Data Mining in Clinical Records to foretell the risk of Osteoporosis

Ms.K.S.Sathawane¹, Prof.Ms.R.R.Tuteja² Computer Science and Engineering Department^{1,2}, P.R.M.I.T.R. Badnera Email: khushboo.sathawane@gmail.com¹, ranu.tuteja@gmail.com²

Abstract- In the healthcare sector quality demands are rising for designing expert systems for medical diagnosis. At the same time growing capture of biological, clinical, administrative data and integration of distributed and heterogeneous databases create a completely new base for medical quality and cost management. Clinical decisions are often made based on doctors' intuition and experience rather than on the knowledge rich data hidden in the database. This practice leads to unwanted biases, errors and excessive medical costs which affects the quality of service provided to patients. Data mining, have the potential to generate a knowledge-rich environment which can help to significantly improve the quality of clinical decisions.

Towards this background we applied intelligent data mining methods for analyzing medical repositories. This paper thus assess the role of the data mining techniques namely apriori algorithm, improved apriori algorithm and Fuzzy Logic rule based classifier in the diagnosis of risk of Osteoporosis in patients. It is based on the statistics already collected about the presence of Osteoporosis from patients data set. Fuzzy logic rule based classifiers can be used an effective tool for accurately diagnosing the severity of Osteoporosis. This paper presents the research in developing data mining techniques for predicting the risk of disease like osteoporosis prevalence. Osteoporosis is a bone disease that commonly occurs among postmenopausal women. Early detection and diagnosis is the key for prevention but are very difficult, without using costly diagnosing devices, due to complex factors involved and its gradual bone lose process with no obvious waning symptoms in particular. Our research aims to develop an intelligent decision support system based on data mining technology to assist General Practitioners in assessing patient's risk of developing osteoporosis. The prediction is based on historical data of the patients. The prediction will identify at-risk patients early and helps in starting the medication before doing the confirmatory test for Osteoporosis.

Index Terms- Datamining, Osteoporosis, Association Rules, Improved Apriori Algorithm, Fuzzy logic

1. INTRODUCTION

Knowledge discovery is the nontrivial extraction of implicit, previously unknown, and potentially useful information from data .In order to get this information, we try to find patterns in the given data set. Patterns that are interesting and certain enough according to the user's measures are called knowledge. The output of a program that discovers such useful patterns is called discovered knowledge. Data mining (DM) is a sub process of Knowledge Discovery in Databases in which the different available data sources are analyzed using various data mining algorithms. Data mining is a logical process that is used to search the relevant data from the large amounts of information or data. The goal of this technique is to find patterns that were previously unknown .Association rules mining is one of the most important tasks used in DM, which can be applied in different domains. Association rule discovery has been widely studied throughout the state-of-the-art techniques .Many data mining researchers had improved upon the quality of association rule for business development by incorporating influential

factors like utility, number of items sold and for the mining of association data patterns We believe that the field of mining frequent patterns and association rules mining is still a research area has raised interest among researchers because the researchers are working to provide effective and efficient methods.

SINCE the advent of advanced computing, doctors have always made use of technology to help them in various possible ways, from surgical imagery to X-ray photography. Unfortunately, technology has always stayed behind when it came to diagnosis and prediction of a risk, a process that still requires a doctor's knowledge and experience to process the sheer number of variables involved, ranging from medical history to climatic conditions, blood pressure, environment, and various other factors. The number of variables counts up to the total variables that are required to understand the complete working of nature itself, which no model has successfully analyzed vet. To overcome this problem, medical decision support systems such as [3] are becoming more and more essential, which will assist the doctors in taking correct decisions. A major challenge facing healthcare organizations (hospitals, medical centers) is the provision of quality services at affordable costs. Quality service implies

diagnosing patients correctly and administering treatments that are effective. Poor clinical decisions can lead to disastrous consequences which are therefore unacceptable. Hospitals must also minimize the cost of clinical tests. They can achieve these results by employing appropriate computer-based information and/or decision support systems.

