Design and Implementation of an Early Screening Application for Dengue Fever Patients Using Android-Based Decision Tree C4.5 Method

Iswatun Hasanah[a,1], Endah Purwanti[a,2], Prihartini Widiyanti[a,b]

[a]Biomedical Engineering Study Program, Departement of Physics, Faculty of Science and Technology, Universitas Airlangga, Surabaya, East Java, Indonesia
E-mail:[1]iswatun.hasanah-13@fst.unair.ac.id; [2]en4hp@gmail.com

[b]Institute of Tropical Disease, Universitas Airlangga, Surabaya, East Java, Indonesia
E-mail: pwidiyanti@fst.unair.ac.id

Abstract— **In Indonesia, dengue fever has been a public health problem for the past 46 years. According to the World Health Organization (WHO), in 2011, an estimated 2.5 billion people, or about 2/5 of the world population living in tropical and subtropical areas, were at great risk of being infected with dengue fever every year. This study aims to take advantage of the development of Android smartphone technology to design a system capable of screening the early stage of dengue fever with four possible classes, namely1[st]Degree, 2nd Degrees, 3[rd]Degrees, and non-dengue using the C4.5 Decision Tree based on patients medical records at Hajj General Hospital (*Rumah Sakit Haji*), Surabaya, East Java. The initial diagnostic decision of this application was determined by ten input parameters, i.e., unstable fever, conjunctival hemorrhage, rash, headache, age, pulse pressure, nausea/vomiting, body temperature, heartburn (abdominal pain), and decreased appetite. To determine diagnosis results in this application, trees formation using the highest gain ratio for each parameter was conducted. This early screening application for dengue fever patients recorded an accuracy score of 95% out of 20 data tested. Evaluation results for this application showed some good ratings by obtaining the average value of the survey in visual design and user interaction test at 8.3 rates; functionality test 8.932; performance and stability at 9.168; and overall satisfaction test at 8.733. This application also recorded a high accuracy level and good application performance.**

Keywords— **dengue fever; decision tree C4.5; android system; dengue fever patients medical records.**

I. INTRODUCTION

Dengue Fever (DHF) is a mosquitoes-infected disease caused by the dengue virus and spread by *Aedes aegypti* and *AedesAlbopictus* [1], [2]. Infected patients experienced several symptoms of mild to high fever, frequent headaches, muscle and joint pain, and spontaneous hemorrhage [3]. Dengue fever has been one of the health issues in the world whose prevalence has been increasing and spreading more widely. An estimated 2.5 billion people or about 2/5 of the world population living in tropical and subtropical areas are at the greatest risk of dengue infections every year [4].

In Indonesia, dengue fever has been a public health problem for the past 46 years. Since 1968, the spread of dengue fever, which was first discovered in Surabaya and Jakarta, has been increasing in all provinces in Indonesia [3]. In 2008, the number of dengue fever cases in Indonesia reached 137,469 people, with 1,187 deaths or Case Fatality Rate (CFR) at 0.86%, while in 2009, there were 154,855 cases where 1,384 people died or CFR at 0.89%. Moreover, the number of dengue fever cases in Indonesia in 2010 was 156,086 cases, with 1,358 death and CFR at 0.87%. Dengue fever is ranked 2nd in the Top 10 Disease, which caused in-patient hospitalization in 2010 [5].

The increasing number of cases and affected areas are caused by increasing migration and new settlement, lack of community awareness to clean mosquitos nests, the presence of mosquito vectors nearly in every corner of this country, and four types of virus cells circulating throughout the year. Furthermore, there are several factors affecting dengue fever incidence, namely host factor, environment factor, clean and healthy lifestyle, and virus factor. The host factor talks about susceptibility and immune response; environmental factors are related to geographic conditions (altitude from sea level, rainfall, wind, humidity, season); and demographic conditions are related to density, mobility, behavior, customs [6].

As the dengue fever rate and its death toll keep escalating, based on the reported cases, it is required more serious handling through early diagnosis or detection so that the

disease can be handled more quickly and, in turn, reduce the death toll. In this modern era, there have been numerous researchers conducting researches on the early detection of diseases using artificial intelligence. Those studies were conducted to help prospective patients of dengue fever to be diagnosed earlier. There is a method called Decision Tree using the C4.5 algorithms, which is an algorithm capable of processing continuous data. It has several advantages, such as fast process, high accuracy, and easily understandable and implementable results [7].

