

A Prediction Model for the Drug Efficacy of Interferon in CHC Patients Based on SNPs

Eugene Lin, Dennis Chen, Yuchi Hwang, Ashely Chang, Z. John Gu
Vita Genomics, Inc.

7 Fl., No. 6, Sec. 1, Jung-Shing Road, Wugu Shiang, Taipei, Taiwan
{eugene.lin, dennis.chen, yuchi.hwang, ashely.chang, john.gu}@vitagenomics.com

Abstract

In the studies of pharmacogenomics, genetic predisposition information, such as single nucleotide polymorphisms (SNPs), can be used to understand the relationship between genetic variations (or population variations) and drug efficacy. In this paper, a prediction model is resulted from analyzing chronic hepatitis C (CHC) patient's SNPs, comparing to control groups, to predict the responsiveness of interferon (IFN) combination treatment. We have developed an advanced methodology with the combination of artificial neural network (ANN) and other algorithms to achieve a prediction with high accuracy among the patients. Filtering through thousands of SNPs of 150 genes, we found nearly 30 SNPs relevant to the responsiveness of IFN. With a statistical analysis of sensitivity (SEN), specificity (SPE), positive prediction value (PPV), and negative prediction value (NPV), our model achieves a higher successful rate of prediction, i.e., > 90% accuracy. This model allows patients and doctors to make more informed decisions based on SNP genotyping data. The data was generated in the high-throughput genomics lab of Vita Genomics, Inc.

1. Introduction

Hepatitis C virus (HCV) is a major cause of acute hepatitis and chronic liver diseases, including cirrhosis and liver cancer. Globally, an estimated 200 million persons are affected with HCV, and 3 to 4 million persons are newly infected each year [2]. About 80% of newly infected patients progress to develop chronic infection. For the patients with chronic infection over a period of 20 to 30 years, about 10% to 20% of these patients will develop to cirrhosis, and 1% to 5% of these patients will result in liver cancers. Antiviral drugs, such as interferon (IFN) taken alone or in

combination with ribavirin, can be used for the treatment of persons with chronic hepatitis C (CHC). But, the high cost of treatment and the associated side-effects are major factors stopping patients pursuing the treatment. Furthermore, the effectiveness of treatment is only about 10% to 20% of patients for IFN alone and is about 30% to 50% of patients for IFN combined with ribavirin. Therefore, it is important to establish models to distinguish responders from non-responders. Recently, a pharmacogenomic approach [2] was introduced to predict clinical outcomes using single nucleotide polymorphisms (SNPs), the genetic markers of human DNA variations, prior to the treatment.

In this work, we have successfully developed a computer algorithm based on an artificial neural network (ANN) model to help predict the treatment responses for CHC patients. The accuracy of this algorithm was studied using a statistical analysis.

This paper is organized as follows: Section 2 describes our prediction model for the drug efficacy of IFN. The simulation results are presented in Section 3, which is followed by the discussion in Section 4.

2. Methods

Artificial neural network algorithms, such as supervised feed-forward neural networks (FFNN), are generally adopted for classification applications [1]. From the structural point of view, an FFNN is a spatially iterative neural network with several layers of hidden neuron units between the input and output neuron layers. The basis function of each neuron is the linear basis function, and the activation is modeled by a non-decreasing and differentiable sigmoid function [3]. In this work, a feed-forward neural network is used for modeling the responsiveness of IFN. Inputs contain the SNPs information of the CHC patients. Outputs contain the information about the responsiveness of IFN.

From the algorithmic point of view, the processing of this multilayer FFNN can be divided into the retrieving and learning phases. In the retrieving phase, the FFNN iterates through all the layers to generate the retrieval response at the output layer according to the inputs and the known weights of the network.

In the learning phase of this multilayer FFNN, a simple gradient descent approach known as the back-propagation algorithm [3] is used for the learning scheme. For the hidden layers, the weight updating adopts the mechanism of back-propagated corrective signals from the output layer. By presenting a pair of input/target training patterns, the goal is to iteratively choose a set of weights for all layers so that the squared error function can be minimized.

Furthermore, an adaptive learning rate is employed to improve the convergence rates of the back-propagation algorithm. The gradient descent method includes a term with a proportion of the last weight change. Hence, the determined learning rate is different for each epoch.

In summary, the FFNN is trained first by repeatedly presenting input-output training pairs and performing the back-propagation learning algorithm. After this training, the FFNN is tested by presenting the inputs of testing pairs (i.e., SNPs) to the network. The forward propagation of the FFNN provides the responsiveness of IFN for a particular patient, giving us a means of inference from cause to effect.

3. Results

Based on the current treatment strategy for CHC patients, we have focused our searching for genetic markers in IFN signaling pathways [2]. Thousands of SNPs in 150 candidate genes were discovered, genotyped, and studied among CHC patients with IFN combination treatment. There were 385 subjects including 250 responders and 135 non-responders. The SNPs genetic markers of these subjects were generated at the high-throughput genomics lab of Vita Genomics, Inc. We used a rotating sampling approach to select six groups of different training and testing sets. For each group, we randomly selected 325 subjects as a training set and 60 subjects as a testing set, respectively.

The 325 input-output training data pairs were used to train a FFNN. Inputs were the 30 SNPs genetic markers. Outputs were the IFN responding status. All these input/output values were normalized to 0.1 and 0.9. Each SNP represented two inputs of the FFNN, for example, {0.9, 0.9} for the minor allele. The responder and non-responder were converted into 0.9 and 0.1, respectively. Using this information, we

trained the FFNN which had 60 inputs, 1 output, and 1 hidden layer. This trained FFNN approximates the model of the responsiveness of IFN among CHC patients. After the network was trained, we used the trained network to find the responsiveness condition corresponding to the 60 subjects of the testing set.

To validate the prediction model, we calculated sensitivity (SEN), the proportion of true predicted responders of all tested responders, and specificity (SPE), the proportion of true predicted non-responders of all the tested non-responders. In addition, we quantified positive prediction values (PPV), the proportion of true predicted responders among all predicted responders, and negative prediction values (NPV), the proportion of true predicted non-responders among all the predicted non-responders.

By using the sensitivity and specificity with different threshold values, receiver operating characteristic (ROC) curves were created to evaluate the results of the prediction model. The overall classification accuracy by the FFNN was 91% with a ROC area of 0.85. The values of SEN, SPE, PPV, and NPV were 1, 0.75, 0.89, and 1, respectively.

4. Discussion

We have developed an advanced methodology to predict the drug efficacy of IFN in CHC patients based on SNPs. Our results have shown that gene polymorphisms were strongly associated with the drug responding status in CHC patients. We demonstrated that a trained FFNN model is a promising method for providing the inference from SNPs to the responsiveness of IFN. Our model achieves a higher successful rate of prediction and allows patients and doctors to make more informed decisions.

References

- [1] J.N. Hwang and E. Lin, "Mixture of discriminative learning experts of constant sensitivity for automated cytology screening," *1997 IEEE Workshop on Neural Networks for Signal Processing*, Amelia Island, Florida, September 1997, pp. 152–161.
- [2] Y. Hwang, K. Chen, C. Chang, and J. Chen, "SNP genotype modeling - pharmacogenomic approach for studying hepatitis C treatment," *International Liver Congress*, Hong Kong, February 2004, pp. 96–97.
- [3] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representation by error propagation," in *Parallel Distributed Processing: Explorations in the Micro-Structure of Cognition, vol.1, Foundations*, Cambridge, MA: MIT Press, 1986.