Various algorithms and techniques like Classification, Clustering, Regression, Artificial Intelligence, Neural Networks, Association Rules, Decision Trees, Genetic Algorithm, Nearest Neighbor method etc., are used for knowledge discovery from databases

In this paper we have proposed an expert system based on clinical data mining which will make use of Association Rule Mining and Improved Apriori Algorithm ans fuzzy logic to find out certain association rules which will be useful in decision making for doctors to predict risk of a specific disease such as Osteoporosis. Osteoporosis is a bone disease that commonly occurs among postmenopausal women. Early detection and diagnosis is the key for prevention but are very difficult, without using costly diagnosing devices There are many significant risk factors that cause the osteoporosis in all over the world. Considering these factors, our system will derive association rules which will lead to statistical Analysis, and this knowledge is further used for decision making for whether or not the person is having Osteoporosis prevalence.

2. LITERATURE REVIEW

Shaikh Abdul Hannan et.al describes aiming to develop a expert system for diagnosing of heart disease using support vector machine and feed-forward backpropagation technique. Now a day's neural network are being used successfully in an increasing number of application areas. This work includes the detailed information about patient and reprocessing was done. The Support Vector Machine (SVM) and feed-forward Backpropagation technique have been applied over the data for the expert system.[5].

In the paper of Agrawal and Srikanth, the Apriori Algorithm [17] is used for Association Rule Mining to find out frequent dataset that satisfy the predefined minimum support and confidence from a given database. As computers are handy now-a-days, a viable CSCP system can be setup to collect large volumes of data and to stored in the database simultaneously. This kind of data includes the transaction records of clinics, hospitals, supermarkets, banks, stock markets, telephone companies and so on.

Wenjia Wang and Sarah Rea in there paper presents the research in developing an ensemble of data mining techniques for predicting the risk of osteoporosis prevalence in women. The paper focuses on investigating the methodologies for constructing effective ensembles, specifically on the measurements of diversity between individual models induced by two types of machine learning techniques, i.e. neural networks and decision tress for predicting the risk of osteoporosis. The constructed ensembles as well as their member predictors are assessed in terms of reliability, diversity and accuracy of prediction.[8]

The national Organisation of Osteoporosis Canada Practice Guidelines for the Diagnosis and Management of Osteoporosis of Canada makes evidence based recommendations to decrease the risk of fractures in the Canadian population. According to their research Patients with fragility fractures are at the highest risk of developing new fractures. Interventions can reduce that risk. They surveyed that Osteoporotic or fragility fractures are extremely common, more common than heart attack, stroke and breast cancer combined, the following statistics (see Figure1) shows this. At least one in three women and one in five men will suffer from an osteoporotic fracture during their lifetime.[13]



Figure 1: Statistic of Osteoporosis Fracture compared with others

3. SYSTEM ARCIETECTURE

Our proposed system architecture consist of the following parts:

3.1. Association Rule Mining

An Association Rule is a rule, which implies certain association relationships among a set of objects in a database. In this process people discover a set of association rules at multiple levels of abstraction from the relevant set(s) of data in a database. For example, one may discover a set of symptoms often occurring together with certain kinds of diseases and further study the reasons behind them. Since finding interesting association rules in databases may disclose some useful patterns for decision support, selective marketing, financial forecast, medical diagnosis, and many other applications, it has attracted a lot of attention in recent data mining research.

Mining association rules may require iterative scanning of large transactions or

relational databases, which is quite costly in terms of processing. Therefore, there is a need to study efficient mining of association rules in long transaction and /or relational databases. Or in other word, it is necessary to evaluate efficiency of association rule mining among different algorithms [17]

3.2. Apriori Algorithm

One of the first algorithms to evolve for frequent itemset and Association rule mining was Apriori. Two major steps of the Apriori algorithm are the join and prune steps. The join step is used to construct new candidate sets. A candidate itemset is basically an itemset that could either

be frequent or infrequent with respect to the support threshold. Higher level candidate itemsets (Ci) are generated by joining previous level frequent itemsets are Li-1 with itself. The prune step helps in filtering out candidate itemsets whose subsets (prior level) are not frequent. This is based on the anti-monotonic property as a result of which every subset of a frequent item set is also frequent. Thus a candidate item set which is composed of one or more infrequent item sets of a prior level is filtered(pruned) from the process of frequent itemset and association mining.

Agarwal and colleagues divided the problem of finding good rules into two phases:

- Find all itemsets with a specified minimal support (coverage). An itemset is just a specific set of items, e.g. {apples, cheese}. The Apriori algorithm can efficiently find all itemsets whose coverage is above a given minimum.
- 2. Use these itemsets to help generate interesting rules. Having done stage 1, we have considerably narrowed down the possibilities, and can do reasonably fast processing of the large itemsets to generate candidate rules.

