International Research Journal of Computer Science (IRJCS) ISSN: 2393-9842
Issue 09, Volume 4 (September 2017) www.irjcs.com

A STUDY ON CATRACT USING ANN WITH DATAMINING ALGORITHM USING (NAIVE BAYES, J48 AND RANDOM TREEE ALGORITHM)

P.Kuppan, Assistant Professor in Computer Science, TVUCAS Thennangur College, INDIA
Dr .N.Manoharan, HOD of Department of Computer Science,Thennangur,INDIA

Manuscript History
Number: **IRJCS/RS/Vol.04/Issue09/SISPCS10101**
Received: 08, September 2017
Final Correction: 13, September 2017
Final Accepted: 20, September 2017
Published: September 2017
Editor: Dr.A.Arul L.S, Chief Editor, IRJCS, AM Publications, India

Abstract-- Data mining refers to extracting knowledge from large amount of data. Real life data mining approaches are interesting because they often present a different set of problems for data miners. The process of designing a model helps to identify which factor is highly infected Cataract Diseases. A cataract can cause a decrease in visual function, which in turn can be classified as a visual disability. Thus, cataract can be defined in three ways. The first definition is an objective lens change. The second is a lens opacity that is associated with a defined level of visual acuity loss. The third relates to the functional consequences of lens opacification. This guideline focuses on the last definition .Taking into account the prevalence of cataract among men and women the study is aimed at finding out the characteristics that determine the presence of cataract and to track the maximum number of men and women suffering from cataract with 500 population using weka tool. In this paper the data classification is cataract patient's dataset is developed by collecting data from hospital repository consists of 500 instances with 11 different attributes. The instances in the Dataset are pertaining to the categories of Age, Smoking and alcohol, obesity, high blood pressure, diabetes, Radiation, Genetics, Skin diseases, Inadequate vitamin C, Medications, Post-operative. W EKA tool is used to classify the data and the data is evaluated using 10 fold cross validation and the results are compared.
Keywords: Data Mining, Classification, Decision Tree Algorithm, Weka tool

I. INTRODUCTION

The main focus of this paper is the classification of different types of datasets that can be performed to determine if a person has cataract disease. For this reason, the goal of this research is classifier in order to correctly classify the datasets. The major motivation for this work is that Cataract affects a large number of the world population and it's a hard disease to diagnose. This Optometric Clinical Practice Guideline for Care of the Patient with Visual Impairment describes appropriate examination and treatment procedures for the evaluation of the visual abilities and eye health of people with visual impairments. It contains recommendations for timely diagnosis, management, and, when needed, referral for consultation with (or treatment by) another health care provider or rehabilitation professional. This Guideline will assist optometrists in achieving the following goals:

- Identify patients with visual impairment(s) who might benefit from low vision care and rehabilitation
- Evaluate visual functioning of a compromised visual system effectively
- Emphasize the need for comprehensive assessment of patients with impaired vision and referral to, and interaction with, other appropriate professionals
- Maintain and improve the quality of eye and vision care rendered to visually impaired patients

- Inform and educate other health care practitioners and the lay public regarding the availability of vision rehabilitation services
- Increase access for the evaluation and rehabilitative care of individuals with visual impairment(s), thereby improving their quality of life.

Common symptoms of cataracts include:

- blurry vision
- trouble seeing at night
- seeing colors as faded
- increased sensitivity to glare
- halos surrounding lights
- double vision in the affected eye
- a need for frequent changes in prescription glasses

II. LITERATURE REVIEW

Endophthalmitis, although rare, is one of the most devastating complications of intraocular surgeries. As cataract surgery consists of a large fraction of ophthalmic operations, the majority of literature reports about the endophthalmitisis centered on cataract surgery.[1] An aging population worldwide necessitates an increase in the number of cataract surgeries, rendering post-cataract surgery endophthalmitis a public health concern.[2]

High morbidity and subsequent medical care expenses are part of this complication.[2] Visual outcomes are not often favorable; about 40% of affected patients sustain severe visual loss (corrected distance visual acuity of less than 20/200), and only one-third of cases reach visual acuity of better than 20/40.[3] Evisceration as a last resort has also been employed in the case of endophthalmitis.[4]

The reported rate of post-operative endophthalmitis varies between a range of 0.04%–0.2%.[5] However, the incidence of post cataract surgery endophthalmitis shows significant changes overtime. By the time of introduction of clear cornea cataract extraction, as opposed to scleral or limbal incisions, an increase in endophthalmitis rate was observed.[6], [7], [8]There are also studies rejecting this hypothesis, with clear corneal technique being even a safer approach.[9], [10].

