Classifying Chief Complaint in Eye Diseases using Data Mining Techniques

Archana L. Rane[1], P. D. Mahajan[2]

[1,2] Department of MCA, K K Wagh Institute of Engineering Education & Research
Hirabai Haridas Vidyanagar, Panchvati, Nashik - 422 003, Maharashtra
E-mail: rane_nehete_archana@yahoo.com[1], [2]punamgf@yahoo.co.in

Abstract— Eyes are the important organ for the vision system. The system itself is very complicated. The clinicians attempt to determine the correct diagnosis using signs, symptoms and test results to formulate the hypothesis of the diagnosis before providing treatments. Most patients in this study have severe illness. Therefore, the clinicians decide to take the treatment by surgery rather than treating the patients with medicine. The result of the classification is very critical for the clinicians to support their diagnosis before giving the surgery to the patients. This study endeavors on using intelligent capability of data mining to discover hidden patterns in the data. Here, Artificial Neural Networks (ANN) and Naïve Bayes are utilized as techniques to classify patients with chief complaints in eye diseases. The results of classifying the eye diseases are very encouraging with the percentage accuracy of 100% for both techniques.

Keywords- classifier, data mining techniques, Artificial Neural Network, Naïve Bayes, eye disease

I. INTRODUCTION

Eyes are organs that detect light and convert it into electro-chemical impulses in neurons. The simplest photoreceptors in conscious vision connect light to movement. In higher organisms the eye is a complex optical system which collects light from the surrounding environment, regulates its intensity through a diaphragm, focuses it through an adjustable assembly of lenses to form an image, converts this image into a set of electrical signals, and transmits these signals to the brain through complex neural pathways that connect the eye via the optic nerve to the visual cortex and other areas of the brain [1]. Human vision is a highly complex activity with a range of physical and perceptual limitations, yet it is the primary source of information for the average person.

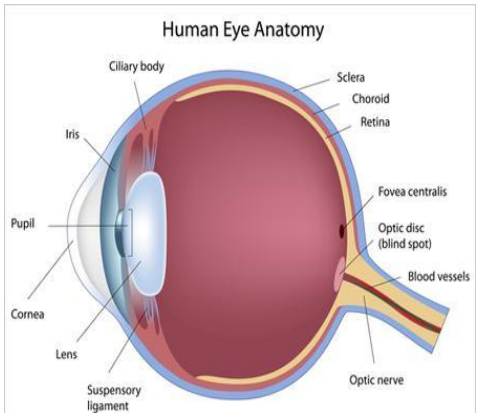
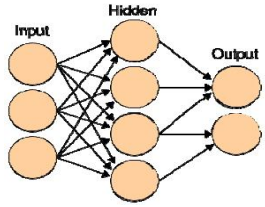


Figure 1: Human Eye Anatomy

Eye disease is referred to as any abnormal thing that

III. METHODOLOGY



A general approach for solving classification can be shown as in Figure 2.

Eye disease is referred to as any abnormal thing that happens tothe blind system. The blind system by itself is very complicated. Figure1 shows the anatomy of the eye. The eye has a number of importance components
The cornea and lens at front of the eye focus the light into a sharp image on the back of the eye, the retina is light sensitive and contains two types of photoreceptors: rods and cones. Rods are highly sensitive to light and therefore allow us to see under a low level of illumination. The eye has approximately 12 millions rods per eye. Cones are less sensitive to light than rods and can therefore tolerate more light. The eye has approximately 6 millions cons and mainly concentrated on fovea which is small area of the retina on which image are fixated.In general, diagnosis of the eye [1] requires special knowledge and skill in addition to common physical examinations. Special inspections and tools are required for diagnosis the eye diseases. Patients will servere eye diseases, who do not get the proper treatment or delay for treatment can lose their vision sense permanently. The objective of this study is to assist medical providers in screening and diagnosis of eye disease patients using patient history. In addition, the symptoms can be identified for diagnosis, treatment planning, including referral to medical specialist accurately and quickly. Artificial neural network (ANN) and Naïve Bayes (NB) illustrate the potential for medical diagnosis [2][3]. The study has the main focus on using them as classifiers for patients with eye diseases. The result can be assist a clinician's decision in giving the right treatments to patients. This paper is structured as follows: section II provides the literature review. Section III describes the methodologies. Experimental results are discussed in Section IV. Finally, discussion, conclusion and future works are given in Section.

