International Journal on
"Technical and Physical Problems of Engineering"
(IJTPE)
Published by International Organization of IOTPE

ISSN 2077-3528
IJTPE Journal
www.iotpe.com
ijtpe@iotpe.com

December 2017 | Issue 33 | Volume 9 | Number 4 | Pages 67-70

DIAGNOSING LIVER DISEASES WITH DECISION TREE ALGORITHM

M.G. Borulday [1] E.G. Yegin [2] P. Mahouti [1] F. Gunes [1]

1. Electronics and Communications Engineering Department, Yildiz Technical University, Istanbul, Turkey
m.gokay@borulday.com, pmahouti@gmail.com, gunes@yildiz.edu.tr
2. Gastroenterology Department, Bozyaka Training and Research Hospital, Izmir, Turkey, drendergunes@hotmail.com

Abstract- Machine learning and its applications emerge as remarkable methods used frequently in medical diagnosis with quite successful results. In this regard, we expect that this study on the diagnosis of various liver diseases using machine learning will facilitate the decision-making process for doctors and help them diagnose the disease in a fast and effective manner.

Keywords: Decision Tree, Liver Diseases, Machine Learning.

I. INTRODUCTION

Diagnosis is one of the primary problems in medicine. There is a large amount of data that doctors need to evaluate, given the wide range of probabilities.

A. Statistics of Liver Diseases

Certain types of liver diseases are highly contagious and quite dangerous. Contagious liver diseases (especially Hepatitis types) are observed worldwide and pose a serious threat to humanity.

In this regard, here is some striking information based on statistics [1][2]:

• Every year 38.170 people die because of liver diseases in the USA.
• Liver cancer is the second most deadly type of cancer after lung cancer.
• 14 million new Hepatitis cases appear each year.
• There are approximately 350 million chronic Hepatitis B patients around the world, and only in Europe 36.000 people die because of chronic Hepatitis B each year.

B. Suggested Solution

Recent literature on machine learning and data mining points to the potential use of these methods in medical diagnosis [3-10].

We have developed an interface which analyzes laboratory test results with machine learning techniques and spots variables that are most useful for an accurate diagnosis.

Thus, a database based on most frequently used data related to liver diseases is created and an interface to facilitate doctor's diagnosis process through "Decision Tree" is developed. In this interface, results can be gathered in the form of percentages after entering the patient's data. After viewing the results in percentages, doctors can make their decisions based on more compelling tests and they can add new test results to the interface.

A decision tree is designed to be used in the diagnosis of 14 different liver diseases, whose schematic is given in (Figure 1). It consists of 71 different analysis results including demographic data, physical examination, laboratory tests, symptoms, radiological imaging, liver biopsy and unclassified data. After the doctor chooses the appropriate options on the interface, results are created from a narrower pool of probabilities.

II. DECISION TREE DATA

Many different factors may cause liver diseases such as alcohol consumption, bacteria or viruses, or immunity system problems. The analysis data chosen to be used for the decision tree is picked among those that are most frequently seen and are highly possible to overlook.

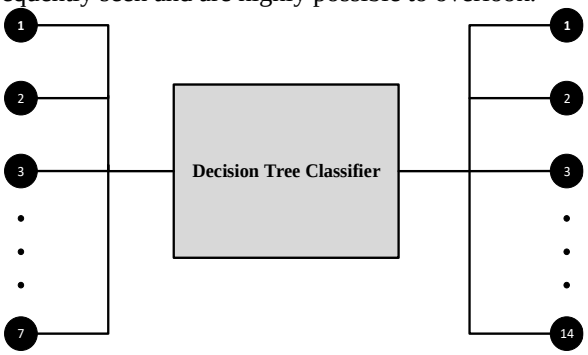

Figure 1. Schematic of decision tree for diagnosing liver diseases

Table 2 shows various data samples that include 5 of the 7 different test categories chosen for 14 different liver diseases with the meanings given in Table 1. On this table, 7 important sets of data are chosen to be shown to present a sample from a pool of 71 sets of data in total. The data on the table is interpreted in the following way: a positive test result shows the influence of the related test on the diagnosis of the disease. For example, "2" signifies the influence of the positive status of "Alt elevation" on "Acute Hepatitis B" disease. 4 different data sets are used to determine the influence. These data sets and their meanings are as Table 1.

