Hepatitis Diagnosis by Training of an 
MLP Artificial Neural Network 

Sara Moein
Computer Engineering Department, Azad University of Najafabad
E-mail: smoein@iaun.ac.ir

Abstract

In this paper we discuss using artificial neural network for Hepatitis diagnosis. In particular we show how our method is accurate in diagnosis by accurate attribute in dataset. A MLP network trained for this approach and handled a dataset according to symptoms and disease features. We compared how changing of some measures improved the final decision of our system in medical diagnosis.

Keywords: Medical diagnosis, Training Cycle, Artificial Neural Network.

1. Introduction

Nowadays medical diagnosis is an art. Why we say this sentence? Because there are many aspects from different range of symptoms per each disease to experience and proficiency of physicians, that should be considered for an appropriate diagnosis. While we have many diseases form different parts of body. So a huge dataset with variety of features has a significant impact in this area.

We need to use an artificial brain for storing and retrieving all symptoms of diseases; of course we should mix proficiency with content of that brain to reach to the best diagnosis. As a result, artificial neural network with an acceptable range of capabilities has been used for medical diagnosis.

2. Medical Diagnosis

The major problem in medical field is to diagnose disease. Human being always make mistake and because of their limitation, diagnosis would give the major issue of human expertise. One of the most important problems of medical diagnosis, in general, is the subjectivity of the specialist. It can be noted, in particular in Pattern recognition activities, that the experience of the professional is closely related to the final diagnosis. This is due to the fact that the result does not depend on a systematized solution but on the interpretation of the patient's signal [3]. Typical diagnostic process is the following. During the interview of the patient the anamnesis data is obtained and immediately afterwards during the preliminary examination of the patient the physician records the status data. Depending on the anamnesis and the status data, the patient takes additional laboratory examinations. The diagnosis is then determined by the physician who takes into account the whole available description of the patient's state of health. Depending on the diagnosis the treatment is prescribed and after the treatment the whole process may be repeated. In each iteration the diagnosis may be confirmed, refined or rejected. [5]

2. MLP Medical diagnosis

What the most challenging part was for us about MLP is its high ability in classifying data related to patients. The first stage for using this algorithm is choosing the best amount for primary parameters. It is clear that appropriate amounts are so effective for faster convergence. Wrong initializing for parameters leads to trapping at start of training. So a competitive effort has been done for initializing of parameters. Providing a good dataset for training is another circumstance. Choosing acceptable and real data is vital, because system has to work with body health. A dataset based on continuous range of numbers is the other issue.

Table 1. shows the used dataset. This dataset provided symptoms of some diseases for 80 patients. Below is a short description about the table's rows and columns.

<table>
<thead>
<tr>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of instance: 80</td>
</tr>
<tr>
<td>Number of attributes: (including the class attribute)</td>
</tr>
<tr>
<td>Attribute information:</td>
</tr>
</tbody>
</table>
1. Class: DIE, LIVE
2. Age: 10, 20, 30, 40, 50, 60, 70, 80
3. Sex: MALE, FEMALE
4. Steroid: NO, YES
5. Antivirals: NO, YES
6. Fatigue: NO, YES
7. Malaise: NO, YES
8. Anorexia: NO, YES
9. Liver big: NO, YES
10. Liver firm: NO, YES
11. Spleen Palpable: NO, YES
12. Spiders: NO, YES
13. Ascites: NO, YES
14. Varices: NO, YES
15. Bilirubin: 0.39, 0.80, 1.20, 2.00, 3.00, 4.00
16. Alk Phosphat: 33, 80, 120, 160, 200, 250
17. Sgot: 13, 100, 160, 200, 250
18. Albumin: 2.1, 3.0, 3.8, 4.5, 5.0, 6.0
19. Protime: 10, 20, 30, 40, 50, 60, 70, 80, 90
20. Histology: NO, YES

