Analysis of Liver Disorder Using Data mining Algorithm

P.Rajeswari[1], G.Sophia Reena[2]

GJCST Classification
H.2.8,J.3

***Abstract*-There are many disorders of the liver that require clinical care by a physician or other healthcare professional. The study of liver development has significantly contributed to developmental concepts about morphogenesis and differentiation of other organs. Knowledge of the the understanding of human congenital diseases. Significantly, much of understanding of organ development has arisen from analyses of patients with liver deficiencies. In this paper the data classification is based on liver disorder the training data set is developed by collecting data from UCI repository consists of 345 instances with 7 different attributes. The instances in the dataset are pertaining to the two categories of blood tests which are thought to be sensitive to liver disorders that might arise from excessive alcohol consumpt mechanisms that regulate hepatic epithelial cell differentiation has been essential in creating efficient cell culture protocols for programmed differentiation of stem cells to hepatocytes as well as developing cell transplantation therapies. Such knowledge also provides a basis for ion, labeled as Low (L), and (H) to represent the profit as 0 and 1 which result in accuracy and time taken to build the algorithm. WEAK tool is used to classify the data and the data is evaluated using 10-fold cross validation and the results are compared.**

Keywords-Naive Bayes, KStar, FT Tree, WEKA tool.

I. Introduction

The liver is the largest glandular organ of the body. It weighs about 3 lb (1.36 kg). It is reddish brown in colour and is divided into four lobes of unequal size and shape. The liver lies on the right side of the abdominal cavity beneath the diaphragm. Blood is carried to the liver via two large vessels called the hepatic artery and the portal vein. The heptic artery carries oxygen-rich blood from the aorta (a major vessel in the heart). The portal vein carries blood containing digested food from the small intestine. These blood vessels subdivide in the liver repeatedly, terminating in very small capillaries. Each capillary leads to a lobule. Liver tissue is composed of thousands of lobules, and each lobule is made up of hepatic cells, the basic metabolic cells of the liver. This paper describes Excessive consumption of alcohol can cause an acute or chronic inflammation of the liver and may even harm other organs in the body, alcohol induced liver disease remains a major problem. This paper also describes the blood test taken when a person is affected to liver disorder such as alkaline phosphotase, alamine aminotransferase, aspartate aminotransferase, gamma-glutamyl transpeptidase.

About[1]M.phil Research Scholar P.S.G.R.Krishnammal College for Women
About[2]-HOD (BCA) Dept, P.S.G.R.Krishnammal College for Women
#pa.raji.87@gmail.com

II. Causes of Liver Disease

When the liver becomes diseased, it may have many serious consequences. Liver disease (also called hepatic disease) is a broad term describing any single number of diseases affecting the liver. Many are accompanied by jaundice caused by increased levels of bilirubin in the system. The bilirubin results from the breakup of the hemoglobin of dead red blood cells; normally, the liver removes bilirubin from the blood and excretes it through bile.

1) *Disease of Liver*

Several diseases states can affect the liver. Some of the diseases are Wilson's disease, hepatitis (an inflammation of the liver), liver cancer, and cirrhosis (a chronic inflammation that progresses ultimately to organ failure). Alcohol alters the metabolism of the liver, which can have overall detrimental effects if alcohol is taken over long periods of time. Hemochromatosis can cause liver problems.

2) *Common Liver Disorder*

Fatty liver (also known as steatorrhoeic hepatosis or steatosis hepatitis) is a reversible condition where large vacuoles of triglyceride fat accumulate in liver cells via the process of steatosis. It can occur in people with a high level of alcohol consumption as well as in people who never had alcohol.

Hepatitis (usually caused by a virus spread by sewage contamination or direct contact with infected body fluids).

Cirrhosis of the liver is one of the most serious liver diseases. It is a condition used to denote all forms of diseases of the liver characterized by the significant loss of cells. The liver gradually contracts in size and becomes leathery and hard. The regenerative activity continues under liver cirrhosis but the progressive loss of liver cells exceeds cell replacement.

