I) Title of the Thesis: Predictive Modeling Approach in Health Care Data Mining

I) Introduction:

Medical data mining has great potential for exploring the hidden pattern in the data sets of the medical domain. These patterns can be utilized for clinical diagnosis. However the available raw medical data are widely distributed, heterogeneous in nature, and voluminous. This data can be integrated to form a hospital information system [HK06]. Due to growing volume of data, medical treatment is facing a challenge of knowledge discovery. Healthcare environment is usually information rich but knowledge poor. However, data mining techniques can be applied to create knowledge rich healthcare environment. For example, applications of data mining techniques to Acquired Immune Deficiency Syndrome (AIDS) datasets can be highly challenging area. The threats to people's health from chronic disease are always existing and increasing gradually. How to decrease the threats is an important issue in Medical treatments [MJ07]. The application of data mining in medical and health researches had proved itself to be effective, showing great development potentialities. Now data mining has become a key method in obtaining information in clinical medicine, biomedicine, pharmacy and public health

Data Mining applied in Public Health

i. Spatio-temporal Data Mining used in infectious diseases monitoring to search for the epidemic rules and distribution characteristics of diseases.

ii. Using Time Series Analysis, Neural Networks to predict the incidence and infant mortality rate of infectious diseases.

iii. Association Rules to discuss the influencing factors of diseases and health seeking behavior.

Data Mining applied in Clinical Research

i. Finding the relationship among diseases.

ii. Searching the rule of disease development and prevalence.

iii. Disease diagnosis and treatment.

iv. Summarizing therapeutic effects.

e.g. using Bayes classification & decision tree classification in disease diagnosis.

Data mining is a process, which involves the application of specific algorithms for extracting patterns (models) from data. New knowledge may be obtained in the process while eliminating one of the largest costs, *viz.*, data collection. Medical data, for example, often exists in vast quantities in an unstructured format.

A new predictive modeling approach known as associative classification, integrating association Mining and classification inside into single system is being discussed as better alternative for predictive analytics [RV08]. Some of the classification techniques presented are CBA[LH98], CMAR[WL01], CPAR[YX03].As discussed in [LH98] it achieves higher classification accuracy than do traditional classification approaches such as C4.5, FOIL, RIPPER. According to [LH98] these traditional classifiers are faster but in many cases accuracy is not so high. Moreover many of the rules found by associative classification method cannot be discovered by traditional classification algorithm [LH98]. Given the readability of the associative classifiers, they are especially fit to applications were the model may assist domain experts in their decisions. Medical field is a good example were such applications may appear. Let us consider an example were a physician has to examine a patient. There is a considerable amount of information associated with the patient (e.g. personal data, medical tests, etc.). A classification system can assist the physician in this process. The system can predict if the patient is likely to have a certain disease or present incompatibility with some treatments. Considering the output of the classification model, the physician can make a better decision on the treatment to be applied to this patient[LA07].

Data mining technique have been successfully applied to several medical problem domain recent examples are as discussed in [ER08, MJ07, LA07, CO06, YJ06HK06, DW04]. Further Predictive Modeling Approach of Data Mining have been systematically integrated with Machine learning in [MJ07, YJ06] for the diagnosis of chronic disease. In this effort the classifiers has been used to discover the hidden rules among health examination data.

II) Brief Review of work already done in the field:

[LH09] Classification rule mining aims to discover a small set of rules in the database that forms an accurate classifier. Association rule mining finds all the rules existing in the database that satisfy some minimum support and minimum confidence constraints. For association rule mining, the target of discovery is not pre-determined. while for classification rule mining there is one and only one predetermined target. This paper, proposes to integrate these two mining techniques. The integration is done by focusing on mining a special subset of association rules, called *class association rules* (CARs). An efficient algorithm is also given for building a classifier based on the set of discovered CARs. Experimental results show that the classifier built this way is, in general, more accurate than that produced by the state-of-the-art classification system C4.5.