Table 1. Parts of Dataset for Hepatitis Patients

<table>
<thead>
<tr>
<th>#</th>
<th>Class</th>
<th>Age</th>
<th>Sex</th>
<th>Steroid</th>
<th>Antivirals</th>
<th>Fatigue</th>
<th>Malaise</th>
<th>Anorexia</th>
<th>Liver big</th>
<th>Liver firm</th>
<th>Spleen Palpable</th>
<th>Spiders</th>
<th>Ascites</th>
<th>Varices</th>
<th>Bilirubin</th>
<th>Alk Phosphat</th>
<th>Sgot</th>
<th>Albumin</th>
<th>Protime</th>
<th>Histology</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>DIE</td>
<td>10</td>
<td>MALE</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>0.39</td>
<td>33</td>
<td>13</td>
<td>2.1</td>
<td>10</td>
<td>NO</td>
</tr>
<tr>
<td>2</td>
<td>LIVE</td>
<td>20</td>
<td>MALE</td>
<td>YES</td>
<td>YES</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>0.80</td>
<td>80</td>
<td>100</td>
<td>3.0</td>
<td>20</td>
<td>YES</td>
</tr>
<tr>
<td>3</td>
<td>DIE</td>
<td>30</td>
<td>MALE</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>1.20</td>
<td>120</td>
<td>160</td>
<td>3.8</td>
<td>30</td>
<td>NO</td>
</tr>
<tr>
<td>4</td>
<td>LIVE</td>
<td>40</td>
<td>MALE</td>
<td>YES</td>
<td>YES</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>2.00</td>
<td>160</td>
<td>200</td>
<td>4.5</td>
<td>40</td>
<td>YES</td>
</tr>
<tr>
<td>5</td>
<td>DIE</td>
<td>50</td>
<td>MALE</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>3.00</td>
<td>200</td>
<td>250</td>
<td>5.0</td>
<td>50</td>
<td>NO</td>
</tr>
<tr>
<td>6</td>
<td>LIVE</td>
<td>60</td>
<td>MALE</td>
<td>YES</td>
<td>YES</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>4.00</td>
<td>250</td>
<td>300</td>
<td>6.0</td>
<td>60</td>
<td>YES</td>
</tr>
<tr>
<td>7</td>
<td>DIE</td>
<td>70</td>
<td>MALE</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>5.00</td>
<td>300</td>
<td>400</td>
<td>7.0</td>
<td>70</td>
<td>NO</td>
</tr>
<tr>
<td>8</td>
<td>LIVE</td>
<td>80</td>
<td>MALE</td>
<td>YES</td>
<td>YES</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>6.00</td>
<td>400</td>
<td>500</td>
<td>8.0</td>
<td>80</td>
<td>YES</td>
</tr>
</tbody>
</table>

3. Network training

A MLP is the particular network for our research. So we trained the network based on the present dataset. Training has done with 1 node in output layer. Netlab is suitable software for training of this dataset. Fig1 to Fig5 shows the result of this training by balancing between two significant measures for training: number of training cycles and number of hidden layer node. In this level, the number of epoch’s impact has been examined. The more number of epochs, the more processing time needed and the better result calculated. On the other hand hidden layer nodes have some effects on accuracy. So trade off between two important measures is to determine number of hidden layer nodes and number of cycles.

6 statuses include of figures give us the chance to discuss about this matter. 6 different states are:

1. Training cycle = 150, hidden layer node = 30
2. Training cycle = 500, hidden layer node = 30
3. Training cycle = 150, hidden layer node = 150
4. Training cycle = 500, hidden layer node = 15
5. Training cycle = 150, hidden layer node = 5
6. Training cycle = 500, hidden layer node = 5

Again we continue our examining with more changes in hidden nodes in output layer and changing in number
of Training cycles. The result has shown in Fig 3, Fig 4 and Fig 5.

Fig 3. Training cycle = 150, hidden layer node = 150

Fig 4. Training cycle = 500, hidden layer node = 30

Fig 5. Training cycle = 150, hidden layer node = 5

What we see is what we need. The error rate is less than 5%. Something acceptable for physicians and intelligent system designers, too. In the other effort, we did the leave one out technique. As before, we used 16 records as testing set and rest of records used for training set. Another confirmation to using neural network for medical diagnosis. See Fig 6, Fig 7, Fig 8.

Fig 6. Stage 1 in leave one out technique

Fig 7. Stage 2 in leave one out technique

Fig 8. Stage 3 in leave one out technique
4. Analyze

That is something great. 18 out of 20 patients have been diagnosed exactly like the physician diagnosis. 1 out of 20 has an error less than 50% and 1 out of 20 has error more than 50%. We can increase the number of training cycles to 1000. Not a better result occurred. But more dataset proved that with increasing the number of training cycles the error becomes less. Totally we can say that the system can do medical diagnosis with more than 90% accuracy.

5. Conclusion

In this paper we tried to offer a system for medical diagnosis, but without concerning about how to calculate the best membership function. Actually neural network solved this problem for us and by trained network we can diagnose disease. For future works we can concentrate on this problem that how we can train our neural network to get the best result for output targets. By using more powerful networks like recursive networks we can improve our system.

References