III. TOOLS AND TECHNIQUE

Weka (pronounced to rhyme with Mecca) contains a collection of visualization tools and algorithms for data analysis and predictive modeling, together with graphical user interfaces for easy access to these functions.[11] The original non-Java version of Weka was a front-end to (mostly third-party) modeling algorithms implemented in other programming languages, plus data preprocessing utilities in C, and a Make file-based system for running machine learning experiments. This original version was primarily designed as a tool for analyzing data from agricultural domains,[12],[13] but the more recent fully Java-based version (Weka 3), for which development started in 1997, is now used in many different application areas, in particular for educational purposes and research. Advantages of Weka include:

- Free availability under the GNU General Public License.
- Portability, since it is fully implemented in the Java programming language and thus runs on almost any modern computing platform.
- A comprehensive collection of data preprocessing and modeling techniques.
- Ease of use due to its graphical user interfaces.

IV. DATA MINING BY USING WEKA TOOLS

Weka supports several standard data mining tasks, more specifically, data preprocessing, clustering, classification, regression, visualization, and feature selection. All of Weka's techniques are predicated on the assumption that the data is available as one flat file or relation, where each data point is described by a fixed number of attributes (normally, numeric or nominal attributes, but some other attribute types are also supported). Weka provides access to SQL databases using Java Database Connectivity and can process the result returned by a database query. It is not capable of multi-relational data mining, but there is separate software for converting a collection of linked database tables into a single table that is suitable for processing using Weka. Another important area that is currently not covered by the algorithms included in the Weka distribution is sequence modeling.

CLASSIFICATION:

Classification is the most commonly applied data mining technique, which employs a set of pre-classified examples to develop a model that can classify the population of records at large. Fraud detection and credit-risk applications are particularly well suited to this type of analysis. This approach frequently employs decision tree or neural network-based classification algorithms. The data classification process involves learning and classification. In Learning the training data are analyzed by classification algorithm. In classification test data are used to estimate the accuracy of the classification rules. If the accuracy is acceptable the rules can be applied to the new data tuples. For a fraud detection application, this would include complete records of both fraudulent and valid activities determined on a record-by-record basis. The classifier-training algorithm uses these pre-classified examples to determine the set of parameters required for proper discrimination. The algorithm then encodes these parameters into a model called a classifier.

Types of classification models:

- Classification by decision tree induction
- Bayesian Classification
- Neural Networks
- Support Vector Machines (SVM)
- Classification Based on Associations

A) Naïve bayes

Naive Bayes is a simple technique for constructing classifiers: models that assign class labels to problem instances, represented as vectors of feature values, where the class labels are drawn from some finite set. It is not a single algorithm for training such classifiers, but a family of algorithms based on a common principle: all naive Bayes classifiers assume that the value of a particular feature is independent of the value of any other feature, given the class variable. For example, a fruit may be considered to be an apple if it is red, round, and about 10 cm in diameter. A naive Bayes classifier considers each of these features to contribute independently to the probability that this fruit is an apple, regardless of any possible correlations between the color, roundness, and diameter features. For some types of probability models, naive Bayes classifiers can be trained very efficiently in a supervised learning setting. In many practical applications, parameter estimation for naive Bayes models uses the method of maximum likelihood; in other words, one can work with the naive Bayes model without accepting Bayesian probability or using any Bayesian methods.