II. LITERATURE REVIEW

Data mining has been recognized as a useful technique in analyzing the data. The technique has been successfully employed in different areas such as business [4] [5] [6] and many other areas [7][8]. J. Víctor Marcos et al [9] classified Obstructive Sleep Apnea (OSA) patients using artificial neural network (ANN) and K- nearest neighbor (KNN). The result shown that the ANN performs better than the KNN with the percentage accuracy 91.30%. Paul S. Heckerling et al [10] utilized the ANN technique to classify symptoms in Pneumonia symptom patients.The training set is employed for building the classification model. The test set is utilized for measuring the accuracy of the model.

Artificial Neural Network (ANN) or neural network for short is inspired by the human brain structure. A neural network can be used to represent a nonlinear mapping between input and output vectors. Neural networks are among the popular signal-processing technologies. In engineering, neural networks serve two important functions: as pattern classifiers and as nonlinear adaptive filters.

Figure 3 shows a typical Multilayer Perceptron ANN architecture (MLP). A Multi-layers Perception (MLP) is a special kind of artificial neural network. Generally, a MLP consists of 3 layers: input layer, one or more hidden layers and output layer. Each layer is composed of neurons, which are interconnected with each other by weights. In each neuron, a specific mathematical function called activation function accepts inputs form previous layer and generates output for the next layer. Here the MPL is trained using a standard back-propagation algorithm [13].

Two techniques utilized in this study are: neural network and Naïve Bayes. The learning algorithm of each technique is used to identify a model that best represent the relationship between the input attributes and the output classes.

A. Artificial Neural Network (ANN)

The idea of neural networks came out many decades ago. Currently, it plays a significant role in classifying, prediction, and analysis. Numerous applications based on neural network have successfully solved the problems in many different fields [12]. There are many parameters involve in training the MLP. One of the important parameter is number of hidden nodes. Choosing the number of hidden units is a important factor. There are several publications discussing the number of hidden units, such as Elisseeff, et al [14]. The equation to estimate the number of hidden unit (H) is given as:

$$H \geq \frac{n - m}{m(k + 6)}$$

Where n is the sample size
m is a constant value, set to 6 in this experiment minimum value of H = 1)

B. Naïve Bayes(NB)

The Naive Bayes (NB) is a probabilistic method based on Bayes's theorem. NB employed a conditional independence assumption that is each feature is conditionally independent on other features [11]. The conditional independence assumption can be stated as:

$$P(X \mid Y = y) = \prod_{i=1}^{d} P(X_i \mid Y = y)$$

Where X denote the attribute set; each attribute set $X = \{X_1, X_2, \ldots, X_d\}$ consists of d attribute
Y denote the class attribute
y is the class label

Because of its simplicity and stability, NB is selected as a classification standard in most data miners.

IV. EXPERIMENTAL DATA AND RESULTS

A. Experimental Data

This study obtained data from the department of eyes, ears, nose, throat (EENT), faculty of medicine, Hospital, Nashik. The patients come from all regions of the country. Generally, patients with symptoms such as red eye, watering of the eye, virus infection, congestive glaucoma, headache, photophobia, history of any injury are considered mild symptoms.
The doctors may only prescribe medicines to patients. In the severe cases, the doctors must use special tools to diagnose the eye and spend longer time to determine the disease.
In this study, patients are over 14 years old. For patients who are younger than 14 years old will be considered as children patients. Too young patients are not able to explain their symptoms to the doctors. The information related to symptoms of these patients is mainly dependant on their parents or guardians. Therefore, only adult patients will be considered in the experiment.
The datasets were collected from June 2010 to December 2011. There are 192 patient records. Each record consists of 15 attributes. One attribute is contributed as the class attribute. Eight different class for eye diseases applied in this study and can be briefly explained as follows [15]:

1. Conjunctivitis is the commonest disease of the conjunctiva. The conjunctiva may be affected by the Exogenous, Endogenous or by local spread of lesion from the surrounding structures like skin, lacrimal apparatus, the eye itself from the affection of cornea, sclera or uveal tract or orbit. Major symptoms of this disease red eye, watering of the eye, difficulty to tolerate light, mistiness of vision, sticking together of the lid margins during sleep due to the discharge, rainbow halo around the light. Treatment is to take a conjunctival swab for culture, to isolate the organism and to find its sensitivity to any antibiotics and then to use that antibiotics.

