Global Journal of Computer Science and Technology
Interdisciplinary
Volume 13 Issue 3 Version 1.0 Year 2013
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Colon Cancer Prediction based on Artificial Neural Network

By Md. Asaduzzaman Sabuj & Priyam Biswas
Chittagong University of Engineering and Technology, Bangladesh

Abstract - Artificial neural networks (ANNs) consists of computational neurons or processing elements are linear mathematical model which abstract away the complex biological model and its aim is good, human like predictive ability. Artificial intelligence tries to simulate some properties of biological neural networks. In this study on the basis of previous dataset the in symptoms data are applied to a supervised back propagation artificial neural network learning process to find out the predictive outcome which is better than logistic regression (LR) process. As in most cases ANN is an adaptive system that changes its structure on the basis of internal and external information, the predictive result is more accurate than any other processes.

Keywords : *artificial neural network, back propagation, colon cancer, supervised learning, prediction.*

GJCST-G Classification: *F.1.1*

Strictly as per the compliance and regulations of:

Colon Cancer Prediction based on Artificial Neural Network

Md. Asaduzzaman Sabuj [α] & Priyam Biswas [σ]

Abstract - Artificial neural networks (ANNs) consists of computational neurons or processing elements are linear mathematical model which abstract away the complex biological model and its aim is good, human like predictive ability. Artificial intelligence tries to simulate some properties of biological neural networks. In this study on the basis of previous dataset the in symptoms data are applied to a supervised back propagation artificial neural network learning process to find out the predictive outcome which is better than logistic regression (LR) process. As in most cases ANN is an adaptive system that changes its structure on the basis of internal and external information, the predictive result is more accurate than any other processes.

Keywords : artificial neural network, back propagation, colon cancer, supervised learning, prediction.

I. Introduction

Colon cancer is the third most commonly diagnosed cancer in the world, but it is more common in developed and developing countries. Around 60% of cases were diagnosed in the world. Most colon cancer occurs due to lifestyle and increasing age with only a minority of cases associated with underlying genetic disorders. It typically starts in the lining of the bowel and if left untreated, can grow into the muscle layers underneath, and then through the bowel wall.

Colon cancer prediction system is designed based on the staging system which has been introduced by American Joint Committee. Colon cancer staging is an estimate of the amount of penetration of a particular cancer. It is performed for diagnostic and research purposes, and to determine the best method of treatment. The systems for staging colon cancers depend on the extent of local invasion, the degree of lymph node involvement and whether there is distant metastasis. The staging system for colon cancer had four categories that are based on tumour-node-metastasis. The stages are I, II, III and IV by the use of T stage (i.e. tumour depth of penetration) and N stage (i.e., number of lymph nodes) and M stage (i.e., metastasis). Total resulting seven stages are I, IIa, IIb, IIIa, IIIb, IIIc and IV.

Here in this article we use the information of surveillance Epidemiology and End result (SEER) program. The percentage of survival rate is collected from SEER database and American society of clinical oncology. In case of supervised learning process these data are used to learn the inputted data and finally to get the predicted result.

Each and every stage included particular tumor grade, specific histology, tumor location, number of positive lymph nodes, and metastases.

Table 1 : Stages as defined by the American joint committee on cancer (ajcc) fifth and sixth edition

Staging system	T stage	N stage	M stage
AJCC fifth edition			
I	T1 or T2	N0	M0
II	T3 or T4	N0	M0
III	Any T	N1	M0
IV	Any T	Any N	M1
AJCC sixth edition			
I	T1 or T2	N0	M0
IIa	T3	N0	M0
IIb	T4	N0	M0
IIIa	T1 or T2	N1	M0
IIIb	T3 or T4	N1	M0
IIIc	Any T	N2	M0
IV	Any T	Any N	M1

*T1= tumour invades submucosa; T2= tumor invades muscularis propria; T3= tumor invades through the muscularis propria into the subserosa or into non-peritonealized pericolic tissues;T4= tumor directly invades other organs or structures and/or perforates visceral peritoneum; N0= no regional lymph node metastasis; N1= metastasis to one to three regional lymph nodes; N2= metastasis to four or more regional lymph nodes; M0= no distant metastasis; M1= distant metastasis.

