Expert Systems for Self-Diagnosing of Eye Diseases Using Naïve Bayes

Rahmad Kurniawan
UIN Sultan Syarif Kasim Riau, Indonesia; [2]Fakulti Teknologi & Sains Maklumat,
Universiti Kebangsaan Malaysia,
43600 Bangi, Selangor, Malaysia;
*Corresponding author:
rahmad_kurniawan@yahoo.co.id

Novi Yanti
UIN Sultan Syarif Kasim Riau, Indonesia
novikaliyee@yahoo.com

Mohd Zakree Ahmad Nazri
Fakulti Teknologi & Sains Maklumat,
Universiti Kebangsaan Malaysia,
43600 Bangi, Selangor, Malaysia;
zakree@ukm.edu.my

Zulvandri
UIN Sultan Syarif Kasim Riau, Indonesia
zulvandri@yahoo.com

Abstract— **The best defense against eye diseases is to have regular checkups. However, in reality, poverty stops people outside the developing world from seeing an eye doctor regularly. Thus, many patients did not get appropriate treatment for their eye disease until it is too late. This paper presents an expert system for diagnosing eye disease based on Naive Bayes. The developed expert system applies Case-Based Reasoning (CBR), which is a paradigm for reasoning from experience while the Naïve Bayes is used as a method for classifying eye diseases by applying Bayes' theorem. The outputs of the expert system are classification of an eye disease and information on the best treatment. The result of this study is obtained by comparing the expert system diagnostic results with an expert diagnostic result. Based on the experimental results, the Naïve Bayes based expert system has been able to obtained 82% accuracy. Thus, it can be concluded that an expert system with Naïve Bayes has the potential to be used effectively by the people but still has plenty room for improvement.**

Keywords—Case-Based Reasoning; Expert System; Eye Disease; Naïve Bayes;

I. INTRODUCTION

Seeing is one the five most important senses which are hearing, smelling, tasting and touching. There are sayings that, if we take good care of our eyes, the eyes will take good care or ourselves. Without a good diet and regular checkups with an eye doctor, eye disorders such as dry eyes, glaucoma and others can develop in one or both eyes. Eyes disorder cause vision problems globally and they are more common among poor people. Poverty stops people outside the developing world from seeing an eye doctor regularly. Thus, many patients did not get the necessary advice from an expert and therefore they do not take their eye disorder seriously. Usually, patients believe that their eye disorder is not permanent. Online self-diagnosing is one of the solutions as the Internet is relatively cheap to get more information and cybercafé is one of the prevalent ways for poor people to search about their predicament. However, self-diagnosis is prone to misdiagnosis that may endanger the users health if users make a wrong decision. Thus, the main goal of this project is not just to develop an online self-diagnosing of eye diseases but to guide users to seek medical experts advise at the nearest hospital so that patients will get attention, medical treatment appropriately and as early as possible.

Nowadays, there are many expert systems being used in many domains such as in healthcare, legal and financial. In the 80s, most of the expert systems have been purely rule-based before a new paradigm for reasoning was introduced at Yale University called the Case-Based Reasoning (CBR) [1].In this study, we used the Case-Based Reasoning (CBR) for reasoning. CBR is a paradigm for reasoning from the solutions of previous cases. CBR has a memory model that organizes and maintain past cases and a process model for retrieving and modifying old cases and assimilating new ones. Thus, the CBR-based expert system can identify the type of disease using the experiences of previous case that stored as a case base.

In recent years, the integrated methods of case-based reasoning have become an increasingly important research issue in the case-based reasoning community [7]. An expert knowledge is an important component in this study, to understand the effectiveness of an expert system when using a Case-Based Reasoning paradigm.