3.2.1 Terminologies:

Support: an itemset has support *s*% if *s*% of the records in the DB contain that itemset.

minimum support: the Apriori algorithm starts with the specification of a minimum level of support, and will focus on itemsets with this level or above

large itemset: doesn't mean an itemset with many items. It means one whose support is at least minimum support.

 L_k : the set of all large *k*-itemsets in the DB.

 C_k : a set of *candidate* large *k*-itemsets. In the algorithm we will look at, it generates this set, which contains all the *k*-itemsets that might be large, and then eventually generates the set above.

3.1.1. Pseudo-Code:

Join Step: Ck is generated by joining Lk-1with itself

Prune Step: Any (k-1)-itemset that is not frequent cannot be

a subset of a frequent k-itemset

 $\begin{array}{l} C_k: \mbox{ Candidate itemset of size k} \\ L_k: \mbox{ frequent itemset of size k} \\ \end{array} \\ L_f = \{\mbox{ frequent items}\}; \mbox{ for } (k = 1; L_k != \emptyset; k^{++}) \mbox{ do begin} \\ C_{k+f} = \mbox{ candidates generated from } L_k; \mbox{ for each transaction } t \mbox{ in database do } \\ \mbox{ increment the count of all candidates in } C_{k+f} \\ \mbox{ that are contained in } t \\ L_{k+f} = \mbox{ candidates in } C_{k+f} \mbox{ with min_support } \\ \mbox{ end} \\ \mbox{ return } \cup_k L_k; \end{array}$

Figure 2: Pseudo Code for Apriori Algorithm

3.2.3 Example: Consider a database, D, consisting of 9 transactions(see Figure 3). Suppose min. support count required is 2 (i.e. min_sup = 2/9 = 22 %) Let minimum confidence required is 70%. We have to first find out the frequent itemset using Apriori algorithm.

Then, Association rules will be generated using min. support & min. confidence.

Step 1: Generating 1-itemset Frequent Pattern: The set of frequent 1-itemsets, L1, consists of the candidate 1-itemsets

satisfying minimum support. In the first iteration of the algorithm, each item is a member of the set of candidate.

Step 2: Generating 2-itemset Frequent Pattern: To discover the set of frequent 2-itemsets, L2, the algorithm uses L1 *Join* L1to generate a candidate set of 2-itemsets, C2.

Step 3: Generating 3-itemset Frequent Pattern: The generation of the set of candidate 3-itemsets, C3, involves use of the Apriori Property.In order to find C3, we compute $L2JoinL2.C3=L2 JoinL2 = \{\{11, 12, 13\}, \{11, 12, 15\}, \{11, 13, 15\}, \{12, 13, 14\}, \{12, 13, 15\}, \{12, 14, 15\}\}$.Now, Join step is complete and Prune step will be used to reduce the size of C3.

Step 4: Generating 4-itemset Frequent Pattern: The algorithm uses L3 *Join*L3to generate a candidate set of 4-itemsets, C4. Although the join results in $\{\{11, 12, 13, 15\}\}$, this itemset is pruned since its subset $\{\{12, 13, 15\}\}$ is not frequent.

Thus, C4= ϕ , and algorithm terminates, having found all of the frequent items.

The following Figure (see figure 4.)shows this.

TID	List of Items
T100	11, 12, 15
T100	12, 14
T100	12, 13
T100	11, 12, 14
T100	I1, I3
T100	12, 13
T100	I1, I3
T100	11, 12 ,13, 15
T100	11, 12, 13

Figure 3: Sample Transaction Database

Step 5:Generating Association Rules from Frequent Itemsets

Procedure:

•For each frequent itemset "*I*", generate all nonempty subsets of *l*.

•For every nonempty subset sof *I*, output the rule "s (l-s)"if support_count(l)/support_count(s)>=min_conf

where min conf is minimum confidence threshold.

•Back To Example:

We had $L = \{\{11\}, \{12\}, \{13\}, \{14\}, \{15\}, \{11,12\}, \{11,13\}, \{11,15\}, \{12,13\}, \{12,14\}, \{12,15\}, \{11,12,13\}, \{11,12,15\}\}$

- -Lets take $l = \{I1, I2, I5\}.$
- -Its all nonempty subsets are {I1,I2}, {I1,I5}, {I2,I5}, {I1}, {I2}, {I2}, {I5}.