Based on the explanation above, it is deemed necessary to create an application for dengue fever early screening. The data can be obtained from patient medical records using Decision Tree C4.5 methods. This study shall identify dengue fever using some of its clinical symptoms and then accumulate it with Decision Tree C4.5 methods to diagnose whether someone is suffering from dengue fever or not, as well as diagnosing his/her severity. Clinical symptoms are detected using some inputs, namely age, rash, conjunctival hemorrhage, unstable fever, decreased appetite, nausea or vomiting, heartburn, headache, pulse pressure, and patient body temperature. This study shall also display the Android application. This application can be used to display dengue fever diagnosis and facilitate user access to the latest technology. Finally, this research is also expected to help early screening of dengue fever for the public, so they can immediately take action.

II. MATERIALS AND METHOD

The research was conducted at the Laboratory of Medical Instrumentation, Biomedical Engineering Program, Faculty of Science and Technology, Airlangga University, as well as in Hajj Public Hospital Surabaya, East Java in approximately six months, from February to June 2017. The data were taken from dengue fever medical records at Haji General Hospital, Surabaya. While the ten parameters used are age, rash, conjunctival hemorrhage, unstable fever, decreased appetite, nausea or vomiting, heartburn, headache, pulse, and body temperature. The application program output consists of 4 classes, namely non-DB, DB degree 1, DB degree 2, and DB degree 3.

This research procedure is titled "Design and Implementation of an Early Screening Application for Dengue Fever Patients using Android-based Decision Tree C4.5 Method" and was conducted in several stages. Figure 1 below shows a flow chart of the research method. A decision tree is the most popular classification method in data mining. This method presents prediction procedures that are visualized in the form of a tree so that the decision making path can be clearly observed.

They have several advantages over other techniques [8]:
- The simplicity of its presentation makes it easy to understand.
- They can work for different types of attributes, nominal or numerical
- They can classify new examples fast.

Decision Tree Learning is identical to the process of training and data testing [9]. There are several steps needed to classify those processes, from the training process to data testing using Decision Tree C4.5 methods.
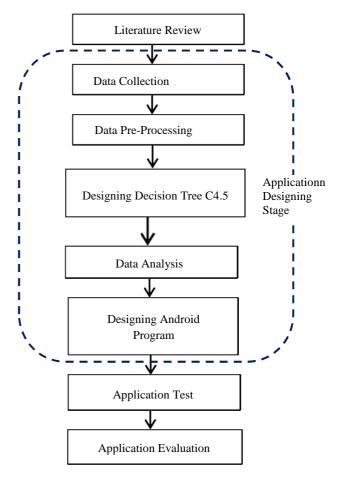

Fig. 1 Research method flow chart

The flow chart of early diagnosis application using Decision Tree C4.5 methods is illustrated in Figure 2. The stages of creating a Decision Tree C4.5 are presented in the following subsections.

A. Select an Attribute as the Root

Attribute selection as the root depends on the highest gain value from an existing attribute. To calculate the gain ratio value of C4.5 algorithms, another measurement called entropy shall be understood. Entropy is a parameter to measure the data sample heterogeneity. If the data pool is heterogeneous, then the entropy value shall be greater. Entropy calculation is mathematically shown in Equation 1 [9].

$$Entropy\ (X) = \sum_{j=1}^{k} p_j * \log_2 \frac{1}{p_j} = - \sum_{j=1}^{k} p_j * \log_2 p_j \qquad (1)$$

Where,
X: a set of cases
k: number of X partitions X
pj: the proportion of Xj to X

B. Attribute Measurement

After obtaining entropy value for a set of data samples, the attribute effectiveness in classifying data can be measured by information gain. Mathematically, information gain from attribute A is given in Equation 2 [10].

$$Gain\ (X, A) = Entropy(X) - \sum_{j=1}^{k} \frac{|X_i|}{X} * Entropy\ (X_i) \qquad (2)$$

