Artificial Neural Network for Prediction of Distant Metastasis in Colorectal Cancer

Akbar Biglarian1*, Enayatollah Bakhshi1, Mahmood Reza Gohari2, Reza Khodabakhshi3

Abstract

Background and Objectives: Artificial neural networks (ANNs) are flexible and nonlinear models which can be used by clinical oncologists in medical research as decision making tools. This study aimed to predict distant metastasis (DM) of colorectal cancer (CRC) patients using an ANN model. Methods: The data of this study were gathered from 1219 registered CRC patients at the Research Center for Gastroenterology and Liver Disease of Shahid Beheshti University of Medical Sciences, Tehran, Iran (January 2002 and October 2007). For prediction of DM in CRC patients, neural network (NN) and logistic regression (LR) models were used. Then, the concordance index (C index) and the area under receiver operating characteristic curve (AUROC) were used for comparison of neural network and logistic regression models. Data analysis was performed with R 2.14.1 software. Results: The C indices of ANN and LR models for colon cancer data were calculated to be 0.812 and 0.779, respectively. Based on testing dataset, the AUROC for ANN and LR models were 0.82 and 0.77, respectively. This means that the accuracy of ANN prediction was better than for LR prediction. Conclusion: The ANN model is a suitable method for predicting DM and in that case is suggested as a good classifier that usefulness to treatment goals.

Keywords: Colorectal cancer - artificial neural network - distant metastasis - prediction - concordance index
Materials and Methods

In this study, we analyzed the data from 1219 patients with CRC were been collected by cancer registry of the Research Center for Gastroenterology and Liver Disease of Shahid Beheshti University of Medical Sciences, Tehran, Iran (Asghari et al., 2009). At first we dropped out those patients with lower than one-month or higher than six years survival time and also who were deaths for other causes. Accordingly, a total of 1007 patients (786 colon cancer and 204 rectum cancer patients) were entered in the study. Observed distant metastasis (metastasis to one organ or site and metastasis in more than one organ/site or the peritoneum) over the follow-up was considered as the outcome variable. The covariates for this outcome was consisted of age at diagnosis (year), sex (female/male), ethnicity (Persian/Kurd/Azeri/Lur/other), marital status (married/other), high risk behavior (i.e. tobacco smoking, or alcohol history, or opium, or IV drug user, or betel use), pathologic stage grouping (primary/advanced), and first treatment (surgery/biopsy/chemotherapy/radiotherapy).

To predict DM by LR, the back-ward stepwise selection method was used to model building based on main effect and all possible interaction terms and the p-value less than 0.05 was considered significant.

In the ANN strategy, at first the data was divided into two subsets: training/learning (70%) and testing/validation (30%) subset. Model building process was made on training dataset based on multilayer perceptron (MLP). Afterward, the model is validated by testing dataset. In this context, the areas under receiver operation characteristic curve (AUROC) and concordance index (C index) were used for comparing the prediction ability of the described models. It is mentioned that the C index estimates the probability of concordance/agreement between predicted and observed responses. Note that, in fitting ANN model we used a three-layer MLP network with 7 variables in input layer, 5 to 16 nodes in middle layer and one node in output layer, the sigmoid transfer function in middle and output layers, a back-propagation learning algorithm.