[WL01] A new associative classification method,, CMAR, i.e., **C**lassification based on **M**ultiple **A**ssociation **R**ules has been presented. This new method extends an efficient frequent pattern mining method, FP-growth to constructs a class distribution-associated FP-tree, and mines large database efficiently. It applies a CR-tree structure to store and retrieve mine association rules efficiently, and prunes rules effectively based on confidence, correlation and database coverage. The classification is performed based on a weighted $\chi 2$ analysis using multiple strong association rules. Experiments on databases from UCI machine learning database repository show that CMAR is consistent, highly effective at classification of various kinds of databases and has better average classification accuracy in comparison with CBA andC4.5. Moreover, the performance study shows that the method is highly efficient and scalable in comparison with other reported associative classification methods.

[YX03] is an extension of PRM which in turn is an extension of FOIL (Quinlan and Cameron-Jones 1993). The distinction between CPAR and PRM is that instead of choosing only the attribute that displays the best gain on each iteration (as in FOIL and PRM), CPAR may choose a number of attributes if those attributes have similar best gain. The distinguished feature of CPAR are

i.Uses Greedy approach and in rule generation

ii.Dynamic programming is used to avoid repeated calculation in rule generation.

iii.Selects multiple literals and builds multiple rules simultaneously to improve the time consuming part of FOIL

iv.Uses Laplace accuracy to evaluate rules and uses the best k rules in prediction.

[FT05] Compare the four recent associative classifiers algorithm(CBA, CMAR, CPAR, MCAR) with the goal to determine accurate classifiers. The result obtained on 12 different benchmark problems revealed that there is a consistency in between the rule learning associative algorithm in terms of predictive power. Also there was no dominant algorithm In terms of classification accuracy. However MCAR classification system have slightly better accuracy than the other ones on six data sets. The reason is that MCAR generates few more rules than the rest.

[FT07] Focuses on surveying and comparing the state-of-the-art associative classification techniques with regards to following aspect:

i. **Rule discovery**: Candidate generation, Frequent Pattern tree, Greedy Approach, Confidence rule, Confidence rule, Multi Support approach, Multipass atomic rules using dynamic support and adaptive confidence, The intersection of training objects, locations.

ii. **Rule ranking**: support confidence, cardinality method, distribution method

iii. **Rule pruning**: $\chi 2$ testing, redundant rule pruning, database coverage, Pessimistic error estimation, lazy pruning , Conflicting rules and impact of Pruning.

iv. **Rule prediction**: Maximum likelihood based prediction, Multiple rule based prediction, Score based prediction methods and laplace based prediction.

v. **Evaluation methods of resulting classifiers**: general evaluation measures in classification, associative classification evaluation measures

vi. Finally, future directions in associative classification, such as incremental learning and mining low-quality data sets (Missing data). Exponential growth of rules problem are also highlighted in this paper.

[RV08] have extremely compared various associative classifiers for their performance evaluations and this work uses LUCA S KDD cup data sets for investigating the performance of CPAR and CMAR. The paper records execution time for increasing number of records in a test set. The resulting graph of this work shows that the CMAR algorithm is slow in terms of execution time as compared to CMAR but it has shown to be more accurate than CPAR.

[KG08] Approaches the problem of generating negative rules from a classification perspective. how to generate a sufficient number of high quality negative rules efficiently so that classification accuracy is enhanced. A simple variant of Apriori algorithm was discussed which show that proposed classifier "associative classifier with negative rules"(ACN) is not only time-efficient but also achieves significantly better accuracy than four other state-of-the-art classification methods by experimenting on benchmark UCI datasets.

[AM04] A new associative classifiers that take the advantage of negative association rule mining and associative classifiers. These are two relatively new domains of research. The paper extends the concept of positive association rule of the form $X \rightarrow Y$ to $\neg X \rightarrow Y$, $X \rightarrow \neg Y$ and $\neg X \rightarrow \neg Y$ with the meaning X is for presence and $\neg X$ is for absence. The algorithm uses correlation coefficient to measure the interestingness of a rule. The algorithm has been tested for UCI datasets and encouraging results are obtained when both positive and negative rules are used for classification. Also accuracy decreases when using only negative rules for classification.

[ZC08] This proposes an associative classification approach, namely Classification with Fuzzy Association Rules (CFAR), where fuzzy logic is used in partitioning the domains.

In doing so, the notions of support and confidence are extended, along with the notion of compact set in dealing with rule redundancy and conflict. Furthermore, the corresponding mining algorithm is introduced and compared with traditional CBA approach on benchmarking datasets. following results are reported Accuracy of CFAR is higher than CBA.