Let minimum confidence thresholdis, say 70%.

The resulting association rules are shown below, each listed with its confidence.

-R1: I1 ^ I2 → I5

•Confidence = $sc\{I1, I2, I5\}/sc\{I1, I2\} = 2/4 = 50\%$

•R1 is Rejected.

 $-R2: I1 ^ I5 \longrightarrow I2$

•Confidence = $sc\{I1,I2,I5\}/sc\{I1,I5\} = 2/2 = 100\%$

- •R2 is Selected.
- -R3: I2 ^ I5 → I1

•Confidence = sc {11,12,15}/sc {12,15} = 2/2 = 100% •R3 is Selected.

- Step 5:Generating Association Rules from
- Frequent Itemsets
- •Confidence = $sc \{I1, I2, I5\}/sc \{I1\} = 2/6 = 33\%$

•R4 is Rejected.
-R5: 12 → 11 ^ 15
•Confidence = sc {11,12,15}/{12} = 2/7 = 29%
•R5 is Rejected.
-R6: 15 → 11 ^ 12
•Confidence = sc {11,12,15}/ {15} = 2/2 = 100%
•R6 is Selected.
In this way, We have found three strong association rules.

3.3 The Improved Apriori Algorithm

In our proposed approach, we enhance the Apriori algorithm to reduce the time consuming for candidates itemset generation. We firstly scan all transactions to generate L1 which contains the items, their support count and Transaction ID where the items are found. And then we use L1 later as a helper to generate L2, L3 ... Lk. When we want to generate C2, we make a self-join L1 * L1 to construct 2itemset C (x, y), where x and y are the items of C2.



Figure 4: Steps for implementing A priori Algorithm

Before scanning all transaction records to count the support count of each candidate, use L1 to get the transaction IDs of the minimum support count between x and y, and thus scan for C2 only in these specific transactions. The same thing for C3, construct 3-itemset C (x, y, z), where x, y and z are the items of C3 and use L1 to get the transaction IDs of the minimum support count between x, y and z, then scan for C3 only in these specific transactions and repeat these steps until no new frequent itemsets are identified.(see Figure 5)

3.3.1 Pseudo Code :

The improvement of algorithm can be described as follows:

```
//Generate items, items support, their transaction ID

(1) L_1 = find_frequent_1_itemsets (T);

(2) For (k = 2; L_{k,1} \neq 0; k++) {

//Generate the Ck from the L_{k,1}

(3) C_k = candidates generated from <math>L_{k-1};

//get the item I_w with minimum support in C_k using L_1, (1 \le w \le k).

(4) x = Get_item_min_sup(C_k, L_1);

// get the target transaction IDs that contain item x.

(5) Tgt = get_Transaction_ID(x);

(6) For each transaction t in Tgt Do

(7) Increment the count of all items in C_k that are found in Tgt;

(8) L_k= items in C_k \ge min_support;

(9) End;

(10) }
```

Figure 5: Pseudo Code for Improved Apriori Algorithm

3.3.2 Example:

Step 1: Suppose we have transaction set D has 9 transactions, and the minimum support = 3.

Step 2: firstly, scan all transactions to get frequent 1-itemset 11 which contains the items and their support count and the transactions ids that contain these items, and then eliminate the candidates that are infrequent or their support are less than the min_sup.

Step3: The next step is to generate candidate 2-itemset from L1. To get support count for every itemset, split each itemset in 2-itemset into two elements then use 11 table to determine the transactions where you can find the itemset in, rather than searching for them in all transactions. for example, let's take the first item (I1, I2), in the original Apriori we scan all 9 transactions to find the item (I1, I2); but in our proposed improved algorithm we will split the item (I1, I2) into I1 and I2 and get the minimum support between them using L1, here i1 has the smallest minimum support. After that we search for itemset (I1, I2) only in the transactions T1, T4, T5, T7, T8 and T9.

Step 4: The same thing to generate 3-itemset depending on L1 table.(see Figure 6)



Figure 6: Steps for Improved Apriori Implementation

3.4 Comparing Algorithms

In the previous example, if we count the number of scanned transactions to get (1, 2, 3)-itemset using the original Apriori and our improved Apriori, we will observe the obvious difference between number of scanned transactions with our improved Apriori and the original Apriori. From the table 1, number of transactions in1itemset is the same in both of sides, and whenever the k of k-itemset increase, the gap between our improved Apriori and the original Apriori increase from view of time consumed, and hence this will reduce the time consumed to generate candidate support count.

