The Diagnosis of Hepatitis diseases by Support Vector Machines and Artificial Neural Networks

Modjtaba Rouhani, IEEE Member,
Islamic Azad University, Gonabad branch
Gonabad, Iran
m.rouhani@ieee.org

Mehdi motavalli haghighi
Islamic Azad University, Gonabad branch
Gonabad, Iran
mehdi.haghighi@yahoo.com

Abstract—in this paper, we use Support Vector Machine (SVM) and artificial neural networks to diagnosis Hepatitis diseases. Furthermore, we use those networks to identify the type and the phase of disease. Considering the most important hepatitis cases leads us to six classes: hepatitis B (two phases), hepatitis C (two phases), non-viral hepatitis and no-hepatitis. For this purpose, we design various networks including RBF, GRNN, PNN, LVQ and SVM. The performance of each of them has studied and the best method is selected for each of classification tasks. The overall accuracy of diagnosis system is near 97%.

Keywords- hepatitis disease, PNN, LVQ, SVM, RBF, GRNN)

I. INTRODUCTION

Hepatitis disease is a fatal and deadly disease and is thought to be the fifth deadly disease worldwide. Hepatitis disease is the inflammation and damage to hepatocytes in the liver and can be caused by infections with viruses, bacteria, fungi, exposure to toxins, alcohol consumption and autoimmunity. Clinical symptoms of hepatitis are nausea, fever, general weakness, and jaundice. Five viruses have been identified and named hepatitis A through E [1,2].

In [1] and [2], the authors use different ANN structures to recognize the hepatitis disease. Reference [10] analyses a large database with hepatitis C virus infected patients. There are made a lot of statistical analyses on the records of this database in order to determine the evolution of biological parameters during the treatment. The results of the statistical analyses and the expert system predictions indicate the use of such a system to facilitate the physician work. In [8], authors use a feature selection (FS) and artificial immune recognition system (AIRS) with fuzzy resource allocation mechanism diagnosis of hepatitis disease with total accuracy of 92.59% on data set of UCI. In [9], ANN has been used to diagnosis chronic hepatitis disease. They report total accuracy of 93%.

In fact, In addition to recognizing the hepatitis cases, it is important to identify the phase and the type of the hepatitis the person caused by, which is the main propose of this paper.

There are fifteen parameters measured for each patient in order to diagnose hepatitis, as follow:
1-sex, 2-age, 3-ALK, 4-AST, SGOT, 5-ALT, SGPT, 6-Bi, T, 7-Bi, D, 8-G.G.T, 9-HBSAg, 10-Alb, 11-LHD, 12-PT, 13-FBS, 14-CHO, and 15-HCVAb.

These experimental data show six cases: 1-non hepatitis person, 2-person who carries hepatitis B (no symptoms), 3-person affected by hepatitis B, 4-person who carries hepatitis C (no symptoms), 5-person affected by hepatitis C and 6-non viral hepatitis.

We prepared a data set of 250 suspicious cases, from patients visited in two major hospitals in Mashhad, Iran, each of which has been carefully checked up by specialists and the diagnose is made. 58 of those patients was diagnosis as non hepatitis, 32 cases for person who carriers hepatitis B, 53 cases for person affected by hepatitis B, 32 specimens for person who carriers hepatitis C, 32 cases for person affected by hepatitis C and 43 specimens for person who affected by non-viral hepatitis.

This paper organized as follow: in the second section a brief introduction to each ANN and SVM used is made. The third section presents the main results and in the forth section a combination of the networks is made which results the best accuracy. Section five concludes the paper.

II. ANN CLASSIFIERS

The performance of five different networks in classification of hepatitis data sets are compared in this article. Those include radial basis functions (RBF), generalized regression neural network (GRNN), Probabilistic neural networks (PNN), LVQ network, and support vector machines (SVM). A brief description for them follows.
A. RBF networks:
Radial basis networks are powerful tools for classification. RBF has a two-layer structure. Neurons in the first layer compute the distance of the input vector to their centers, so the first layer determines which neuron is more similar (closer) to input vector. The second layer has linear activation function and its output is weighted sum of output of first layer. Learning of RBF is much easier and faster than BP networks as network parameters in first layer (centers) are set randomly and weights of the second layer are calculated to minimize the classification error. The activation function of hidden layer neurons is usually Gaussian function which is shown in equation (1).

\[ \psi_k(x-c_k) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-c_k)^2} \]  

Where \( c_k \) is the center of radial basis function of neuron \( k \)[4,3].

B. GRNN networks:
A generalized regression neural network (GRNN) is often used for function approximation. It has a radial basis