Liver cancer. The risk of liver cancer is higher in those who have cirrhosis or who have had certain types of viral hepatitis; but more often, the liver is the site of secondary (metastatic) cancers spread from other organs.

3) *Symptoms of Liver Disorder*

The external signs include a coated tongue, itchy skin, excessive sweating, offensive body odor, dark circles under the eyes, red swollen and itchy eyes, acne rosacea, brownish spots and blemishes on the skin, flushed facial appearance or excessive facial blood vessels.[1] Other symptoms include jaundice, dark urine, pale stool, bone loss, easy bleeding, itching, small, spider-like blood vessels visible in the skin, enlarged spleen, and fluid in the abdominal cavity, chills,

pain from the biliary tract or pancreas, and an enlarged gallbladder.[2] The symptoms related to liver dysfunction include both physical signs and a variety of symptoms related to digestive problems, blood sugar problems, immune disorders, abnormal absorption of fats, and metabolism problems.[1] Nervous system disorders include depression, mood changes, especially anger and irritability, poor concentration and "foggy brain", overheating of the body, especially the face and torso, and recurrent headaches (including migraine) associated with nausea.The blood sugar problems include a craving for sugar, hypoglycaemia and unstable blood sugar levels, and the onset of type 2 diabetes.Abnormalities in the level of fats in the blood stream, whether too high or too low levels of lipids in the organism. Hypercholesterolemia: elevated LDL cholesterol, reduced HDL cholesterol, elevated triglycerides, clogged arteries leading to high blood pressure heart attacks and strokes, build up of fat in other body organs (fatty degeneration of organs), lumps of fat in the skin (lipomas and other fatty tumors), excessive weight gain (which may lead to obesity), inability to lose weight even while dieting, sluggish metabolism, protuberant abdomen (pot belly), cellulite, fatty liver, and a roll of fat around the upper abdomen (liver roll) etc.[1] Or too low levels of lipids: hypocholesterolemia: low total cholesterol, low LDL, VLDL cholesterol and low triglyderides.

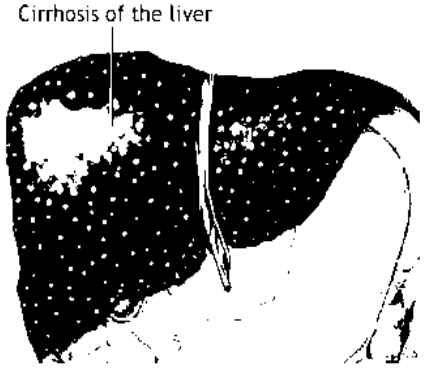


Fig.1 Snapshot of Cirrhosis of the Liver

Symptoms may include:

1) Jaundice
2) Tendency to bruise or bleed easily
3) ascites
4) Impaired brain function
5) General failing health

III. Analysis of Liver Disorder

1) *Risk Factors*

Hepatitis is an inflammation of the liver that can be caused by a virus, inherited disorders, and sometimes by certain medications or toxins such as alcohol and drugs. Scientists have identified four main types of viral hepatitis: hepatitis A, hepatitis B, and hepatitis C, and hepatitis D. A fifth type, hepatitis E, is generally not found in North America.

Hepatitis A is waterborne and spread mainly via sewage and contaminated food and water. Hepatitis B is transmitted by contact with infected semen, blood, vaginal secretions and from mother to newborn. Hepatitis B is most commonly spread by unprotected sex and by sharing of infected needles (including those used for tattooing, acupuncture, and ear piercing). Hepatitis C spreads via direct blood-to-blood contact. Hepatitis D is spread by infected needles and blood transfusions.

Improved screening of donated blood has greatly reduced the risk of catching hepatitis B or C from blood transfusions. Both hepatitis B and C can be spread through sharing of razors, toothbrushes, and nail clippers.

The main cause of cirrhosis is chronic infection with the hepatitis C virus. Other causes include:

1) Long-term, excessive alcohol consumption
2) Chronic infection with hepatitis B virus
3) Inherited disorders of iron and copper metabolism
4) Severe reactions to certain medications
5) Fatty liver caused by obesity
6) Infections from bacteria and parasites usually found in the tropics
7) Repeated episodes of heart failure with liver congestion and bile-duct obstruction

With cirrhosis, the liver tissue is irreversibly and progressively destroyed as a result of infection, poison or some other disease. Normal liver tissue is replaced by scars and areas of regenerating liver cells.

