

Automatic Rule Generator via FP-Growth for Eye Diseases Diagnosis

Rahmad Kurniawan^{#1}, Mohd Zakree Ahmad Nazri^{#2}, Siti Norul Huda Sheikh Abdullah^{#3}, Jemaima Che Hamzah^{*}, Rado Yendra⁺, Westi Oktaviana⁺

[#]Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, Bangi, Selangor, 43600, Malaysia
E-mail: ¹rahmadkurniawan@uin-suska.ac.id; ²zakree@ukm.edu.my; ³snhsabdullah@ukm.edu.my

^{*}Faculty of Medicine, Universiti Kebangsaan Malaysia Medical Centre, Jalan Yaacob Latif, Bandar Tun Razak, Cheras, 56000, Malaysia
E-mail: ⁴jemaima@ppukm.ukm.edu.my

⁺ Faculty of Science and Technology Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, 28293, Indonesia
E-mail: rado.yendra@uin-suska.ac.id; westioktaviana@gmail.com

Abstract— The conventional approach in developing a rule-based expert system usually applies a tedious, lengthy and costly knowledge acquisition process. The acquisition process is known as the bottleneck in developing an expert system. Furthermore, manual knowledge acquisition can eventually lead to erroneous in decision-making and function ineffective when designing any expert system. Another dilemma among knowledge engineers are handing conflict of interest or high variance of inter and intrapersonal decisions among domain experts during knowledge elicitation stage. The aim of this research is to improve the acquisition of knowledge level using a data mining technique. This paper investigates the effectiveness of an association rule mining technique in generating new rules for an expert system. In this paper, FP-Growth is the machine learning technique that was used in acquiring rules from the eye disease diagnosis records collected from Sumatera Eye Center (SMEC) Hospital in Pekanbaru, Riau, Indonesia. The developed systems are tested with 17 cases. The ophthalmologists inspected the results from automatic rule generator for eye diseases diagnosis. We found that the introduction of FP-Growth association rules into the eye disease knowledge-based systems, able to produce acceptable and promising eye diagnosing results approximately 88% of average accuracy rate. Based on the test results, we can conclude that Conjunctivitis and Presbyopia disease are the most dominant suffering in Indonesia. In conclusion, FP-growth association rules are very potential and capable of becoming an adequate automatic rules generator, but still has plenty of room for improvement in the context of eye disease diagnosing.

Keywords—association rules; eye diseases; FP-Growth; knowledge base.

I. INTRODUCTION

The development of the rule-based expert system has been continuously developed as reported [1]–[3]. The strength of a rule-based expert system is, it can be developed quickly. However, a study proved that a rule-based is one of "barrier in the process of development of expert systems" is knowledge acquisition [4]. Making a mistake during eliciting knowledge from various sources may eventually lead to inaccuracies in decision-making and cause expert system becomes ineffective. On top of that, time and ambiguity resources of knowledge may also affect the achievement of expert systems before rule error. Apart from that, mistakes may also occur due to conflict of interest and insufficient time spending while interviewing with the expert domain [2].

There are two most important persons in the process of acquiring knowledge is the programmer and domain experts. With proper skillset, a programmer can obtain extensive

information and knowledge from experts during interview sessions. The programmer should also be well-trained in determining the quality and effectiveness of an expert system [5]. Even though, a standard framework in the process of acquiring knowledge such a guided interview has been established [6], it solely relies on the transfer of manual knowledge from expert to a programmer. Hence, the achievement of the knowledge elicitation depends absolutely on the ability of a programmer to obtain information.

Furthermore, personal issues and conflicts of interest are also common problems in the rule-based expert system development. Lack of communication and less preparation between programmers and experts may also encounter inadequate and vague knowledge during the knowledge acquisition phase. Sometimes, many terms from a particular domain, e.g., medical terms, can create some obstacle in understanding and remarking immediate and significant knowledge. Apart from confusing terms, personality issues

such as misunderstanding and low self-esteem may arise uneasy moment between programmers and experts that may contribute to a severe obstacle to the knowledge acquisition phase [4]. In some case, experts are not familiar with information technology, and they refuse to take part in expert system development efforts. Sometimes, they reject applying automated system for supporting neither easy nor complicated problem-solving. Moreover, certain experts think that the expert system can weaken their role and make their work no longer useful, so knowledge acquisition is impossible to be conducted [2].