	Origional Apriori	Improved Apriori
1-itemset	45	45
2-itemset	54	25
3-itemset	36	14
sum	135	84

3.4 Data mining in Clinical Records

The World Health Organization (WHO) defines the criteria, shown in table-2 and established the definition of osteoporosis based on BMD as "faulty and weakened bone structure due to low amount of bone minerals per unit volume"4, The proposed system does not need of BMD, as calculating BMD from DEXA is an expensive scan and common people cannot afford this scan.

There are many significant risk factors for osteoporosis; the table-3 shows some of the factors that cause the osteoporosis in all over the world The multilayer perceptron network is the most often used in the medical diagnosis systems. We also proposed same network for predicting the future fracture risk. The attributes taken for diagnoses are; Age (months), Sex (male/female), Height (inch), weight (kg), Years since menopause (months), Heredity (Y/N), Cigarettes (per day), Alcohol (unit/day), eight bearing exercise (Y/N), Calcium in diet (mg/day), Low back pain (Y/N), Fracture (Y/N), Height loss (cm), Inactivity (Y/N), and Glucocorticoid (Y/N). These factors have significant impact on bone mass and trabecular microarchitecture. So, the listed attributes efficiently diagnose the fracture risk.

3.4.1 Clinical records are collected and stored in the large Database in $SQL/Access\,$

3.4.2 Pre-processing in performed and necessary data related to Osteoporosis is Extracted with the help of ETL process

3.4.3 With the help of Data mining software such as R-Tool, data mining algorithm such as improved apriori algorithm is implemented.

3.4.4.This algorithm generates association rules which can be used to develop patterns and gives statistical output, which is used for decision making.

3.4.5 In our paper, data of particular person is input to a form, from where our system will perform a decision making of whether or not Osteoporosis Prevalence is there in that patient.(see Figure 7).

Table 2: Reference values defined by WHO

Diagnosis	T-score Relative to Bone Mineral Density
Normal	BMD value with in 1 SD, (T-score -1)
Osteopenia	BMD value more then 1 SD below the mean and less then 2 SD below the mean, (-1 > T-score > -2.5)
Osteoporosis	BMD value 2.5 SD or more below the mean, $(T-score \le -2.5)$
Severe	BMD value 2.5 SD or more below the mean
Osteoporosis	with fragility fracture, (T-score \leq -2.5)

Table 3: Risk factors for Osteoporosis

abic 5. Risk lace	
General Factors	Family history of osteoporosis, Personal history of fragility fracture as an adult, Postmenopausal, Age 65 and above
Lifestyle Factors	Smoking, Obesity, Alcohol intake more then 2 units per day, Inadequate physical activity, Eating disorders
Nutritional Factors	Excess caffeine intake, Low calcium intake, Vitamin D deficiency, Low body weight (<127 pounds)
Medical Conditions	Weight loss surgery, Bone marrow transplantation, Female athlete triad, Hemophilia, Mastocytosis, Spinal cord transsection, Stroke, Thalassemia, Dialysis
Medications	Long-term glucocorticoid and heparin use, Anti-seizure medication, Aromatase therapy for breast cancer, Aluminum, Cytotoxic drugs, Proton pump inhibitors, Tamoxifen (premenopausal)



Fig 7: System Architecture for Data mining in Clinical Records

3.4.1 A Quantitative Approach:

Problem of classical association rules is that not every kind of data can be used for mining. Rules can only be derived from data containing binary data, where an item either exists in a transaction or it does not exist. When dealing with a quantitative database, no association rules can be discovered. This fact led to the invention of quantitative association rules, where the quantitative attributes are split into intervals and the single elements are either members or nonmembers of those intervals. With this approach, a binary database can be constructed out of a quantitative one.

The quantitative approach allows an item either to be member of an interval or not. This leads to an under- or overestimation of values that are close to the borders of such "crisp" sets. To overcome this problem, the approach of fuzzy association rules has been developed. It allows the intervals to overlap, making the set fuzzy instead of crisp. Items can then show a partial membership to more than one set, overcoming the above addressed, so-called "sharp boundary problem". The membership of an item is defined by a membership function and fuzzy set theoretic operations are incorporated to calculate the quality measures of discovered rules. Using this approach, rules can be discovered that might have got lost with the standard quantitative approach[see figure 8].

