2017 5th International Conference on Electrical, Electronics and Information Engineering (ICEEIE)

An Intelligent Support System for Diagnosing Dehydration in Children

Maulana Miftakhul Faizin, Subiyanto, Ulfah Mediaty Arief
Universitas Negeri Semarang, UNNES
Semarang, Indonesia
maulanafaizin.a6.19@gmail.com, subiyanto@mail.unnes.ac.id, ulfahmediatyarief@mail.unnes.ac.id

***Abstract*—This paper present the implementation of artificial intelligent on medical decision support system for diagnosing childrens dehydration. In this study, the intelligent system constructed using decision tree method with C4.5 algorithm and pruned with REP (Reduced Error Pruning) method. This study was a collaboration between the doctor and the hospital in order to analyze the dataset of children dehydration in indonesia. The number of 92 medical data was recorded for dataset and divided into two subsets: trainingset (57 records) and testset (35 records). The medical symptoms of dehydration that used for Input variables are general appearance, eyes, respirations, turgor and mucous membranes, while the output variable is the severity of dehydration that classified into three categories: severe dehydration, some dehydration and no dehydration. The validation was done by comparing the classification performance of the intelligent system and the doctor diagnose. The confusion matrix was used for mapping the classification performance of intelligent system and evaluated by using accuracy and the value of error rate. The result show that, the implementation of artificial intelligent on medical decision support system have an accuracy of 91% and the error rate value of 0.085714286. From the result it can be concluded that the implementation of artificial intelligent on medical decision support system can be use for supporting dehydration diagnostics in children.**

Keywords— decision tree; C4.5 algorithm; reduced error pruning; dehydration.

I. INTRODUCTION

Diarrhea is most common leading cause of death in children worldwide and kills 760,000 children each year. The most dangerous threat that can be caused by diarrhea is dehydration [1]. Dehydration is often cited to determine the severity of diarrhea in children aged under 5 years and as a reference to provide treatment against diarrhea [2]. An accurate assessment of the dehydration severity is an important step to prevent the occurrence of morbidity and to provide an appropriate treatment for patients with dehydration [3]. There is a lot of interest in the pediatrics world to develop a simple and non-Invasive diagnosis system to assess the severity of dehydration in children [4].

Implementation of decision support systems in medical diagnosis is very important. Approximately 11% of the expert system was implemented in the medical world that is used to diagnose, and about 21% of paper which uses implementation of the method used to handle problem in the medical world [5]. The concept of developing a decision support system that is simple and allows to perform automatic learning is the criterion of decision support systems that needed in the medical world, and decision tree is the most appropriate candidate to perform this task. Decision tree is one of the decision-making techniques that are reliable and effective, and produce high classification accuracy with a simple representation of the knowledge. When using a decision tree, decision-making process can be easily validated by experts. Because of that reason, decision tree is used to make decision support in the medical world [6].

In the last few decade decision tree has been widely implemented and proven successful in many areas of life, especially in the medical world [6]. Decision tree has been widely used to predict appropriate treatment for various problems in the medical world [7]. In a previous study, decision tree has been implemented to diagnose various diseases and other medicals field. Somaya Hashem et al (2016) applied the decision tree method to predict liver fibrosis in Egyptian patients with hepatitis C and the results showed that the prediction model has an accuracy rate of 86.2% [8]. Decision tree is a method that uses a learning algorithm to build the tree [9]. From some of the learning algorithm used in the decision tree, one of the most popular and commonly used for classification is C4.5 algorithm [10]. C4.5 algorithm is an algorithm proposed by J.R Quinlan in 1993 and was the development of the previous algorithm namely ID3[11]. C4.5 algorithm is an effective and efficient algorithm for classification [12]. In medical world, C4.5 algorithm has been used to make a system for diagnosis disease. In a previous study C4.5 algorithm has been implemented in the medical world. Phong D Tong et al (2016) on their study implemented C4.5 algorithm to diagnose dengue fever and the model reached accuracy of 98% [13].