Table 1. General Characteristics of the Patients with Colorectal Cancer

<table>
<thead>
<tr>
<th>Variable</th>
<th>Rectum Cancer</th>
<th>Colon Cancer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>Male</td>
<td>Female</td>
</tr>
<tr>
<td></td>
<td>128</td>
<td>76</td>
</tr>
<tr>
<td></td>
<td>62.7%</td>
<td>37.3%</td>
</tr>
<tr>
<td></td>
<td>484</td>
<td>302</td>
</tr>
<tr>
<td></td>
<td>61.6%</td>
<td>38.4%</td>
</tr>
<tr>
<td>Marital status</td>
<td>Married</td>
<td>Other</td>
</tr>
<tr>
<td></td>
<td>185</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>92.5%</td>
<td>7.5%</td>
</tr>
<tr>
<td></td>
<td>699</td>
<td>87</td>
</tr>
<tr>
<td></td>
<td>92.8%</td>
<td>7.2%</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>Persian</td>
<td>Azeri</td>
</tr>
<tr>
<td></td>
<td>95</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>48.2%</td>
<td>17.3%</td>
</tr>
<tr>
<td></td>
<td>397</td>
<td>171</td>
</tr>
<tr>
<td></td>
<td>52.6%</td>
<td>22.6%</td>
</tr>
<tr>
<td></td>
<td>Kurd</td>
<td>Lur</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>9.1%</td>
<td>12.7%</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>6.9%</td>
<td>6.9%</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td></td>
</tr>
<tr>
<td></td>
<td>25</td>
<td></td>
</tr>
<tr>
<td></td>
<td>12.7%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>85</td>
<td></td>
</tr>
<tr>
<td></td>
<td>11.3%</td>
<td></td>
</tr>
<tr>
<td>High risk behavior</td>
<td>Has</td>
<td>Hasn’t</td>
</tr>
<tr>
<td></td>
<td>82</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>42.5%</td>
<td>57.5%</td>
</tr>
<tr>
<td></td>
<td>272</td>
<td>466</td>
</tr>
<tr>
<td></td>
<td>36.9%</td>
<td>63.1%</td>
</tr>
<tr>
<td>Type of first</td>
<td>Surgery</td>
<td>Biopsy</td>
</tr>
<tr>
<td>treatment</td>
<td>187</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>91.7%</td>
<td>3.4%</td>
</tr>
<tr>
<td></td>
<td>576</td>
<td>108</td>
</tr>
<tr>
<td></td>
<td>73.6%</td>
<td>13.8%</td>
</tr>
<tr>
<td>Pathologic stage</td>
<td>Primary</td>
<td>Advanced</td>
</tr>
<tr>
<td></td>
<td>101</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td>51.5%</td>
<td>48.5%</td>
</tr>
<tr>
<td></td>
<td>282</td>
<td>387</td>
</tr>
<tr>
<td></td>
<td>42.2%</td>
<td>57.8%</td>
</tr>
</tbody>
</table>

Table 2. ANN and LR Modeling Results to Determine The Important Factors on DM in CRC

<table>
<thead>
<tr>
<th>Ordered variables</th>
<th>ANN model</th>
<th>LR model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pathologic stage</td>
<td>0.469</td>
<td>0.616</td>
</tr>
<tr>
<td>First Treatment</td>
<td>0.250</td>
<td>0.106</td>
</tr>
<tr>
<td>Sex</td>
<td>0.126</td>
<td>0.092</td>
</tr>
<tr>
<td>Age at diagnosis</td>
<td>0.067</td>
<td>0.073</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>0.061</td>
<td>0.049</td>
</tr>
<tr>
<td>Marital status</td>
<td>0.019</td>
<td>0.038</td>
</tr>
<tr>
<td>High risk behavior</td>
<td>0.008</td>
<td>0.026</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ordered variables</th>
<th>Colon</th>
<th>P_value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pathologic stage</td>
<td>First Treatment</td>
<td>0.008</td>
</tr>
<tr>
<td>First Treatment</td>
<td>Sex</td>
<td>0.129</td>
</tr>
<tr>
<td>Sex</td>
<td>Pathologic stage</td>
<td>0.358</td>
</tr>
<tr>
<td>Age at diagnosis</td>
<td>Ethnicity</td>
<td>0.511</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>Age at diagnosis</td>
<td>0.623</td>
</tr>
<tr>
<td>Marital status</td>
<td>Marital status</td>
<td>0.727</td>
</tr>
<tr>
<td>High risk behavior</td>
<td>High risk behavior</td>
<td>0.757</td>
</tr>
</tbody>
</table>

