International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-8 Issue-4, February 2019

Associative Analysis among Attribute of ILPD Medical Datasets Using ARM

Ramjeevan Singh Thakur

Abstract*: *Early detection of liver disease plays a major role in efficient diagnosis the disease. It significantly increases the chance of effective treatment. The liver is one of the largest organs in the human body. It plays an important role in digestion, as detoxifying chemicals in the digestion process. A dreadful fact of liver disease is that, the liver maintains a normal functionality even after partially damage. The major challenge in liver disease is to find the hidden patterns of liver disorder. The proposed approach analysis the patterns on the selected features using association rule mining (ARM) technique. The performance of the proposed approach is tested on the well-renowned ILPD dataset from the UCI repository. ILPD dataset consists of different clinical examination parameter like total bilirubin, direct bilirubin, SGPT, SGOT, alkphos, total protein, albumin etc. The proposed approach selected the essential features from ILPD and ARM is applied to find the association among attributes to detect pattern.

Index Terms*: *Indian Liver Patient Datasets, Association rule mining, Liver Disorder, Associative Analysis.

I. INTRODUCTION

Liver is one of the important gland in human and play a vital role in functioning of metabolism and other different crucial functioning including red blood cells decomposition, detoxification, digestion and removing of different hazardous elements from human body[1, 2].

Liver disorder is one of the most widespread problems in human body due to industrialization. There is a number of liver disorders that require early detection and clinical care of physician. Liver disorder is caused by damaged hepatocyte which is infected due to bacteria, fungi or viruses, and intake of toxins like alcohol. Liver disease has been consistently listed as one of the top ten fatal diseases in worldwide. A dreadful fact is that discovering of liver disease is a complex task, due to capability of liver that maintain a normal functionality even after partially damage. Early detection of live disorder is required for effective diagnosis [1-4].

In recent years, the rapid developments in information technology have increased the data communication around the world. And different devices make it possible to communicate access and manage clinical data. Different data mining techniques can be applied on these data to extract useful and valuable hidden patterns [3, 5-11].

Various medical examinations, namely, health, biochemical and lifestyle are done by researchers to generate the clinical data for liver disorder. This clinical data have been used in the proposed approach and hidden patterns are found for early diagnosis of liver disease.

Manuscript published on 28 February 2019.
***Correspondence Author(s)
Ramjeevan Singh Thakur, Department of Computer Applications, MANIT, Bhopal, India.

The outcome of the work can be assist by the physician to enhance the quality of medical diagnosis and reduce the uncertainties and potential risk [9].

Association rule mining is a technique of extracting relationship between different variable in a large database. An association rule consists of if/then statements to explore the relationships between unrelated data in a database or other data repository. It is used to find the association among variables that occur frequently. Association rules are used in many different application areas including medical diagnosis, bioinformatics, intrusion detections, security applications and many more [12-20].

II. LITERATURE REVIEW

As per literature, there are number of studies have been conducted on Liver data and Indian Liver Patient Datasets (ILPD) to predicted liver disorder in human being. Most of the work are dedicated to liver data classification using varies classifier and prediction methods [1, 3, 5-11]. The related work on different liver datasets is summarized in Table 1; here works are elaborated at different parameters such as research area, principles task, techniques that is used findings, datasets used and issues etc.

There is significantly less work available on association rule mining for Indian Liver Patient Datasets, so this issue is considered in this paper and various formations of attributes are explored to predict the perfect cause of liver disorder.

III. RULES MINING PROCESS IN ILPD DATASET

The proposed methodology is based on a very first approach of Association rule mining i.e. Apriori algorithms, where every k-itemsets is formed by frequent k-1-itemsets and all subsets of k-1-itemsets must be frequent[12, 21]. In this analysis ILPD datasets is used for finding association rules among attributes of dataset. Each attribute treated as item and each patient record as transaction in datasets. The whole process of rule mining in ILPD dataset is shown in Figure 1, where it is divided into six steps namely Data collection, Data cleaning, Data discretization, Feature selection, Rule mining and Pattern evaluations[21].

A. Data Collection

ILPD Data is collected from UCI repository; it consists of 10 attributes with class label of liver disorder and normal liver case. It contains total 583 instances including 416 liver patient and 167 non-liver patient records. The complete description of each attribute is shown in Table 2.