Decision tree often occurs misclassification during the learning process and has possibility to generating a decision tree with large size [14]. The Large size of the decision tree and misclassification during the learning process can lead to over fitting [15]. Trees pruning could be used to address the problems over fitting in decision tree [16]. There are various methods used in pruning and one of pruning method is reduced error pruning. Reduced error pruning or REP is a pruning method proposed by J.R. Quinlan (1987). REP is the simplest pruning method, easily to understand and most efficient technique to improve the accuracy of decision tree [17]. REP has been implemented in several previous studies. In the study from Atul Kumar Pandey et al (2013), C4.5 algorithm and REP is used to diagnose heart disease with the result that decision tree with no pruning has a truth percentage of 72.82%, while decision tree that pruned with REP has a percentage of truth 75,73% [18]. Hoon Jin et al (2016) in his research comparing the classification performance of decision tree with J48 algorithm and decision tree that use REP. The results showed that decision tree with J48 algorithm has an accuracy of 94%, while the decision tree that uses the REP has the accuracy of 96%, higher 2% compared to the accuracy of the decision tree with J48 algorithm [19].

According to the explanation, decision tree with C4.5 and REP method has been widely implemented in medical field for diagnosing disease. Altough the implementation of decisison tree with C4.5 and REP method for diagnosing dehydration in children still rare. Therefore in this study, decision tree with C4.5 algorithm and REP method will be implemented on medical intelligent system to produce high classification performance for diagnosing dehydration in children.

II. Method

A. Dataset

In this study, we collaborated with a hospital in Indonesia to collected medical record of children with dehydration that used as dataset. We also collaborated with a doctor from Indonesia to analize the dataset that was recorded. The dataset that used is the medical records of children aged 0 – 14 years with dehydration from February 2015 until November 2016. A total of 104 medical records were recorded as dataset. After cleaning process, the final data is 92 data and split into two subsets, trainingset (57 data) and testset (35 data).

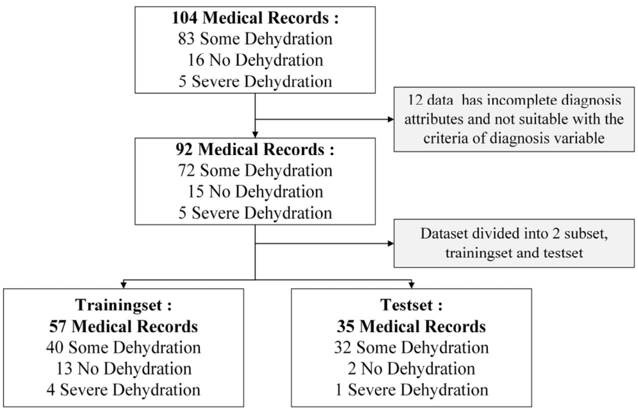

Fig. 1. Data preparation diagram

Before the data is processed in the classification process, the data must first be cleaned through the cleaning process data. Cleaning the data is one important step in the data mining process to determine incomplete and inconsistent dataset. After cleaning process, the final data obtained as many as 92 data that can be used in the classification process. A total of 11 dumped data is data that has incomplete diagnosis attributes and not suitable with the criteria of diagnosis variable, so the data should be discarded because they can not be used in classification process. The process of dataset preparation showed on Figure 1.

B. Input And Output Variable

Input variables used in this study are based on the diagnostics variables used in A. Levin et al study [20]. But not all the diagnosis variables on A. Levin et al study used in this study. This is because the three variables, heart rate, quality of pulse and tears have a value that is not suitable with the criteria condition and have incomplete data. Because of the condition of dataset and with the recomendation of all collaborated doctor so, the input variables used in this study is general appearance, respiratory, eye, turgor, and mucous membranes (showed on Table I). Although the diagnosis variables that are not used in this study may be the diagnosis variables that have a strong influence, but the variables that difficult to obtain are not likely to be a variable that efficient to used [20]. While the output variables used in this study is the severity of dehydration that divided into three categories, no dehydration, some dehydration, and severe dehydration.

