARDS, eds. ML92: Proceedings of the Ninth International Conference on Machine Learning. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., pp. 249–256.

[10] JOHN, George H., Ron KOHAVI, and Karl PFLEGER, 1994. Irrelevant featuresand the subset selection problem. In: William W. COHEN and Haym HIRSH, eds. Machine Learning: Proceedings of the Eleventh International Conference. San Francisco, CA: Morgan Kaufmann Publishers, pp. 121–129.

[11] PUDIL, P., J. NOVOVI˙COV´A, and J. KITTLER, 1994. Floating search methods
in feature selection. Pattern Recognition Letters, 15(11), 1119–1125.

[12] KOLLER, Daphne, and Mehran SAHAMI, 1996. Toward optimal feature selection, Proceedings of the Thirteenth International Conference on Machine Learning. Morgan Kaufmann, pp. 284–292.

[13] JAIN, Anil, and Douglas ZONGKER, 1997. Feature selection: Evaluation, application, and small sample performance. IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(2), 153–158.

[14] DASH, M., and H. LIU, 1997. Feature selection for classification. Intelligent Data Analysis, 1(1–4), 131–156.

[15] WESTON, Jason, et al., 2001. Feature selection for SVMs. In: Todd K. LEEN,
Thomas G. DIETTERICH, and Volker TRESP, eds. Advances in Neural Information Processing Systems 13. Cambride, MA: The MIT Press, pp. 668– 674.

[16] Ilczuk, G. and Wakulicz-Deja, A.: Attribute Selection and Rule Generation Techniques for Medical Diagnosis Systems. In: RSFDGrC 2005, Regina (2005) 352–361

[17] Ziarko, W.: The discovery, analysis and representation of data dependencies in databases. In G. Piatesky-Shapiro andW. J. Frawley, editors, Knowledge Discovery in Databases. MIT Press, (1991)

[18] Modrzejewski, M.: Feature selection using rough sets theory. In Pavel B. Brazdil, editor, Proceedings of the European Conference on Machine Learning (ECML-93) 667 (1993) 213–226

[19] Everitt, B. S.: The analysis of contingency tables. Chapman and Hall, London (1977)

[20] Rich, E. and Knight, K.: Artificial Intelligence. McGraw-Hill Science, New York (1990)

[21] ^ "Trauma". *Dictionary.com*. Dictionary.com, LLC. 2010. Retrieved 2010-10-31

[22] Cem Kaner, Florida Institute of Technology, *Quality Assurance Institute Worldwide Annual Software Testing Conference*, Orlando, FL, November 2006

[23] Shafer, J., Agrawal, R., and Mehta, M. (1996). Sprint: A scalable parallel classifier for data mining. Proceedings of the 22nd international conference on very large data base. Mumbai (Bombay), India

[24] Matthew N. Anyanwu & Sajjan G. Shiva, Comparative Analysis of Serial Decision Tree Classification Algorithms International Journal of Computer Science and Security, (IJCSS) Volume (3) : Issue (3)

[25] Srivastava, A., Singh, V., Han, E., and Kumar, V .(1997). An efficient, scalable, parallel classifier for data mining . University of Minnesota, Computer Science, technical report, USA.

[26] Hunt, E.B., Marin. and Stone,P.J. (1966). Experiments in induction, Academic Press, New York.

[27] Podgorelec, V., Kokol, P., Stiglic, B., Rozman, I. (2002). Decision trees: an overview and their use in medicine, Journal of Medical Systems Kluwer Academic/Plenum Press, vol.26, Num. 5, pp.445-463.

[28] Quinlan, J. R. (1987). Simplifying decision trees, International Journal of Machine Studies, number27, pp. 221-234.

[29] Quinlan, J. R. (1986). Induction of decision trees. Machine Leaning, vol (1), pp.81-106

[30] Quinlan, J. R. (1993). C45: Programs for Machine Learning. Morgan Kaufmann, San Mateo, CA.

[31] Breiman, L., Friedman, J., Olshen, L and Stone, J. (1984). Classification and Regression trees. Wadsworth Statistics/Probability series. CRC press Boca Raton, Florida, USA.

[32] Lewis, R.J. (200). An Introduction to Classification and Regression Tree (CART) Analysis. 2000 AnnualMeeting of the Society for Academic Emergency Medicine, Francisco, California

[33] Anyanwu, M., and Shiva, S. (2009). Application of Enhanced Decision Tree Algorithm to Churn Analysis. 2009 International Conference on Artificial Intelligence and Pattern Recognition (AIPR-09), Orlando Florida.

[34] Mohammad Hossein Saraee, Zahra Ehghaghi, Hoda Meamarzadeh, Bahare Zibanezhad Islamic University of Najaf Abad Isfahan, Iran , "Applying Data Mining In Medical Data With focus on mortality related to accident in children" Proceedings of the 12th IEEE International Multitopic Conference, December 23-24, 2008

[35] G.Richards ,V.J Rayward Smith, P.H Sonksen, S.Carey, C.Weng ,"Data mining for indicators of early mortality in a database of clinical records" Elsevier, Artificial intelligence in medicine 22 (2001) 215

[36] SPSS Inc . "Clementine 12.0 Algorithms Guide". http://www.spss.com.

[37] Breiman, Friedman, Olshen, and Stone" C&RT".

[38] A. B. M. S. Ali and S. A. Wasimi, Data Mining: Methods and Techniques, Thomson Publishers,Victoria, Australia, 2007.

[39] M. Singh, How to Handle Missing Values, Articlebase, viewed on Oct 2009, at http://www.articlesbase.com/information-technologyarticles/ how-to-handle-missing-values-538449.html#.

[40] A. B. M. S. Ali and K. A. Smith, On learning algorithm for classification, *Applied Soft Computing*, Dec 2004. pp. 119-138.

[41] MOHAMMED M MAZID, A B M SHAWKAT ALI, KEVIN S TICKLE School of Computing Science Central Queensland University AUSTRALIA, "Improved C4.5 Algorithm for Rule Based Classification", RECENT ADVANCES in ARTIFICIAL INTELLIGENCE, KNOWLEDGE ENGINEERING and DATA BASES.

[42] Daria Prilutskya,b,c, BorisRogachevd, RobertS.Marksc,e,f,∗, LeslieLobela, MarkLastb, "Classification of infectious diseases based on chemiluminescent signatures of phagocytes in whole blood", Accepted 18 April 2011.