A Neural Network Approach For Heart Disease Prediction

Tanu Verma¹, R. K. Srivastava²
¹ Mewar University
² Dept. of Computer Science, Bareilly College Bareilly, UP, India

Abstract
Artificial Neural Network for intelligent medical diagnoses has been shown to be an interesting topic. In this paper we have used a neural network technique to predict a system, which can detect heart disease from their physical symptoms. We have used a prediction method with the help of the back propagation algorithm to train the networks. The actual procedure of medical diagnosis that usually is employed by physicians was analyzed and converted to a computer program implemented format. After selecting some symptoms a data set contains the information of fifty two cases was configured and applied to a M L P neural network.

Keywords: artificial neural network, back propagation algorithm

I. Introduction
Neural Network have been used in the field of artificial intelligence, preferred as these use the relation of dependency, a generation of function. The model and language of neural networks are more mathematically formulated hence most of the doctors avoid to use neural networks for prediction of disease. A prediction of heart disease system usually starts with patients complaints and the physician learn more about the patients situations interactively during the diagnosis as well as by measuring some metrics such as blood pressure, hemoglobin, s. urea, s. creatin, FBS(mg/d), PPb(mg/d), RBS etc. The quantity of examples is playing an important role for training purpose but examples need to be selected carefully for the reliability and efficient system. I have considered the indirect relation of various parameters for data and presume that their relations are time invariant in view of the patient pathological test reports. The proposed method is more efficient and provide better forecast. The forecasted values obtained through the back propagation algorithm process have been compared with the observed productivity and their robustness has been examined.

II. Artificial Neural Network
Back Propagation through time is a powerful tool of artificial neural network with application to many areas as pattern recognition, dynamic modeling and nonlinear systems. Back propagation algorithm(BPA) provides an efficient way to calculate the gradient of the error function using chain rule of differentiation. The error after initial computation in the forward pass is propagated backward from the output units, layer by layer. BPA, a generalized Delta rule is commonly used algorithm for supervised training of multi layer feed forward artificial neural network. In supervised learning, we try to adapt an artificial network so that the actual outputs (Ŷ) come close to some target outputs(Y) for a training set, which contains T patterns. The goal is to adapt the parameters of network so that it performs well for pattern from outside the training set.

2.1 Back propagation Algorithm : We have proposed a neural network that will combine the features of multi perceptron concept of both feed forward part of back propagation algorithm and Let the training set be \{x(k),d(k)\} \( k = 1 \). Where x(k) is the input pattern vector to the network and d(k) is the desired output vector for the input pattern x(k).The output of the jth output unit is denoted by \( y_j \), connections weights from the ith unit in one layer to the jth unit in the layer above are denoted by \( w_{ij} \). If \( m \) be the no. of output units and \( d_j \) is the desired output from the jth output unit whose actual output in response to the kth input exemplar \( x(k) \) is \( y_j \), for \( j = 1, 2, 3, \ldots m \). The sum of squares of the error over all the output unit for this kth exemplar by

\[
E(k) = \frac{1}{2} \sum_{j=1}^{m} [y_j(k)-d_j]^2
\]

Error \( E(k) \) is affected by the output from unit \( j \) at the output layer and is determined by

\[
\frac{\partial E(k)}{\partial y_j} = y_j - d_j
\]

The net input to output layer is of the form

\[
S_j = \sum_{i} w_{y_j} \in_{y_{ij}} y_i^{(1)}
\]