TABLE I. Input Variable Of Intelligent System For Diagnosing Dehydration In Children

Atribut	Value
General appearance	Normal
	Lethargic
Respirations	Normal
	Deep
Eyes	Normal
	Sunken
Turgor	Normal
	Slow
Mucous membranes	Moist
	Dry

C. Decision Tree Building Process

The building process of decision tree is divided into two steps, first steps is building full decision tree with C4.5 algorithm and the second steps is pruning decision tree using REP method to produce final tree. Process of building decision tree showed in Figure 2, flowchart of C4.5 algorithm and REP method.

The learning process in C4.5 algorithm begins with calculating entropy of each of target attribute, and then followed with calculating the gain, split info and gain ratio for each attribute. Choose attributes with the highest gain ratio as root node. Then the calculation of entropy (equation 1), gain (equation 2), split info (equation 3) and gain ratio (equation 4) will repeated to get attribute with the highest gain ratio that will be used as internal node under root node. Calculation process will stop when the attribute with the lowest gain ratio obtained and used as final internal node.

$$Entropy(D) = \sum_{i=1}^{3} -p_i * \log_2 p_i \tag{1}$$

$$gain(D, A) = Entropy(D) - \sum_{i=1}^{3} \frac{|A_i|}{|D|} * Entropy(A_i) \tag{2}$$

$$SplitInfo(Ti) = \sum_{i=1}^{3} -\left(\frac{Ti}{T}\right)^{*} \log_2\left(\frac{Ti}{T}\right) \quad (3)$$

$$gain\ Ratio(A) = \frac{gain\ (A)}{SplitInfo(A)} \quad (4)$$

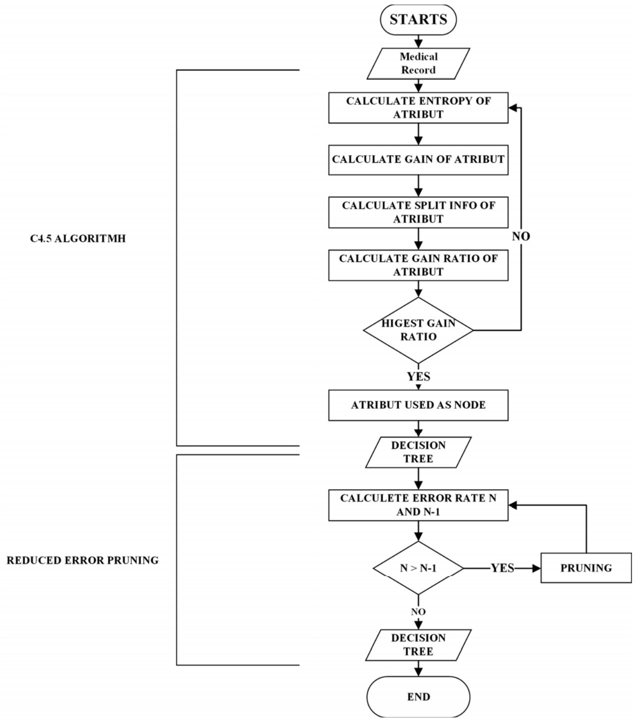

Fig. 2. Flowchart of decision tree construction process

After the decision tree is built, the next stage is pruning the decision tree. In this study, REP used as pruning method. REP will pruned the branch with highest error rate starting from the bottom to the top. Pruning is applied by replacing the internal attribute node with the leaf node that have maximum frequency the class.

The first step in REP method is to let the induction algorithm built the full decision tree. Then calculate the value of the error rate of each node using test set. After that, prune the decision tree by making internal node becomes leaf node and labeled with most dominant target attribute in the class. After decision tree is pruned then calculate the error rate of the branch. Then compare the error rates value of unpruned decision tree and error rates value of pruned decision tree. If the error rates value of pruned decision tree is less than error rates value of unpruned decision tree, then prune the decision tree Based on figure 2 at the stage of reduced error pruning, unpruned decision tree branch denoted as N-1 while the pruned decision tree branches denoted by N.