Table 1: Summary of related work

S. No.	Author & Year	Area	Principle Task	Technique used	Findings	Dataset used	Issues
1.	Avi Ben-Cohen, 2018[1]	Liver Lesion Detection	Classification	Fully convolutional network	Better detection capabilities, including of small lesions,	CT Examinations	Critical cancer diagnosis
2.	Rong-Ho Lin Chun-Ling Chuang,2010[2]	Liver Diseases prediction	Classification	Hybrid diagnosis model	Diagnose the existence of disease with greater accuracy and confidence	Liver patient data	Efficiency
3.	Maayan Frid-Adar et al., 2018[3]	Liver lesion classification	Classification	Generative Adversarial Networks (GANs)	Improve performance on a medical problem with limited data	Synthetic data	Inaccuracy
4.	CEZARY Z. JANIKOW, 1996[4]	Liver Diseases predicti-on	Optimization	Fuzzy decision tree	Optimizing Fuzzy Decision Trees	-	Biasness
5.	Tapas RanjanBaitharua, Subhendu Kumar Pani, 2016[5]	Decision Support System for Healthcare	Classification	J48, Naive Bayes, ANN, ZeroR, 1BK and VFI	Comparative analysis of data classification approaches using Liver disorder data	Synthetic data	Inaccuracy
6.	Yugal Kumar* and G. Sahoo, 2013[6]	liver diseases prediction	Classification	SVM, RI, DT, NB, ANN	Rule based classification model	Synthetic data	Improper classificati-on
7.	S. Dhamodharan 2014[7]	Liver diseases prediction	Classification	Bayesian Classification	Classification accuracy better using Naïve Bayes algorithm gives better accuracy for classification	Synthetic data	Improper classificati-on
8.	Nazmun Nahar and FerdousAra, 2018[8]	Liver disease prediction	Classification	Decision Tree	Decision Stump provides the highest accuracy	UCI repository	Inaccuracy
9.	P. Rajeswari, G.SophiaReena, 2010[9]	Analysis of Liver Disorder	Classification	FT Tree	Better accuracy shows by FT tree	UCI repository	Inaccuracy
10.	Ayesha Pathan et al., 2018[10]	Comparative Study of Different Classificat-ion algorithms	Classification	Classification Algorithms	Random Forest algorithm provides best performance	ILPD Dataset	Inaccuracy
11.	P. Thangaraju, R. Mehala, 2015[11]	Liver Diseases Classificat-ion	Classification	PSO-KStar Classifier	Enhanced the performance of PSO-KStar Classifier	UCI repository	Inaccuracy

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

B. Data cleaning

One of the attribute albumin and globulin ratios having some of missing values. To find missing values, we applied replace by mean method and replaces all missing values by mean value as about 80% record belongs to liver disorder patient, so mean value also should not fall into normal range of albumin and globulin ratio as per medical standard norms.

C. Data Discretization

As attribute of dataset is characterized as multivariate with integer and real values. Further it is not possible to generate patterns followed by rules with the present situation[21]. So, attribute values are converted as some categorical values, such as age is converted in three values i.e. child, adult and old, gender already given in Male and Female category, for remaining eight attributes (TB, DB, ALP, SGPT, SGOT, TP, ALB,A/G Ratios) related data is converted as low, normal and high range values. For categorizing into lower, normal and high range value of each attributes the norms provided by Gastrocare Liver & Digestive Disease Centre Bhopal are followed because different pathology center follows slightly different standards. After data discretization sample of datasets is shown in Table 3.

D. Feature Selection

The consideration of all 10 attributes to generate candidate itemsets lead to complex and time consuming task because some of the attributes are not important candidates. Most of them may infrequent in further stages. To reduce this type of infrequent candidate feature selection is one of the important methods. Attributes those are not much dependent on targeted attribute, may be removed from input datasets. There are number of approaches available for this purpose. We have used chi-square technique to remove comparatively less important attributes. After applying Chi-square technique the rank of attributes are obtained [21]. The Direct Bilirubin got rank I, Alamine Aminotransferase rank II, Aspartate Aminotransferase rank III, Total Bilirubin rank IV, Alkaline phosphotase rank V, Albumin rank VI, Age rank VII, Gender rank VIII, Total protein IX and Albumin and Globulin ratio X. So, out of 10 attributes only first six attributes i.e. Direct Bilirubin, Alamine Aminotransferase, Aspartate Aminotransferase, Total Bilirubin, Alkaline phosphotase and Albumin are selected for rule mining.

E. Rule Generation

Apriori algorithm is used to generate frequent patterns and rule generation from ILPD datasets. The Apriori algorithm is a basic technique to find frequent itemsets from a set of data by applying candidate generation method. Items that occur often together can be associated to each other, these together occurring items form a frequent itemsets. It apply an iterative approach known as a level-wise search as the k-itemsets is used to find the (k + 1)-itemsets [12, 21]. This search begins with search of 1-itemsets denoted F_1. F_1 is then used to find the set of frequent 2-iemsets, F_2. F_2 is then used to find F_3 and so on. This iteration is repeated until no more frequent k-itemsets can be found.

An association rule is a pattern of the form X & Y $\Rightarrow$ Z (*support, confidence*), where X, Y, and Z are items in the dataset. Usefulness of a rule is measured with a minimum support and minimum confidence threshold value. It measure how many events have such itemsets.

***Support*:** For an itemset X, this relation compares the number of events containing itemset X to number of all events in database [21].

Confidence: Certainty of a rule is measured with this threshold value. It lets to measure how often an event's itemset that matches the left side of the implication in the association rule also matches for the right side [21].