- CFAR generated fewer rules than CBA, providing better understandability performance.
- Because of smooth boundaries, a fuzzy rule can cover more transactions than a crisp one with the same original attributes.

Sunita Soni, Jyothi Pillai, **A Case-Based Expert System For Weight Management Counseling To Obese Children**, Proceedings of International Conference on Resource Utilizat ion & Intelligent Systems, INCRUIS'08, Kongu Engineering College, Erode, 3rd- 5th Jan.'08. An expert system for weight management counseling to obese children integrates Data Mining and CBR techniques. The case retrieval method of CBES-WMC system uses Euclidean distance method to retrieve the most similar case to new case. from the case library.[INCRUIS-08]

Sunita Soni, Jyothi Pillai, **An Expert Case-Based System using Decision Tree Induction for Weight Management Counseling To Obese Children**, International Journal of Computer Science and Applications, IJCSA, Vol. 1, No. 2, 2008 and Proceedings of International Conference on Advanced Computing, ICAC'08 Decision tree analysis has long been used when a multi-stage decision process is involved. An expert case-based system using decision tree induction for weight management counseling to obese children uses ID3 to analyze the data set for representing relationship among data.[International Journal for Computer Science and Applications, IJCSA and ICAC-08]

Soni S., Pillai J., **Usage of Nearest Neighborhood, Decision Tree and Bayesian Classification Techniques in Development of Weight Management Counseling System,** Proceedings of **International Conference on Emerging Trends in Engineering and Technology-ICETET'08, G.H.Raisoni College, Nagpur** July 2008

and **IEEE-Xplore, IEEE Computer Society On-line store, IEEE Computer Society(CSDL) Digital Libraries, IEE INSPEC, EI(Compendex), Thomson ISI**Three Data Mining techniques, Nearest Neighborhood, Decision Tree and Bayesian Classification, were applied on distributed case bases for Case retrieval and Case adaptation phase of Weight Management Counseling system. On comparing the techniques, ID3 and Naïve Bayesian were found to be giving more precise results as compared to Euclidean Distance method. [On Line IEEE Digital Library & ICETET-08]

III) Note worthy contribution:

[MJ07] The paper suggest a model of a chronic disease prognosis and diagnosis system integrating Data Mining(DM) and case-based reasoning(CBR). And makes following critical contributions:

i. Adapting the data mining techniques to discover the implicit meaningful rules from Health Examination Data.

ii. Using the extracted rules for the specific chronic disease prognosis

iii. Results proved that the CDPD system can act as a medical expert system for discovering the useful rules from health examination data for supporting chronic diseases prognosis and diagnosis.

[CO06] Defined utilizes following constraints to reduce and generate relevant rules for heart disease prediction :

i. Item filtering,
ii. Attribute grouping,
iii. Maximum item size
iv. Antecedent / consequent rule filtering.

Association rules represent a promising technique to improve heart disease prediction. Unfortunately, when association rules are applied on a medical data set, they produce an extremely large number of rules. Most of such rules are medically irrelevant and the time required to find them can be impractical. A more important issue is that, in general, association rules are mined on the entire data set without validation on an independent sample. To solve these limitations, an algorithm that uses search constraints to reduce the

number of rules, searches for association rules on a training set, and finally validates them on an independent test set has been proposed. The medical significance of discovered rules is evaluated with support, confidence, and lift. Association rules are applied on a real data set containing medical records of patients with heart disease. In medical terms, association rules relate heart perfusion measurements and risk factors to the degree of disease in four specific arteries. Search constraints and test set validation significantly reduce the number of association rules and produce a set of rules with high predictive accuracy. The paper exhibit important rules with high confidence, high lift, or both, that remain valid on the test set on several runs. These rules represent valuable medical knowledge.

