

# Prediction of hepatitis disease based on principal component analysis and artificial immune recognition system

Kemal Polat \*, Salih Güneş

*Selcuk University, Department of Electrical and Electronics Engineering, 42035 Konya, Turkey*

---

## Abstract

In this study, prediction of hepatitis disease, which is a very common and important disease, was conducted with principal component analysis (PCA) and artificial immune recognition system (AIRS). The proposed approach consists of two stages. Firstly, the feature number of hepatitis disease dataset was reduced to 5 from 19 by principal component analysis (PCA). Secondly, hepatitis disease dataset is normalized in the range of [0, 1]. Normalized input values is classified by using AIRS classifier system. We took the dataset used in our study from the UCI Machine Learning Database. The obtained classification accuracy of our system was 94.12% using 10-fold cross-validation and it was very promising with regard to the other classification applications in Literature for this problem. Testing results were found to be compliant with the expected results that are derived from the physician's direct diagnosis. The end benefit would be to assist the physician to make the final decision without hesitation. This result is for hepatitis disease but it states that this method can be used confidently for other medical diseases diagnosis problems, too.

© 2006 Elsevier Inc. All rights reserved.

*Keywords:* Artificial immune system; AIRS; Principal component analysis (PCA); Hepatitis disease prediction

---

## 1. Introduction

Most of the time the hepatitis prediction is made by a routine blood testing or during a blood donation. The hepatitis is a viral infection that also was transmitted by blood or blood products in the past, when there was no test available to screen for this infection. Risk factors are as follows: blood transfusions, tatoos and piercing, drug abuse, hemodialysis, health workers, sexual contact with hepatitis carrier [1].

The use of classifier systems in medical diagnosis is increasing gradually. There is no doubt that evaluation of data taken from patient and decisions of experts are the most important factors in diagnosis. But, expert systems and different artificial intelligence techniques for classification also help experts in a great deal. Classification systems, helping possible errors that can be done because of fatigued or inexperienced expert to be minimized, provide medical data to be examined in shorter time and more detailed.

---

\* Corresponding author.

*E-mail addresses:* [kpolat@selcuk.edu.tr](mailto:kpolat@selcuk.edu.tr) (K. Polat), [sgunes@selcuk.edu.tr](mailto:sgunes@selcuk.edu.tr) (S. Güneş).

In this study a new medical diagnosis method was proposed to be used hepatitis disease diagnosis problem as a classifier. This method involves a two-stage system in which a principal component analysis (PCA) and artificial immune recognition system are hybridized. The first stage of the whole system is used to feature selection for feature reduction in hepatitis disease using principal component analysis (PCA). In the second stage, artificial immune recognition system as classification method is used to medical decision making for hepatitis disease.

The used data source is taken from the University of California at Irvine (UCI) Machine Learning Repository [2]. This dataset is commonly used among researchers who use Machine Learning (ML) methods for hepatitis disease classification and so it provides us to compare the performance of our system with other conducted studies related with this problem.

The performance of the system was analyzed with regard to the classification accuracy, sensitivity and specificity. The values of these performance criterions were obtained via 10-fold cross-validation, which is a very common performance evaluation method in ML literature. Our proposed system reached 94.12% classification accuracy in test phase for hepatitis disease and this result is the highest one among the studies applied for hepatitis disease classification problem so far.

The rest of the paper is organized as follows. Section 2 gives the background information including hepatitis disease classification problem, previous research in corresponding area and brief introduction to natural and artificial immune systems. We explained the method in Section 3 with subtitles of proposed a new medical diagnosis method and Measures for Performance Evaluation. In each subsection of that section, the detailed information is given. The results obtained in applications are given in Section 4. This section also includes the discussion of these results in specific and general manner. Consequently in Section 5, we conclude the paper with summarization of results by emphasizing the importance of this study and mentioning about some future work.

## 2. Background

### 2.1. Hepatitis disease dataset

Hepatitis B is caused by a virus that attacks the liver. The virus, which is called hepatitis B virus (HBV), can cause lifelong infection, cirrhosis (scarring) of the liver, liver cancer, liver failure, and death. In 2003, an estimated 73,000 people were infected with HBV. People of all ages get hepatitis B and about 5000 die per year of sickness caused by HBV. HBV is spread when blood from an infected person enters the body of a person who is not infected. Healthcare personnel who have received hepatitis B vaccine and developed immunity to the virus are at virtually no risk for infection. For a susceptible person, the risk from a single needlestick or cut exposure to HBV-infected blood ranges from 6% to 30%. The annual number of occupational infections has decreased 95% since hepatitis B vaccine became available in 1982, from >10,000 in 1983 to <400 in 2001 [3].

This hepatitis disease dataset requires determination of whether patients with hepatitis will either live or die. It was donated by Jozef Stefan Institute, Yugoslavia. The used data source in this study was taken from UCI Machine Learning Repository. The purpose of the dataset is to predict the presence or absence of hepatitis disease given the results of various medical tests carried out on a patient. This database contains 19 attributes, which have been extracted from a larger set of 155. Hepatitis dataset contains 155 samples belonging to two different classes (32 “die” cases, 123 “live” cases). There are 19 attributes, 13 binary and 6 attributes with 6 to 8 discrete values. Attributes of symptoms that are obtained from patient are as follows [2]:

1. Age: 10, 20, 30, 40, 50, 60, 70, 80
2. Sex: Male, Female
3. Steroid: No, Yes
4. Antivirals: No, Yes
5. Fatigue: No, Yes
6. Malaise: No, Yes
7. Anorexia: No, Yes
8. Liver Big: No, Yes

Download English Version:

<https://daneshyari.com/en/article/4635515>

Download Persian Version:

<https://daneshyari.com/article/4635515>

[Daneshyari.com](https://daneshyari.com)