Table 1. Data set meanings

Data Set	Meanings
No Effect	No Effect
0	None Existent
1	Disease
2	Probable Disease

In addition, the difficulty level of each test category and a coefficient determined according to the difficulty level are presented. The influence of the test results on disease diagnosis is determined according to these coefficients. Since liver biopsy is the most difficult and expensive compared to the laboratory tests, the disease diagnosis coefficients will have strong influence on the accuracy of the disease diagnosis. A coefficient is assigned to each category after all data on the table is evaluated in terms of processing difficulty and expenses.

Table 2. Sample data

	Laboratory Tests		Signs and Symptoms		Physical Examination	Radiological Imaging	Liver Biopsy
	ALT elevation	Hemolysis	Ulcerative colitis	Cholangitis	Ascites	Hepatic steatosis by radiologic imaging	Pericentral steatosis in liver biopsy
Acute Hepatitis B	2	0	0	0	0	no effect	0
Chronic Hepatitis B	2	0	0	0	2	no effect	0
Chronic Hepatitis C	2	0	0	0	2	no effect	0
Chronic Hepatitis B+D	2	0	0	0	2	no effect	0
Acute Hepatitis A	2	0	0	0	0	no effect	0
Autoimmune Hepatitis	2	0	0	0	2	no effect	0
Nonalcoholic Fatty Liver Disease	2	0	0	0	2	1	1
Alcoholic Liver Disease	2	0	0	0	2	1	1
Hemochromatosis	2	0	0	0	2	no effect	0
Wilson Disease	2	2	0	0	2	no effect	0
Primary Biliary Cirrhosis	2	0	0	0	2	no effect	0
Secondary Biliary Cirrhosis	2	0	0	1	2	no effect	0
Primary Sclerosing Cholangitis	2	0	1	1	2	no effect	0
Cirrhosis	2	no effect	2	2	1	2	no effect

A. Main Reasons for Using Decision Tree Algorithm

Decision Tree algorithm is used to make diagnosis easier. The main reasons for using this algorithm are:

• It is simple to understand and interpret. A simple explanation is enough for people to understand the results of decision tree learning.

• It requires a quick preprocessing. The data can become ready for use with much less processing compared to many other alternative techniques. The preprocessing phase takes less time and is simpler compared to other alternatives.

• It can be used to process both quantitative and categorical data. Most machine learning algorithms are useful either for quantitative applications or for categorization problems whereas decision tree learning can be used in both fields.

• It uses white-box model. In the white-box model, which is an approach used in software engineering, each step can be monitored and interpreted, whereas in black-box model, which is also a software engineering application, machine learning is based more on the artificial neural network. Although the latter allows for the interpretation of input and output, it makes it impossible to monitor and interpret the internal dynamics of the system step by step.

• It offers low computational complexity. It can process a large amount of data more quickly because of its simplicity and its speed compared to the alternative methods, which makes it much more preferable when the amount of data to be processed is larger.

III. DECISION TREE INTERFACE

The data set that is used to create the decision tree is presented to the doctor through the interface. The aim is to help the doctor to diagnose the disease fast and accurately, starting with the simplest and the most cost-efficient procedure. The ranking below is created taking into account the cost and difficulty of the application with its effect on the result:

- Demographic Data Entry
- Physical Examination
- Determining the Symptoms
- Laboratory Tests
- Radiological Imaging
- Liver Biopsy
- Unclassified Data

The doctor is expected to start from the simplest procedure and move to more difficult ones. The coefficients of the tests and the physical examination are ranked according to their difficulty levels.

In the light of the data that is entered, a graphic chart which lists possible diseases with their probability rate is created. By adding new and more efficient test results to the data, it becomes possible to narrow down probable diseases and increase the chance of reaching the correct diagnosis.

To present an example for the application of this method, the diagnosis process of a patient is broken down to steps. As the first step, a male patient under 40 showed no symptom of disease in the physical examination. It is decided to run laboratory tests, and "HFE Gene Mutation" is found.

Once the related data is entered in the interface (Figure 2), it is shown that there are two probable diseases: Cirrhosis, with the probability rate of 66.7%, and Hemochromatosis, with the probability rate of 33.3% (+).