3.5 Classifiers

Classification is a technique that predicts categorical class labels. It classifies data (constructs a model) based on the training set and the values (class labels) in a classifying attribute and uses it in classifying new data. Classification is a two – step process consisting of Model Construction and Model Usage. Model Construction is defined as a process of describing a set of predetermined classes whereas Model Usage is helpful for classifying future or unknown objects. Formerly, three Classifiers Decision Tree, Classification via Clustering and Naive Bayes can be used to diagnose the presence of particular disease in patients[21].

Fuzzy logic is used to convert quantitative values of attributes to categorical ones so as to eliminate any loss of information arising due to sharp partitioning(using ranges) and then generate fuzzy association rules.

"Fuzzy logic may be viewed as an extension of multivalued logic. Its uses and objectives, however, are quite different. Thus, the fact that fuzzy logic deals with approximate rather than precise modes of reasoning implies that, in general, the chains of reasoning in fuzzy logic are short in length and rigor does not play as important a role as it does in classical logical systems. In a nutshell, in fuzzy logic everything, including truth, is a matter of degree."

Fuzzy logic deduces its greater expressive power from including probability theory and probabilistic logic.



Fig. 8: Fuzzy logic classifiers

4. OSTEOPOROSIS RISK FACTORS

The Osteoporosis Risk Factors are categorized into following two types:

Uncontrollable Risk Factors

- 1. Being over age 50.
- 2. Being female.
- 3. Menopause.
- 4. Family history of osteoporosis.
- 5. Low body weight/being small and thin.
- 6. Broken bones or height loss.

Controllable Risk Factors

- 1. Not getting enough calcium and vitamin D.
- 2. Not eating enough fruits and vegetables.
- 3. Getting too much protein, sodium and caffeine.
- 4. Having an inactive lifestyle.
- 5. Smoking.
- 6. Drinking too much alcohol.
- 7. Losing weight.

According to our algorithm, the following sample database is of around 200 patients and their Osteoporosis risk factors, as shown in Figure 9. The No(%) shows the minimum support count.

By applying our Algorithm and setting minimum support, we can get various Association Rules in order to predict the prevalence of Osteoporosis

Characteristics	No. (%)
Descriptive	
Age, y	
45-64	79 (39)
≥65	123 (61)
Weight, kg	
<60	58 (29)
60-69	74 (37)
≥70	70 (35)
Osteoporosis risk factors	
Estrogen use	
Never	120 (59)
Ever	82 (41)
Current	43 (21)
Fracture history*	
Yes	42 (21)
No	160 (79)
Smoking status	
Never	118 (58)
Ever	84 (42)
Family history of fracture+	
Yes	13 (6)
No	189 (94)
BMD	
Femoral neck T score±	
Greater than -2.0	95 (47)
-2.0 or less	107 (53)
-2.5 or less	69 (34)
Age 45-64 y	11 (5)
Age ≥65 v	58 (29)
Lumbar spine T score‡	()
-2.5 or less	15 (7)
Age 45-64 y	3 (1)
Age ≥65 v	12 (6)

Figure 9: Summary of descriptive characteristics Osteoporosis risk factors of around 200 postmenopausal Women

5.IMPLEMENTATION

Our Clinical Descision Support System looks like this: Step 1: First step is to open the database of clinical history of patients. Browsing through the folder where data is kept. (See Figure 10)



Figure 10: Opening Database

Clinical history of 100s of patients is stored in .xls file, the records contains relevant factors viz. uncontrollable parameters such as Patient ID, Age>65,Gender,Whether menopaused, BMI<18, History of Osteoporosis (Family), Fragility Fracture(Personal), and Controllable factors such as Estrogen Intake,Smoking, Alcoholic Intake > 2units/D, Inactive Lifestyle, and whether suffering from Osteoporosis. The Data is feed in Yes/No form.