Before the issues as mentioned earlier, the development of a rule-based expert system faces a higher risk of creating 'noise rules' in coincidence. Therefore, this study aims to substitute traditional knowledge acquisition method into automatic knowledge acquisition based on the FP Growth method by using historical medical records.

Kurniawan has built the case based reasoning expert system for diagnosing eye diseases using the Naïve Bayes algorithm, whereby it capable of handling 140 cases [7]. The expert system via case-based reasoning is claimed to be faster in the knowledge acquisition process. Developing the knowledge model in the rule-based expert system is the easiest, but maybe tiresome and pricey because of an expert expressed inconsistently and piece by piece of valuable knowledge and information [8] to the knowledge engineer. Indeed, the rule-based expert system is becoming more challenging in the knowledge acquisition processes.

Millette overcomes the problem mentioned above of knowledge acquisition during rule-based expert system development by using data mining techniques and data warehouse techniques. Millette obtained the rule base essentially by generating a set of data using the association rule technique to address the problem of knowledge acquisition on expert systems. In correspondence, association rules technique has become more popular to overcome the weaknesses of the rule-based expert system [3], [9].

On the other hand, association rules have been used to generate the rules [10], [11]. Karabatak has combined clustering based learning, namely Apriori algorithm and supervised learning namely Neural Networks, to produce rules automatically for an expert system in the field of a breast cancer diagnosis. They were able to produce better result via association rules by reducing rules' dimensions. This hybridization method proved that breast cancer expert system is more effective rather than applying Neural Networks alone. Another study by Weng, Liu and Wu used data mining techniques and Bayesian Network for making the best product recommendations [12]. Similarly, the study of Ikram has used association rules on expert systems for earthquake prediction [13].

Association rules algorithm is a data mining technique for finding associative rules between combinations of items on an extensive database. Association rules is an alternative technique when statistical approach result is not satisfactory when applying to real cases [14]. In this case, the association rules are defined as an implication of the form $A \rightarrow B$, with A and B being an item. The form of rule $A \rightarrow B$ can be interpreted as if there is item A then raised item B. The parameters, i.e. support and confidence influenced the

formation of an association rule. Support presents a combination of items in the database, while confidence is the probability of an item purchased along with others.

$$\{A \cap B\} \rightarrow \{C\} \text{ or } \{A, B\} \rightarrow \{C\} \quad (1)$$

Association rule is a first algorithm used to analyze supermarkets problem [15], to find out the potential customers. For example, who buys coffee and sugar? With this knowledge, supermarket owners can place correlated products nearby. Supermarket owners can also determine the best way to market their products or give discount coupons to attract customers. Data mining for generating rules in the knowledge base has shown good potential to apply. However, presently, the FP-Growth algorithm has not yet been explored its effectiveness as an automatic rule generator.

Based on the literature review, there were also many researchers [2], [3], [9]–[11], [13] used data mining techniques to obtain essential or significant rules. However, most of them used the basic techniques of association rules [2], [3], [9], [11], [13] and Apriori [10] algorithms. Apriori scans the dataset repeatedly to generate frequent itemsets candidate item sets [16]. Overcoming previous automatic rule generators' weaknesses, we continue exploring better design for fully automatic knowledge acquisition by using primary and explicit performing algorithms. Therefore, this study is seeking a better way technique for generating rules.

In this research, we employ FP-Growth, an improvement of the Apriori algorithm. It is notable that the FP-Growth algorithm can fix the lacking of Apriori algorithm. Eventually, their frequent item set search process between are unique and diverse [17]. Nowadays, the FP-Growth is one of the quickest algorithms among the rules of association [18]. Frequent Pattern Growth is one of the alternative methods to determine the most frequent item set in a data set. The main feature of the FP-Growth algorithm is FP-Tree treated as its data structures. We organized the rest of this paper as follows. Section II described overview of the materials and methods of this study. Section III presents the implementation of the FP-Growth algorithm, experimental results, and discussion. We present the conclusions in the final section.