Where,
A: Attribute
i: declare a possible value for attribute A

$|X_i|$: Number of samples for i value

$|X|$: total number of data samples

Entropy (X_i): entropy for samples having i value

In C4.5 tree construction, in each tree node, the attribute with the highest gain ratio value is selected as a split attribute for the node. The formula of the gain ratio is given in Equation 3 [11].

$$Gain\ ratio\ (X,A) = \frac{Gain\ (X,A)}{SplitInfo\ (X,A)} \quad (3)$$

Where *Gain(X, A)* is gain information of attribute A for X sample set. At the same time, *SplitInfo (X, A)* denotes the potential information obtained at X sub-division X into n sub-sets based on a study on Attribute A. S*plitInfo (X, A)* is formulated in Equation 4 [11].

$$SplitInfo(X,A) = -\sum_{j=1}^{k} \frac{|X_i|}{X} * \log_2\left(\frac{|X_i|}{|X|}\right) \quad (4)$$
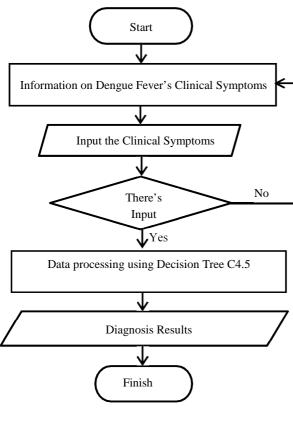

Fig. 2 Program flow chart

III. RESULTS AND DISCUSSION

A. *Design Results of Decision Tree C4.5*

Decision Tree Learning is identical to the process of training and data testing as a way to classify [9]. The training process to design a tree used up 100 amount of data. The symbols are given at each variable as program input and

output. The symbols for parameter input is presented in table 1 and table 2.

TABLE I
SYMBOLS OF THE INPUT PROGRAM

Input Data	Input Code
Age	X1
Rash	X2
Conjunctival hemorrhage	X3
Decreased appetite	X4
Nausea/Vomiting	X5
Heartburn	X6
Heartburn	X7
Unstable Fever	X8
Body Temperature	X9
Pulse	X10

TABLE II
SYMBOLS OF THE OUTPUT PROGRAM

Output Data	Output Code
1st Degree of DF	G1
2nd Degree of DF	G2
3rd Degree of DF	G3
Non-DF patients	G4

The design results of Decision Tree C4.5 are illustrated in Figure 3**.**
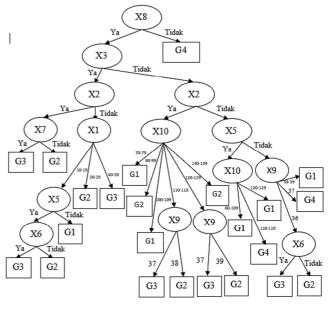

Fig. 3 Design results of decision tree C4.5

Making this decision tree starts with selecting the input parameters that will be used as the root node. The selection of the root node is based on the highest gain value of the existing parameters. To calculate the gain ratio used formula 3. From the results of the calculations performed, it was found that the largest gain ratio value is the heat input parameters up and down, which is equal to 0.62939. Therefore, the heat up and down parameters are chosen as the root node, for the next branch, recursively calculated the gain ratio in such a way that a decision tree is obtained in Figure 3.

The decision tree results show that the heat up and down parameters are the most important parameters in determining

whether a person is diagnosed with dengue fever. If the user answer is "no" then the program output will immediately indicate not dengue and if the user answer is "yes" then the tree will check the conjunctival bleeding parameters. Decision making to determine the program output then follows the path in the decision tree that has been built.