Durin II	run mi	5 m	Chinica	ii nec	orus to) 10r	eten tu	e risk	or st	eem	c Disea	ise	
				(Oste	oporos	is) P	revaler	ice					
					Input D	ata For	Association	Rule M	ning				
Open Dataset	The?	Same :	Cruisers/Hitestopicdata.com										
0		1	2	3	4	5	6	T	5	2	10	11	
Prepressions case	1	Patient k	Age+65	Gender	Meropaused	DIII < 12	Hadory of Oa.	Fragility Frac	Estropen Inte.	Smeking	Alcoholic Inte.	inactive Lifes.	Oate
	2	P001	Yes	Male	Yes	No	Yes	Yes	Vez	Yes	Yes	168	yes
Apply Apron Algo	3	P002	Yes	Female	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No.	Yes
	4	P003	Yes	Female	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Yes
Anda Train	5	P004	No	Female	No	Yes	No	Yes	Yes	No	Yes	No.	No
	6	P005	Yes	Female	Yes	No	Yes	Yes	Yes	No	Yes	No.	Yes
	7	P006	Yes	Female	Yes	No	Yes	Yes	Yes	No	Yes	No.	Yes
Clessify		POOT	Yes	Female	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Yes
	9	P005	Yes	Female	Yes	Yes	Yes	Yes	Yes	No	Yes	50	Yes
	20	P009	No	Fenale	No	Yes	79.9	Yes	Yes	No	Yes	760	No
	22	P010	Yes	Mole	No	Yes	Yes	Yes	Yes	No	199	52	No
	22	P011	Yes	Fenale	Yes	Yes	Tres	No	Yes	No	Yes	50	No
	13	P012	Yes	Fenale	Yes	190	Yes	Yes	No	No	Yes	52	Yes
	- 54	P013	Yes	Fenale	Yes	Yes	Yes	Yes	Yes	No	Yes	59	Yes
	15	P014	Yes	Fenale	140	Yes	759	Yes	Yes	140	Yes	59	No
	36	P015	Yes	Fenale	140	140	Yes	Yes	Yes	140	Yes	59	No
	17	P016	Yes	Male	140	Yes	Yes	Yes	Yes	140	Yes	59	No
		96.47	Vee	Excepte	Yes	No	Yes	Yes	Yes.	No	Yes	59	Vec

Figure 11: Clinical database

Step 2: Preprocessing the Data: Here the textual data is converted into numbers.



Figure 12: Preprocessed data

Step 3: Applying our improved apriori algorithm for finding association rules, providing Minimum Support Threshold = 0.03 and Minimum Confidence Threshold = 0.01



Figure 13: Applying Improved Apriori Algorithm

Step4:Processing dataset with minimum support threshold = 0.03 Displaying Frequent Items found above Threshold Support and the rules generated,



Figure 14: Diplaying Rules

Step 5:Performing ANFIS Training and the Output of ANFIS Training



Figure 15: Output 1 of ANFIS Training



Figure 16: Output 2 of ANFIS Training

6. CONCLUSION

The development of Information Technology has generated large amount of databases and huge data in various areas. The research in databases and information technology has given rise to an approach to store and manipulate this precious data for further decision making.

This suggestion is promising as data modeling and analysis tools, e.g., data mining, have the potential to generate a knowledge-rich environment which can help to significantly improve the quality of clinical decisions.

In this paper, an improved Apriori is proposed through reducing the time consumed in transactions scanning for candidate itemsets by reducing the number of transactions to be scanned. This algorithm scans the transaction database D only when L1-candidates is produced, the rest scan the preceding result instead of the

transaction database, greatly reducing I/O load and increase efficiency. The interesting association rule can be accessed more effectively by the improved Apriori algorithm. Existing electronic medical details obtained from hospitals are utilized as training data set for analysis. Algorithm is used in mining the training data set, which is turn produces association rule and with the help of classifiers such as fuzzy logic, it discovers implicit and potential useful knowledge from large preprocessed databases.

In other words, the objective of this paper is to develop a decision support system that identify persons who are susceptible to Osteoporosis and raise an awareness among health care providers and patients. This will be useful in primary health centers where equipments for measuring bone strength are not easily available.

ACKNOWLEDGMENTS

Our thanks to the experts who have contributed towards development of the paper.