[ER08] This paper focuses a new algorithm called BitArrayNegativePos that mines both positive and negative rules from the real time
surveyed medical database. Association rules are defined as implication of the form A→B where A and B are frequent itemsets in a transaction database. This new algorithm extends this definition to include association rules of forms A → ^B, ^A→ B and ^A → ^B, which indicate negative associations between item sets is called negative rules. Negative rules are generated from infrequent itemsets. Rules of the form A→ B are called positive rules Negative rules are very useful in association analysis although they are hidden and different from positive rules. BitArrayÉclat algorithm extracts only positive rules. BitArrayNegativePos algorithm is able to find all valid rules in a support-confidence framework Experimental results show the efficiency of our new algorithm

[LA07] proposed the integration of new types of association rules and new methods to reduce the number of rules in the model. This research work studied the behavior of associative classifiers when negative association rules. Many applications can benefit from a good classification model. Given the readability of the associative classifiers, they are especially fit to applications were the model may assist domain experts in their decisions. **Medical field** is a good example was such applications may appear. Consider an example were a physician has to examine a patient. There is a considerable amount of information associated with the patient (e.g. personal data, medical tests, etc.). A

classification system can assist the physician in this process. The system can predict if the patient is likely to have a certain disease or present incompatibility with some treatments. Considering the output of the classification model, the physician can make a better decision on the treatment to be applied to this patient.

[HK06] Identify a few areas of healthcare where data Mining techniques can be applied for knowledge discovery from healthcare databases. This paper briefly examine the impact of data mining techniques, including Artificial Neural Networks on medical diagnostics. Data mining techniques can help in answering several critical questions. Following are few illustrations:

- Given the records of dialysis patients, what can be done to improve the treatment of these patients.
- Given the historical patient records on cancer, should the treatment include chemotherapy alone, or radiation alone or both chemotherapy and radiation.
- Can human DNA databases be characterized as genetic coding models.

A simple Structured Query language (SQL) can be to a dataset to obtain lot of information but not the hidden data. One can start with some simple statistical techniques such as averages. For example,

What is the average age of persons suffering from diabetes

- What is the average of persons suffering from heart disease
- What is the average period of survival after angioplasty/ heart operation/ cancer treatment
- What is the average age of person admitted to a hospital suffering from a heart problem / cancer/ diabetes
- What is the average hemoglobin of women of a particular community

It is interesting to see how these averages change when focus is on different disease or when focus is only on males or only on females or persons belonging to a particular area or particular regions. Statistics provide a strong fundamental background for quantification and evaluation of results. However, algorithms based on statistics need to be modified and scaled before they are applied to data mining.

IV) Proposed methodology during the tenure of research work:

Data Mining Techniques

Associative classification mining is a promising approach in data mining that utilizes the association rule discovery techniques to construct classification systems. It is a three-step process.

i. Generate the set of association rules from the training set with certain support and confidence thresholds as candidate rules.
ii. Pruning the set of discovered rules to weed out those rules that may introduce over fitting
iii. Classification Phase is the step to make a prediction for a new object.

Class Association Rule

For Associative Classification it is assumed that in training data set all continuous attributes (if any) have been discredited as a preprocessing step. For all attributes, all the possible values are mapped to a set of consecutive positive integers. With these mappings, a data case can be treated as a set of (attribute, integer-value) pairs and a class label. Each (attribute, integer-value) pair is called an item. Let D be the dataset. Let I be the set of all items in D, and Y be the set of class labels. We say that a data case $d \subseteq D$ contains $X \subseteq I$, a subset of items, if $X \subseteq d$. A classification rule (CAR) is an implication of the form $X \rightarrow y$, where $X \subseteq I$, and $y \subseteq Y$. A rule $X \rightarrow y$ holds in D with confidence c if c% of cases in D that contain X are labeled with class y. The rule $X \rightarrow y$ has support s in D if s% of the cases in D contain X and are labeled with class y.

Positive and Negative Association Rules and Interestingness measure

Over the years, a number of associative classifiers that utilizes advanced Association Rule Mining for example Association rule mining using Positive and Negative Rules. A new associative classifiers that take the advantage of negative association rule mining and associative classifiers. These are two relatively new domains of research. The concept of positive association rule of the form $X \rightarrow Y$ is extended to $\neg X \rightarrow Y$, $X \rightarrow \neg Y$ and

$\neg X \rightarrow \neg Y$ with the meaning X is for presence and $\neg X$ is for absence. The algorithm uses correlation coefficient to measure the interestingness of a rule.

