RESEARCH COMMUNICATION

Artificial Neural Network for Prediction of Distant Metastasis in Colorectal Cancer

Akbar Biglarian[1]*, Enayatollah Bakhshi[1], Mahmood Reza Gohari[2], Reza Khodabakhshi[3]

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

Introduction

Colorectal cancer (CRC) is one of the most common cancer, major malignancies and health problem worldwide (Boyle and Langman, 2000; Parkin, 2004; Moradi et al., 2009). The higher incidence rate was reported in Australia, North America and Northern and Western Europe and the lower rates were reported in Asia and particularly Africa (Ferlay et al., 2010). however, CRC incidence rate is increasing in Asia (Song et al., 2005; Missaoui et al., 2010; WHO, 2010). In Iran, CRC is third common cancer in females and fifth in males (Iranian annual national cancer registration report, 2007) and five year survival probability of CRC patients were reported 0.45 and 0.39 for women and men respectively (Moradi et al., 2009).

Distant metastasis (DM) in CRC patients is as metastasis to one organ or site and metastasis in more than one organ/site or the peritoneum (American cancer society, 2011). Main sites of metastasis of colon cancer are liver, peritoneum and lungs and for rectum cancer is liver, lungs and adrenal gland (Disibio and French, 2008; Talmadge and Fidler, 2010; Coghlin and Murray, 2010; National cancer Institute, 2010). The treatments for DM cancer are made on with systemic therapy (chemotherapy, biological therapy, targeted therapy, and hormonal therapy that is not usual), local therapy (surgery, radiation therapy) or a combination of them. In addition, 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).

Many different studies were conducted to determine the risk factors of CRC and their importance (Boyle and Langman, 2000; Giovannucci, 2002; Bingham et al., 2003; Corrao et al., 2004; American Joint Committee on Cancer, 2010; Moghimi et al., 2008; Asghari et al., 2009; Kelsall et al., 2009). The analyzing of CRC data and prediction of CRC was made on statistical methods and recently artificial neural network (ANN) (Bottaci et al., 1997; Anand et al., 1999; Grumett et al., 2003; Lee et al., 2004; Kyung-Joong and Sung-Bae, 2004; Bittern et al., 2005; Ahmed, 2005; Alladi et al., 2008; Fathy, 2011; Gohari et al., 2011). ANN models are flexible and nonlinear methods that allow better fit to the data and leads to accurate prediction (Bishop, 1997). However, DM prediction in CRC patients was not done in these studies. In the present study, the ANN model was used to predict DM of CRC patients and then its accuracy was compared to logistic regression (LR) model.

[1]*Department of Biostatistics, Pediatric Neurorehabilitation Research Center, University of Social Welfare and Rehabilitation Sciences, [2]Department of Biostatistics, Hospital Management Research Center, Tehran University of Medical Sciences, [3]Radiation Oncology Department, Fayazbakhsh Hospital, Tehran, Iran *For correspondence:abiglarian@gmail.com, abiglarian@uswr.ac.ir*

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. data were analyzed using R 2.14.1 software (available at: http://cran.r-project.org/).

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 55.11±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

Table 1. General Characteristics of the Patients with Colorectal Cancer

Variable		Rectum cancer No	%	Colon cancer No	%
Sex	Male	128	62.7	484	61.6
	Female	76	37.3	302	38.4
Marital status	Married	185	92.5	699	92.8
	Other	19	7.5	87	7.2
Ethnicity	Persian	95	48.2	397	52.6
	Azeri	34	17.3	171	22.6
	Kurd	18	9.1	52	6.9
	Lur	25	12.7	50	6.9
	Other	25	12.7	85	11.3
High risk behavior	Has	82	42.5	272	36.9
	Hasn't	11	57.5	466	63.1
Type of first treatment	Surgery	187	91.7	576	73.6
	Chemotherapy /Radiotherapy	10	4.9	99	12.6
	Biopsy	7	3.4	108	13.8
Pathologic stage	Primary	101	51.5	282	42.2
	Advanced	95	48.5	387	57.8