II. MATERIALS AND METHOD

We obtained eye disease data from Medical Record in Sumatera Eye Center (SMEC) Hospital Pekanbaru. The medical records were collected only for the year 2016 containing 1600 patients with six attributes (No Medical Record, Date, Gender, Age, Symptoms, and Diagnosis). The strategy used in this study is both expert reviews (opinions) and extensive experiments in accordance to research methodology standard of the procedure [19]. We need the opinions and assessments of medical experts as a strategy in conducting research. Furthermore, this opinion strategy is also conducted by interviewing experts, Delphi, and brainstorming. Our experiments were conducted through simulation and analysis of experimental results. Analysis is the stage of understanding of a problem before taking action and decisions. The system testing and validation based on support, the confidence of rules and an expert assessment of the rules.

There are four steps to designing experiments using the FP-Growth algorithm. The first step is the selection. We perform attribute selection by removing unnecessary attributes. In the second step, we impose data pre-processing by deleting duplicate and deleting incomplete of eye disease data. The third step is data transformation by changing and storing eye disease data into a specialized format, i.e. attribute-relation file format. At this stage, we initialize attributes and modifies the type of data, such as conversion of numeric data types into binominal. The last step is an experiment to get rules based on the FP-Growth algorithm.

In Figure 1 illustrates a cycle of knowledge acquisition process, it starts with generating the rules based on FP Growth association rule algorithm then the final stage is comparing expert decision results and our validated eye disease rules. We carry out an analysis procedure for eye disease rule validation as the following:

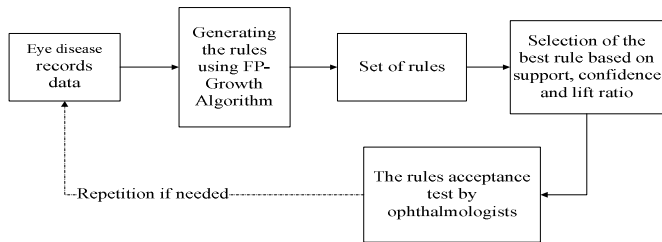


Fig. 1 The proposed cycle of knowledge acquisition process

The following section of the complete attributes of the eye disease data covering age, gender, symptoms and disease, and its data representation for $age \in \{U1, \dots, U8\}$, $gender \in \{L, P\}$, $symptoms \in \{G1, G18\}$, $diseases \in \{P1, P9\}$ are described as accordance to Table I.

TABLE I
MEDICAL RECORD ATTRIBUTES

No	Attributes	Attributes code	Description
1	Age	U1	Baby (18 or 24 months)
		U2	Child (25 months - 5 years)
		U3	Child (6 - 11 years)
		U4	Teen (12 - 15 years)
		U5	Teen (16 - 19 years)
		U6	Adult (20 - 30 years)
		U7	Adult (31 - 59 years)
		U8	Adult (>60)
2	Gender	P	Female
		L	Male
3	Symptoms	G1	Blurred vision, distance
		G2	Blurred vision, near
		G3	Glare
		G4	Cloudy Vision
		G5	Fatty eye
		G6	Dizzy
		G7	Discomfort
		G8	Watery eyes
		G9	Itchy eyes
		G10	Sticky eyes
		G11	Red eye
		G12	Swollen eyes
		G13	Eye pain
		G14	Tired Eyes
		G17	Dry Eyes
		G18	Trauma
4	Diseases	P1	Astigmatism or Myopia
		P2	Conjunctivitis
		P3	Dry Eye

No	Attributes	Attributes code	Description
		P4	Immature senile cataracts
		P5	Mature senile cataracts
		P6	Myopia
		P7	Presbyopia
		P8	Pterygium
		P9	Trauma

A. Selection

We found that medical record data in that hospitals are kept in a manual form. Then, we digitised those manual eye disease medical data using data processing tools. We exclude the *medical record number*, *name*, and *date*, from the six attributes the patient's medical report form such as *gender*, *age*, *symptom* and *diagnosis* of the disease. Next, our task is to find the relationship among age, gender, symptom and disease.

B. Pre-processing

The next process is data pre-processing whereby we eliminate noise or clean the missing data value, inconsistent data and outliers. The data extraction process is challenging due to irrelevant, redundant, noisy, unreliable attribute and unsuitability of a method to obtain unidentified patterns from data may effect to an inaccurate of result [20]. The data pre-processing stage purposes to enhance machine-learning performance. We use several techniques to address missing values, such as replace missing values with their mean value and delete rows with more than one missing values. In the case of the age attribute, we assume that missing values distributed similarly to existing values. This technique was called Missing Completely at Random (MCAR). Due to the missing value in the eye disease dataset, we removed records with the missing value. The remaining records in the dataset are approximately 1214 records which are used in the data mining process for generating rules.