2. Chalazion is the chronic inflammatory granuloma of the meibomian gland. Major symptoms of this disease depending on the size of the Chalazion, there is sense of heaviness in the lid and mild irritation, small, cystic or hard swelling of a pea on the lid. Treatment is a course of antibiotic ointment and for bigger chalazia apply an analgesics for calm down then can be treated by surgery.

3. Myopia is an error of refraction in which the parallel rays of light infinity come to focus in front of retina. There are three types of myopia. a) Axial myopia: when the antero-posterior length of the eyeball is more than normal. b) Curvature myopia: when the curvature of the cornea or the lens is more than normal. c) Index myopia: when the refractive indexes of different media, particularly of the lens is more than normal. Retinoscopy examination required to confirm the disease. No treatment for pathological changes but prescribing concave spherical glasses. Surgery like keratotomy is fast becoming popular.

4. Trachoma is a kind of kerato-conjunctivitis, with simultaneous affection of the cornea and conjunctiva, the causative agent being Chlamydia or Bedsonia group of organisms. This disease is more common in countries with dry and dusty weather. A local treatment is susceptible to sulphonamides, sulphacetamide drop instilled into the eye for some time. (Four times a day for one month) or take the tablets for some days (one gram three times a day by mouth for ten days). If any complications then plastic operations on lids for correction of entropion and trichiasis.

5. Blepharitis is a subacute or chronic inflammation of the lid margin. Common symptoms such as external irritants, unhygienic condition, nature of skin or allergic. Treatment for the disease is judicious and balanced diet plan or antibiotic ointment or removal of scales.

6. External Hordeolum or Stye is a suppurative inflammation of the follicle of the eyelash including the glands of Zies. Common symptoms such as acute pain in the lid margin with a sense of heaviness and heat, redness, swelling, marked oedema of the affected lid. Treatment for the disease is hot compress two-three times a day, eye drop, tablets or small surgery.

7. Pterygium is a triangular fold of conjunctiva, encroaching the cornea, in the horizontal meridian, in the palpebral fissure either form the nasal side or from the temporal side of the bulbar conjunctiva. Generally no symptom unless it encroaches the papillary area, when there is a vision disturbance or may be diplopia or limitation of movement of the eyeball. If the Pterygium is stationary in the atrophic stage no treatment is necessary but if it is progressive then operate.

8. Cataract is very common eye disease and generally happens in adults. Treatment depends upon the amount of vision disturbance and it can be treat by only surgery.

All patients are pass through the admissions patient department. Therefore, they have been physically examined,including eye disease examination with special tools. Then doctors diagnose for patients eye and plan the treatment by surgery. It is important to note that in surgical treatment planning, patients are carefully examined before clinicians come out with the decisions and plan for the surgical treatments. As a result, data obtained in the study are highly reliable.

B. Results: from training the network

The multilayer perceptron is trained under supervision using the back-propagation algorithm on WEKA [16], a well-known machine learning software and 192 patient records.

The number of parameters relating to training the network needs to be analyzed. Thus, many experiments are run to find the optimal result. The ranges of examined parameters are: the learning rate is between 0.1 to 0.5, the hidden nodes are between 1 to 10 nodes, and the momentum value is from to 0.5. All combinations of these parameters are investigated.

		Prediction Result							
		A	B	C	D	E	F	G	H
Diahnostic Results	A	36	0	0	0	0	0	0	0
	B	0	24	0	0	0	0	0	0
	C	0	0	30	0	0	0	0	0
	D	0	0	0	12	0	0	0	0
	E	0	0	0	0	12		0	0
	F	0	0	0	0	0	30	0	0
	G	0	0	0	0	0	0	18	0
	H	0	0	0	0	0	0	0	30

Table 1Confusion Matrix for neural network

Table 1 illustrates the confusion matrix for the experiment neural network. The confusion matrix demonstrates information about the actual diagnostic and the predicted class from the network. In the matrix, each row represents the actual diagnostics, whereas, each column of the matrix represents the predicted results from the network. The labels A, B, C, D, E, F, G and H denote eye disease classes. For example, A represents conjunctivitis , B is Chalazion and H is Cataract vision loss. Performance of the neural network approach can be evaluated using data in the matrix. The examples of interpretations are:

-Horizontal reading from diagnostic results for eye disease class A is 36 and the predicted result from the network is also 36. All records are predicted correctly. Thus, the accuracy percentage is 100%

-Horizontal reading from diagnostic results for eye disease class C is 30 and the predicted result from the network is also 30. All records are predicted correctly. Thus, the accuracy percentage is 100%. It can be seen that the network correctly classified 192 records from a total test data set 192 records. The accuracy percentage is 100 %. As mentioned earlier, the data employed is in this study is highly reliable due to the patients being carefully examined.