F. Pattern Evaluation & Presentation

This is last step of process, where rules are evaluated based on interestingness of decision maker. In general, association rules with low support and high confidence value are accepted [21]. In this analysis 0.1 to 0.4 minimum support value is used to generate the rules and 91% to 99% confidence is considered for rule acceptance. Finally, rules and patterns can be represented in more interpretable form like histogram, bar chart, pie chart etc.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

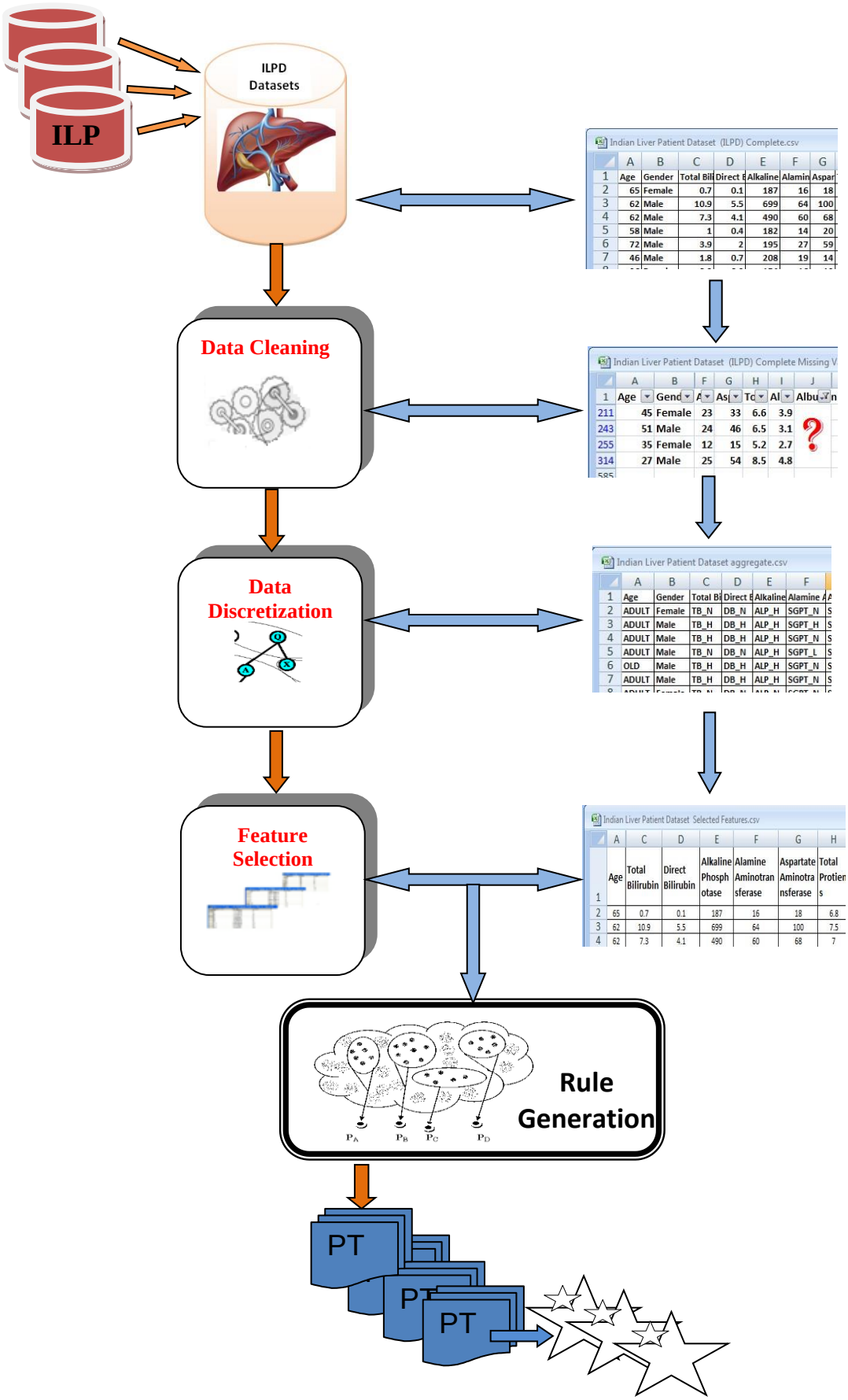

Fig. 1: Rule Mining process in ILPD Dataset

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

Table 2: ILPD Datasets with attribute description

S. No.	Attribute Name	Attribute Description	Attribute Value	
			Min.	Max.
1	Age	Patient age in years	04	90
2	Gender	Patient Gender	Male & Female	
3	Total Bilirubin	Total Bilirubin (TB) in **mg/dL**	0.4	75.0
4	Direct Bilirubin	Direct Bilirubin (DB) in **mg/dL**	0.1	19.7
5	Alkaline Phosphotase	Alkaline Phosphotase (Alkphos) **in U/L**	63	2110
6	Alamine Aminotransferase	Alamine Aminotransferase-ALT level (SGPT) **in U/L**	10	2000
7	Aspartate Aminotransferase	Aspartate Aminotransferase- AST level (SGOT) **in U/L**	10	4929
8	Total Protiens	Total Protiens (TP) **in gms/dL**	2.7	9.6
9	Albumin	Albumin (ALB) **in gms/dL**	0.9	5.5
10	Albumin and Globulin Ratio	Albumin/ Globulin Ratio (A/G Ratio)	0.3	2.8
11	Pora	Data is split into two sets	1- Liver disorder 2- Normal Liver	