D. Validation of Classification Performance

In this study, the validation was done by comparing classification performance of the medical decision support system with the doctor diagnose. Presentation of correct prediction measured using classification accuracy, that can be calculated using equation (5). While the presentation of incorrect prediction measured using error rate value, that can be calculated using equation (6). Classification performance of model was mapped by using confusion matrix tables (Table II).

TABLE II. CONFUSION MATRIX TABLE

Confusion matrix table		**Predictif class**	
		Class = 1	***Class = 0***
Actual class	Class = 1	$f11$	$f10$
	Class = 0	$f01$	$f00$

$$accuracy = \frac{number\ of\ correct\ predictions}{total\ number\ of\ predictions} \quad (5)$$

$$error\ rate = \frac{number\ of\ wrong\ predictions}{total\ number\ of\ predictions} \quad (6)$$

III. RESULT AND DISCUSSION

After learning process on trainingset using the C4.5 algorithm and pruned with REP method, it will produce the final decision tree showed in Figure 3. The final decision tree produced a number of rules that can implemented for diagnosing dehydration on children.

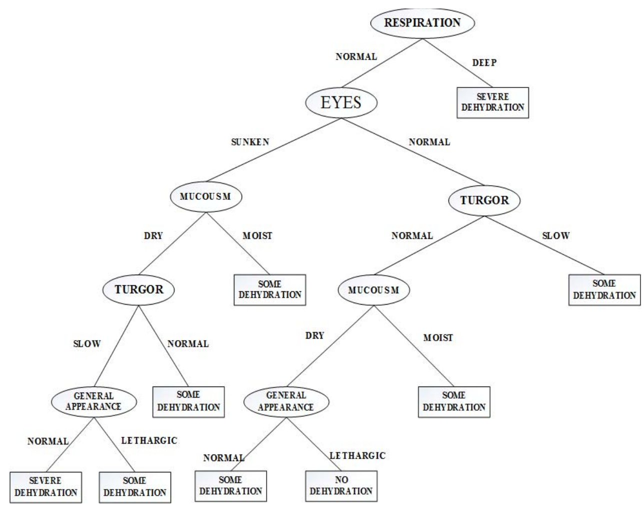

Fig. 3. Final constructed decision tree

After the final decision tree was constructed then, the model will test using testset to see the classification performance of the decision tree. In this study, the implementation of decision tree with C4.5 algorithm and REP method on intelligent system for diagnosing dehydration in childern produce good classification performance. From the total of 35 medical revord that used as testset, the intelligent system can correctly predict 32 cases consist of 31 cases of some dehydration and 1 case of severe dehydration ased on the doctor diagnose. The intelligent system, just got 3 predicted cases with uncorrect result that consist of 2 cases of some dehydration based on doctor diagnose. The classification performance of intellegent system for diagnosing dehydration mapped on confusion matrix table that shown on Table III.

TABLE III. THE CONFUSION MATRIX TABLE OF DECISION TREE.

		Model prediction			**Total**
		No dehydration	***Some dehydration***	***Severe dehydration***	
Doctor diagnose	***No dehydration***	0	2	0	2
	Some dehydration	1	31	0	32
	Severe dehydration	0	0	1	1
Total		1	33	1	35

In medical world, the implementation of decision for diagnosing disease has been proved give a better performance. In previous study, some researcher have been implementing decision tree for diagnosing disease and proved have a good result, showed with the performance of classification that can be see on the value of accuracy and error rate.

In this study the implementation of decision tree with C4.5 algorythm and REP method can produce a high classification performance. According to the result from the test, the intelligent system have the high accuracy. The over all accuracy of the intelligent system reach 91%, also have a high accuracy in diagnosing certain cases of dehydration severity, especially in diagnosing some dehydration cases that reach 96%. Not only have high accuracy, but the intelligent system also have a low error rate value of 0.085714286. The over all accuracy, the error rate value and the number of rules of the intelligent system showed in Figure 4.