One of the advantages in using CBR in this project is that the expert system will have a substantial expert knowledge in the knowledge based. Moreover, it tries to mimic one of the method for doctors to diagnose their patients as reasoning with cases is important in medical practice of all kinds. The doctor is trained to use facts, knowledge and hands-on experiences. They were trained by studying real life cases so that later they will be able to recognize disorders or common combinations of disorders, how the disorders were treated and later make inferences (drawing a conclusion). In CBR research, there are many approaches for modeling the prior and making inferences. The most common approaches in CBR are statistically-oriented, model-based, planning/design-oriented, exemplar-based, and adversarial or precedent-based. In this work, we used statistical-oriented approach by using the Naïve Bayes for making inference.

Naïve Bayes method is used for to determined which type of eye disease has the highest probability to be effecting a patient. Naïve Bayes calculate the chances based on existing symptoms and the system will later on be able to appropriate preventive and medical solutions to the type of illness accordingly.

II. RELATED WORK

Case-Based Reasoning (CBR) which is introduced in the 80s, is one of the methods that have been applied in Problem Solving Environment (PSE) where each problem and solutions are represented in the form of cases [15].In 2001, Phuong et al [8] proposed an approach that combines CBR with Rule Based reasoning for lung disease diagnosis. Their study have applied fuzzy set theory for Case Based Reasoning and Rule Based Reasoning. This is one of the earliest work in using CBR for a diagnose system.

Ibrahim [2] developed an expert system to diagnose numerous types of eye disease in Malaysia. However, the system does not shows how the conclusion was made as such, by stating or ranking the probability for each eye disease that the user could have suffered. The reason is, although the users have different symptoms, a patient may suffer the same disease. Thus, it is very important to know how the system come to conclusion or make inference as both two cases, have different percentage of probability.

Neural Network and Decision Tree are two popular methods that have been used for reasoning in diagnosing eye diseases [3]. Syiam [4] also depends on Neural Network to diagnose eye diseases. The system built by Syiam is based on a multilayer feed forward networks with a single hidden layer.

Chandra and Gupta [5] suggested a novel method to overcome the limitations of expert system during the inference process using a robust function for estimating probabilities by using Naïve Bayes. Naïve Bayes also has been used for text classification in machine learning based on the conditional probability of features belonging to a class, which the features are chosen by feature selection methods.

III. NAÏVE BAYES

Naïve Bayes or better known as Naïve Bayes classifier is a classification with a simple method of probability and statistics that was created by the British scientist named Thomas Bayes. The Bayes theorem is combined with the "Naïve" in which the every attribute conditions are assumed independent [9]. In addition, the Naïve Bayes Classifier is suitable when the dimension has high input. Despite its simplicity, Naïve Bayes can often do better than more sophisticated classification methods [10]. Naïve Bayes also allows any attributes to contribute to the final decision of the other independent attributes [11].

Naïve Bayes is widely used for classification in machine learning. Naïve Bayes also used for many classification problems because it is simpler and gives better accuracy than other supervised learning methods [5]. Naïve Bayes is a simple probabilistic classifier that learns from training data and then predicting the test instance class uses the highest posterior probability [12]. Let C be the random variable that denotes the class of an instance and let $X\ (X_1, X_2,..., Xm)$ be a vector of random variables denoting the observed attribute values. Let c_j represent jth class label and let $x\ (x_1, x_2, \ldots ,xm)$ represent a particular observed attribute value vector. To predict the class of a test instance x, Bayes theorem used to compute the probability as follows:

$$p(C = cj \mid X = x) = \frac{p(C = cj)p(X = X \mid C = cj)}{p(X = x)} \tag{1}$$

Naïve Bayes known with the advantages of high efficiency and good classification accuracy and they have been widely used in many domains. However, the classifiers need complete data. Leng et al. [13] compared Naïve Bayes with common methods that have been used in dealing with missing data. Their research found that the Naïve Bayes method is more efficient and reliable [13]. Naïve Bayes is one of the most effective and efficient classification algorithms. It is a simple probabilistic classifier based on applying Bayes' theorem with strong (naive) independence assumptions. Naïve Bayes classifier is a straightforward, frequently used method for supervised learning.