Where \( y_i^{(1)} \) is the output from the \( i^{th} \) unit in the first layer below the output layer, \( w_{ij} \) is connection weight multiplying \( y_i^{(1)} \) and \( \in_{y_{ij}} \) is the threshold of unit \( j \). The negative of threshold is defined to be the bias.
2.2 Computer code in C Language to trained the Network: Let us consider the components x[0][0] to x[ntmax][ni] so that there are “ntmax * ni” inputs to the network and y[0] to y[ntmax] outputs. The value “ntmax * ni” decides how many neurons in the network, “net” represents the total level of existing a neuron and y[nt][k] represents the intensity of resulting output from the neuron or activation level. We assume the full range of allowed connections, simply for the sake of generality.

```c
#include<stdio.h>
#include<math.h>
#include<conio.h>

void main()
{
  double
  x[52][13],h[13][13],y[13],yd[13],dy[13][13],net,whi[13][13],woh[13];
  double
  xmax[700],xmin[700],ydmax[13],ydmin[13],e,dnet,dw
  int ni=13,ntmax=52,ncmax,nh,i,j,k,nt,nc;
  float
  eta;
  FILE *f,*ee,*o;
  clrscr();
  printf("type no of iterations");
  scanf("%d",&ncmax);
  printf("type no. of hidden neurons");
  scanf("%d",&nh);
  printf("type value of learning rate");
  scanf("%f",&eta);
  for(i=0;i<=ni;++i)
    for(j=0;j<=nh;++j)
      whi[i][j]=0;
  for(i=0;i<=ni;++i)
    woh[i]=0;
  f=fopen("data1","r");
  o=fopen("out","w");
  ee=fopen("error","w");
  for(nt=0;nt<=ntmax;++nt)
    fscanf(f,"%ld",&yd[nt]);
  for(nt=0;nt<=ntmax;++nt)
    for(i=0;i<=ni;++i)
      fscanf(f,"%ld",&x[nt][i]);
  for(nc=0;nc<=ncmax;++nc)
    {for(nt=0;nt<=ntmax;++nt)
      net=0.0;
      for(i=0;i<=ni;++i)
        for(j=0;j<=nh;++j)
          net+=whi[i][j]*x[nt][i];
      h[nt][j]=1.0/(1.0+exp(-net));
    }for(nt=0;nt<=ntmax;++nt)
    net=0.0;
    for(i=0;i<=ni;++i)
      net+=whi[i][j]*h[nt][j];
  y[nt]=1.0/(1.0+exp(-net));
  e=0.0;
  for(nt=0;nt<=ntmax;++nt)
    {dnet=y[nt]-yd[nt];
      e+=0.5*pow(dnet,2);
      dy[nt]=dnet*y[nt]*(1.0-y[nt]);
    }
  printf("nc=%d,e=......%f",nc,e);
  fprintf(ee,"%d%f
",nc,e);
  for(nt=0;nt<=ntmax;++nt)
    for(j=0;j<=nh;++j)
      dnet=0.0;
      dnet+=dy[nt]*woh[j];
      dh[nt][j]=dnet*h[nt][j]*(1.0-h[nt][j]);
    }
  for(i=0;i<=nh;++i)
    woh[i]=woh[i]+dy[nt]*woh[i];
  for(j=0;j<=nh;++j)
    dh[nt][j]=dh[nt][j]+h[nt][j]*dy[nt];
  for(i=0;i<=ni;i++)
    for(j=0;j<=nh;j++)
      whi[i][j]=eta*dh[nt][j];
  for(j=0;j<=nh;++j)
    woh[j]=eta*woh[j];
  fclose(f);
  getch();
}

III. Computation of ANN forecasted value

Considering the patient pathological and other test reports related to heart disease as input x(k) and particular S. No. of patient data set to be predicted as desired output d(k) after applying the BPA, the calculated output is treated as predicted value of cor. Angio. The steps adapted for calculation of predicted cor. Angio. value i.e. output through BPA is as follows:
Step 1: Data set of patient S. No. 1 to 46 as input set and Coronary Angiogram value of patient S. No. 46 as desired output.
Step 2: Data set of patient S. No. 1 to 47 as input set and Coronary Angiogram value of patient S. No. 47 as desired output.
Step 3: Data set of patient S. No. 1 to 48 as input set and Coronary Angiogram value of patient S. No. 48 as desired output.
Step 4: Data set of patient S. No. 1 to 49 as input set and Coronary Angiogram value of patient S. No. 49 as desired output.