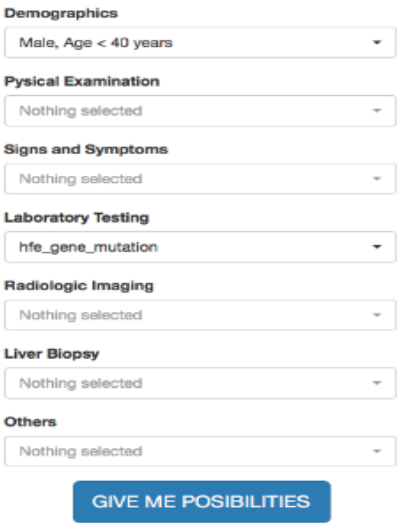

Figure 2. Use of application step 1

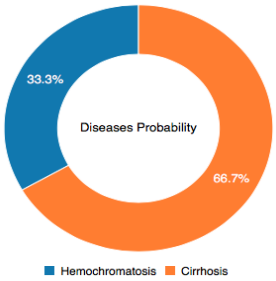

Figure 3. Use of application step 1 result

As the second step, it is decided to do "Radiological Imaging", and "Hepatic steatosis" is found (Figure 4). Once the related data is entered into the interface, it is shown that there are two probable diseases: Cirrhosis, with the probability rate of 73.5%, and Hemochromatosis, with the probability rate of 26.5% (Figure 5).

As the third step, it is decided to run a "Liver Biopsy" for a more reliable diagnosis, and "Pericentral steatosis" is found (Figure 6). Once the related data is entered in the interface, it is shown that the disease is Cirrhosis with a probability rate of 100% (Figure 7).

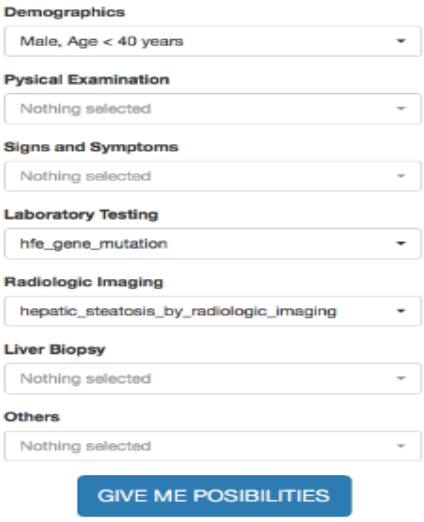

Figure 4. Use of application step 2

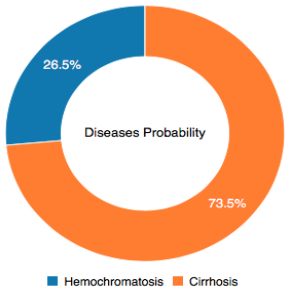

Figure 5. Use of application step 2 result

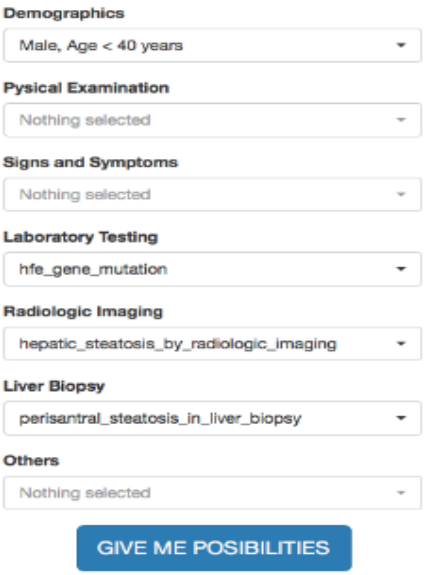

Figure 6. Use of application step 3

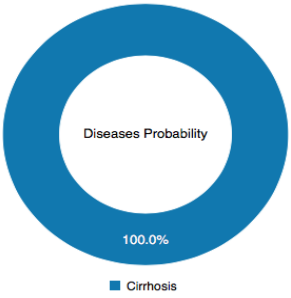

Figure 7. Use of application step 3 result

IV. CONCLUSION

Symptoms of liver diseases are difficult to recognize and identify. More and more people suffer from liver diseases, yet they may not be aware of it. They are unable to identify the symptoms. This makes accurate diagnosis difficult for the doctor, which may result in incorrect treatment. Hence, an accurate diagnosis is essential to provide the correct and necessary treatment.

This study has used machine learning as an analytical tool for the purposes of medical diagnosis and yielded effective positive results. It offers doctors considerable help in diagnosis while at the same time reducing the need for additional medical procedures.