C. Results from training the Naïve Bayes

Now, Naïve Bayes is utilized for classifying the eye diseases. The simulation is run under Weka software.

Diahnostic Results	Prediction Result							
	A	B	C	D	E	F	G	H
A	36	0	0	0	0	0	0	0
B	0	24	0	0	0	0	0	0
C	0	0	30	0	0	0	0	0
D	0	0	0	12	0	0	0	0
E	0	0	0	0	12		0	0
F	0	0	0	0	0	30	0	0
G	0	0	0	0	0	0	18	0
H	0	0	0	0	0	0	0	30

Table2 Confusion Matrix for Naïve Bayes

There are 192 records as the training set.

Table 2 illustrates the confusion matrix from the NB method.

Figure 4: Results from training the Naïve Bayesusing WEKA

Here, the same notation as in table 1 is used. For example, the results can be interpreted as follows:

-Horizontal reading from diagnostic results for eye disease class B is 24 and the predicted result from the network is also 24. All records are predicted correctly. Thus, the accuracy percentage is 100%.

-Horizontal reading from diagnostic results for eye disease class E is 12 and the predicted result from the network is also 12. All records are predicted correctly. Thus, the accuracy percentage is 100%. It can be seen that Naïve Bayes correctly classified 192 records from a total test data set 192 records. The accuracy percentage is 100 %. The result is the same as that from the neural network as well as in tree structure as shown in figure 5 and figure 6

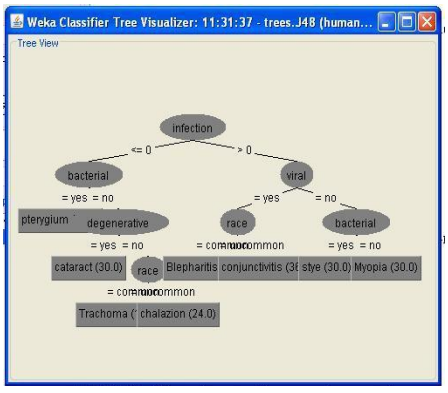
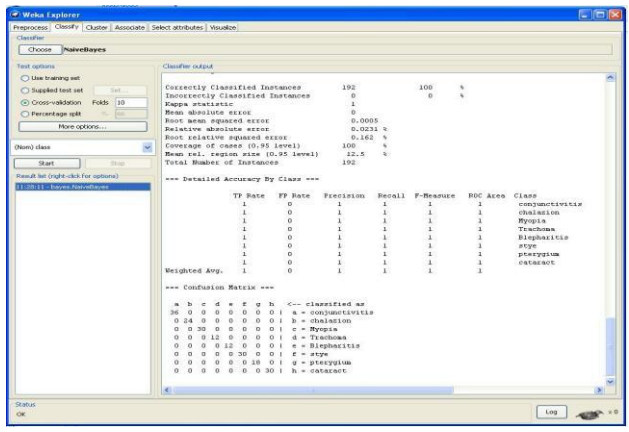


Figure 5: Result of Neural Network



V. DISCUSSION, CONCLUSION AND FUTURE WORK

As mentioned earlier, the data is highly reliable so there is no surprise that the accuracy percentage from both techniques reach 100.Clinicians attempt to determine the correct diagnosis using signs, symptoms and test results to formulate the hypothesis of the diagnosis before providing treatments. The misdiagnosis may lead to wrong treatments that might risk patient lives.

The research on chief complaints in the eye diseases is still at its beginning. This study is only applied in adult patients. The work on classifying eye diseases in children should also be conducted. This can be very challenging since very young patients are not able to talk or clearly explain their symptoms. A special examinations and tools are needed for diagnosis. Some of the immediate future work along this line is to increase the number of class categories and utilize other techniques in the experiments for the comparison. Even develop a new and more specific technique to classify the eye disease in patients.