B. Interface Design

The application program design consists of 5 parts, namely splash screen, MyIntro, Main menu, Diagnostic window, and Information window. Splash screen (Fig.4) is a page designed to display the beginning of using an application program. MyIntro (Fig.5) is a page that informs users about application program specifications. The main menu (Fig.6 and Fig.7) is the main page as the center of user interaction with the program. On this page, there are buttons "about the application", "DHF info", "diagnosis," and "exit". The diagnosis window (Fig.8 and Fig.9) is a window for DHF screening with ten input parameters in the form of clinical symptoms experienced, and the program output is Non-DHF, DHF 1, DHF 2, DHF 3, and DHF 4. Information window (Fig.10, 11, and 12) is a window that displays information related to dengue fever, namely the definition of DHF, DHF symptoms, the severity of DHF, and DHF prevention.
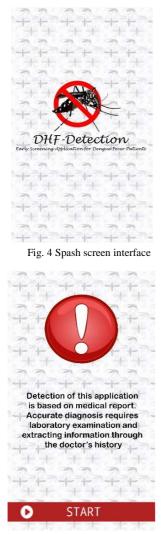

Fig. 4 Spash screen interface
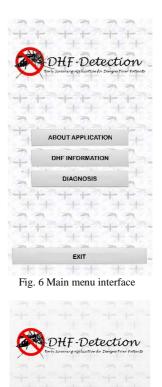

Fig. 5 My intro interface

Fig. 6 Main menu interface

Fig. 7 Fragment interface for exit button

(a)   (b)
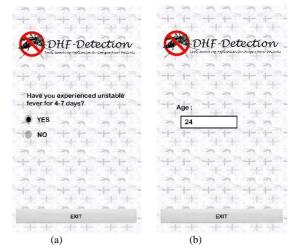Fig. 8 (a) Diagnose interface with radio button; (b) Diagnose interface with edit text

C. Application Test Results

Before a finished program is launched to users, according to Google Play Developer Guidelines, there are several stages of program readiness test on the Android system.

2240

1) Visual design and user interaction test: For visual design and user interaction test, the created application was expected to fulfill Android Design Guidelines, such as not using system icons which deviate far from application function, as well as the existence of button, back, and other features. The test results are displayed in Table 3 [12].

TABLE III
THE RESULT OF THE VISUAL DESIGN AND USER INTERACTION TEST

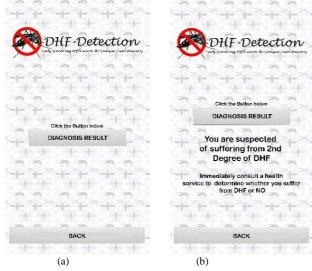| Area	ID	Test-ID	Result
Standard Design	UX-B1	CR-all	Passed
Navigation	UX-N1	CR-3	Passed
	UX-N2	CR-3	Passed
	UX-N3	Cr-1	Passed
Notifications	UX-S1	CR-11	Passed
Notifications	UX-S2	CR-11	Passed

Fig. 9 (a) Diagnose interface prior to diagnosis results; (b) diagnose interface after diagnose results
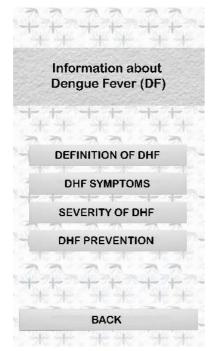

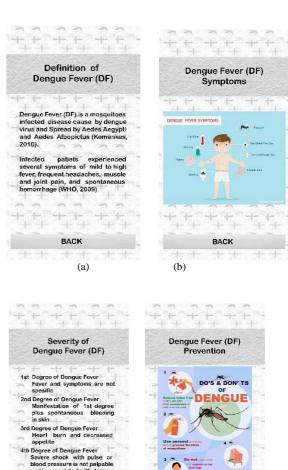Fig. 10 Interface of information on dengue fever

Fig. 11 Some fragments of dengue fever information (a) Definition of dengue fever; (b) Dengue fever symptoms; (c) Dengue fever severity level; (d) Dengue fever prevention

Fig. 12 The interface on application description

2) Functionality Test: During the functionality test, the created application was expected to run and function correctly on a targeted device (minimal). Some of the other tests revolved around not to leave the application on the background after it is closed, enabling the application to start, and restart at the same state. The test results are displayed in Table 4 [12].

TABLE IV
FUNCTIONALITY TEST RESULTS

Area	ID	Test-ID	Result
Permission	FN-P1	CR-11	Passed
	FN-P2	CR-11	Passed
Install Location	FN-L1	SD-1	Passed
Audio	FN-A1	Cr-7	Unknown
Audio	FN-A2	CR-8	Unknown
Audio	FN-A3	CR-1, CR-2	Unknown
Audio	FN-A	CR-1, CR-8	Unknown
UI and Graphics	FN-U1	CR-5	Passed
UI and Graphics	FN-U2	CR-5	Passed
UI and Graphics	FN_U3	CR-5	Passed
User/App State	FN-S1	CR-6	Passed
User/App State	FN-S2	CR-6	Passed

For this test stage, several audio test results were unknown. It was because the created application did not contain file, command, player, and notification in audio.