REFERENCES

- K.S.Kavitha, K.V.Ramakrishnan, Manoj Kumar Singh "Modeling and design of evolutionary neural network for heart disease detection"IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 5, September 2010 ISSN (Online): 1694-0814 www.IJCSI.org 272
- [2] Abdul Basit Shaikh, Muhammad Sarim, Sheikh Kashif Raffat, Kamran Ahsan, Adnan Nadeem and Muhammad Siddiq "Artificial Neural Network: A Tool for Diagnosing Osteoporosis" Research Journal of Recent Sciences ISSN 2277-2502Vol. 3(2), 87-91, February (2014) Res.J.Recent Sci. International Science Congress Association 87 Available online at: www.isca.in, www.isca.me
- [3] R. A. Miller, "Medical diagnostic decision support systems—Past, present, and future—A threaded bibliography and brief commentary,"J. Amer. Med. Inf. Assoc., vol. 1, pp. 8–27, 1994.
- [4] Chen, J., Greiner, R.: Comparing Bayesian Network Classifiers. In Proc. of UAI-99, pp.101–108,1999.
- [5] Shaikh Abdul Hannan, V. D. Bhagile R. R. Manza, R. J. Ramteke, —Diagnosis and Medical Prescription of Heart Disease Using Support Vector Machine and Feed forward Back propagation techniquel, International Journal on Computer Science and Engineering, pp 2150-2159, 2010.
- [6] Mrs. S. V. Shinde, Dr. U.V. Kulkarni "Mining Classification Rules from Database by Using Artificial Neural Network" IIJACKD JOURNAL OF RESEARCH | VOL 1 | ISSUE 1 | FEBRUARY 2012
- [7] Ghazi Johnny "Developing A-priori Algorithm for Fast Mining Association Rules+" http://www.iasj.net/iasj?func=fulltext&aId=32331
- [8] Wenjia Wang and Sarah Rea "Intelligent Ensemble System Aids Osteoporosis Early Detection"[Proceedings of the 6th WSEAS Int. Conf. on EVOLUTIONARY COMPUTING, Lisbon, Portugal, June 16-18, 2005 (pp123-128)
- [9] Mahmood A. Rashid1, Md Tamjidul Hoque2, Abdul Sattar1 "Association Rules Mining Based Clinical Observations"paper available at cornell university library site 1Institute for Integrated and Intelligent Systems (IIIS), 2Discovery Biology, Eskitis Institute for Cell & Molecular Therapies, Griffith University Nathan, QLD, Australia

- [10] Babak Taati, Jasper Snoek, Dionne Aleman, and Ardeshir Ghavamzadeh "Data Mining in Bone Marrow Transplant Records to Identify Patients With High Odds of Survival" IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS, VOL. 18, NO. 1, JANUARY 2014
- [11] Yanchang Zhao "Book on R and Data Mining: Examples and Case Studies 1"yanchang@rdatamining.com http://www.RDataMining.com August 7, 2014
- [12] Professor Anita Wasilewska Lecture Notes on "APRIORI Algorithm" http://

www3.cs.stonybrook.edu/~cse634/lecture_notes/07apri ori.pdf

- [13] National Institute of Osteoporosis Canada "OSTEOPOROSIS Towards a Fracture-Free Future" Osteoporosis Canada <u>www.osteoporosis.ca</u>]
- [14] Ruijuan Hu "Medical Data Mining Based on Association Rules"-Published by Canadian Center of Science and Education www.ccsenet.org/cis Computer and Information Science Vol. 3, No. 4; November 2010 108 ISSN]
- [15] Mohammed Al-Maolegi , Bassam Arkok "AN IMPROVED APRIORI ALGORITHM FOR ASSOCIATION RULES" International Journal on Natural Language Computing (IJNLC) Vol. 3, No.1, February 2014
- [16] Xiang Fang An Improved Apriori Algorithm on the Frequent Itemse International Conference on Education Technology and Information System (ICETIS 2013)
- [17] Agrawal and Srikant, Fast algorithm for Mining Association Rules, in The 20th VLDB conference Santiago, Chile, 1994.
- [18] Jiao Yabing Research of an Improved Apriori Algorithm in Data Mining Association Rules International Journal of Computer and Communication Engineering, Vol. 2, No. 1, January 2013
- [19] Shruti Aggarwal, Ranveer Kaur Comparative Study of Various Improved Versions of Apriori Algorithm International Journal of Engineering Trends and Technology (IJETT) - Volume4Issue4- April 2013
- [20] Wenjia Wang and Sarah Rea, Intelligent Ensemble System Aids Osteoporosis Early Detection Proceedings of the 6th WSEAS Int. Conf. on EVOLUTIONARY COMPUTING, Lisbon, Portugal, June 16-18, 2005 (pp123-128)
- [21] Heydar Jafarzadeh*, Mehdi Sadeghzadeh, ImprovedApriori Algorithm Using Fuzzy Logic International Journal of Advanced Research in Computer Science and Software Engineering Volume 4, Issue 6, June 2014 ISSN: 2277 128X