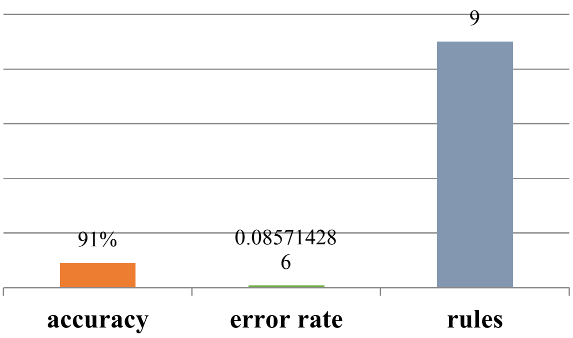

Fig. 4. The over all accuracy, the error rate value and the number of rules of the intelligent system

The final decisision tree of the intelligent system produce a simple rule. The final rules that generated from the decision tree just 9 rules (showed on Table IV). The output generated from 9 rules consists of 2 outputs severe dehydration, 1 output of no dehydration and 6 outputs of some dehydration. The low number of generated rules make decision tree become easily to understand by user.

TABLE IV. DECISION TREE RULES

No	**Medical Symptoms**					**Diagnosis**
	General Appearance	***Eyes***	***Mucous Membranes***	***Turgor***	***Respiration***	
1	-	-	-	-	Deep	Severe Dehidration
2	-	Sunken	Moist	-	Normal	Some Dehidration
3	-	Sunken	Dry	Normal	Normal	Some Dehydration
4	Normal	Sunken	Dry	Slow	Normal	Severe Dehydration
5	Lethargic	Sunken	Dry	Slow	Normal	Some Dehydration
6	-	Normal	-	Slow	Normal	Some Dehidration
7	-	Normal	Moist	Normal	Normal	Some Dehidration
8	Lethargic	Normal	Dry	Normal	Normal	No Dehydration
9	Normal	Normal	Dry	Normal	Normal	Some Dehydration

IV. CONCLUSION

In this study, decision tree using C4.5 algorithm with REP method has implemented on intelligent system for diagnosing dehydration in children. The developed system has a very good classification performance that showed from the accuracy and number of error rate value. Moreover the deeveloped system also produce a simple rule that easy to use for diagnosing dehydration in children. For the next study, expected to use a larger number of dataset so can generated decision tree with better classification performance.

ACKNOWLEDGMENT

This article was supported by Department of Electrical Engineering, Universitas Negeri Semarang, UEESRG, and Tugurejo hospital of Semarang, Indonesia.

REFERENCES

[1] WHO. 2013. *"Diarrhoeal Disease",* WHO Media Center.

[2] Laura M Lamberti, Christa L Fischer Walker, Robert E Black. 2012. "Systematic Review Of Diarrhea Duration And Severity In Children And Adults In Low- And Middle-Income Countries", BMC Public Health 2012, 12:276.

[3] Teuta Hoxha, Luan Xhelili, Mehmedali Azemi. 2015. "Performance of Clinical Signs in the Diagnosis of Dehydration in Children with Acute Gastroenteritis", Med Arh, 2015 Feb, 69(1): 10-12.

[4] Ron K Tam, Hubert Wong, Amy Plint, Nathalie Lepage, Guido Filler. 2014. *" Comparison Of Clinical And Biochemical Markers Of Dehydration With The Clinical Dehydration Scale In Children: A Case Comparison Trial",* BMC Pediatrics 2014, 14:149

[5] Robert Burduk, Michal Wozniak. 2012. "Different Decision Tree Induction Strategies for a Medical Decision Problem", Central European Journal of Medicine, 7(2), 2012, 183-193

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

[7] P Venkatesan, N R Yamuna. 2013. "Treatment Response Classification in Randomized Clinical Trials: A Decision Tree Approach", Indian Journal of Science and Technology, Vol: 6 Issue: 1 January 2013

[8] Somaya Hashem, Gamal Esmat, Wafaa Elakel, Shahira Habashy, Safaa Abdel Raouf, Samar Darweesh, Mohamad Soliman, Mohamed Elhefnawi, Mohamed El-Adawy, and Mahmoud ElHefnawi1. 2016.