Each tumor stage was coded according to the TNM stage organization for each edition (T1= tumor invades submucosa; T2=tumor invades muscularis propria; T3= tumor invades through the muscularis propria into the subserosa or into nonperitonealized pericolic tissues; T4= tumor directly invades other organs or structures or perforates visceral peritoneum; N0= no regional lymph node metastasis; N1= metastasis to one to three regional lymph nodes; N2= metastasis to four or more regional lymph nodes; M0= no distant metastasis; M1= distant metastasis). TNM stage was determined by SEER's extent of disease (for T stage and M stage) and number of lymph nodes (for N stage) coding schemes. All patients were included in both analyses of survival for both staging

Author α : Dept. of CSE, CUET. Chittagong, Bangladesh.
E-mail : sabuj_asaduzzaman@yahoo.com
Author σ : Lecturer, Dept. of CSE, CUET. Chittagong, Bangladesh.

systems. Tumor grade was categorized as low grade, low grade and others. High grade tumors are well or moderately differentiated and low grade tumors are poorly differentiated. Tumor location was categorized as right (cecum, ascending colon, hepatic flexure), transverse, left (splenic flexure, descending colon), and sigmoid colon. The numbers of positive lymph nodes were also categorized. Dukec and Macd are also categorized as A, B and C. Histologic subtypes are categorized Adenocercinomous, Mucinous adenocercinomous and Signet ring cercinomous. The purpose of this study was to determine the presence or absence of colon cancer using ANN. An ANN technology was chosen as an analysis tool primarily because of its demonstrated accuracy in a wide variety of situations.

A logistic regression analysis was chosen as a comparison primarily because it is an accepted standard. Artificial neural networks (ANNs) grew out of attempts to mimic the fault tolerance and capacity to learn of biological nervous systems. The ANNs do this by modeling the low level structure of the brain. A biological nervous system is composed of a very large number of neuron cells, massively interconnected to one another. Each neuron is a specialized entity that can propagate an electrochemical signal. Each neuron has branching input structures called dendrites and branching output structures called axons. The axons of one cell are connected to the dendrites of other cells by synapses. Signals are propagated throughout this complex organism, regulated primarily by the synapses.

In like manner, a typical ANN consists of computational neurons or processing elements connected by weighted signal pathways. They typically have a much simpler architecture, with many fewer neurons and connections, than a biological nervous system has. An artificial neuron receives a number of inputs, either from data entering the network or as output from other neurons. Each input comes via a pathway connection that has strength or, in terms of ANNs, weight. These weights correspond to synaptic strength in biological systems. Each neuron also has a single threshold value. The activation of this artificial neuron is composed of the weighted sum of its inputs less the threshold value. This activation signal is transformed through an activation or transfer function to produce the output of the neuron. The transfer function is generally a nonlinear, continuously differentiable function that may not have a direct biological equivalent. Artificial neural networks consist of input elements that bring in signals from the outside world in a manner somewhat similar to biological sensory nerves from, for example, the eye. The input signals are fed to one or more layers of neurons through the weighted pathway connections. These hidden neurons process the signals and produce another set of signals that are sent to an output layer of neurons through weighted pathway connections. The output neurons generate a signal to the outside world that is somewhat similar to biological motor nerves connected, for example, to the hands.

II. Survival Analysis

5-year survival was 65.2%. According to stages defined by the AJCC fifth edition system, 5-year stage-specific survivals were 93.2% for stage I, 82.5% for stage II, 59.5% for stage III, and 8.1% for stage IV. According to stages defined by the AJCC sixth edition system, 5-year stage-specific survivals were 93.2% for stage I, 84.7% for stage IIa, 72.2% for stage IIb, 83.4% for stage IIIa, 64.1% for stage IIIb, 44.3% for stage IIIc, and 8.1% for stage IV. Under the sixth edition system, 5-year survival was statistically significantly better for patients with stage IIIa colon cancer (83.4%) than for patients with stage IIb disease (72.2%) (P<.001).

a) Survival by Histologic Subtype

Among patients in the entire cohort, 87.4% had adenocarcinomas, 11.6% had mucinous adenocarcinomas, and 1.0% had signet ring cell carcinomas. Among the entire cohort, a worse 5-year survival was statistically significantly associated with signet ring cell carcinomas (36.0%) than with adenocarcinomas (65.9%) or with mucinous adenocarcinomas (61.8%). When we further stratified data in each stage (as defined by the fifth edition system) by histologic subtype, we observed similar survival distributions in stages II, III, and IV, but not in stage I. For example, in stage III, the 5-year survival was 36.6% for signet ring cell carcinomas, 60.1% for adenocarcinomas, and 58.7% for mucinous adenocarcinomas (P=.001). For stage I, however, the 5-year survival was 100.0% for signet ring cell carcinomas, 93.3% for adenocarcinomas, and 92.0% for mucinous adenocarcinomas; these values were not statistically significantly different from each other [2].