Table 3: ILPD Dataset after data discretization process

Age	Gender	Total Bilirubin	Direct Bilirubin	Alkaline Phosphotase	Alamine Aminotransferase	Aspartate Aminotransferase	Total Protiens	Albumin	Albumin and Globulin Ratio
ADULT	Female	TB_N	DB_N	ALP_H	SGPT_N	SGOT_N	TP_N	ALB_N	AGR_L
ADULT	Male	TB_H	DB_H	ALP_H	SGPT_H	SGOT_H	TP_N	ALB_N	AGR_L
ADULT	Male	TB_H	DB_H	ALP_H	SGPT_N	SGOT_H	TP_N	ALB_N	AGR_L
ADULT	Male	TB_N	DB_N	ALP_H	SGPT_L	SGOT_N	TP_N	ALB_N	AGR_L
OLD	Male	TB_H	DB_H	ALP_H	SGPT_N	SGOT_H	TP_N	ALB_L	AGR_L
CHILD	Male	TB_N	DB_N	ALP_N	SGPT_N	SGOT_N	TP_N	ALB_N	AGR_L
ADULT	Male	TB_N	DB_N	ALP_H	SGPT_N	SGOT_H	TP_N	ALB_N	AGR_L

IV. EXPERIMENTAL RESULTS AND ANALYSIS

For experimental purpose, we performed association rule mining using Weka tool. All experiments are done on ILPD data collected from UCI repository [22]. Once data is collected, missing value in some attributes are filled by mean value of remaining data of particular attribute. Further, data categorization is done to convert data in proper format for applying association rule mining. As the number of features in the database in very large, only selected number of features are used to generate the fruitful result. Experiments are performed separately on complete database (i.e. consists of normal and liver disorder data) and partial database (i.e. only liver disorder data). For generating frequent pattern/rules minimum support value is considered 0.1 to 0.4 in case of complete database. In case of only liver disorder data the minimum support value is 0.2 to 0.3, and we got some of the best rules at each case. Figure 2 is showing number of frequent itemsets generated at various support values like at min. support = 0.15, 1-itemset are 12, 2-itemset are 43, 3-itemset are 57, 4-itemset are 28 and 5-itemset are 04.

Similarly at min. support = 0.35 the value of 1-itemset are 10, 2-itemset are 15, 3- itemset are 02 and no further itemsets.

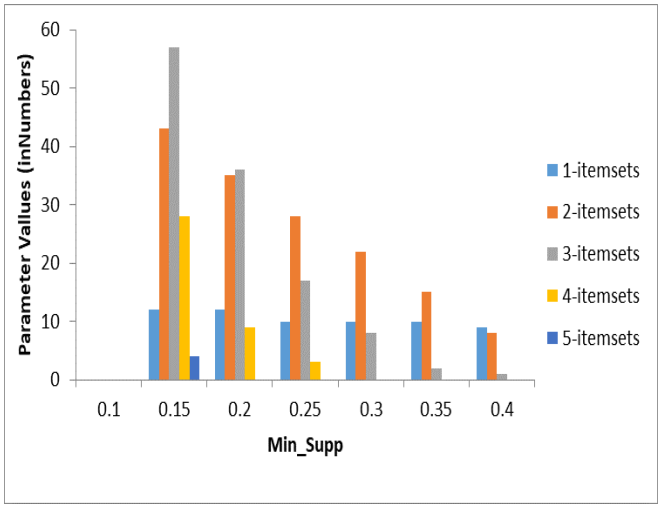

Fig 2: No. of generated items in both normal and liver disorder case

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

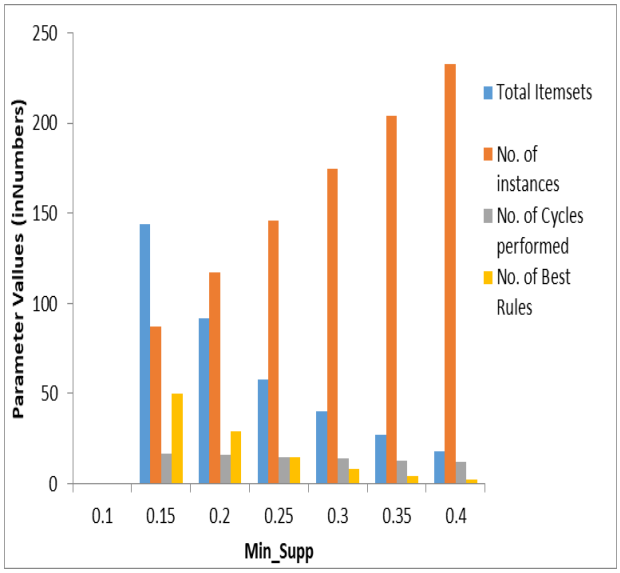

Fig 3: No. of best rules with other parameter in both normal and liver disorder case in ILPD.