3) Performance and Stability Test. During the performance and stability test, this application was expected to run and function well on a targeted device (minimal). Some tests conducted during this stage were possible lag, crash, freeze, and forced close, which might happen when the application ran under both normal and strict modes. Those test results are displayed in Table 5 below [12].

TABLE V
PERFORMANCE AND STABILITY TEST RESULTS

Area	ID	Test-ID	Result
Stability	PS-S1	CR-all, SD-1, HA-1	Passed
Performance	PS-P1	CR-all, SD-1	Passed
	PS-P2	PM-1	Passed
Battery	PS-B1	BA-1	Passed
Media	PS-M1	CR-all, SD-1, HA-1	Passed
Visual Quality	PS-V1	CR-all	Passed
Audio	PS-V2	CR-all	Passed

4) User Satisfaction Test: For this user satisfaction test, this application was given to some users to be tested and surveyed. It was also conducted to discover user satisfaction. Besides, it could be used to discover any bugs and ask for user feedback. These test results are illustrated in Table 6 [12].

TABLE VI
AVERAGE SURVEY RESULTS

Num.	Aspects	Average Value
1.	Visual Design and User Interaction	8.3
2.	Functionality	8.932
3.	Performance and Stability	9.168
4.	Overall Satisfaction	8.733

In Table 6, all average values from each aspect are displayed on the dengue fever detection application. For visual design and user Interaction, the average value was 8.3.

It shows good user satisfaction in its design and interaction. For the functionality test, the average value was 8.932. It proves that the program functions well. Moreover, the performance and stability aspect recorded an average value of 9.168. It shows that the program performs well without any disruptions, such as lag, forced close, and many others. Finally, overall satisfaction, which was a test for the overall program, obtained 8.733 average value.

D. Evaluation of the Application

The test phases used as many as 20 test data. Those tests were conducted based on Decision Tree C4.5 methods. Determining the optimal accuracy level for early screening of dengue fever by comparing the results of classification output between the test results using the Decision Tree C4.5 method and classification results from a physician. The test results using Decision Tree C4.5 methods are presented in Table 7.

TABLE VII
THE TEST RESULTS

Data	Application Diagnose Results	Expert Diagnose Results	Notes
1	NON-DF	NON-DF	Matching
2	NON-DF	NON-DF	Matching
3	NON-DF	NON-DF	Matching
4	1st Degree of DF	1st Degree of DF	Matching
5	1st Degree of DF	1st Degree of DF	Matching
6	1st Degree of DF	1st Degree of DF	Matching
7	1st Degree of DF	1st Degree of DF	Matching
8	1st Degree of DF	1st Degree of DF	Matching
9	1st Degree of DF	1st Degree of DF	Matching
10	1st Degree of DF	1st Degree of DF	Matching
11	1st Degree of DF	1st Degree of DF	Matching
12	1st Degree of DF	1st Degree of DF	Matching
13	2nd Degree of DF	2nd Degree of DF	Matching
14	2nd Degree of DF	2nd Degree of DF	Matching
15	2nd Degree of DF	2nd Degree of DF	Matching
16	2nd Degree of DF	2nd Degree of DF	Not-Matching
17	2nd Degree of DF	2nd Degree of DF	Matching
18	3rd Degree of DF	3rd Degree of DF	Matching
19	3rd Degree of DF	3rd Degree of DF	Matching
20	3rd Degree of DF	3rd Degree of DF	Matching

Based on Table 4.11, it is proven that the accuracy level of early screening application for dengue fever by comparing the results of expert diagnosis and application diagnosis reached 95%. Basically, in order to provide a diagnosis, doctors firstly ask about the patients symptoms. Furthermore, the diagnosis is then reinforced with a laboratory test, namely about patient thrombocyte level and many other things. However, in this study, the diagnostic process, which was conducted through an application and used a data set, is considered sufficient by some doctors. This research can help users perform the initial screening process so that they can find out about their illness soon and get it handled earlier, as well. However, in countries with endemic diseases, it is necessary to perform laboratory

service as a control strategy for a patient disease [10]. Dengue is a re-emerging disease that often becomes recurrent epidemics. The initial detection of clinical manifestation is frequently confusing with other febrile states of another disease [12]. The identification of algorithms that differentiate dengue from other febrile symptoms would be very beneficial to perform primary care to the suspected patient and protect them fall into the worse condition.