Stage	0 m0		30 m0			60 m0		
	Survival (%)	N	Survival (%)	N	P	Survival (%)	N	P
I	100	14500	96.1	8591	-	93.2	4515	-
II	100	34361	89.2	19492	<.0001	82.5	10105	<.0001
III	100	26949	72.7	12192	<.0001	59.5	5514	<.0001
IV	100	20802	17.3	1832	<.0001	8.1	432	<.0001

Figure 1 : Five-year survival by American Joint Committee on Cancer fifth edition system stages I–IV. *P* value determined with the log-rank test refers to the corresponding stage and the stage in the row above. All statistical tests were two-sided

Stage	0 m0		30 m0			60 m0		
	Survival (%)	N	Survival (%)	N	P	Survival (%)	N	P
I	100	14500	96.1	8591	-	93.2	4515	-
IIa	100	28535	91.0	2105	<.001	84.7	8494	<.001
IIb	100	5826	80.2	3060	<.001*	72.2	1611	<.001*
IIIa	100	1989	91.4	1120	NS+	83.4	551	NS+
IIIb	100	15946	77.3	7786	<.001+	64.4	3579	<.001+*
IIIc	100	8600	59.1	3039	<001	44.3	1220	<.001
IV	100	20802	17.3	1832	<.001	8.1	432	<.001

Figure 2 : Five-year survival by the American Joint Committee on Cancer sixth edition system stages I–IV. P value determined by the log-rank test refers to the corresponding stage and the stage in the row above, unless otherwise indicated. All statistical tests were two-sided. *= IIIa versus IIb; + = IIa versus IIIa; +* = IIb versus IIIb; NS = not statistically significant

b) Survival by Tumor Grade

We next used colon cancer stages as defined by the AJCC fifth edition system and stratified data in each stage further by other factors to assess their prognostic value. Among all patients evaluated in the cohort, 67.8% (n= 81 493) had low-grade tumors, 19.4% (n= 23 287) had high-grade tumors, and 12.8% (n= 15 343) had tumors whose grade was unknown. For those patients whose tumor grade (high versus low) was known (n= 104 780), tumor grade was statistically significantly associated which is shown in figure 3.

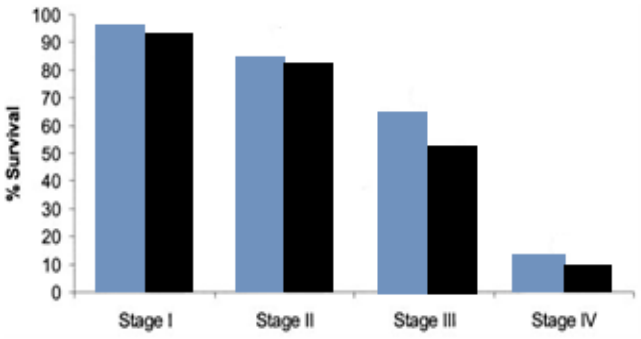

Figure 3 : Five-year survival for American Joint Committee on Cancer fifth edition by grade. Solid bars, low-grade tumors; shaded bars, high-grade tumors. Star, P=.001, log-rank test. All statistical tests were two-sided

c) Survival by Tumor Location

Among patients in the entire cohort, 44.6% had tumors in the right colon, 9.4% had tumors in the transverse colon, 10.4% had tumors in the left colon, 31.6% had tumors in the sigmoid colon, and 4.0% had tumors whose location was unknown. Among the overall cohort, a better 5-year survival was statistically significantly associated with tumors located in the sigmoid colon (69.8%) than with tumors located in the right colon (63.7%) (*P*=.001), in the transverse colon (65.0%) (*P*=.001), and in the left colon (65.1%) (*P*=.001). When we further stratified each stage (as defined by the fifth edition system) by these tumor locations, we observed similar survival distributions in stages I, III, and IV, but not in stage II (Fig. 5). For example, in stage III, 5-year survival was 64.3% for sigmoid lesions, 57.0% for right colon lesions (*P*=.001), 57.9% for transverse (*P*=.001), and 60.2% for left-colon lesions (*P*=.001),whereas in stage II, 5-year survival was 83.6% and 83.7%, respectively, for rightand transverse-colon lesions, 81.5% for the left colon, and 80.7% for sigmoid lesions[1].