The no. of best rules at various support value is generated in complete ILPD dataset is shown in fig. 3. As shown in figure at min. support 0.15, no. of best rules generated are 50, total no. of items (include all length) are 144. A number of 17 cycles were required to generate the best rules, and total no. of 87 instances were participated for best rule generation.

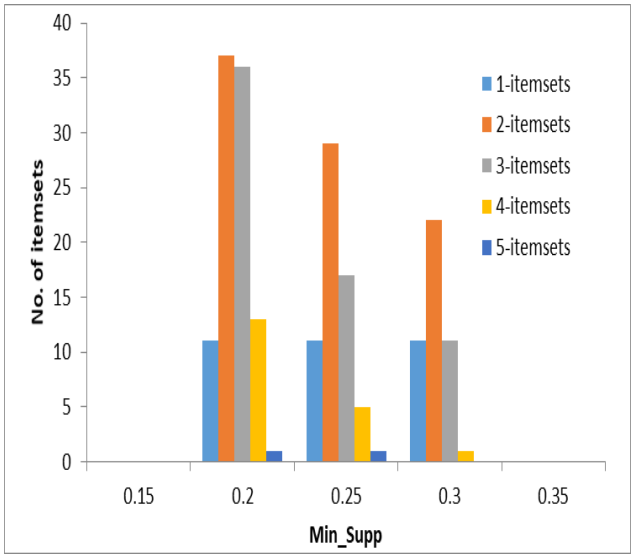

Fig 4: No. of generated items in liver disorder case only

Figure 4 is showing number of frequent itemsets generated at various support value like at min. support = 0.20, the number of 1-itemset are 11, 2-itemset are 37, 3- itemset are 36, 4-itemset are 13 and no further itemset. Similarly, at min. support = 0.30 the number of 1-itemset are 11, 2-itemset are 22, 3-itemset are 11, 4-itemset is 01 and no further itemset.

In all cases maximum length of itemset (parameter) is 4 i.e. one parameter called Alamine Aminotransferase (SGPT) was not frequent and it did not play a role in detection of liver disorder.

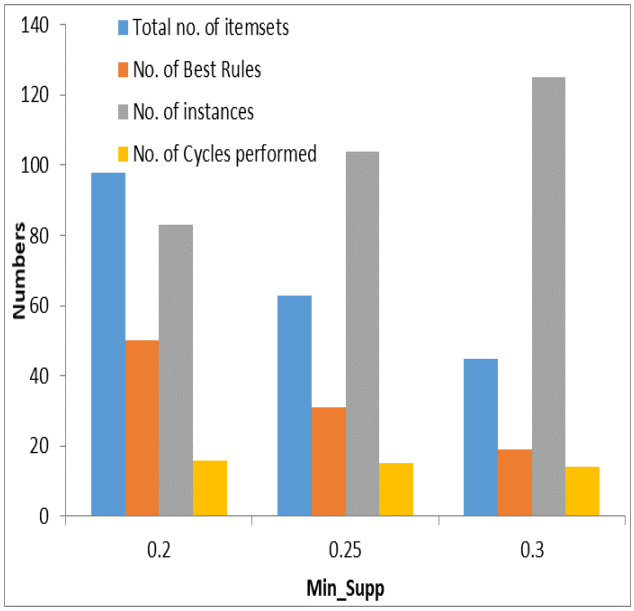

Fig 5: No. of rules with other parameter in liver disorder case only

The no. of best rules at various support value is generated in complete ILPD dataset is shown in fig. 5. As shown in figure at min. support 0.20, no. of best rules generated are 50, total no. of item (include all length) are 98. A number of 16 cycles were required to generate the best rule, and total no. of 83 instances were participated for best rule generation. In the case when we consider only liver disorder data in ILPD dataset, with minimum support = 0.2, best 50 rule generated which is shown in Fig. 6.

As per analysis, mostly in all rules the values of total bilirubin (TB) and direct bilirubin (DB) is high with 99% confidence, it means that these two parameters are very crucial to detect a liver disorder. As per rule no. 14,16,18,19, 47 and 49, albumin (ALB) value is low but other parameter like TB, DB, ALP, SGOT, SGPT value is high with 99% confidence, it focus on the specific situation where patient may be considered as special case for diagnosis due to lower albumin value. The same situation is also verified at minimum support value of 0.25 and 0.3 at 99% confidence.

Further, with min support of 0.2 and 99% confidence, rule 23,27,43 and 48 show that parameter SGPT is normal but other parameter like TB, DB, ALP are high that means only SGPT will not make decision in liver disorder always. Similarly, rule 34 shows that value of SGOT and SGPT is normal despite of high value of ALP with confidence of 93%. It means high value of ALP may cause liver disorder only.