Step 5: Data set of patient S. No. 1 to 50 as input set and Coronary Angiogram value of patient S. No. 50 as desired output.
Step 6: Data set of patient S. No. 1 to 51 as input set and Coronary Angiogram value of patient S. No. 51 as desired output.
Step 7: Data set of patient S. No. 1 to 52 as input set and Coronary Angiogram value of patient S. No. 52 as desired output.

The algorithm has been implemented through C programming language, considering two hidden layers and computations have been made by various iterations levels like: 100, 200, 500 & 1000. Out of these, the best suitable forecasted values have been obtained by model with 1000 iterations. The result so obtained has been illustrated in Table 1 as follows:

### Table 1

<table>
<thead>
<tr>
<th>S.No</th>
<th>NAME</th>
<th>AGE (Yrs)</th>
<th>Hb (gm/dl)</th>
<th>S.UR EA (mg/dl)</th>
<th>S.CR EATI N (gm/dl)</th>
<th>FBS (mg/dl)</th>
<th>PPBS (mg/dl)</th>
<th>RBS (mg/dl)</th>
<th>TG (mg/dl)</th>
<th>S.CHOLE S (mg/dl)</th>
<th>LDL (mg/dl)</th>
<th>VLDL (mg/dl)</th>
<th>HDL (mg/dl)</th>
<th>FASTING INSULIN (mU/l)</th>
<th>I.R. (mmol/l)</th>
<th>COR. ANG. (O.E.)</th>
<th>ANN. Forecasted Cor. Angiogram value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>H.C. YADAV</td>
<td>63</td>
<td>12.1</td>
<td>33</td>
<td>1.2</td>
<td>93.85</td>
<td>134.5</td>
<td>158</td>
<td>122</td>
<td>206</td>
<td>130</td>
<td>26</td>
<td>50</td>
<td>7.34</td>
<td>1.3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>B.M PANDEY</td>
<td>65</td>
<td>12</td>
<td>20</td>
<td>0.8</td>
<td>72.28</td>
<td>118.6</td>
<td>132</td>
<td>117.14</td>
<td>145.5</td>
<td>83.23</td>
<td>23.43</td>
<td>38.9</td>
<td>12.2</td>
<td>2.18</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>K.P YADAV LALMANI SAROJ KAVITA PURVAR</td>
<td>54</td>
<td>13</td>
<td>24</td>
<td>1.12</td>
<td>76.4</td>
<td>114.8</td>
<td>148</td>
<td>114.6</td>
<td>138</td>
<td>84</td>
<td>22</td>
<td>32</td>
<td>10.2</td>
<td>1.92</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>55</td>
<td>12</td>
<td>24</td>
<td>1.23</td>
<td>70.2</td>
<td>116.8</td>
<td>98</td>
<td>117</td>
<td>136</td>
<td>76</td>
<td>24</td>
<td>36</td>
<td>8.13</td>
<td>1.4</td>
<td>3</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>38</td>
<td>12</td>
<td>20.9</td>
<td>0.98</td>
<td>74.81</td>
<td>136.1</td>
<td>124</td>
<td>115.75</td>
<td>124.97</td>
<td>71.03</td>
<td>23.15</td>
<td>30.79</td>
<td>10.1</td>
<td>1.86</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>KALI PRASAD</td>
<td>62</td>
<td>13.7</td>
<td>23</td>
<td>1.27</td>
<td>92.25</td>
<td>124.6</td>
<td>118</td>
<td>121.34</td>
<td>124.41</td>
<td>76.42</td>
<td>24.27</td>
<td>23.72</td>
<td>16</td>
<td>3.64</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>RAIPAT</td>
<td>60</td>
<td>11.4</td>
<td>24</td>
<td>1.1</td>
<td>68.69</td>
<td>94.7</td>
<td>96</td>
<td>105.45</td>
<td>103.5</td>
<td>48.21</td>
<td>21.09</td>
<td>34.2</td>
<td>2.07</td>
<td>0.35</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>SI KANDEV</td>
<td>60</td>
<td>8.9</td>
<td>33.6</td>
<td>1.37</td>
<td>69.4</td>
<td>126.1</td>
<td>103</td>
<td>73.76</td>
<td>89.76</td>
<td>23.51</td>
<td>14.75</td>
<td>51.5</td>
<td>2.12</td>
<td>0.36</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>RAIDEV</td>
<td>45</td>
<td>13.9</td>
<td>41.3</td>
<td>1.13</td>
<td>78.87</td>
<td>142.4</td>
<td>134</td>
<td>152.1</td>
<td>163.2</td>
<td>82.58</td>
<td>50.2</td>
<td>50.2</td>
<td>13.9</td>