In addition, Naïve Bayesian classifier is one of the most effective and efficient classification algorithms. The simplicity and apparent accuracy of Naïve Bayes even when the independence assumption violated, fosters the on-going interest in the model[14]. The Naïve Bayes method is more popular because it produces better accuracy, the attributes are mutually independent and getting the output is more efficient than other methods [10].

IV. RESEARCH METHODOLOGY

The following are the research methodology of this study. The construct stages of this study are as follows:

A. Data Collection /Knowledge Acquisition

In this stage, an interview with an expert as references that will serve as a knowledge base to develop of the system.

B. Identify the Problem

The second stage is to identify problems by determining the required data and information to analyze and build an expert system.

C. Method

In this stage, we selected the appropriate method to address the problem. The Naïve Bayes method used in this study for calculating the possibility of patients affected by the eye disease. On the other hand, Case-Based Reasoning (CBR) used for reasoning.

D. Analysis

In this study, analysis has become the most important step. Analysis was performed by analyzing the Naïve Bayes and CBR, knowledge base, and an inference engine. Analysis carried out in order to the system worked accordance with expected. There are several information which became the focus of this step:

a) Reasoning : Case-Based Reasoning
b) Inference : Naïve Bayes

E. Implementation and Testing

We have conducted this stage after the analysis and design phase is completed. Implementation and testing are conducted for developing a system using an appropriate programming language as a web based.The system validation using black box method and User Acceptance Test (UAT). UAT is done by testing the system on patients who suffer the eye diseases. The diagnosis result of expert system will be assessed by an expert.

F. Conclusion and Documentation.

Conclusion has included the overall findings and limitations. The purpose of the study and limitations are to help improve the quality for the next research. Furthermore, the research was documented from start to finish testing.

V. EXPERT SYSTEM ANALYSIS

Here are some of the stages of the analysis which has been carried out:

A. Analysis of Inference

Inference used for the construct of this expert system, which is Naïve Bayes method in thediagnosis process of eye diseases.

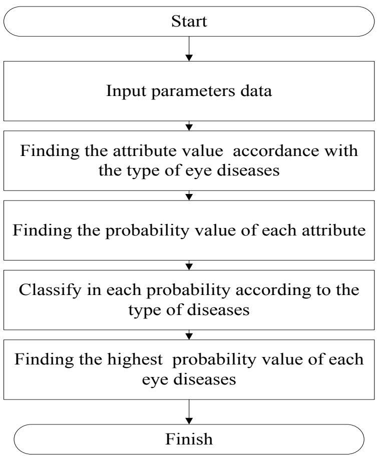

Figure 1. Naïve Bayes Process

a) Naïve Bayes provided the diagnostic results use the case base as the data learning. The process was performed using the Naïve Bayes method and calculating the probability of each symptom from users. Subsequently, the final diagnosis based on the greatest chances value. The overview of Naïve Bayes process is shown in Figure. 1.

B. Reasoning Analysis

Based on the advantages of Case-Based Reasoning (CBR), that the CBR model is used to solve problems and generate the results based on historical eye disease cases. Several processes have to be completed in order to apply CBR for diagnosis which are Retrieve, Reuse, Revise, and Retain.

1) Retrieve

Retrieve is conducted bya process called the initiation of the problem (case). This process includes a process to identify new problems. In particular, patients that shows similar symptoms with the symptoms that exists in the previous case. The following tables are some examples from the available cases.

TABLE I. THE NEW CASES

Symptoms	Options
Red eyes	No
Feels pain in eyes	No
Feels like there is dust in the eyes	No
The eyes sticky in the morning	Yes
Often the squinted	No

2) Reuse

The system will search in the database of historical cases. Then, the similar cases will be used to reasons for classifying new case. In this process, the system search and use the classified disease which is inferred from previous cases.