REFERENCES

[1] Narin Watanasin, Siripun Sanguansintukul, ―Classifying Chief Complaint in Ear Diseases using Data Mining Techniques‖, in in IEEE Conference, 12 September 2011,pp-149-153

[2] Sofianita Mutalib, Nor Azlin Ali, Shuzlina Abdul Rahman and Azlinah Mohamed, ―An Exploratory Study in Classification Methods for Patients' Dataset,‖ in IEEE Conference on Data Mining and Optimization, 2009, pp. 79-83.

[3] Plamena Andreeva, Maya Dimitrova and Petia Radeva, ―Data Mining Learning Models and Algorithms for Medical ,‖ in Benvinguts alCentre de Visio per Computador.CVC,http://www.cvc.uab.es/~petia/maya%20Saer3.pdf

[4] Fei fei Wang, Siripun Sanguasintukul and Chidchanok Lursinsap (2008). Curl Forecasting for Paper Quality in Paper

Making Industry.International Conference on Scientific Computing (ICSC 2008). 10-12 October, 2008, Beijing, China. Page 1281-1286.

[5] Naroumon Yordphet and Siripun Sanguansintukul, ―Safety Stock Based On Consumption Forecast by the Artificial Neural Network‖. The International Conference on Software and Computing Technology (ICSCT 2010), Kunming, China, ,October 18-19th,2010.

[6] Yi Wang, Siripun Sanguasintukul and Chidchanok Lursinsap (2008) The Customer Lifetime Value Prediction in Mobile Telecommunications, IEEE International Conference on Management of Innovation & Technology Publication (ICMIT 2008). 21-23 September, 2008, Bangkok, Thailand.

[7] Malinee Wongruen, Chidchanok Lursinsap, Siripun Sanguansintukul and Krung Sinapiromsaranm, ―Mixed Integer Linear Programming Model for Workload Leveling on Offshore Platform of Petroleum Company‖, The International Conference on Intelligent Computation Technology and Automation (ICICTA 2010), Changsha, Hunan, China, May 11-12, 2010 (pp.683-686).

[8] S. Areerachakul, S. Sanguansintukul, ―Clustering Analysis of Water Quality for Canals in Bangkok,‖ The 2010 Internaitonal Conference on Computational Science and ITS Applications. 2009 Eigth IEEE/ACIS International; Conference on Computer and Information Science. LNCS SPRINGER, Fukuoka, Japan. March 23-26, 2010.

[9] J. Víctor Marcos, Roberto Hornero, Daniel Álvarez, Félix Del Campo and Miguel López, ―Applying Neural Network Classifiers in the Diagnosis of the Obstructive Sleep Apnea Syndrome from Nocturnal Pulse Oximetric Recordings,‖ in IEEE Computer Society, 2007, pp. 5174-5177.

[10] Paul S. Heckerling, Ben S. Gerber, Thomas G. Tape and Robert S. Wigton, ―Prediction of Community-Acquired Pneumonia Using Artificial Neural Networks,‖ in Medical Decision Making, 2003, pp. 112-121.

[11] Written Pang-Ning Tan, Michael Steinbach and Vipin Kumar.General Approach to Solving a Classification Problem, Introduction to Data Mining, Boston : Pearson Addison-Wesley, 2005.

[12] C. Thitipornvanid and S.Sanguansintukul ―Predicition of a human facial image by ANN using image data and its content on web pages. International Conference on CESSE 2009 : "International Conference on Computer, Electrical, and Systems Science,and Engineering", Singapore. August 26-28, 2009.

[13] Robert Hecht-Nielsen, ―Theory of the Backpropagation Neural Network,‖ International Conference (IJCNN), Washington DC., USA, 1989, pp. I-593 - I-605

[14] Isabelle Guyon and Andre Elisseeff, ―An Introduction to Variable and Feature Selection,‖ Journal of Machine Learning Research 3, 2003, pp. 1158 - 1182.

[15] Writen Theeraporn Ratana-anakchai and Supaporn Srirompothong. Otolaryngology Book, Khon-Kaen : Klangnanawittaya, 2008.

[16] Ian H.Witten and Eibe Frank, ―Data Mining: Practical machine Learning Tools and Techniqueues‖, Morgan Kaufmann Publishers, San Francisco, 2nd edition, 2005.

[17] http://en.wikipedia.org/wiki/Diagnosis