d) Lymph Nodes

Among patients in the entire cohort, 32.5% had positive lymph nodes. When we used a histogram analysis of the number of positive lymph nodes, we found that the N stage could be stratified into the following four categories: N1 (one to three positive lymph nodes), N2 (four or five positive lymph nodes), N3 (six to eight positive lymph nodes), and N4 (nine or more positive lymph nodes). We used the proposed N stages in combination with the AJCC sixth edition staging system as a new staging system (Table 2). In this new system, stages I, IIa, IIb, IIIa, and IIIb are the same as corresponding stages in the sixth edition system, but the new stages IIIc, IIId, and IIIe are stratified by categories N2, N3, and N4, respectively, as defined above. The 5-year survival by these proposed stages is 93.2% for stage I, 84.7% for stage IIa, 72.2% for stage IIb, 83.4% for stage IIIa, 64.1% for stage IIIb, 52.3% for stage IIIc, 43.0% for stage IIId, 26.8% for stage IIIe, and 8.1% for stage IV [2]. Corresponding Kaplan–Meier survival curves for this system are shown in Fig. 4.

Stage	0 m0		30 m0			60 m0		
	Survival (%)	N	Survival (%)	N	P	Survival (%)	N	P
I	100	14500	96.1	8591	-	93.2	4515	-
IIa	100	28535	91.0	2105	<.001	84.7	8494	<.001
IIb	100	5826	80.2	3060	<.001*	72.2	1611	<.001*
IIIa	100	1989	91.4	1120	NS+	83.4	551	NS+
IIIb	100	15946	77.3	7786	<.001+*	64.4	3579	<.001+*
IIIc	100	4092	67.1	3039	<.001	52.3	725	<.001
IIId	100	2655	57.3	908	<.001	43.0	384	<.001
IIIe	100	1853	43.1	434	<.001	26.8	141	<.001
IV	100	20802	17.3	1832	<.001	8.1	432	<.001

Figure 4 : Survival by American Joint Committee on Cancer sixth edition staging with proposed lymph node (N) stages. **, P* values determined by the log-rank test refers to the corresponding stage and the stage in the row above, unless otherwise indicated. * = IIIa versus IIb; + = IIa versus IIIa; +*= IIb versus IIIb; NS = not statistically significant. All statistical tests were two-sided

III. Methodology

A back-propagation (BP) neural network is a multi-layer network and the layers are fully connected that is every neuron in each layer is connected to every other neuron in the adjacent forward layer. In a back-

propagation neural network, learning algorithm has two phases. First, a training input pattern is presented to the network input layer. The network then propagates the input pattern from layer to layer until the output pattern is generated by output layer. If this pattern is different from the desired output, an error is calculated and then propagated backwards through the input layer. The weights are modified as the error is propagated.

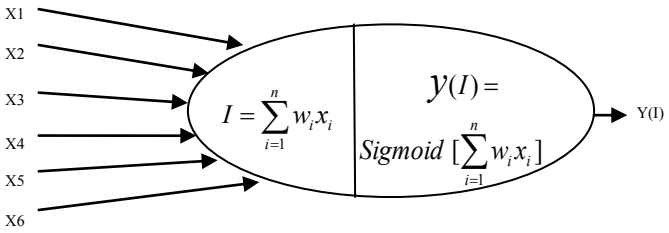

Figure 5 : Neuron Output Determination in Backpropagation NN

The feed forward BP MLP can be viewed basically as a set of equations that are linked together through shared variables in a formation diagramed as a set of interconnected nodes in a network capable of general functional approximation that provides learning capabilities. Variables for inclusion in the final network architecture are usually chosen by a sensitivity analysis method, which tests each input variable by dropping it from the input list and determining the resulting loss of predictive accuracy. Only variables that result in a significant loss of accuracy when dropped are retained in the final network's architecture. Classification tasks like tumor staging, diagnosis, or predicting survival can be performed by FFANNs. FFANN is typically organized as a set of interconnected layers of artificial intermediate (hidden) nodes depicted as a row or collection of nodes, each receiving input from other nodes, connected together to form the network. The MLP has an associated output activation level known as a "squashing" or "activation" function; the most popular is the sigmoid function [f(I)] expressed as:

$$f(I) = 1/[1 + \exp(-I)]$$

Step 1 : Initialization

Set initial weights w_{ij}, w_{jk}, [i=1...n], [j=1..m], [k=1..l], threshold values θ_j, θ_k and learning rate with random number within the range [-2.4/ F_i, +2.4/ F_i] where F_i = maximum no. of inputs connected to the single neuron.