IV. Extraction of Liver Disease Datawarehouse

The liver disorder data warehouse contains the screening the data of liver disorder patients. Initially, the data warehouse is pre-processed to make the mining process more efficient. In this paper WEKA tool is used to compare the performance accuracy of data mining algorithms for diagnosis liver disease dataset. The pre-processed data warehouse is then classified using WEKA tool. The feature selection in the tool describes the attribute status of the data present in the liver disease. Using supervised machine learning algorithm such as Naive Bayes, FT Tree and Kstar and the result are compared.WEKA is a collection of machine learning algorithms for data mining tasks. The algorithms can be applied directly to a dataset. WEKA contains tools for data classification, Associate, clustering and visualization. It is also well suited for developing new machine learning schemes. This paper concentrates on functional algorithms like Naive Bayes, FT Tree and Kstar.

1) *Classification*

The basic classification is based on supervised algorithms. Algorithms are applicable for the input data. Classification is done to know the exactly how data is being classified. The Classify Tab is also supported which shows the list of machine learning algorithms. These algorithms in general operate on a classification algorithm and run it multiple times manipulating algorithm parameters or input data weight to increase the accuracy of the classifier. Two learning performance evaluators are included with WEKA.

The first simply splits a dataset into training and test data, while the second performs cross-validation using folds. Evaluation is usually described by the accuracy. The run information is also displayed, for quick inspection of how well a classifier works.

2) *Manifold machine learning algorithm*

The main motivation for different supervised machine learning algorithms is accuracy improvement. Different algorithms use different rule for generalizing different representations of the knowledge. Therefore, they tend to error on different parts of the instance space. The combined use of different algorithms could lead to the correction of the individual uncorrelated errors. As a result the error rate and time taken to develop the algorithm is compared with different algorithm.

3) *Algorithm selection*

Algorithm is selected by evaluating each supervised machine learning algorithms by using supervised learning assessment (10-fold cross-validation) on the training set and selects the best one for application on the test set. Although this method is simple, it has been found to be highly effective and comparable to other methods. Several methods are proposed for machine learning domain. The overall cross validation performance of each algorithm is evaluated.The selection of algorithms is based on their performance, but not around the test dataset itself, and also comprising the predictions of the classification models on the test instance. Training data are produced by recording the predictions of each algorithm, using the full training data both for training and for testing. Performance is determined by running 10-fold cross-validations and averaging the evaluations for each training dataset. Several approaches have been proposed for the characterization of learning domain. The performance of each algorithm on the data attribute is recorded. The algorithms are ranked according to their performance of the error rate.

4) *Manuscript details*

This paper deals with Naive Bayes, KStar, and FT Tree algorithm. Experimental setup is discussed using 345 data and the results are compared. The performance analysis is done among these algorithms based on the accuracy and time taken to build the model.

V. EXPERIMENTAL SETUP

The data mining method used to build the model is classification. The data analysis is processed using WEKA data mining tool for exploratory data analysis, machine learning and statistical learning algorithms. The training data set consists of 345 instances with 7 different attributes. The instances in the dataset are representing the results of different types of testing to predict the accuracy of liver disease. The performance of the classifiers is evaluated and their results are analysed. The results of comparison are based on 10 ten-fold cross-validations. According to the attributes the dataset is divided into two parts that is 70% of the data are used for training and 30% are used for testing.

1) *Learning Algorithms*

This paper consists of three different supervised machine algorithms learning algorithms derived from the WEKA data mining tool. Which include:

- Naive Bayes,
- KStar
- FT Tree

The above algorithms were used to predict the accuracy of liver disease.