V. CONCLUSION

Liver disease is cause of near two million deaths per year worldwide, out of which one million due to liver cirrhosis and another one million due to viral hepatitis and hepatocellular carcinoma.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

1. Total Bilirubin =TB_N Aspartate Aminotransferase =SGOT_N 102 ==> Direct Bilirubin=DB_N 102 conf:(1)
2. Total Bilirubin =TB_H Alamine Aminotransferase =SGPT_H 97 ==> Aspartate Aminotransferase =SGOT_H 97 conf:(1)
3. Total Bilirubin =TB_N Alamine Aminotransferase =SGPT_N Aspartate Aminotransferase =SGOT_N 90 ==> Direct Bilirubin=DB_N 90 conf:(1)
4. Total Bilirubin =TB_H Alkaline Phosphotase=ALP_H Alamine Aminotransferase =SGPT_H 90 ==> Aspartate Aminotransferase =SGOT_H 90 conf:(1)
5. Direct Bilirubin=DB_H Alamine Aminotransferase =SGPT_H 85 ==> Total Bilirubin =TB_H 85 conf:(1)
6. Direct Bilirubin=DB_H Alamine Aminotransferase =SGPT_H 85 ==> Aspartate Aminotransferase =SGOT_H 85 conf:(1)
7. Direct Bilirubin=DB_H Alamine Aminotransferase =SGPT_H Aspartate Aminotransferase =SGOT_H 85 ==> Total Bilirubin =TB_H 85 conf:(1)
8. Total Bilirubin =TB_H Direct Bilirubin=DB_H Alamine Aminotransferase =SGPT_H 85 ==> Aspartate Aminotransferase =SGOT_H 85 conf:(1)
9. Direct Bilirubin=DB_H Alamine Aminotransferase =SGPT_H 85 ==> Total Bilirubin =TB_H Aspartate Aminotransferase =SGOT_H 85 conf:(1)
10. Direct Bilirubin=DB_H Alkaline Phosphotase=ALP_H 177 ==> Total Bilirubin =TB_H 176 conf:(0.99)
11. Direct Bilirubin=DB_H Aspartate Aminotransferase =SGOT_H 168 ==> Total Bilirubin =TB_H 167 conf:(0.99)
12. Direct Bilirubin=DB_H Alkaline Phosphotase=ALP_H Aspartate Aminotransferase =SGOT_H 155 ==> Total Bilirubin =TB_H 154 conf:(0.99)
13. Total Bilirubin =TB_N Alkaline Phosphotase=ALP_H 136 ==> Direct Bilirubin=DB_N 135 conf:(0.99)
14. Direct Bilirubin=DB_H Albumin =ALB_L 132 ==> Total Bilirubin =TB_H 131 conf:(0.99)
15. Alamine Aminotransferase =SGPT_H 126 ==> Aspartate Aminotransferase =SGOT_H 125 conf:(0.99)
16. Direct Bilirubin=DB_H Alkaline Phosphotase=ALP_H Albumin =ALB_L 119 ==> Total Bilirubin =TB_H 118 conf:(0.99)
17. Alkaline Phosphotase=ALP_H Alamine Aminotransferase =SGPT_H 117 ==> Aspartate Aminotransferase =SGOT_H 116 conf:(0.99)
18. Direct Bilirubin=DB_H Aspartate Aminotransferase =SGOT_H Albumin =ALB_L 116 ==> Total Bilirubin =TB_H 115 conf:(0.99)
19. Direct Bilirubin=DB_H Alkaline Phosphotase=ALP_H Aspartate Aminotransferase =SGOT_H Albumin =ALB_L 105 ==> Total Bilirubin =TB_H 104 conf:(0.99)
20. Total Bilirubin =TB_N Alkaline Phosphotase=ALP_H Alamine Aminotransferase =SGPT_N 100 ==> Direct Bilirubin=DB_N 99 conf:(0.99)
21. Direct Bilirubin=DB_H 195 ==> Total Bilirubin =TB_H 193 conf:(0.99)
22. Total Bilirubin =TB_N Albumin =ALB_N 95 ==> Direct Bilirubin=DB_N 94 conf:(0.99)
23. Direct Bilirubin=DB_H Alkaline Phosphotase=ALP_H Alamine Aminotransferase =SGPT_N 95 ==> Total Bilirubin =TB_H 94 conf:(0.99)
24. Total Bilirubin =TB_N 184 ==> Direct Bilirubin=DB_N 182 conf:(0.99)