IV. CONCLUSIONS

Decision Tree C4.5 method based on patients' medical records discovered that unstable fever was the first node influencing the early screening application. It is followed by conjunctiva hemorrhage as the first branch node, rash as the second and third branch node, headache as the fourth branch node, age as the fifth branch, nausea/vomiting as the sixth branch, heartburn as the seventh branch, pulse as the eighth branch, body temperature as the ninth and tenth branch, nausea/vomiting symptom as the eleventh branch, pulse symptom as a twelfth branch, and heartburn symptom as the last branch.

This application and its Decision Tree C4.5 method recorded 95% of accuracy. The accuracy results were obtained by comparing the number of correct data from test results and dividing it with the total amount of test data. The results of Android-based early screening applications for dengue fever patients show good respond. The application obtained 8.3 rates for the visual design and user interaction test; 8.932 for the functionality test; 9.168 for the performance and stability test; and 8.733 rates for overall satisfaction.

ACKNOWLEDGMENT

REFERENCES

[1] Center for Disease Control and Prevention. US Department of Health and Human Services. https://www.cdc.gov/dengue/epidemiology/, p 1-2J.

[2] Doggett, Stephen L. NSW Arbovirus Surveillance & Vector Monitoring Program: Mosquito Photos. 2003. http://medent.usyd.edu.au/arbovirus/mosquit/photos/mosquitophotos.htm

[3] WHO. Guidelines for Diagnosis, Treatment, Prevention and Control. WHO Library Cataloguing in Publication Data. 2009. p. 4.

[4] World Health Organization. World Health Statistics, 2011. WHO Library Cataloguing in Publication Data. 2011.

[5] Kementrian Kesehatan. ProfilKesehatan Indonesia Tahun 2010. Jakarta: Pusat Data dan Informasi. 2011.

[6] Pusat Data dan Surveilans Epidemiologi Kementrian Kesehatan Republik Indonesia. Demam berdarah dengue. Buletin jendela epidemiologi 2010;2

[7] Kusrini, Luthfi, E. T. Algoritma Data Mining. Edisi 1. 2009.Yogyakarta. Andi Offset.

[8] W. Häm̈al̈ainen and M. Vinni, "Classifiers for educational data mining," *Handbook of Educational Data Mining*, 2010.

[9] Suyanto. Artificial Intelligence: Searching, Reasoning, Planning, and Learning. Bandung. Informatika. 2012.

[10] Prasetyo, Eko. Data Mining: Mengolah Data Menjadi Informasi Menggunakan Matlab. Yogyakarta: Andi Publisher. 2014.

[11] Badr, H., Merbouha, A., Ezzikouri, H., Erritali, M. A Comparative Study Of Decision Tree ID3 and C4.5.Morocco: TIAD laboratory, Computer Sciences Department, Faculty of sciences and techniques, Sultan Moulay Slimane University. 2014.

[12] Developer Android. Core App Quality. [cited : 2016 Oktober 16]. Available from: URL: https://developer.android.com/distribute/essentials/quality/core.html

[13] Lukas Tanner, Mark Schreiber, Jenny G. H. Low, Adrian Ong, Thomas Tolfvenstam, Yee Ling Lai, Lee Ching Ng, Yee Sin Leo, Le Thi Puong, Subhash G. Vasudevan, Cameron P. Simmons, Martin L. Hibberd, Eng Eong Ooi, "Decision Tree Algorithms Predict the Diagnosis and Outcome of Dengue Fever in the Early Phase of Illness" Plos Neglected Tropical Diseases, vol. 2, issue 3, e196,pp. 1-9, Mar 2008