2) *Performance study of Algorithms*

The table I consists of secondary values of different classifications. According to these values the accuracy is calculated and analysed. It has 7 attributes for classification. Each one has a distinct value. Performance can be determined based on the evaluation time of calculation and the error rates. Comparison is made among these classification algorithms out of which the FT Tree algorithm is considered as the better performance algorithm. Because it takes only some time to calculate the accuracy than other algorithms.

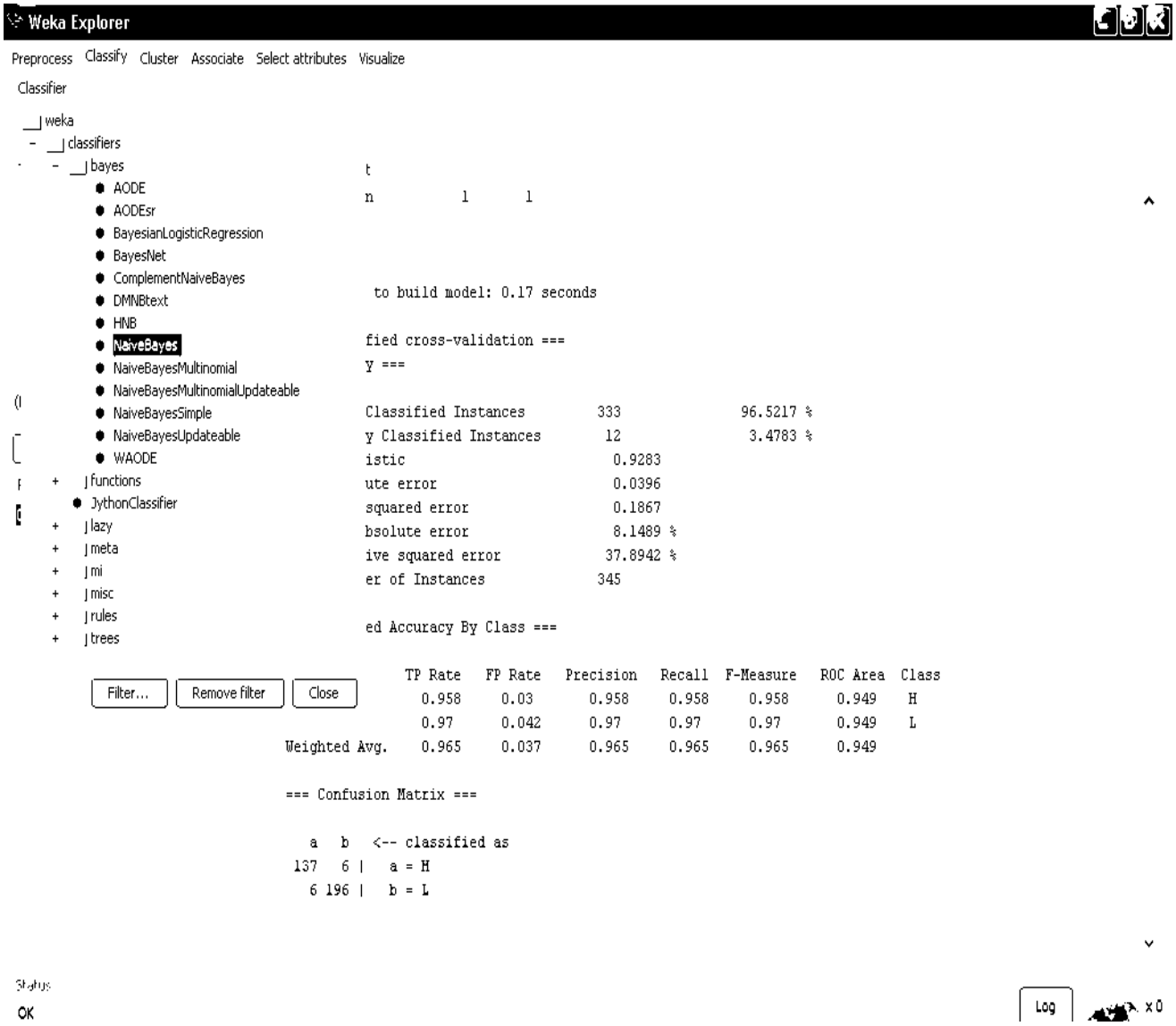
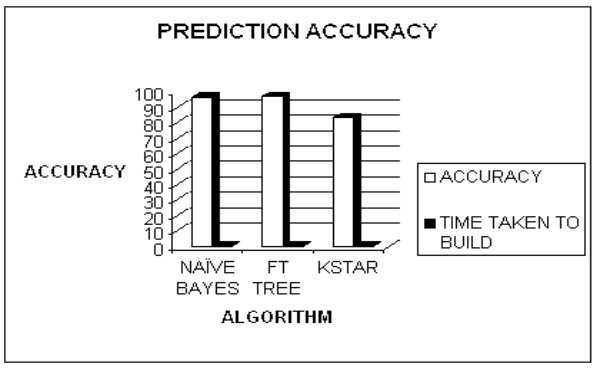