TABLE II. PROBABILITY OF EVERY DISEASE

Eye Diseases	Probability Value
Blepharitis	9/140
Dacryocystitis	9/140
Episcleritis	11/140
Glaucoma	7/140
Hordeolum	8/140
Cataracts	12/140
Keratitis	12/140
Conjunctivitis	14/140
Myopia	11/140
Scleritis	12/140
Corneal ulcers	14/140
Uveitis	10/140

The system look for the probability of each symptom experienced and determine the greatest probabilities of disease by calculating the probability of each symptom based on the type of eye diseases. This is one example for Blepharitis:

P(Symptoms|Blepharitis)=
P(S1=No|Blepharitis)*P(S2=Yes|Blepharitis)*
P(S3=No|Blepharitis)*P(S4=No|Blepharitis)*
P(S5=Yes|Blepharitis)*P(S6=Yes|Blepharitis)*
P(S7=Yes|Blepharitis)*P(S8=Yes|Blepharitis)*
P(S9=No|Blepharitis)*P(S10=Tidaak|Blepharitis)*
P(S11=Yes|Blepharitis)*P(S12=No|Blepharitis)*
P(S13=No|Blepharitis)*P(S14=No|Blepharitis)*
P(S15=No|Blepharitis)*P(S16=No|Blepharitis)*
P(S17=Yes|Blepharitis)*P(S18=No|Blepharitis)*
P(S19=No|Blepharitis)*P(S20=No|Blepharitis)*
P(S21=Yes|Blepharitis)*P(S22=No|Blepharitis)*
P(S23=Yes|Blepharitis)*P(S24=Yes|Blepharitis)*
P(S25=No|Blepharitis)*P(Blepharitis)
= 0.22*0.89*0.89*0.89*0.89*0.89* 0.89*0.78*
0.78*0.89*0.11*0.89*0.89*0.89*0.89*0.89* 0.11*0.78*
0.89*0.89*0.11*0.89*0.89*0.11*0.11*0.06
=2.60561E-07 ≈ **0.00269**

Based on the calculation, the obtained probability value for Blepharitis is P(Symptoms|Blepharitis) = 2.60561E-07 ≈ **0.00269**. The next step is to classify (make inference) from the variety of eye disease. Table III shows a sample of probability based on a given case.

TABLE III. PROBABILITY VALUES FOR ALL DISEASES

Eye Diseases	Probability Values	
Blepharitis	2.60561E-07	0.00269000
Dacryocystitis	7.42895E-09	0.00007900
Episcleritis	2.84135E-11	0.00000000
Glaucoma	1.46455E-09	0.00001600
Hordeolum	4.9132E-13	0.00000049
Cataracts	3.10216E-09	0.00003900
Keratitis	1.88651E-12	0.00000000
Conjunctivitis	2.2588E-09	0.000025
Myopia	3.20242E-10	0.00000300
Scleritis	1.88651E-12	0.00000000
Corneal ulcers	7.29735E-13	0.00000000
Uveitis	2.82111E-13	0.00000000

3) Revise

Before the final step of this proposed method, the system will look for the highest probability value on each eye diseases. An eye disease with the highest probability will be inferred (concluded) as the best prediction. Based on table III as an example of a case, it can be concluded that a user is diagnosed with ***Blepharitis***with probability of 0.00269.

4) Retain

In the end of the process, the system will save the new case into the knowledge base so that it can used to solve future cases. The recent case that have been saved on the existing case base will be used to calculate the new posterior probabilities.

VI. IMPLEMENTATION

Implementation is the stage of developed the system based on the results of the previous design that has been designed, so that it can be used in the real situation.

VII. EXPERT SYSTEM TESTING

A. System Testing

System testing and evaluation is the final stage in this research. System testing and evaluation is ran to evaluate the effectiveness of the system. The develop system has to produce the desired outcome in accordance with its objectives with high accuracy without any systemic error. Testings are conducted using the black box method.

B. User Acceptance Test

User acceptance test was conducted to validate the system output directly to the experts and end-users. 11 cases have been selected and assessed and tested by an ophthalmologist, namely Dr. R. Handoko Pratomo, Sp.M. The following table compares diagnostic results by expert systems and the human expert.