The objective is to make the necessary improvements and additions so that the application and the interface can be used in real patient cases. The decision tree uses the 71 tests and narrows down the 14 probable diseases. It gives the most probable disease as the output.

This application is aimed to be used in real patient cases, and is the first step towards improving the success rate of diagnosis. It is also the first step towards the possibility of interpreting a broad spectrum of disease and test data as a whole.

REFERENCES

[1] http://www.who.int/gho/alcohol/harms_consequences /deaths_liver_cirrhosis/en/.
[2] http://2016.ilc-congress.eu/wpcontent/ uploads/2016/ 04/Liver-disease-backgrounder.pdf.
[3] H. Ayeldeen, O. Shaker, G. Ayeldeen, K.M. Anwar, "Prediction of Liver Fibrosis Stages by Machine Learning Model: A Decision Tree Approach", Third World Conference on Complex Systems (WCCS), pp. 1-6, Morocco, 2015.
[4] H. Ayeldeen, O. Hegazy, A.E. Hassanien, "Case Selection Strategy Based on K-Means Clustering", 2nd International Conference on Information Systems Design and Intelligent Applications, Springer, India, 2015.
[5] A. Zidan, N.I. Ghali, A.E. Hassanien, H. Hefny, J. Hemanth, "Level Setbased CT Liver Computer Aided Diagnosis System", Journal of Intelligent and Robotic Systems, Issue on Practical Perspective of Digital Imaging for Computational Applications, Vol. 9, Iss. 1, 2013.
[6] H. Ayeldeen, O. Shaker, O. Hegazy, A.E. Hassanien, "Distance Similarity as a CBR Technique for Early Detection of Breast Cancer: An Egyptian Case Study", Information Systems Design and Intelligent Applications, Vol. 340, pp. 449-456, 2015.
[7] K.K.A. Ghany, H.A. Hefny, A.E. Hassanien, N.I. Ghali, "A Hybrid Approach for Biometric Template Security", IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012), pp. 941-942, 2012.
[8] Y. Fanid Fathabad, M.A. Balafar, "Application of Content Based Image Retrieval in Diagnosis Brain Disease", International Journal on Technical and Physical Problems of Engineering (IJTPE), Issue 13, Vol. 4, No. 4, pp. 133-138, December 2012.
[9] Sh. Akbarpour, "A Review on Content Based Image Retrieval in Medical Diagnosis", International Journal on Technical and Physical Problems of Engineering (IJTPE), Issue 15, Vol. 5, No. 2, pp. 148-153, June 2013.
[10] S. Salimi, M. Sabbagh Nobarian, S. Rajebi, "Skin Disease Images Recognition Based on Classification Methods", International Journal on Technical and Physical Problems of Engineering (IJTPE), Issue 22, Vol. 7, No. 1, pp. 78-85, March 2015.

BIOGRAPHIES

Muhammet Gokay Borulday received his undergraduate degree in Electronics and Communication Engineering from Yildiz Technical University, Istanbul, Turkey. He has been currently working as a software developer on his own company. His research interests are in the areas of neural networks, machine learning algorithms and different kind of computer programming languages.

Ender Gunes Yegin received her M.D. degree from Faculty of Medicine, Istanbul University, Istanbul, Turkey. She completed her internal medicine residency and fellowship in gastroenterology and hepatology at Faculty of Medicine, Marmara University, Istanbul, Turkey. She is currently an Associate Professor in Izmir Bozyaka Training and Research Hospital, Izmir, Turkey. Her clinical and research interests cover digital image analysis in liver fibrosis, cirrhosis, and endoscopic ultrasonography.

Peyman Mahouti received his Ph.D. degree in Electronics and Communication Engineering from Yildiz Technical University, Istanbul, Turkey in 2016. He has been currently working as a Teaching Assistant in Department of Electronic and Communication Engineering of the same university. His main research areas are optimization of microwave circuits, broadband matching circuits, and device modelling, and computer-aided circuit design, and microwave amplifiers.

Filiz Gunes received her M.Sc. degree in Electronics and Communication Engineering from Istanbul Technical University, Istanbul, Turkey. She attained her Ph.D. degree in Communication Engineering from Braivendford University in 1979. She is currently a Full Professor in Yildiz Technical University, Istanbul, Turkey. Her research interests are in the areas of multivariable network theory, device modelling, computer aided microwave circuit design, monolithic microwave integrated circuits, and antenna designs.