C. Example to Generate Rules Using FP-Growth

Here, we illustrate in the case of 20 sample data of medical record on eye disease to find the relationship between age, gender, symptoms and kinds of eye disease.

TABLE II
SAMPLE DATA OF MEDICAL RECORD

TID	Age	Gender	Symptoms	Diseases
1	U5	P	G8	P4
2	U5	P	G8	P1
3	U4	P	G8	P1
4	U7	P	G8, G11, G15	P3
5	U7	P	G8, G11, G15	P3
6	U3	P	G8, G11, G15	P3
7	U7	P	G8, G11, G15	P3
8	U7	L	G8, G11, G15	P3
9	U6	L	G8, G11, G15	P3
10	U7	P	G8, G11, G15	P3
11	U7	L	G8, G11, G15	P3
12	U5	L	G8, G11, G15	P3
13	U5	P	G8, G11, G15	P3
14	U5	L	G8, G11, G15	P3
15	U7	L	G8, G11, G15	P3
16	U2	L	G8, G11, G15	P3
17	U5	L	G8, G11, G15	P3
18	U5	L	G8, G18	P4
19	U5	L	G8, G18	P4
20	U7	L	G8, G9, G12	P2

In this sample, we assume that the support count $\geq 10\%$. This support count value affects FP-tree construction. The support count 10% as the threshold for filtering the item set, i.e. only the item set have support count is equal, or more than 10% appeared. We illustrate the searching pattern process as shown in Table III. We can observe that items U3, U6, U2, G9, G12, P2 and U4 have not appeared in the table. The best item based on minimum support program works as follow:

Input: Eye-diseases-data

Output: The best item based on minimum support

Procedure best-itemset

- 1 Scan eye-diseases-data to find support
- 2 For each support do
- 3 Sort (support) order by support desc;
- 4 If item set support $< 10\%$ then
- 5 discard item set

TABLE III
THE BEST ITEM BASED ON MINIMUM SUPPORT

Itemset	Frequent	Support	Support (%)
G8	20	$20/20 = 1$	100
G11	14	$14/20 = 0.70$	70
G15	14	$14/20 = 0.70$	70
P3	14	$14/20 = 0.70$	70
L	11	$11/20 = 0.55$	55
P	9	$9/20 = 0.45$	45
U5	8	$8/20 = 0.40$	40
U7	8	$8/20 = 0.40$	40
P4	3	$3/20 = 0.15$	15
G18	2	$2/20 = 0.10$	10
P1	2	$2/20 = 0.10$	10

The next step is to scan the data and re-arrange data by the minimum support. The confidence and support value are parameters for measuring the best rules. The data does not meet the minimum support and does not contain antecedent and consequent will removed from the transaction. The attributes become antecedent are *age*, *gender*, *symptoms*, while the consequence is *the disease*.

The next step is FP-Tree construction based on the transaction scanned. Based on the sample data provided, P1, P2, P3, P4, P5, P6, P7, P8 and P9 are the consequence, but only P3 and P4 meet the minimum support. The next step is to get a frequent item set by generating a conditional pattern base then create the rules by calculating the confidence of each rules combination. The FP-Tree construction works as follow:

Input: The best item based on minimum support

Output: FP-Tree mapped

Procedure FP-tree

- 1 Read transaction1: {U5, P, G8, P4}.
- 2 Create 4 nodes U5, P, G8 and P4 and the path $null \rightarrow U5 \rightarrow P \rightarrow G8 \rightarrow P4$.
- 3 Set counts of U5, P, G8 and P4 to 1.
- 4 Continue to read the transactions until last transactions, and FP-Tree mapped.