TABLE IV. EXPERT SYSTEM RESULTS

Cases	Diagnosis by Expert System	Diagnosis by Human Expert
1	Glaucoma	Glaucoma,
2	Blepharitis	Blepharitis
3	Conjunctivitis	Conjunctivitis
4	Uveitis	Uveitis
5	Corneal ulcers	Corneal ulcers
6	Keratitis	Conjunctivitis
7	Conjunctivitis	Conjunctivitis
8	Corneal ulcers	Corneal ulcers
9	Dacryocystitis	Dacryocystitis
10	Scleritis	It cannot be determined
11	Cataracts	Cataracts

Regarding the system result, it has similarities diagnoses with an expert by 82%. Thus the system may used and the outcome of a system according to the expected.

VIII. CONCLUSION

We have presented a Web-based expert system for self-diagnosis of eye diseases. The system, which employs the CBR and Naïve Bayes, has proved its capability to handle 140 cases. Based on the experimental results, it can be concluded that the integration of Case-Based Reasoning and Naïve Bayes in the expert system shows a promising result. However, much work still needs to be done before a full-size expert system is constructed with standard expert systems components..

ACKNOWLEDGMENT

We would like to thank Dr. R. Handoko Pratomo, Sp.M who has guided this study and provided precious data on eye diseases.

REFERENCES

[1] Roger Schank, Dynamic Memory: *A Theory of Learning in Computers and People* (New York: Cambridge University Press, 1982)..

[2] F. Ibrahim., J. B. Ali, F. JaaisandM. N. Taib,“Expert System for Early Diagnosis of Eye Diseases Infecting the Malaysian Population,” IEEE Catalogue No. 01 CH37239, 2001, pp. 430–432.

[3] L. G. Kabari,., andE. O. Nwachukwu, "Neural Networks and Decision Trees For Eye Diseases Diagnosis." 2012.

[4] M. M. Syiam, "A Neural Network Expert System for Diagnosing Eye Diseases,"Proceedings of the Tenth Conference on Artificial Intelligence for Applications,1994, pp.491–492.

[5] B. Chandra, andM. Gupta,“Expert Systems with Applications Robust Approach For Estimating Probabilities In Naïve – Bayes Classifier for Gene Expression Data,”Expert Systems With Applications, 2011, pp. 1293–1298.

[6] W. Zhang and F. Gao, “Procedia Engineering An Improvement to Naive Bayes for Text Classification.” 2011.

[7] Z. Ni, S. Yang, L. LizandR. Jia, “Integrated Case-Based Reasoning,” November 2003.

[8] N. H. Phuong, N. R. Prasad, D. H.HungandJ. T.Drake, “Approach to Combining Case Based Reasoning with Rule Based Reasoning for Lung Disease Diagnosis,” 2001, pp. 883–888.

[9] P. Bhargaviand S. Jyothi. “Applying Naive Bayes data mining technique for classification of agricultural land soil,”International Journal Of Computer Science and Network Security, 2009, pp. 117-122.

[10] G. Subbalakshmi, K. Ramesh, and M. Chinna Rao,“Decision Support In Heart Disease Prediction System Using Naive Bayes.”Indian Journal Of Computer Science And Engineering(Ijcse), 2011, pp. 170-176.

[11] F. I. Putra, A. Rahman, andA. Sudiarno, “Perancangan Knowledge Management System Dalam Penanganan Penyakit Tropis Dengan Pemenuhan Prinsip Ergonomi Kognitif.”

[12] N. Friedman, D. GeigerandM. Goldszmidt, “Bayesian Network Classifiers,” Machine Learning, 1997, 29, pp. 131-161.

[13] C. Leng, S. Wangand H. Wang, “Learning Naive Bayes Classifiers with Incomplete Data,” 2009, pp. 350–353.

[14] K. M. Al-aidaroos, A. A. Bakar, and Z. Othman,“Naïve Bayes Variants in Classification Learning,”2010, pp. 276–281.

[15] J. Jauhari,“Perancangan Problem-Solving Environment Menggunakan Case-Based Reasoning,”2011,pp. 1–4.