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

ANN model				LR model			
Colon		Rectum		Colon		Rectum*	
Ordered variables	Importance (gain value)	Ordered variables	Importance (gain value)	Ordered variables	P_value	Ordered variable	P_value
Pathologic stage	0.469	First Treatment	0.616	First Treatment	0.008	-	-
First Treatment	0.250	Ethnicity	0.106	Sex	0.129	-	-
Sex	0.126	Pathologic stage	0.092	Pathologic stage	0.358	-	-
Age at diagnosis	0.067	Age at diagnosis	0.073	Ethnicity	0.511	-	-
Ethnicity	0.061	Sex	0.049	Age at diagnosis	0.623	-	-
Marital status	0.019	High risk behavior	0.038	Marital status	0.727	-	-
High risk behavior	0.008	Marital status	0.026	High risk behavior	0.757	-	-

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

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

Type of cancer	Distant Metastasis	Observed No	True prediction by ANN	True prediction by LR
Colon	Hasn't	117	107 (91.4)	108 (92.3)
	Has	37	18 (48.6)	12 (32.4)
	Total	154	125 (81.2)	120 (77.9)
Rectum	Hasn't	28	24 (85.7)	-
	Has	9	4 (44.4)	-
	Total	37	28 (75.7)	-

* 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

References

Ahmed (2005). Artificial neural networks for diagnosis and survival prediction in colon cancer. *Mol cancer*, **4**, 29. Available at http://creativecommons.org/licenses/by/2.0

Alladi SM, Santosh S, Ravi V, et al (2008). Colon cancer prediction with genetic profiles using intelligent techniques. *Bioinformation*, **3**, 130-3.

American Cancer Society. Cancer Facts and Figures 2011. Available at http://www.cancer.org/acs/groups/content/@epidemiologysurveilance/documents/document/acspc-029771.pdf

American Joint Committee on Cancer: AJCC 7th edition (2010). Colon and Rectum Cancer Staging poster. available at www.cancerstaging.org/staging/posters/colon24x30.pdf

Anand SS, Smith AE, Hamilton PW, et al (1999). An evaluation of intelligent prognostic systems for colorectal cancer. *Artif Intell Med*, **15**, 193-214.

Asghari-Jafarabadi M, Hajizadeh E, Kazemnejad A, et al (2009). Site-specific evaluation of prognostic factors on survival in Iranian colorectal cancer patients, A competing risks survival analysis. *Asian Pac J Cancer Prev*, **1**, 815-22.

Baltzer PAT, Dietzel A, Vag T, et al (2010). Application of artificial neural networks for the prediction of lymph node metastases to the ipsilateral axilla – initial experience in 194 patients using magnetic resonance mammography. *Acta Radiol*, **51**,851-8.

Bingham SA, Day NE, Luben R, et al (2003). Dietary fiber in food and protection against colorectal cancer in the European Prospective Investigation into Cancer and Nutrition (EPIC), an observational study. *Lance*t, **361**, 1496–501.

Bishop CM, Neural Networks for pattern recognition. Oxford University Press, New York, 1997.

Bittern R, Cuschieri A, Dolgobrodov SD, et al (2005). An artificial neural network for analyzing the survival of patients with colorectal cancer. ESANN'2005 proceeding-Uropean symposium on artificial neural networks, *Bruges (Belgium)*, 27-9, ISBN 2-930307-05-6.

Bollschweiler EH, Mönig SP, Hensler K, et al (2004). Artificial neural network for prediction of lymph node metastases in gastric cancer, a phase II diagnostic study. *Ann Surg Oncol*, **11**, 506-11.

Bottaci L, Drew PJ, Hartley JE, et al (1997). Artificial neural networks applied to outcome prediction for colorectal cancer patients in separate institutions. *Lancet*, **350**, 469-72.

Boyle P, Langman JS (2000). ABC of colorectal cancer. *BMJ*, **321**, 805-8.