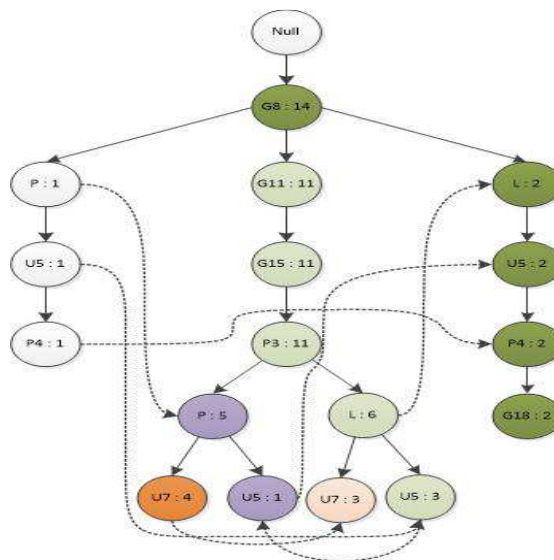


Fig. 2 Example of FP-Tree Constructions

After the rules obtained from frequent itemset, then the next step is to evaluate the strong association rules obtained by using the lift ratio. Lift ratio is the ratio between the confidence of a rule and the benchmark confidence value. Benchmark Confidence is a comparison between the numbers of consequent with the total of transactions. We consider lift ratio more than 1 indicating as possible rules. Below is the expected confidence equation,

$$\text{Expected Confidence} = \frac{NC}{N} \quad (2)$$

NC= the numbers of consequent

N= total of transactions.

Then, we calculate a lift ratio as below:

$$\text{Lift Ratio} = \frac{\text{Confidence}(A,B)}{\text{Expected Confidence}(A,B)} \quad (3)$$

TABLE IV
THE STRONG ASSOCIATION RULES

No	If	Then	SP	CF	LR
1	$L \wedge U5 \wedge P4$	G18	14%	100%	1
2	G18	$L \wedge U5 \wedge P4$	14%	100%	7.14
3	$U5 \wedge P4 \wedge G18$	L	14%	100%	1.75

381	$U5 \wedge G11 \wedge P3$	$G8 \wedge G15$	28%	100%	1.28

SP: Support

CF: Confidence

LR: Lift Ratio

III. RESULTS AND DISCUSSION

The implementation performed on minimum support and minimum confidence parameters. The minimum support used 3%, 10% and 20% while for the minimum confidence used 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% and 100%. The minimum support and minimum confidence affect the rules generated. We initiate with a high value in minimum support and confidence, and then we decrease the values until we find a value that generates enough patterns.

The value of minimum confidence and support are high then the rules are getting less than shown on Table V.

TABLE V
THE NUMBER OF RULES GENERATED BY MINIMUM SUPPORT AND MINIMUM CONFIDENCE PARAMETERS

<i>Min sup</i>	3%	10%	20%	30%-100%
<i>Min conf</i>				
10%	13173	1208	54	0
20%	11962	1206	54	0
30%	10430	1167	54	0
40%	8298	1126	54	0
50%	6459	1038	51	0
60%	4815	815	50	0
70%	3895	732	44	0
80%	2870	459	38	0
90%	2208	359	24	0
100%	0	0	0	0

Concerning performance, when minimum support is higher, it obtains less pattern, and the algorithm is faster. In the meanwhile, for minimum confidence when it is set higher, the number of patterns are lesser, but it may not be faster because many algorithms do not use minimum

confidence to prune the search space. Naturally, setting these parameters also depends on the desire on how many rules seem justifiable. We obtained 2208 rules of FP-Growth association rule mining test. All the rules have qualified the minimum support and the minimum confidence with the lift ratio > 1 so that the resulting rule is feasible to use. The value of lift ratio gives us evidence about the increase in the probability of the then consequent given the antecedent.

Furthermore, there are 17 rules have qualified the antecedent and consequent with antecedent is gender, age, and symptoms while the consequent is eye disease. In this case, there are nine kinds of eye disease, but only two diseases that meet the requirements of frequent itemset namely Presbyopia and Conjunctivitis. We have done the testing with the minimum support of 3% and minimum confidence of 90%.

Association rule testing accomplished in two ways: a test based on minimum support, minimum confidence and lift ratio. We used the lift ratio to evaluate the association rule. Ophthalmologists perform the second tests. We showed the rules generated by FP-Growth for the feasibility of rules. We have tested the rules obtained automatically by FP-Growth with expert knowledge. Here are the expert test results.