Like positive rules, a negative rule $A \rightarrow \neg B$ also has a measure of its strength. conf. defined as the ratio, $supp(A \rightarrow \neg B)/supp(A)$.

By extending the definition in [RA94], negative association rule discovery seeks rules of the form $A \rightarrow \neg B$ with their support and confidence greater than, or equal to, user-specified minimum support and minimum confidence thresholds respectively, where

- A and B are disjoint itemsets, that is, $A \cap B = \emptyset$
- $supp(A) >= ms$, $supp(B) >= ms$ and $supp(A \cup B) < ms$;
- $supp(A \rightarrow \neg B) = supp(A \cup \neg B)$;
- $conf(A \rightarrow \neg B) = supp(A \cup \neg B)/supp(A) >= mc.$, then the rule $A \rightarrow \neg B$ is referred to as an interesting negative rule.

Pruning Technique: Identifying Interesting Item sets

As we have seen, there can be an exponential number of infrequent item sets in a database, and only some of them are useful for mining association rules of interest. Therefore, pruning is critical to efficient search for interesting item sets.

Piatetsky-Shapiro [1991]

Interestingness function $interest(X, Y) = |supp(X \cup Y) - supp(X)supp(Y)|$ and a threshold *mi* (minimum interestingness) can be used to defined interestingness.

If $interest(X,Y) >= mi$, the rule $X \rightarrow Y$ is of potential interest, and $X \rightarrow Y$ is referred to as a **potentially interesting item set**.

Using this approach, an effective pruning strategy for efficiently identifying all frequent item sets of potential interest in a database can be developed .

Medical Database

Data mining provides automatic pattern recognition and attempts to uncover patterns in data that are difficult to detect with traditional statistical methods. Without data mining it is difficult to realize the full potential of data collected within an organization as data under analysis is massive, highly dimensional, distributed and uncertain.

Medical data are characterized by there heterogeneity with respect to data type. These data may be noisy with erroneous or missing values. Below are some of the medical data types:

Treatment records

The records of millions of patients can be stored and computerized.

Pathology Reports

A pathology report is a written medical document which describes the analysis of specimens by the pathologist. These specimens are sent to the laboratory which contains the information about name, date of birth or age, patient ID number, the date and type of the procedure by which the specimen was obtained (blood sample, surgery, biopsy, etc.), and medical history and current clinical diagnosis of the patient.

ECG, X ray, MRI data

There is a need to develop methods for mining different type of data including X-ray, MRI images, electrocardiogram ECG signals and cholesterol level.

Patient history

Medical history of patients is an important factor in determining the nature of treatment for the patient. Following shows frame work for diagnostic tool.