$$y_j(p) = sigmoid[\sum_{i=1}^{n} x_i(p)\, w_{ij}(p) - \theta_j] \quad (1)$$

Step 2 : Activation

Calculate the actual output of neuron of hidden layer.

$$y_k(p) = sigmoid[\sum_{j=1}^{m} x_j(p)\, w_{jk}(p) - \theta_k] \quad (2)$$

Here n is the no. of input layer neurons connected to hidden layer neuron j. Calculate the actual output of neuron of output layer.

Where m is the no. of hidden layer neurons connected to output layer neuron k.

Step 3 : Weight Update

Update the weights in the network.

Hidden layer weight update:

$$w_{jk}(p+1) = w_{jk}(p) + \Delta w_{jk}(p) \quad (3)$$

$$\Delta w_{jk}(p) = \alpha y_j(p) \delta_k(p) \quad (4)$$

$$\delta_k(p) = y_k(p)[1 - y_k(p)] e_k(p) \quad (5)$$

$$e_k(p) = y_{d,k}(p) - y_k(p) \quad (6)$$

Input layer weight update:

$$w_{ij}(p+1) = w_{ij}(p) + \Delta w_{ij}(p) \quad (7)$$

$$\Delta w_{ij}(p) = \alpha x_i(t) \delta_j(p) \quad (8)$$

$$\delta_j(p) = y_i(p)[1 - y_i(p)] \sum_{k=1}^{l} \delta_k(p) w_{jk}(p) \quad (9)$$

Step 4 : Iteration

Increase iteration p by one, go back to Step 2. Process is repeated until the error reduces to zero or closer to zero. Computing the output result and comparing it with the expected one find out the error and if the error is very higher than expected then error reduction process is applied here to reduce it. Each and every iteration comparing with the expected result the weight values are updated back propagating from end from the final layer to first layer. Then again using those weight values we will get the next result which has less error than before. In the same way after some iteration we will get more closer result than before and finally when the result is closest and has the least error then it is defined as the final result.

IV. Conclusions

To aid clinicians in the diagnosis of colon cancer, recent research has looked into the development of computer aided diagnostic tools. Various techniques have been widely used for colon cancer diagnosis. In this paper we have discuss some of effective techniques that can be used for colon cancer determination. The predicting outcome is found based on comparing with previous dataset value. It is proved that in this process the outcome is more accurate than any other process.

References Références Referencias

1. American Joint Committee on Cancer. Missions and objectives. Available at: http://www.cancerstaging.org.

2. Surveillance, Epidemiology, and End Results (SEER) Program Public-Use Data (1973-2000), National Cancer Institute, DCCPS, Surveillance Research Program, Cancer Statistics Branch, submission. Available at: http://www.seer.cancer.gov.
3. Greene FL. TNM staging for malignancies of the digestive tract: 2003 changes and beyond. Semin Surg Oncol 2003; 21:23–9.M. Wegmuller, J. P. von der Weid, P. Oberson, and N. Gisin, —High resolution fiber distributed measurements with coherent OFDR, in Proc. ECOC'00, 2000, paper 11.3.4, p. 109.
4. AJCC cancer staging manual, 6th ed. New York (NY): Springer, 2002.
5. Carcinoma of the colon and rectum. Ann Surg 1954; 139:846–52.
6. Astler VB, Coller FA. The prognostic significance of direct extension of carcinoma of the colon and rectum. Ann Surg 1954; 139:846–52. (2002) The IEEE website. [Online]. Available: http://www.ieee.org/
7. Swanson RS, Compton CC, Stewart AK, Bland KI. The prognosis of T3N0 colon cancer is dependent on the number of lymph nodes examined.
8. North American Association of Central Cancer Registries. Available at: http://www.naaccr.org. [Last accessed: August 6, 2004.]
9. Compton CC, Fielding LP, Burgart LJ, Conley B, Cooper HS, Hamilton SR, et al. Prognostic factors in colorectal cancer. College of American Pathologists Consensus Statement 1999. Arch Pathol Lab Med 2000; 124: 979–94.

Global Journals Inc. (US) Guidelines Handbook 2013

www.GlobalJournals.org