TABLE VI
THE BEST RULE GENERATED BASED ON SUPPORT, CONFIDENCE, LIFT RATIO, AND USER ACCEPTANCE TEST

No	Antecedent	Consequent	SP	CF	LR	Expert I			Expert II		
						Agree	Disagree	CL	Agree	Disagree	CL
1.	P∧U7∧G7 (Female ^ 31 - 59 years ^ Discomfort)	P7 Presbyopia	4.53	1.0	7.27		√	2	√		3
2.	L∧U7∧G9 (Male ^ 31 - 59 years ^ Itchy eyes)	P2 Conjunctivitis	3.54	1.0	5.4	√		8	√		8
3.	L∧U7∧G10 (Male ^ 31 - 59 years ^ Sticky eyes)	P2 Conjunctivitis	3.13	1.0	5.4	√		8	√		9
4.	L∧U7∧G2 (Male ^ 31 - 59 years ^ Blurred vision, near)	P7 Presbyopia	4.28	1.0	7.27			8	√		9
5.	L∧U7∧G7 (Male ^ 31 - 59 years ^ Discomfort)	P7 Presbyopia	3.29	1.0	7.25		√	2	√		3
6.	P∧U7∧G6∧G2 (Female ^ 31 - 59 years ^ Dizzy ^ Blurred vision, near)	P7 Presbyopia	4.94	1.0	7.27	√		8	√		7
7.	P∧U7∧G6∧G7 (Female ^ 31 - 59 years ^ Dizzy ^ Discomfort)	P7 Presbyopia	4.36	1.0	7.27		√	2	√		5
8.	P∧U7∧G2∧G7 (Female ^ 31 - 59 years ^ Blurred vision, near ^ Discomfort)	P7 Presbyopia	4.36	1.0	7.27	√		8	√		9
9.	L∧U7∧G6∧G2 (Male ^ 31 - 59 years ^ Dizzy ^ Blurred vision, near)	P7 Presbyopia	3.45	1.0	7.27	√		8	√		8
10.	L∧U7∧G6∧G7 (Male ^ 31 - 59 years ^ Dizzy ^ Discomfort)	P7 Presbyopia	3.21	1.0	7.25		√	2	√		5
11.	L∧U7∧G11∧G9 (Male ^ 31 - 59 years ^ Red-eye ^ Itchy eyes)	P2 Conjunctivitis	3.13	1.0	5.4	√		7	√		9
12.	L∧U7∧G11∧G10 (Male ^ 31 - 59 years ^ Red-eye ^ Sticky eyes)	P2 Conjunctivitis	3.13	1.0	5.4	√		8	√		9
13.	L∧U7∧G8∧G9 (Male ^ 31 - 59 years ^ Watery eyes ^ Itchy eyes)	P2 Conjunctivitis	3.54	1.0	5.4	√		7	√		9
14.	L∧U7∧G8∧G10 (Male ^ 31 - 59 years ^ Watery eyes ^ Sticky eyes)	P2 Conjunctivitis	3.13	1.0	5.4	√		8	√		9
15.	L∧U7∧G9∧G10 (Male ^ 31 - 59 years ^ Itchy eyes ^ Sticky eyes)	P2 Conjunctivitis	3.13	1.0	5.4	√		8	√		9
16.	L∧U7∧G2∧G7 (Male ^ 31 - 59 years ^ Blurred vision, near ^ Discomfort)	P7 Presbyopia	3.21	1.0	7.27	√		8	√		8
17.	P∧U7∧G6∧G2∧G7 (Female ^ 31 - 59 years ^ Dizzy Blurred vision, near ^ Discomfort)	P7 Presbyopia	4.36	1.0	7.27	√		8	√		7

SP: Support; CF: Confidence, LR: Lift Ratio, CL: Confidence Level

Based on an ophthalmologist's acceptance test, the average accuracy of the FP-Growth association rules algorithm as rules automatic generator is 88%.

We have also conducted experiments by examining the suitability of Lift Ratio and ophthalmologist's Confidence Level. The Confidence Level usually used as a reference to assess the strength of a rule. Eye diseases diagnosis should use the accurate rules and acceptable. The ophthalmologists inspected the results from automatic rule generator. In this study, we have obtained the rules 6, 9, 16 and 17 have the suitability between Confidence Level of ophthalmologists and Lift Ratio values.