Coghlin C, Murray GI (2010). Current and emerging concepts in tumour metastasis. *J Pathol*, **222**, 1-15.

Corrao G, Bagnardi V, Zambon A, et al (2004). A meta-analysis of alcohol consumption and the risk of 15 diseases. *Prev Med*, **38**, 613-9.

Darby E, Nettimi T, Kodali S, et al (2005). Head and neck cancer metastasis prediction via artificial neural networks. This paper appears in, Computational Systems Bioinformatics Conference, Workshops and Poster Abstracts. IEEE: 43 – 4.

Disibio G, French SW (2008). Metastatic patterns of cancer, results from a large autopsy study. *Arch Pathol Lab Med*, **132**, 931-9.

Fathy SK (2011). A predication survival model for colorectal cancer. published in: proceeding AMERICAN-MATH'11/CEA'11 Proceedings of the 2011 American conference on applied mathematics and the 5th WSEAS international conference on Computer engineering and applications, 36-42.

Ferlay J, Parkin DM, Steliarova-Foucher E (2010). GLOBOCAN 2008, cancer incidence and mortality worldwide. IARC cancer base No. 10, Lyon: International Agency for Research on Cancer. Available at http://globocan.iarc.fr.

Giovannucci E (2002) Modifiable risk factors for colon cancer. *Gastroenterol Clin North Am*, **31**, 925-43.

Gohari MR, Biglarian A, Bakhshi E, et al (2011). Artificial neural network to determine the prognostic factors in colorectal cancer patients. *Asian Pac J Cancer Prev*, **12**, 1469-72.

Grumett S, Snow P, Kerr D (2003). Neural networks in the

prediction of survival in patients with colorectal cancer. *Clin Colorectal Cancer*, **2**, 239-44.

Islamic Republic of Iran, Ministry of Health and Medical Education, Office of Deputy , Center for Diseases Control, Cancer office. Iranian Annual National Cancer Registration Report 2005-2006. March 2007.

Kan T, Shimada Y, Sato F, et al (2004). Prediction of lymph node metastasis with use of artificial neural networks based on gene expression profiles in esophageal squamous cell carcinoma. *Ann Surg Oncol*, **11**, 1070-8.

Kelsall HL, Baglietto L, Muller D, et al (2009). The effect of socioeconomic status on survival from colorectal cancer in the Melbourne Collaborative Cohort Study. *Soc Sci Med*, **68**, 290-7.

Kyung-Joong Kim, Sung-Bae Cho (2004). Prediction of colon cancer using an evolutionary neural network. *Neurocomputing*, **61**, 361-79.

Lee SM, Kang JO, Suh YM (2004). Comparison of hospital charge prediction models for colorectal cancer patients, neural network vs. decision tree models. *J Korean Med Sci*, **19**, 677-81.

Missaoui N, Jaidaine L, Ben Abdelkader A, et al (2010). Clinicopathological patterns of colorectal cancer in Tunisia. *Asian Pacific J Cancer Prev*, **11**, 1719-22.

Moghimi-Dehkordi B, Safaee A, Zali MR (2008). Prognostic factors in 1,138 Iranian colorectal cancer patients. *Int J Colorectal Dis*, **4**, 683–8.

Moradi A, Khayamzadeh M, Guya MM, et al (2009). Survival of colorectal cancer in Iran. *Asian Pacific J Cancer Prev*, **10**, 583-6.

National Cancer Institute (2010). Available at http://www.cancer. gov/cancertopics/factsheet/Sites-Types/metastatic.

Parkin DM (2004). International variation. *Oncogene*, **23**, 6329-40.

Sung JJ, Lau JY, Goh KL, et al (2005). Increasing incidence of colorectal cancer in Asia, implications for screening. *Lancet Oncol*, **6**, 871-6.

Talmadge JE, Fidler IJ (2010). AACR centennial series, the biology of cancer metastasis, historical perspective. *Cancer Res*, **70**, 5649–69.

World Health Statistics Annual (2010). WHO Databank, 2010. Available at http://www-dep.iarc.fr/.