SOUTHEAST UNIVERSITY - MONASH UNIVERSITY

JOINTGRADUATE SCHOOL (SUZHOU)

# An Application of the Neural Network Model to Diagnosis of Hepatitis

A Project Overview of the Hepatitis Diagnosis Using the Neural Network Model

Student name: Huang Hanlin ID: 2731****

Student name: Yu Yunxiu ID: 2731****

30 May 2016

# Abstract

The application of neural networks (NNs) in disease diagnosis has been widely adopted in last few decades. One of its application area is in the hepatitis diagnosis. The diagnosis for patients to find whether their hepatitis is life-threatening or not has a significant meaning and many researchers have implemented related NN models. These NN models differ from each other in architecture, training algorithm or activation function but they all provide a relatively high accuracy in diagnosis. In this paper, our research focuses on building a high accuracy model and finding the influential input features which will impact the diagnosis result. This research has not been done by former researchers as far as we know. So, this research project that aims to reveal the influential input features in hepatitis diagnosis model based on a high accuracy NN model is of great need and has realistic meaning.


***Keywords:*** *Hepatitis Diagnosis, Neural Network, Influential Features*

## 1. Introduction

Hepatitis has been one of the worldwide major public health concerns and will cause the damage to the liver cells and impairs the function of the liver. Patients who get infected with this disease may recover within 1 or 2 months or will die with complications in few years later. Since this disease is one of the serious diseases which demands expensive treatment and severe side effects appear very often, it is important to identify those patients' disease type and give appropriate treatment to relieve their pain. Traditionally this diagnosis are done by the experienced doctors. However, it is usually difficult and not efficient.

Since 1990s, the artificial neural network (ANN) techniques has gained widely acceptance, its applications ranges from signal processing in communications to pattern recognition in Business, Engineering and Medicine. Also ANN has been widely applied to medical diagnosis. So, the automatic diagnosis using neural networks will solve the limitation of the traditional methods of diagnosing hepatitis. Numerous research have been conducted and many NN models have been implemented to provide high accuracy classification results. The data set which is most commonly used by researchers to train and test their NN model is from UCI machine learning database. This data set consist of 155 samples of hepatitis patients with 19 physical and medical features. And 32 samples of them belongs to class 1-die while 123 samples of them belongs to class 2-live. Ozyilmaz and Yildirim (2003) conduct study of the Multilayer Perceptron (MLP) structure trained by the standard backpropagation (BP) algorithm and the Radial Basis Function (RBF) network structure trained by the Ordinary Least Square (OLS) algorithm. And the MLP model reaches 81.375% average accuracy and the RBF model reaches 85% accuracy. Ansari et al. (2011) use the Generalized Regression Neural Network (GRNN) model to diagnose hepatitis disease. The GRNN is a variant of the RBF network and in their study their model reaches the diagnosis accuracy of 92%.

Bascil and Temurtas (2011) develop the multilayer neural network (MLNN) for diagnosing hepatitis which uses the Levenberg Morquardt (LM) algorithm as the training algorithm to update the weight of the network. According to their study result, they get 91.87% classification accuracy. Çetin et al. (2015) make an improvement on the NN model in the study of Bascil and Temurtas (2011) which uses the LM algorithm for learning. In their study, they introduce simple polynomial function, piecewise function and Taylor series expansion as approximations of sigmoid activation function. According to their study result, compared with NN models using the sigmoid function which reach 93% classification accuracy, the first approximation sigmoid function have 91.8% accuracy, the piecewise linear approximation function reaches 92.5% accuracy and the Taylor series expansion approximation function reaches 93.1% accuracy. Although lots of work has been done on the building of NN model, we find none of them analyses the influential input features that have impact on the final diagnosis result. It has very practical meaning in that it will build a better and more efficient NN model which will reduce the examine procedures and provide a high diagnosis result at the same time.

## 2. Objectives

The purpose of our project is to find the influential input features in hepatitis diagnosis NN model and their correlations. Once the influential features are confirmed, it will narrow down the input features without impeding the accuracy in hepatitis diagnosis .So related medical examination procedures will be simplified and it can give patients advice in preventing hepatitis becoming more serious.

## 3. Methodology

There exits several NN models with different algorithms used in hepatitis diagnosis. Among this models, the MLNN model with backpropagation (BP) algorithm is most commonly used in classification. Our research project is to build a 3 layers MLNN model with BP algorithm to diagnose the hepatitis and find influential features and correlations among them. In this model, the outputs of the units in one layer form the inputs to the next layer and the weight of network calculated through the training of network by the BP algorithm. The major problem of our study is how to design a model with high accuracy and how to find the influential features and correlations among them. Several methods are used to solve these problems. The first one is doing several rounds of experiments to collect the data. The second one is comparing and analyzing the data. The third one is make deduction according to the data and proving it. Finally, we get our conclusion. By building our own model and finishing all of the experiments, the project is thought to make an improvement on previous research which have not focus on the influential features and correlations among them.