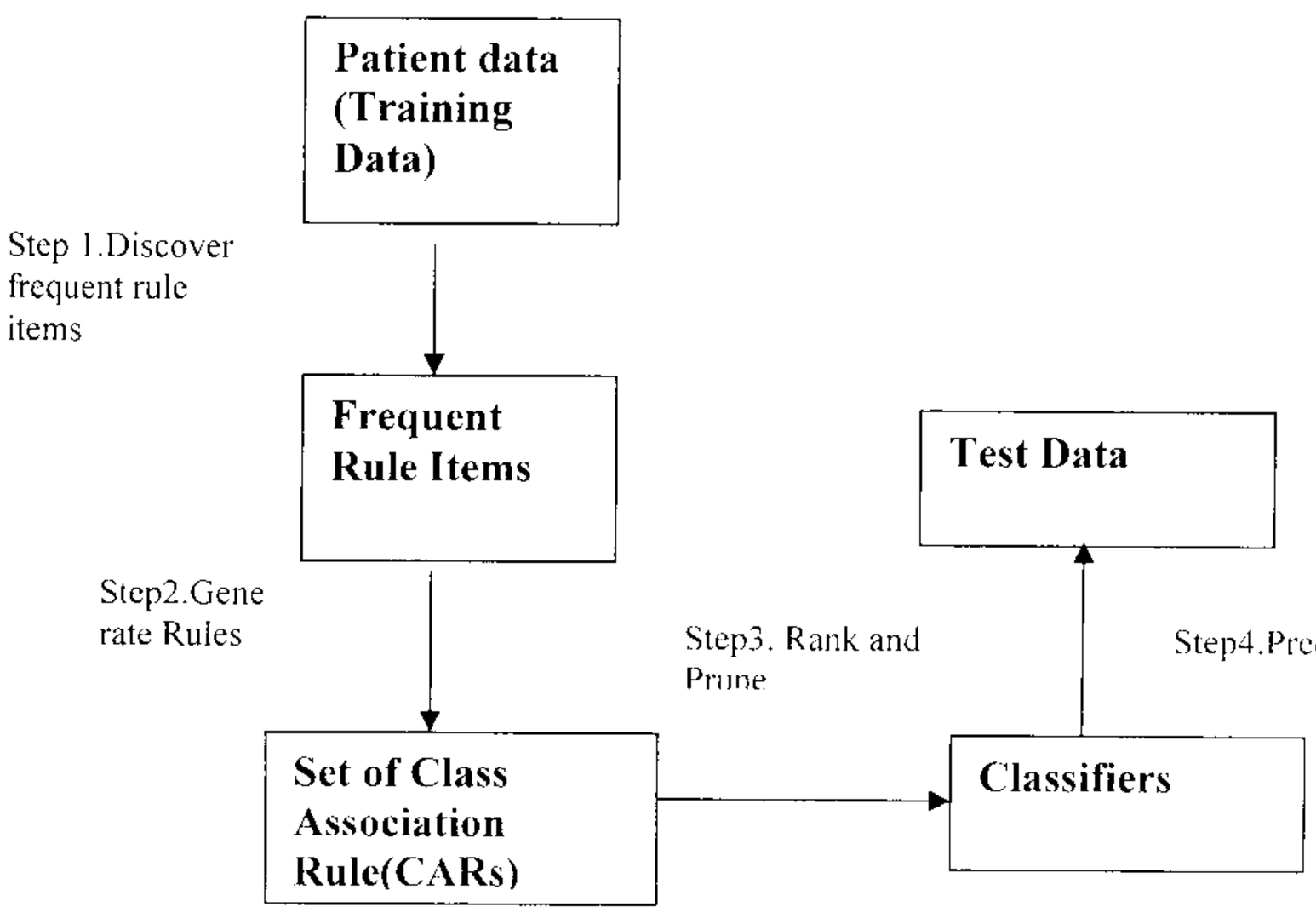

Fig1: Associative Classifier for Medical data

During the tenure of research work the following step will be followed.

- Data collection regarding patient (e.g. personal, Medical tests etc.), data preprocessing and data transformation will be first step to mine the real word health examination data. In case of not availability of real world data the work will be carried out on already available data like **adult.D131.N48842.C2.num, breast.num, heart.D53.N303.C5.num, hepatitis.D56.N155.C2.num**
- Factors like size of dataset, attribute types, number of attributes, etc. will be taken into consideration for analysis of performance existing data mining algorithms. If required the quantitative data may be fuzzified to improve the accuracy.
- Associative classifiers capable of generating both Positive and Negative Rules will be developed and applied on above data to find interesting rules.
- Pruning strategy will apply to get the relevant rules and to reduce the over fitting of data.
- The comparative study of proposed algorithm with other already available predictive mining system will be performed as a technique could work good in one Data set but may not for the others.
- The above rule and relationship will be used for predicting the chance of chronic diseases.
- A new algorithm to generate positive and negative association rule is proposed to be developed.

V) Expected Outcome of the proposed research work:

The proposed research work is expected to yield some significant findings and insight in the medical diagnosis and prognosis by using the appropriate data mining techniques on the Health Examination data. The expected outcome of the proposed work may be as follows –

- The proposed associative classifiers algorithm will improve the classification accuracy with respect to conventional classification approaches.

- The proposed algorithm will generate positive as well as negative rule to find interesting rules in the training data.

- The new algorithm will improve the performance of prediction/ Counseling system in medical domain, where high accuracy is required.

- This approach will open up a new avenue towards building highly specialized machine learning system.

- The proposed work will be very important and helpful for further improvement in the field of the healthcare.