"Accurate Prediction of Advanced Liver Fibrosis Using the Decision Tree Learning Algorithm in Chronic Hepatitis C Egyptian Patients", Gastroenterology Research and Practice Volume 2016

[9] Yan-Yan Song, Ying Lu. 2015. "Decision Tree Methods: Applications For Classification And Prediction", Shanghai Archives of Psychiatry, 2015, Vol. 27, No. 2]

[12] V.Shankar sowmien, V.Sugumaran, C.P.Karthikeyan, T.R.Vijayaram. 2016. "Diagnosis of Hepatitis using Decision tree algorithm", International Journal of Engineering and Technology (IJET), Vol 8 No 3 Jun-Jul 2016

[13] Phong D. Tong, Vinh D. Le, Hieu N. Duong, Hien T. Nguyen, Vaclav Snasel. 2016. "Decision trees for dianosis of dengue fever", International Conference on Information and Convergence Technology for Smart Society Jan. 19-21, 2016 in Ho Chi Minh, Vietnam.

[14] W. Nor Haizan W. Mohamed, Mohd Najib Mohd Salleh, Abdul Halim Omar. 2012. "A Comparative Study of Reduced Error Pruning Method in Decision Tree Algorithms", IEEE International Conference on Control System, Computing and Engineering, 23 - 25 Nov. 2012, Penang, Malaysia

[15] Shefali Chopra. 2016. "Improved Approach And Implementation For Decision Tree Classification Using Reduced Error Pruning", International Journal of Computing and Corporate Research, Volume 6 Issue 3 May 2016

[16] Rinkal Patel, Rajanikanth Aluvalu. 2014." A Reduced Error Pruning Technique for Improving Accuracy of Decision Tree Learning", International Journal of Engineering and Advanced Technology (IJEAT), Volume-3, Issue-5, June 2014

[17] Mradul Dhakar, Akhilesh Tiwari. 2013. " A New Model for Intrusion Detection based on Reduced Error Pruning Technique", I. J. Computer Network and Information Security, 2013, 11, 51-57.

[18] Atul Kumar Pandey, Prabhat Pandey, K.L. Jaiswal, Ashish Kumar Sen. 2013. " A Heart Disease Prediction Model using Decision Tree", IOSR Journal of Computer Engineering (IOSR-JCE), Volume 12, Issue 6 (Jul. -Aug.2013),PP83-8.

[19] Hoon Jin, Yong Gyu Jung. 2016. "Performance Comparison of Decision Trees of J48 and Reduced-Error Pruning", International Journal of Advanced Smart Convergence Vol.5 No.1 30-33 (2016)

[20] Adam C Levine, Justin Glavis-Bloom, Payal Modi, Sabiha Nasrin, Soham Rege, Chieh Chu, Christopher H Schmid, Nur H Alamb. 2015. "Empirically Derived Dehydration Scoring and Decision Tree Models for Children With Diarrhea: Assessment and Internal Validation in a Prospective Cohort Study in Dhaka, Bangladesh", Global Health: Science and Practice 2015, Volume 3, Number 3.

[10] Anna Kasperczuk, Agnieszka Dardzińska. 2016. " Comparative Evaluation Of The Different Data Mining Techniques Used For The Medical Database", acta mechanica et automatica, vol.10 no.3, 2016

[11] Ali Idria, Ilham Kadi. 2016. "Evaluating a decision making system for cardiovascular dysautonomias diagnosis", SpringerPlus (2016) 5:8