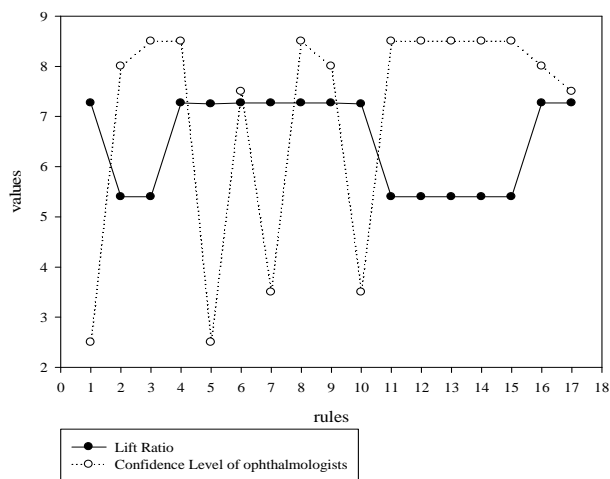


Fig. 3 The relationship between lift ratio and ophthalmologist's confidence level

We have obtained the strong relationship between symptom and disease based on the ophthalmologist's assessment. We can use accurate rules for diagnose eye diseases. Fig. 4 shows the relationship of symptoms and diseases that have been generated automatically using FP-Growth.

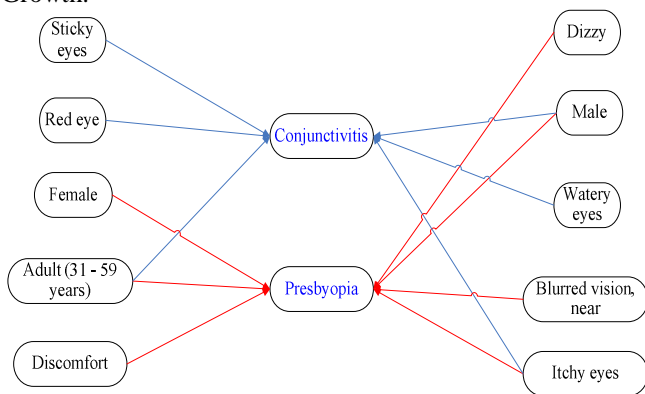


Fig. 4 The relationship between symptoms and eye diseases using FP-Growth algorithm

IV. CONCLUSION

We have successfully conducted research using association rules as a knowledge base in expert systems of eye disease. We used medical record data from Sumatera Eye Center (SMEC) Hospital in Pekanbaru, Riau, Indonesia. Based on the experiments conducted, the determination of

minimum support and minimum confidence in FP-Growth should be appropriate for the outcome of the rules. We assigned the minimum support 3% and minimum confidence 90% to attained 2208 rules.

We have successfully employed FP-Growth algorithm as a rules producer with good accuracy 88%. FP-Growth algorithm fast to create rules in medical records data that only twice database scan. FP-growth constructs a highly compressed FP-tree, which is substantially reduced the original database.

Based on the test results, we obtained that *Conjunctivitis* and *Presbyopia* disease is the most dominant suffered from 225 and 167 cases. Male or female aged 31-59 years are very dominant suffering *Conjunctivitis* and *Presbyopia*. Based on minimum support score concluded if female, aged 31-59 years, dizzy, near blurred vision then suffering presbyopia with support score 4.9% as well as this rule appear more often together than expected with the highest lift 7.27%. It means that the occurrence of this symptom has a positive effect on the occurrence of diseases, i.e. symptoms positively correlated with diseases.

There are three critical symptoms for presbyopia: *blurred vision (near)*, *dizziness*, and *discomfort on eyes*, while for conjunctivitis there are four influencing symptoms: *watery eyes*, *itchy eyes*, *sticky eyes*, and *red eyes*, but the most important symptom is the *sticky eyes*. FP-Growth could be generated rules as a knowledge base in rule-based expert systems. This algorithm considered to be a fast technique in knowledge acquisition, even replace of traditional knowledge acquisition.

However, we need an enlargement study to produce the completely automatic rules as knowledge base expert system in future works. In future, we have planned to get more data from kinds of eye diseases and to develop an actual expert system.

ACKNOWLEDGEMENT

We would like to thank Sumatera Eye Center (SMEC) Hospital in Pekanbaru, Riau, Indonesia that provided precious data on eye diseases and Universiti Kebangsaan Malaysia Hospital has guided this study on acceptance test. This research was supported by research grant PRGS/1/2016/ICT02/UKM/02/1 and AP-2017-005/2.