This project has some positive influences because it focus on increasing the accuracy and researching more specific features which effect the result most. It has practical

contribution to the patients because it helps the patients know their disease state better and can get relevant treatment in time. Besides, the finding of the influential features that affect the diagnosis result and correlations among these features gives suggestion to the hepatitis patients in preventing hepatitis becoming more serious. However it also has some limitations. The limitations of our work lies in 3 aspects. First, the size of the data set is small and many of the samples in the data set are incomplete which result in that smaller data set can be used for training and test of the model. Second, the BP algorithm used in our experiment doesn't have the best performance compared with other algorithms. At last, in our experiment, not all possibility have been tested. In the future studies, more experiments are needed to be done to ensure a better result of recognizing the influential features of the hepatitis diagnosis model.

The following part is the experiment steps that show how we build our model, find key features and correlations among them. Our data set is from the UCI machine learning database. All samples in the data set have 19 features which are age, sex, steroid, antivirals, fatigue, malaise, anorexia, liver big, liver firm, spleen palpable, spiders, ascites, varices, bilirubin, alk phosphate, sgot, albumin, protime and histology. We did 5 rounds of experiments which include 30 experiments. In order to form the test set, 20 % of the data set has been randomly extracted. Round 1 includes experiment 1-3 and its purpose is to change parameters to build a high accuracy model. Round 2 includes experiment 4-5 and its goal is to decide the number of output neurons of the model. Round 3 includes experiment 6-24 and this round is to decide if there is a influential feature which individually affect the result most. Round 4 includes experiment 25 and its purpose is to prove if there is disturbance features. Round 5 includes experiment 26-30. Goal of this round is to verify if there is correlations among some features. In round 2-5, the hidden layer, the learning rate and the momentum are fixed to 30, 0.9 and 0.6 respectively. In round 3-5, the output neuron is fixed to 1 (i.e., Class die is represented as 1, Class live as 2).

## 4. Novelty

NN models have been widely used in hepatitis diagnosis. However previous research have not pay attention on finding the influential features and analyzing joint effect among these features. Our research project has new findings mainly in two aspect. One is that there is no feature that individually greatly affect the diagnosis result. Another is that there do exits four joint features influence the result to an extremely extent only when they combine together. The four features are anorexia, liver big, alk phosphate and protime. This finding inspires us that we don't need to take so many features in to consideration when diagnosing hepatitis. In addition, using the four features when testing the model, the diagnosis accuracy achieved the perfect situation. It also gives patients suggestions that they should pay more attention to the healthcare of the four physical indicators to prevent the hepatitis becoming more serious.

## 5. Conclusion

Our research project aims at finding influential features in hepatitis diagnosis NN model. A 3-layer MLNN model is built and trained by BP algorithm. Through the adjustment of hidden layer number and parameters of the model. We firstly get a high accuracy classification model. Based on this model, we have done a series of experiments trying to find the influential features. After all experiments have been done, we find that there is no single feature that individually affects the result while using the four features anorexia, liver big, alk phosphate and protime as inputs will results in a perfect diagnosis accuracy. Hence, we conclude that these joint four features are key features and influence the classification result most. However, due to that the size of the data set is too small and the best classification algorithm has not been chosen, the results of these experiments may not be accurate enough. In spite of these limitations, we still can conclude that this project has significance in showing the feasibility of doing classification using Neural Networks. It reveals the influential features in hepatitis diagnosis which provides suggestion to the hepatitis patients in preventing hepatitis becoming more serious by paying more attention to the healthcare of this four physical indicators. Moreover, it will simplify the necessary medical examination without losing the diagnosis accuracy. In general, it has realistic meaning in benefiting both the patients and the doctors.

## References

Ansari, S., Shafi, I., Ansari, A., Ahmad, J., & Shah, S. I. (2011). Diagnosis of liver disease induced by hepatitis virus using artificial neural networks. Proceedings of the 2011 IEEE 14th International Multitopic Conference (INMIC), 8-12.

Bascil, M. S., & Temurtas, F. (2011). A study on hepatitis disease diagnosis using multilayer neural network with Levenberg Marquardt training algorithm. Journal of Medical Systems, 35(3), 433-436.

Çetin, O., Temurtaş, F., & Gülgönül, Ş. (2015). An application of multilayer neural network on hepatitis disease diagnosis using approximations of sigmoid activation function. Dicle Tıp Dergisi, 42(2), 150-157.

Gori, M., & Tesi, A. (1992). On the problem of local minima in backpropagation. IEEE Transactions on Pattern Analysis & Machine Intelligence, (1), 76-86.

Ozyilmaz, L., & Yildirim, T. (2003). Artificial neural networks for diagnosis of hepatitis disease. Proceedings of the 2003 IEEE International Joint Conference on Neural Networks, 1, 586-589.