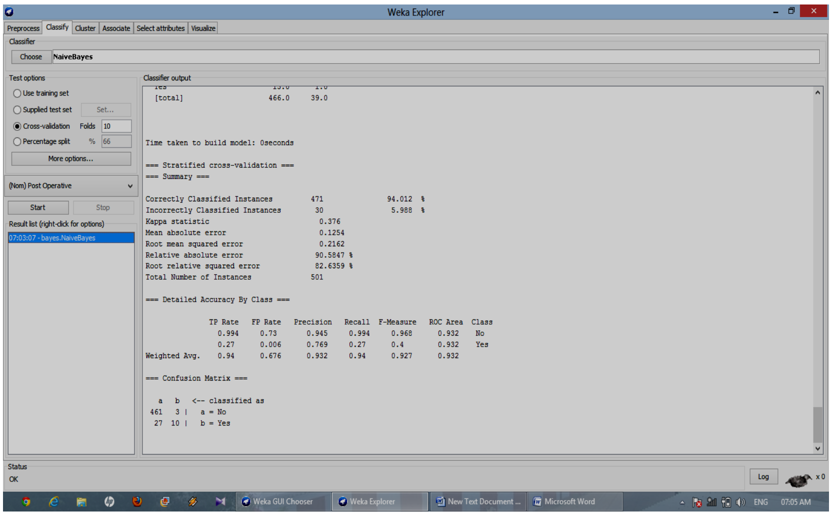

Fig1.This Screen shows the Naive bayes Algorithm result

B). J48 Algorithm

C4.5 is an algorithm used to generate a decision tree developed by Ross Quinlan.[1] C4.5 is an extension of Quinlan's earlier ID3 algorithm. The decision trees generated by C4.5 can be used for classification, and for this reason, C4.5 is often referred to as a statistical classifier. Authors of the Weka machine learning software described the C4.5 algorithm as "a landmark decision tree program that is probably the machine learning workhorse most widely used in practice to date".[2]

At each node of the tree, C4.5 chooses the attribute of the data that most effectively splits its set of samples into subsets enriched in one class or the other. The splitting criterion is the normalized information gain (difference in entropy). The attribute with the highest normalized information gain is chosen to make the decision. The C4.5 algorithm then recurs on the smaller sub lists.

This algorithm has a few base cases.

- All the samples in the list belong to the same class. When this happens, it simply creates a leaf node for the decision tree saying to choose that class.
- None of the features provide any information gain. In this case, C4.5 creates a decision node higher up the tree using the expected value of the class.
- Instance of previously-unseen class encountered. Again, C4.5 creates a decision node higher up the tree using the expected value.

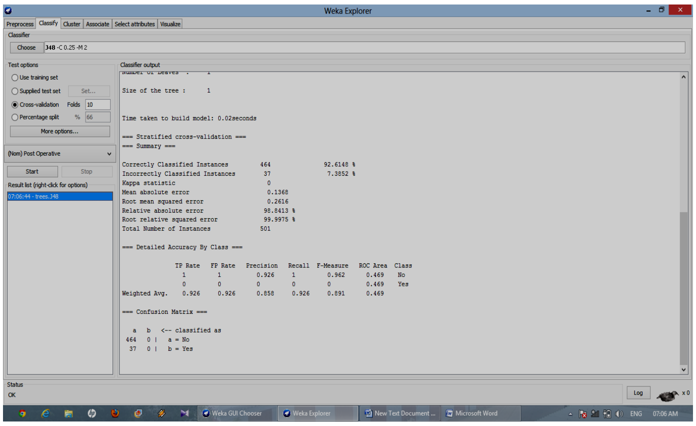

Fig2.This Screen shows the J48 Algorithm result

C). Random tree

In mathematics and computer science, a random tree is a tree or arborescence that is formed by a stochastic process. Types of random trees include: Class for constructing a tree that considers K randomly chosen attributes at each node. Performs no pruning. Also has an option to allow estimation of class probabilities based on a hold-out set

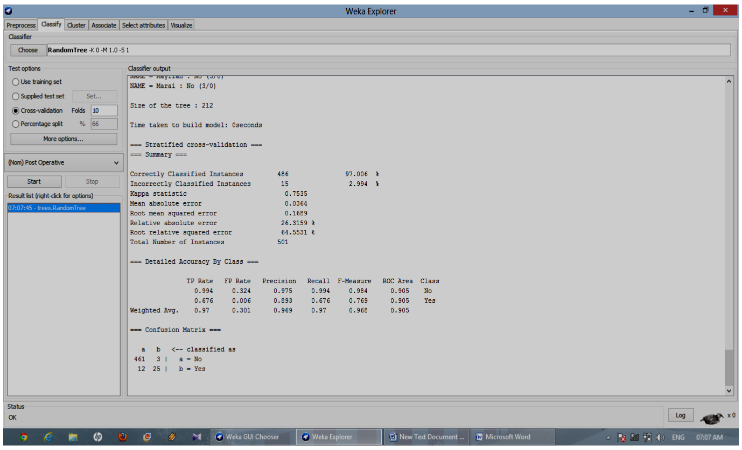

Fig3.This Screen shows the Random tree Algorithm result

TABLE1. RESULT FOR NAÏVE BAYES, RANDOM TREE AND J48 ALGORITHM

Algorithm	Correctly classified instance	Mean absolute Error
Naïve Bayes	94.12%	0.1254
J48	92.62%	0.1368
Random Tree	97.00%	0.364