REFERENCES

- [1] A. Dong and A. M. Agogino, "Text analysis for constructing design representations," *Artificial Intelligence in Engineering*, vol. 11, pp. 65-75, 1997.
- [2] L. Millette, "Improving the Knowledge-Based Expert System Lifecycle," Master's Thesis, University of North Florida, U.S., 2012.
- [3] W. Rui and L. Duo, "The study on the construction of knowledge base of grinding expert system based on data mining," in *International Conference on Mechatronic Science, Electric Engineering and Computer (MEC)*, 2011, p. 845-848.
- [4] M. Golabchi, "A knowledge-based expert system for selection of appropriate structural systems for large spans," *Asian Journal of civil engineering (Building and Housing)*, vol. 9, pp. 179-191, 2008.
- [5] S. Mertens, M. Rosu, and Y. Erdani, "An intelligent dialogue for online rule-based expert systems," in *Proceedings of the 9th international conference on intelligent user interfaces*, 2004, p. 280-282.
- [6] W. P. Wagner, "Issues in knowledge acquisition," in *Proceedings of the 1990 ACM SIGBDP conference on Trends and directions in expert systems*, 1990, p. 247-261.

- [7] R. Kurniawan, N. Yanti, and M. Z. Nazri, "Expert systems for self-diagnosing of eye diseases using Naïve Bayes," in *International Conference of Advanced Informatics: Concept, Theory and Application (ICAICTA)*, 2014, p. 113-116.
- [8] R. Mokhtar, N. A. M. Zin, and S. N. H. S. Abdullah, "Rule-based knowledge representation for modality learning style in AIWBES," in *Knowledge Management International Conference*, 2010.
- [9] D.-L. Ma, W.-J. Zhang, B. Dong, P. Yang, and H.-X. Lu, "Establishing knowledge base of expert system with association rules," in *International Conference on Machine Learning and Cybernetics*, 2008, p. 1785-1788.
- [10] M. Karabatak and M. C. Ince, "An expert system for detection of breast cancer based on association rules and neural network," *Expert systems with Applications*. vol. 36, pp. 3465-3469, 2009.
- [11] S. M. Fakhrahmad, M. H. Sadreddini, and M. Zolghadri Jahromi, "A proposed expert system for word sense disambiguation: deductive ambiguity resolution based on data mining and forward chaining," *Expert Systems.*, vol. 32, pp. 178-191, 2015.
- [12] S.-S. Weng, S.-C. Liu, and T.-H. Wu, "Applying Bayesian network and association rule analysis for product recommendation," *International Journal of Electronic Business Management.*, vol. 9, pp. 149, 2011.
- [13] A. Ikram and U. Qamar, "A rule-based expert system for earthquake prediction," *Journal of Intelligent Information Systems.*, vol. 43, pp. 205-230, 2014.
- [14] Z. A. Othman, N. Ismail, and M. T. Latif, "Association rules of temperature towards high and low ozone in Putrajaya," in *6th International Conference on Electrical Engineering and Informatics (ICEEI)*, 2017, p. 1-5.
- [15] M. F. M. Mohsin, A. A. Bakar, and M. H. A. Wahab, "A Comparative Study of Apriori and Rough Classifier for Data Mining," *Asia-Pacific Journal of Information Technology and Multimedia*. vol. 5, 2008.
- [16] A. S. Saabith, E. Sundararajan, and A. A. Bakar, "Parallel implementation of apriori algorithms on the Hadoop-MapReduce platform-an evaluation of literature," *Journal of Theoretical and Applied Information Technology.*, vol. 85, pp. 321, 2016.
- [17] N. G. Noma and M. K. A. Ghani, "Discovering pattern in medical audiology data with FP-growth algorithm," in *Conference on Biomedical Engineering and Sciences (IECBES)*, 2012, p. 17-22.
- [18] A. S. Hoque, S. K. Mondal, T. M. Zaman, P. C. Barman, and M. A.-A. Bhuiyan, "Implication of association rules employing FP-growth algorithm for knowledge discovery," in *14th International Conference Computer and Information Technology (ICCIT)*, 2011, p. 514-519.
- [19] J. W. Buckley, M. H. Buckley, and H.-F. Chiang, *Research methodology and business decisions*, 1976.
- [20] R. Kurniawan, M. Z. A. Nazri, M. Irsyad, R. Yendra, and A. Aklima, "On machine learning technique selection for classification," in *International Conference on Electrical Engineering and Informatics (ICEEI)*, 2015, p. 540-545.