VI) List of Publications

1) Sunita Soni, Jyothi Pillai, O.P.Vyas, The Predictive Analytics using Innovative Data Mining approach, **International Conference on Advanced data Analysis, Business Analytics and Intelligence, IIM, Gujarat** 6th-7th June'09. [Accepted]

2) Sunita Soni, Jyothi Pillai, Multi-Relational Classifiers: A survey, Proceedings of **Colloquium on Mobile Communication and Data and Knowledge Engineering, SOS in Computer Science and IT, Pt. R.S.S.U., Raipur**, 2nd -3rd , Jan.'09.[Accepted]

3) Sunita Soni, Jyothi Pillai, Case Representation and Retrieval in Weight Management Counseling to obese children: A Semi-Fuzzy Approach, Proceedings of **National Conference on Managing Complexity in Informative World, BITCON'08, BIT, Durg,** 7th-8th Nov.'08.[Published]

4) Soni S., Pillai J., Usage of Nearest Neighborhood, Decision Tree and Bayesian Classification Techniques in Development of Weight Management Counseling System, Proceedings of **International Conference on Emerging Trends in Engineering and Technology-ICETET'08, G.H.Raisoni College, Nagpur** on 16th – 18th July 2008 and **IEEE-Xplore, IEEE Computer Society On-line store, IEEE Computer Society(CSDL) Digital Libraries, IEE INSPEC, EI(Compendex), Thomson ISI.** [Published]

5) Sunita Soni, Jyothi Pillai, An Expert Case-Based System using Decision Tree Induction for Weight Management Counseling To Obese Children, **International Journal of Computer Science and Applications, IJCSA**, Vol. 1, No. 2, 2008 and Proceedings of **International Conference on Advanced Computing, ICAC'08**. [Published]

6) Sunita Soni, Jyothi Pillai, A Case-Based System using Naive Bayesian Classification for Treatment of Pediatric Obesity, Proceedings of **National**

Conference on Information Security : Practices and Challenges, ISAC'08. 9th-10th Feb.'08. [Published]

7) Sunita Soni, Jyothi Pillai, A Case-Based Expert System For Weight Management Counseling To Obese Children, Proceedings of **International Conference on Resource Utilization & Intelligent Systems, INCRUIS'08, Kongu Engincering College, Erode,** 3rd- 5th Jan.'08. [Published] [

8) Sunita Soni, Jyothi Pillai, A Case Based Framework for weight management counseling to obese children, Proceedings of **National Conference on Technological Advancements in Processing and Decision Making, BITCON'07, BIT, Durg,** on 16th -17th March' 2007. [Published]

9) Sunita Soni and M.V.Padmavati, Pattern Recognition with 3D Array Contextual Grammar" Proceedings of CMASM 2007 FIM XIV (COMPUTATIONAL, MATHEMATICAL AND STATISTICAL METHODS) The Fourteenth International Conference of the Forum for Interdisciplinary Mathematics from 06-08 January 2007 on "Formal Languages and Automata Theory" IIT Chennai. [Published]

10) Sunita Soni and M.V.Padmavati, Petrinets –A protocol Modeling System Proceedings of **National Conference on ,ACT, BIT, Durg,** on 2006. [Published]

VII) Bibliography:

[ER08]. E.Ramaraj, N. Venkatesan Positive and Negative association Rule Analysis in Health care, IJCSNS International Journal of Computer Science and Network Security, VOL.8 No.10, October 2008.

[ZC08] Zuoliang Chen, Guoqing Chen, **Bulding an Associative Classifiers based on Fuzzy Association Rules** International Journal of Computational Intelligence Systems, Vol.1, No. 3 (August, 2008), 262 - 273

[RV08] Ranjana Vyas Lokesh Kumar Sharma, L.K. Vyas, O.P. Scheider, S. **Associative Classifiers for Predictive Analytics: Comparative Performance Study,** Computer Modeling and Simulation, 2008. EMS '08. Second UKSIM European Symposium on 8-10 Sept. 2008. pp289-294.

[KG08] Kundu, G. Islam, M.M. Munir, S. Bari, M.F. **ACN: An Associative Classifier with Negative Rules. CSE '08. 11th IEEE International Conference on Computational Science and Engineering,** 16-18 July 2008.page(s): 369-375

[LA07] Luiza Antonie, University of Alberta, **Advancing Associative Classifiers - Challenges and Solutions,** Workshop on Macine Learning, Theory, Applications, Experiences 2007

[MJ07] Mu-Jung Huang, Mu-Yen Chen, Show-Chin Lee, **Integrating data mining with case-based reasoning for chronic diseases prognosis and diagnosis.** Expert Systems with Applications 32 (2007) 856–867, Science Direct.