<td>2.7</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>V.D. SINGH</td>
<td>65</td>
<td>12.4</td>
<td>25.4</td>
<td>1.48</td>
<td>88.03</td>
<td>116.7</td>
<td>128</td>
<td>95.68</td>
<td>155.6</td>
<td>34.56</td>
<td>19.14</td>
<td>51.9</td>
<td>22.2</td>
<td>4.8</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>C.B. SINGH LALMANNI</td>
<td>57</td>
<td>13</td>
<td>27.7</td>
<td>1.4</td>
<td>80.2</td>
<td>136.6</td>
<td>150</td>
<td>96</td>
<td>248.5</td>
<td>180.2</td>
<td>19.26</td>
<td>49</td>
<td>39.3</td>
<td>7.78</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>61</td>
<td>10.4</td>
<td>31.1</td>
<td>1.24</td>
<td>98.86</td>
<td>126.7</td>
<td>142</td>
<td>86.65</td>
<td>118.26</td>
<td>49.63</td>
<td>17.33</td>
<td>51.3</td>
<td>14.2</td>
<td>3.46</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>CHAMPA DEVI</td>
<td>50</td>
<td>11.5</td>
<td>29</td>
<td>1.7</td>
<td>78</td>
<td>97.85</td>
<td>132</td>
<td>114</td>
<td>160</td>
<td>104</td>
<td>16</td>
<td>40</td>
<td>2</td>
<td>0.38</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>RAM DEV</td>
<td>50</td>
<td>15</td>
<td>30.5</td>
<td>1.37</td>
<td>84.6</td>
<td>116.6</td>
<td>102</td>
<td>115.9</td>
<td>261.2</td>
<td>191.1</td>
<td>23.18</td>
<td>46.9</td>
<td>13.9</td>
<td>2.9</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>HARISHANKER CHHOTAK SINGH BANKEYLAL YADAV</td>
<td>57</td>
<td>13.1</td>
<td>28.4</td>
<td>1.23</td>
<td>88.68</td>
<td>121.4</td>
<td>90</td>
<td>125.7</td>
<td>138</td>
<td>73</td>
<td>25</td>
<td>40</td>
<td>4.5</td>
<td>0.98</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>65</td>
<td>13.9</td>
<td>42.5</td>
<td>1.9</td>
<td>72.2</td>
<td>120.7</td>
<td>157</td>
<td>132.3</td>
<td>182.6</td>
<td>112.8</td>
<td>26.46</td>
<td>43.4</td>
<td>10.4</td>
<td>1.85</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>61</td>
<td>7.6</td>
<td>44.9</td>
<td>1.11</td>
<td>76.6</td>
<td>114.9</td>
<td>72.6</td>
<td>80.6</td>
<td>116.7</td>
<td>54.38</td>
<td>16.12</td>
<td>46.2</td>
<td>11.6</td>
<td>2.19</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>B.M PANDEY</td>
<td>40</td>
<td>12</td>
<td>27</td>
<td>89</td>
<td>79.8</td>
<td>119.8</td>
<td>82</td>
<td>167.9</td>
<td>175.8</td>
<td>117</td>
<td>31</td>
<td>46</td>
<td>13.8</td>
<td>2.6</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>K.P YADAV LALMANI SAROJ</td>
<td>64</td>
<td>8.9</td>
<td>31.4</td>
<td>0.93</td>
<td>80.6</td>
<td>134.1</td>
<td>81.9</td>
<td>150.95</td>
<td>138.75</td>
<td>59.59</td>
<td>30.19</td>
<td>42.2</td>
<td>8.6</td>
<td>1.71</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>65</td>
<td>6.8</td>
<td>21.9</td>
<td>0.9</td>
<td>74.8</td>
<td>125.8</td>
<td>84.6</td>
<td>99.65</td>
<td>119.87</td>
<td>63.14</td>
<td>19.03</td>
<td>36.14</td>
<td>10.4</td>
<td>1.92</td>
<td>3</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Name</td>
<td>1</td>
<td>10</td>
<td>20</td>
<td>30</td>
<td>40</td>
<td>50</td>
<td>60</td>
<td>70</td>
<td>80</td>
<td>90</td>
<td>100</td>
<td>110</td>
<td>120</td>
<td>130</td>
<td>140</td>
<td>150</td>
<td>160</td>
</tr>
<tr>
<td>--------------</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>Name</td>
<td>1</td>
<td>10</td>
<td>20</td>
<td>30</td>
<td>40</td>
<td>50</td>
<td>60</td>
<td>70</td>
<td>80</td>
<td>90</td>
<td>100</td>
<td>110</td>
<td>120</td>
<td>130</td>
<td>140</td>
<td>150</td>
<td>160</td>
</tr>
</tbody>
</table>
IV. Result & Conclusion