25. Total Bilirubin =TB_N Albumin =ALB_L 87 ==> Direct Bilirubin=DB_N 86 conf:(0.99)
26. Total Bilirubin =TB_N Alamine Aminotransferase =SGPT_N 140 ==> Direct Bilirubin=DB_N 138 conf:(0.99)
27. Direct Bilirubin=DB_H Alamine Aminotransferase =SGPT_N 105 ==> Total Bilirubin =TB_H 103 conf:(0.98)
28. Total Bilirubin =TB_H Alamine Aminotransferase =SGPT_N Albumin =ALB_L 90 ==> Alkaline Phosphotase=ALP_H 84 conf:(0.93)
29. Alamine Aminotransferase =SGPT_H 126 ==> Alkaline Phosphotase=ALP_H 117 conf:(0.93)
30. Alamine Aminotransferase =SGPT_H Aspartate Aminotransferase =SGOT_H 125 ==> Alkaline Phosphotase=ALP_H 116 conf:(0.93)
31. Total Bilirubin =TB_H Alamine Aminotransferase =SGPT_H 97 ==> Alkaline Phosphotase=ALP_H 90 conf:(0.93)
32. Total Bilirubin =TB_H Alamine Aminotransferase =SGPT_H Aspartate Aminotransferase =SGOT_H 97 ==> Alkaline Phosphotase=ALP_H 90 conf:(0.93)
33. Total Bilirubin =TB_H Alamine Aminotransferase =SGPT_H 97 ==> Alkaline Phosphotase=ALP_H Aspartate Aminotransferase =SGOT_H 90 conf:(0.93)
34. Alkaline Phosphotase=ALP_H Aspartate Aminotransferase =SGOT_N 108 ==>Alamine Aminotransferase =SGPT_N 100 conf:(0.93)
35. Total Bilirubin =TB_H Alamine Aminotransferase =SGPT_N Aspartate Aminotransferase =SGOT_H 91 ==> Alkaline Phosphotase=ALP_H 84 conf:(0.92)
36. Direct Bilirubin=DB_H Aspartate Aminotransferase =SGOT_H 168 ==> Alkaline Phosphotase=ALP_H 155 conf:(0.92)
37. Total Bilirubin =TB_H Direct Bilirubin=DB_H Aspartate Aminotransferase =SGOT_H 167 ==> Alkaline Phosphotase=ALP_H 154 conf:(0.92)
38. Total Bilirubin =TB_H Aspartate Aminotransferase =SGOT_H 192 ==> Alkaline Phosphotase=ALP_H 177 conf:(0.92)
39. Alamine Aminotransferase =SGPT_H 126 ==> Alkaline Phosphotase=ALP_H Aspartate Aminotransferase =SGOT_H 116 conf:(0.92)
40. Direct Bilirubin=DB_H Aspartate Aminotransferase =SGOT_H 168 ==> Total Bilirubin =TB_H Alkaline Phosphotase=ALP_H 154 conf:(0.92)
41. Total Bilirubin =TB_H Aspartate Aminotransferase =SGOT_H Albumin =ALB_L 130 ==> Alkaline Phosphotase=ALP_H 119 conf:(0.92)
42. Total Bilirubin =TB_H Albumin =ALB_L 149 ==> Alkaline Phosphotase=ALP_H 136 conf:(0.91)
43. Total Bilirubin =TB_H Direct Bilirubin=DB_H Alamine Aminotransferase =SGPT_N 103 ==> Alkaline Phosphotase=ALP_H 94 conf:(0.91)
44. Total Bilirubin =TB_H Direct Bilirubin=DB_H 193 ==> Alkaline Phosphotase=ALP_H 176 conf:(0.91)
45. Direct Bilirubin=DB_H 195 ==> Alkaline Phosphotase=ALP_H 177 conf:(0.91)
46. Total Bilirubin =TB_H 232 ==> Alkaline Phosphotase=ALP_H 210 conf:(0.91)
47. Direct Bilirubin=DB_H Aspartate Aminotransferase =SGOT_H Albumin =ALB_L 116 ==> Alkaline Phosphotase=ALP_H 105 conf:(0.91)
48. Direct Bilirubin=DB_H Alamine Aminotransferase =SGPT_N 105 ==> Alkaline Phosphotase=ALP_H 95 conf:(0.9)
49. Total Bilirubin =TB_H Direct Bilirubin=DB_H Aspartate Aminotransferase =SGOT_H Albumin =ALB_L 115 ==> Alkaline Phosphotase=ALP_H 104 conf:(0.9)
50. Direct Bilirubin=DB_H 195 ==> Total Bilirubin =TB_H Alkaline Phosphotase=ALP_H 176 conf:(0.9)