V. RESULTS & DISCUSSIONS

This research is a starting attempt to use data mining functions to analyze and evaluate Cataract patient data. We classify data set with three different classification algorithms: Naïve Bayes, J48 and Random Tree. In our study we are going to compare the correctly classified instances as well as mean absolute error with different algorithms they are, Naive Bayes, J48, and Random Tree. The following table shows this comparison. This will help to improve the performance of such cataract patient in the early stage. The future work will be focused on using the other classification algorithms of data mining. It is a known fact that the performance of an algorithm is dependent on the domain and the type of the data set.

VI. CONCLUSION

This paper illustrates how well different classification techniques are used as predictive tools in the data mining domain and after comparing their performances. From the results it is proven that Random Tree algorithm is most appropriate for cataract performance. Random Tree gives 97% which is relatively higher than other algorithms. This study is an attempt to use classification algorithms for the cataract performance and comparing the performance of Naive Bayes, J48, and Random Tree

REFERENCES

1. Sandvig K.U., Dannevig L. Postoperative endophthalmitis: establishment and results of a national registry. J Cataract Refract Surg. 2003;29:1273–1280.
2. Jabbarvand M., Hashemian H., Khodaparast M., Jouhari M., Tabatabaei A., Rezaei S. Endophthalmitis occurring after cataract surgery: outcomes of more than 480 000 cataract surgeries, epidemiologic features, and risk factors. Ophthalmology. 2016;123(2):295–301.
3. Friling E., Lundstrom M., Stenevi U., Montan P. Six-year incidence of endophthalmitis after cataract surgery: Swedish national study. J Cataract Refract Surg. 2013;39:15–21.
4. Al-Mezaine H.S., Kangave D., Al-Assiri A., Al-Rajhi A.A. Acute-onset nosocomial endophthalmitis after cataract surgery: incidence, clinical features, causative organisms, and visual outcomes. J Cataract Refract Surg. 2009;35:643–649.
5. Packer M., Chang D.F., Dewey S.H. Prevention, diagnosis, and management of acute postoperative bacterial endophthalmitis. J Cataract Refract Surg. 2011;37:1699–1714.
6. Taban M., Behrens A., Newcomb R.L. Acute endophthalmitis following cataract surgery: a systematic review of the literature. Arch Ophthalmol. 2005;123:613–620.
7. West E.S., Behrens A., McDonnell P.J., Tielsch J.M., Schein O.D. The incidence of endophthalmitis after cataract surgery among the U.S. Medicare population increased between 1994 and 2001. Ophthalmology. 2005;112:1388–1394.
8. Lundstrom M., Wejde G., Stenevi U., Thorburn W., Montan P. Endophthalmitis after cataract surgery: a nationwide prospective study evaluating incidence in relation to incision type and location. Ophthalmology. 2007;114:866–870.
9. Colleaux K.M., Hamilton W.K. Effect of prophylactic antibiotics and incision type on the incidence of endophthalmitis after cataract surgery. Can J Ophthalmol. 2000;35:373–378.
10. Fintelmann R., Naseri A. Prophylaxis of postoperative endophthalmitis following cataract surgery. Drugs. 2010;70:1395–1409.
11. Ian H. Witten; Eibe Frank; Mark A. Hall (2011). "Data Mining: Practical machine learning tools and techniques, 3rd Edition". Morgan Kaufmann, San Francisco. Retrieved 2011-01-19.
12. G. Holmes; A. Donkin; I.H. Witten (1994). "Weka: A machine learning workbench" (PDF). Proc Second Australia and New Zealand Conference on Intelligent Information Systems, Brisbane, Australia. Retrieved 2007-06-25.
13. S.R. Garner; S.J. Cunningham; G. Holmes; C.G. Nevill-Manning; I.H. Witten (1995). "Applying a machine learning workbench: Experience with agricultural databases" (PDF). Proc Machine Learning in Practice Workshop, Machine Learning Conference, Tahoe City, CA, USA. pp. 14–21. Retrieved 2007-06-25.