The proposed artificial neural network technique has been implemented to have diagnosis of heart disease. We have considered the indirect relation of various parameters for time series data and presume that their relations are time invariant. The motivation of the study is to diagnose the heart disease, data are collected through various sampling techniques and obtained the diagnostic result through ANN using back propagation algorithm. A network requires information only on the input variable for diagnostic system. As values on test data are comparatively less the diagnostic model is reliable. The availability of data have tremendous amount of imprecision and uncertainty due to test reports of the patients are based on involvement of different electronic and mechanical equipments. Network performance could have been further improved by providing more training data. Moreover the considered connectionist models are robust, capable of handling the approximate data and therefore should be more reliable in worse situation. Optimal result will depend on the selection of parameters which is based on test results and symptoms of the patients. It is evident through the study that neural network model is even suitable over human diagnostic system

Heart disease prediction is of much use to the heart patient having pathological test data tending towards expected position of heart to avoid the heart attack. The motivation of the study is that the pathological test data are collected through various sampling and based on the reading of electronic machines, involving vagueness. Comparison of

<table>
<thead>
<tr>
<th>S No.</th>
<th>Name</th>
<th>COR. ANGIO.</th>
<th>ANN Forecasted COR. ANGIO.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>GUDDU YADAV</td>
<td>1</td>
<td>0.9865</td>
</tr>
<tr>
<td>2</td>
<td>SAUGAR LAL</td>
<td>2</td>
<td>2.2376</td>
</tr>
<tr>
<td>3</td>
<td>NISHATH FATIMA</td>
<td>0</td>
<td>0.0083</td>
</tr>
<tr>
<td>4</td>
<td>ABDUL SALAAM</td>
<td>0</td>
<td>0.0063</td>
</tr>
<tr>
<td>5</td>
<td>RAM SUMAN</td>
<td>1</td>
<td>1.1765</td>
</tr>
<tr>
<td>6</td>
<td>RAM RAJ</td>
<td>1</td>
<td>1.0925</td>
</tr>
<tr>
<td>7</td>
<td>KAZIM</td>
<td>2</td>
<td>1.9742</td>
</tr>
</tbody>
</table>

Table-2

Comparison of

<table>
<thead>
<tr>
<th>Patient's S. No.</th>
<th>Cor.Angio. Values</th>
<th>ANN Forecasted COR. ANGIO.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8.4</td>
<td>0.9865</td>
</tr>
<tr>
<td>2</td>
<td>12</td>
<td>2.2376</td>
</tr>
<tr>
<td>3</td>
<td>157</td>
<td>0.0083</td>
</tr>
<tr>
<td>4</td>
<td>13</td>
<td>1.1765</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>0.0063</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>1.0925</td>
</tr>
<tr>
<td>7</td>
<td>6</td>
<td>1.9742</td>
</tr>
</tbody>
</table>

Fig.- Actual Cor.Angio. Vs Forecasted Cor.Angio. Values
In the study the target patient S.NO. 46 to 52 for the prediction of coronary angiogram value computed through the ANN method are quite impressive by comparison with actual value. Further, the computations shows that the predicted data through ANN method provides much better, suitable and reliable forecast for the heart disease patients.

References
[2] Ferdous Ahmed Sohel, A new Neural Network with Fuzzy Technique, Disease diagnosis, A Case study,Dept. of Computer Science & Engg.,International Islamic University-Dhaka
[6] S Moein,a Monadjemi[2006],A Novel Fuzzy-Neural Based Medical Diagnosis System,volume 1, Number 3