Fig 6: Best 50 rules at min. support of 0.2 (liver disorder data only)

The rate of death due to liver disease is also increasing gradually. The increase in mortality rate due to liver disorder is a major concern for India and worldwide. It becomes essential to use different mining techniques for accurately and early prediction of liver disorders to avoid the death. The rule mining approach is applied on ILPD datasets to demonstrate behavior of parameter in liver disorder case. Data mining techniques like association rule mining facilitates intensification of some hidden association rule between different parameters and liver condition of a patient. The purpose of this study is to identify some rare rules that can contribute to identification of liver disorder. Experimentally, we have found that few rules mislead the detection of liver disorder. It means, the most of the attribute values are normal, but the patient suffers from liver disorders. It indicates that mere one or two parameters can play an important role to detect liver disorder in a human body. A thorough analysis of mined rules has been provided that will assist the medical experts in the identification and diagnosis of liver disorder. In case, when values of attributes are slightly lower or slightly higher than the normal value range, this analysis will not give sufficient information and further work is required in such cases in the future.

REFERENCES

1. Ben-Cohen, E. Klang, A. Kerpel, E. Konen, M. M. Amitai, and H. Greenspan, "Fully convolutional network and sparsity-based dictionary learning for liver lesion detection in CT examinations," Neurocomputing, vol. 275, pp. 1585-1594, 2018.
2. R.-H. Lin and C.-L. Chuang, "A hybrid diagnosis model for determining the types of the liver disease," Computers in Biology and Medicine, vol. 40, no. 7, pp. 665-670, 2010.
3. M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan, "GAN-based Synthetic Medical Image Augmentation for increased CNN Performance in Liver Lesion Classification," arXiv preprint arXiv:1803.01229, 2018.
4. Z. Janikow, "Fuzzy decision trees: issues and methods," IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), vol. 28, no. 1, pp. 1-14, 1998.
5. T. R. Baitharu and S. K. Pani, "Analysis of Data Mining Techniques for Healthcare Decision Support System Using Liver Disorder
6. Dataset," Procedia Computer Science, vol. 85, pp. 862-870, 2016.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

7. Y. Kumar and G. Sahoo, "Prediction of different types of liver diseases using rule based classification model," Technology and Health Care, vol. 21, no. 5, pp. 417-432, 2013.
8. S. Dhamodharan, "Liver disease prediction using bayesian classification," in 4th National Conference on Advanced Computing, Applications & Technologies, 2014, pp. 1-3.
9. N. Nahar and F. Ara, "Liver disease prediction by using different Decision Tree techniques," International Journal of Data Mining & Knowledge Management Process (IJDKP), vol. 8, no. 2, 2018.
10. P. Rajeswari and G. S. Reena, "Analysis of liver disorder using data mining algorithm," Global journal of computer science and technology, vol. 10, no. 14, pp. 48-52, 2010.

A. Pathan, D. Mhaske, S. Jadhav, R. Bhondave, and K. Rajeswari, "Comparative Study of Different Classification Algorithms on ILPD Dataset to Predict Liver Disorder.", IJRASET, vol. 06, pp. 388-394, 2018P.

11. Thangaraju and R. Mehala, "Performance Analysis of PSO-KStar Classifier over Liver Diseases," International Journal of Advanced Research in Computer Engineering, vol. 04, no. 07, pp. 3132-3137, 2015.
12. R. Agrawal and R. Srikant, "Fast algorithms for mining association rules in large databases, In Proc. of the 20th VLDB Conference, 1994, pp. 487-499.
13. R. Srikant and R. Agrawal, "Mining generalized association rules," Future generation computer systems 13, no. 2-3, pp.161-180, 1997.
14. R. Srikant, "Fast algorithms for mining association rules and sequential patterns," PhD diss., University of Wisconsin, Madison, 1996.
15. R. V. Priya, A. Vadivel, and R. Thakur, "Frequent pattern mining using modified CP-tree for knowledge discovery," in International Conference on Advanced Data Mining and Applications, 2010, pp. 254-261: Springer.
16. Sabnis, N. Khare, R. Thakur, and K. Pardasani, "Karnaugh Map Model for Mining Association Relationships in Web Content Data: Hypertext," Data Mining and Knowledge Engineering, vol. 4, no. 11, pp. 579-587, 2012.
17. V. Tiwari and R. S. Thakur, "P²MS: a phase-wise pattern management system for pattern warehouse," International Journal of Data Mining, Modelling and Management, vol. 7, no. 4, pp. 331-350, 2015.
18. V. Tiwari and R. S. Thakur, "Towards important issues of pattern retrieval: pattern warehouse," International Journal of Data Science, vol. 2, no. 1, pp. 1-14, 2017.
19. V. Tiwari and R. Thakur, "A Level Wise Tree Based Approach for Ontology-Driven Association Rules Mining," Data Mining and Knowledge Engineering, vol. 4, no. 5, pp. 252-259, 2012.
20. S. Rajput, R. S. Thakur, and G. S. Thakur, "An Integrated Approach and Framework for Document Clustering Using Graph Based Association Rule Mining," in Proceedings of the Second International Conference on Soft Computing for Problem Solving (SocProS 2012), December 28-30, 2012, pp. 1421-1437: Springer.
21. J. Han, J. Pei, and M. Kamber, Data mining: concepts and techniques. Elsevier, 2011.
22. ILPD Dataset: https://archive.ics.uci.edu/ml/datasets/ILPD+(Indian+Liver+Patient+Dataset).

AUTHORS PROFILE

Ramjeevan Singh Thakur is an Associate Professor in the Department of Computer Applications at Maulana Azad National Institute of Technology, Bhopal, India. He is a Teacher, Researcher and Consultant in the field of Computer Science and Information Technology. He earned his Master Degree from Samrat Ashok Technology Institute, Vidisha (M.P.) in 1999 and Ph.D. Degree (Computer Science) From Rajiv Gandhi Proudyogiki Vishwavidyalaya, Bhopal (M.P.) in 2008.

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)