[FT07] Fadi Thabtah, **A review of associative classification mining,** The Knowledge Engineering Review, Volume 22 , Issue 1 (March 2007),Pages 37-65, 2007.

[SY06] Sun, Y. Wang, Y. Wong, A.K.C., **Boosting an associative classifier, Knowledge and Data Engineering**, IEEE Transactions on Volume 18, Issue 7, July 2006 Page(s): 988 – 992

[CO06] Carloz Ordonez, **Association Rule Discovery with Train and Test approach for heart disease prediction,** IEEE Transactions on Information Technology in Biomedicine, Volume 10, No. 2, April 2006.pp 334-343.

[YJ06] Yoon-Joo Park, Byung-Chun Kim and Se-Hak Chun, **New knowledge extraction technique using probability for case-based reasoning: application to medical diagnosis,** Expert Systems, February 2006, Vol. 23, No. 1.

[HK06] Harleen Kaur , Siri Krishan Wasan and Vasudha Bhatnagar, **THE IMPACT OF DATA MINING TECHNIQUES ON MEDICAL DIAGNOSTICS,** Data Science Journal, Volume 5, 19 October 2006 pp119-126

[OP05] O.P.Vyas , Lokesh Sharma, Ranjana Vyas, U.S. Tiwary. **A Novel Approach of Multilevel Positive and Negative Association Rule Mining for Spatial Databases** LNAI 3587, Springer-Verlag Berlin Heidelberg 2005, pp. 620 - 629

[FT05] F Thabtah, P Cowling, Y Peng **A Study of Predictive Accuracy for Four Associative Classifiers** - JOURNAL OF DIGITAL INFORMATION MANAGEMENT, 2005 pp 209-212.

[AM04] Antonie, M. & Zaïane, O. **An associative classifier based on positive and negative rules.** In Proceedings of the 9th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery.

[DW04] D. Wu, Rosina Weber, and Fredric D. Abramson **:A case based-frame Framework for Leveraging Neutrigenomics Knowledge and Personalized Nutrition Counseling.** The 2004 European Conference in Case-based Reasoning.

[YX03] Yin, X. & Han, J. **CPAR: Classification based on predictive association rule.** In Proceedings of the SIAM International Conference on Data Mining. San Francisco, CA: SIAM Press, 2003.pp. 369-376.

[WL02] W.Li, J. Han, J.Pei , **CMAR- Classification based on Multiple Association Rules**, ICDM'01, , San Jose, CA, Nov. 2001. pp. 369-376

[LB97] Liu,B Hsu. W. Ma, **Integrating Clasification and association rule mining** . Proceeding of the KDD, 1998(CBA) pp 80-86.

[RA94] R. Agrawal and R. Srikant. **Fast algorithms for mining association rules**. In VLDB'94, , Santiago, Chile, Sept. 1994. pp. 487-499

1. Jiawei Han, Micheline Kamber, Book:"Data Mining Concept & Technique", 2001
2. Timothy J. Ross, a book on "Fuzzy Logic with Engineering Applications", McGraw Hill International Editions.
3. Arun K. Pujari, a book on "Data Mining Techniques", Universities Press.
4. Alex Berson, Stephen J. Smith, a book on "Data Warehousing, Data Mining and OLAP", Tata Mcgraw Hill Publishing Company Limited.
5. K.P. Soman, Shayam Diwakar,V .Ajay, a book on "Insight into Data Mining:Theory and Practice, PHI Private Limited.

Signature of Supervisor

(Dr. O. P. Vyas)

Signature of Candidate

(Sunita Soni)

Pt. Ravishankar Shukla University
RAIPUR. (C.G.)

Forwarded

Head
S.O.S in Computer Science
Pt. Ravishankar Shukla University,
Raipur (C.G.)

HEAD,
S. o. S. IN COMPUTER SCIENCE
Pt. RAVISHANKAR SHUKLA UNIVERSITY
RAIPUR (C. G.)