Fig .2 Selection of Naïve Bayes from the Classify Tab

TABLE I
PERFORMANCE STUDY OF ALGORITHM

Algorithm Used	Accuracy	Time Taken
Naive Bayes	**96.52%**	**0 sec**
FT Tree	97.10%	0.2 sec
KStar	83.47%	0 sec

TABLE 11
PERFORMANCE STUDY OF ALGORITHM



VI. Conclusion

Data mining in health care management is unlike the other fields owing to the fact that the data present are heterogeneous and that certain ethical, legal, and social constraints apply to private medical information. Health care related data are voluminous in nature and they arrive from diverse sources all of them not entirely appropriate in structure or quality. These days, the exploitation of knowledge and experience of numerous specialists and clinical screening data of patients gathered in a database during the diagnosis procedure, has been widely recognized. This paper deals with the results in the field of data classification obtained with Naive Bayes algorithm, FT Tree algorithm and KStar algorithm, and on the whole performance made known FT Tree Algorithm when tested on liver disease datasets, time taken to run the data for result is fast when compared to other algorithms. It shows the enhanced performance according to its attribute. Attributes are fully classified by this algorithm and it gives 97.10% of accurate result. Based on the experimental results the classification accuracy is found to be better using FT Tree algorithm compare to other algorithms. From the above results FT Tree algorithm plays a key role in shaping improved classification accuracy of a dataset.

VII. References

1) Duggan JM, Duggan AE. Systematic review: The liver in celiac disease. *Aliment Pharmacol Ther* 2005; 21: 515-518.
2) Cassagnou M, Boruchowicz A, Guillemot F *et al.* Hepatic steatosis revealing celiac disease: a case complicated by transitory liver failure. *Am J Gastroenterol* 1996; 91 : 1291-1292.
3) Austin A, Campbell E, Lane P, Elias E. Nodular regenerative hyperplasia of the liver and coeliac disease: potential role of IgA anti-cardiolipin antibody. *Gut* 2004; 53 : 1032-1034.
4) Hagander B, Brandt L, Sjolund K, Berg NO, Nordén Å Stenstam M. Hepatic injury in adult coeliac disease. *Lancet* 1977; 2 : 270-272.
5) Bonamico M, Pitzalis G. Culasso F *et al.* A. Il danno epatico nella malattia celiaca del bambino. *Minerva Pediatr* 1986; 38: 959-962.
6) Leonardi S, Bottaro G, Patané R, Musumeci S. Hypetransaminasemia as the first symptom in infant celiac disease. *J Pediatr Gastroenterol Nutr* 1990; 11 : 404-406.
7) Maggiore G, De Giacomo C, Scotta MS, Sessa F. Celiac disease presenting as chronic hepatitis in girl. *J Pediatr Gastroenterol Nutr* 1986; 5 : 501-503.
8) Vajro P, Fontanella A, Mayer M *et al.* Elevated Serum minotransferases activity as an early manifestation of glutensensitive enteropathy. *J Pediatr* 1993; 122 : 416-419.
9) Volta U, De Franceschi L, Lari F, Molinaro N, Zoli M, Bianchi FB. Coeliac disease hidden by cryptogenic hypertransaminasemia. *Lancet* 1998; 352 : 26-29.