* For the rectum cancer data, LR model did not fit to the data

Results

Of the 204 rectum cancer patients 128 (62.7%) were men and others were women. 92.5% of these patients were married. The mean±SD of age at diagnosis (in year) for men and women were 51.47±14.3 and 53.96±12.3, respectively. In addition, the 25, 50 and 75 percentiles of age at diagnosis for men were 43, 52, and 63 year and for women were 44, 54, and 66 year, respectively. Of the 788 colon cancer patients 484 (61.6%) were men and others were women. 92.8% of these patients were married. The mean±SD of age at diagnosis for men and women with this cancer were 51.5±14.2 and 55.1±14.6, respectively. In addition, the mean±SD of age at diagnosis for patients with distant metastasis was 52.88±13.99 (49.73±15.23 for women and 54.31±13.07 for men). Most of CRC patients had a surgery as the first treatment (91.7% for rectum and 73.6% for colon). Only 42.5% of rectum and 36.9% of colon cancer patients have had at least one high risk behavior. The advanced stage of tumor for rectum and colon cancer was 48.5% and 57.8%, respectively (Table 1).

Based on validation set, the NN model was used to determine the important factors. Based on importance analysis in ANN strategy, pathologic stage grouping, first treatment, sex, age at diagnosis, ethnicity, marital status, and high risk behavior variables were determined as ordered important factors for colon cancer. First treatment, ethnicity, pathologic stage grouping, age at diagnosis, sex, high risk behavior, and marital status variables were


DOI:http://dx.doi.org/10.7314/APJCP.2012.13.3.927

Artificial Neural Network for Prediction of Distant Metastasis in Colorectal Cancer

Table 3. Classification Accuracy of ANN and LR Models for DM in Validation set of CRC

<table>
<thead>
<tr>
<th>Type of cancer</th>
<th>Metastasis</th>
<th>Observed</th>
<th>True prediction by ANN</th>
<th>True prediction by LR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colon Hasn’t</td>
<td>117</td>
<td>107 (91.4)</td>
<td>108 (92.3)</td>
<td></td>
</tr>
<tr>
<td>Has</td>
<td>37</td>
<td>18 (48.6)</td>
<td>12 (32.4)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>154</td>
<td>125 (81.2)</td>
<td>120 (77.9)</td>
<td></td>
</tr>
<tr>
<td>Rectum Hasn’t</td>
<td>28</td>
<td>24 (85.7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Has</td>
<td>9</td>
<td>4 (44.4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>37</td>
<td>28 (75.7)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* For the rectum cancer data, LR model did not fit to the data.

determined as ordered important factors for rectal cancer (Table 2).

Model vs. LR model for colon cancer data was calculated 91.4% vs. 92.3% and 48.6% vs. 32.4%, respectively. This means that, the ability of ANN and LR predictions to identify patients without DM is similar but the ability of the ANN predictions to identify patients with DM is better than LR predictions. For rectum cancer data, specificity and sensitivity of the ANN model was calculated 85.7% and 44.4%, respectively (Table 3).

Discussion

After primary therapy including main treatment and adjuvant treatment; the patients usually put under follow up schedule. Sometimes we loss the patients during the follow up; so if we have an ability to define high risk patients we could concentrate our program for detecting distant failure in proper time. Such a prediction could increase our insight in future. Researchers are studying new ways to treat metastasis cancer (Disibio and French, 2008; Talmadge and Fidler, 2010; Coghlin and Murray, 2010; National cancer Institute, 2010). Published related studies have reported the ANN prediction of lymph node metastasis was more accurate in esophageal cancers (Kan et al., 2004), gastric cancer (Bollschweiler et al., 2004), head and neck cancer (Darby et al., 2005), breast cancer (Baltzer et al., 2010). However, our findings in the present study showed that ANN strategy is more accurate than LR model to predict DM in CRC patients. It is obvious that, true prediction of DM may be improving CRC care and may be affecting the survival of the patients. In conclusion, the ANN model is suggested to predict DM in CRC patients as a suitable tool and also may possibly be applied clinically in the future.

Acknowledgements

We wish to express our special thanks of all colleagues at Research Center for Gastroenterology and Liver Disease in Shahid Beheshti University of Medical Sciences.

References


Available at http://creativecommons.org/licenses/by/2.0


