International Journal of Engineering Sciences Paradigms and Researches (IJESPR)
(Vol. 16, Issue 01) and (Publishing Month: August 2014)
(An Indexed, Referred and Impact Factor Journal)
ISSN (Online): 2319-6564
www.ijesonline.com

Colon Cancer Prediction using Neural Network

Anjali[1] and Dr. Rajesh Gargi[2]

[1]M.Tech. Scholar, Geeta Engineering College, Noltha Panipat, Haryana (India)
anjalisangwan15@gmail.com

[2]Professor, CSE Deptt., Geeta Engineering College, Noltha Panipat, Haryana (India)
directorengg@geetainstitutes.com

Abstract

Cancer is a group of more than 100 different diseases. They affect the body's basic unit, the cell. Cancer occurs when cells become abnormal and divide without control or order. Like all other organs of the body, the colon and rectum are made up of many types of cells. Normally, cells divide to produce more cells only when the body needs them. This research improves the classification accuracy to detect the colon cancer by using the neural network. The pattern recognition neural network is used to recognize the pattern of the attributes that will decide whether a gene is normal or not. The data set used is arranged with the help of UCI repository.

Keywords: *Colon Cancer, Neural, Fuzzy.*

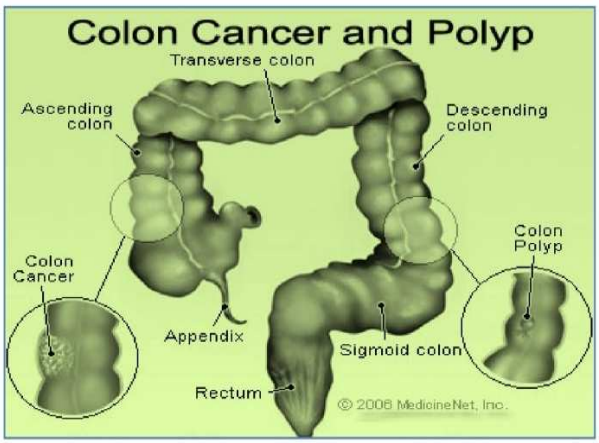
I. Introduction

Colon cancer is a malignant cancer in which the tissue in the walls of colon/rectum develops abnormally. It is the third common type of cancer in both sexes. This cancer forms in the tissues of the colon (the longest part of large intestinal).Colon cancer is usually caused due to a diet high in fat, red meat but low in fruits and vegetables, high caloric intake, low levels of physical activity, obesity, smoking and excessive alcohol intake etc. Prediction of this colon cancer will help to prevent it in its early stage. Data mining techniques have been usually applied in this field (or) other aspects of medical science [1]

The colon is the part of the digestive system where the waste material is stored. The rectum is the end of the colon adjacent to the anus. Together, they form a long, muscular tube called the large intestine (also known as the large bowel). Tumors of the colon and rectum are growths arising from the inner wall of the large intestine. Benign tumors of the large intestine are called polyps. Malignant tumors of the large intestine are called cancers. Benign polyps do not invade nearby tissue or spread to other parts of the body. Benign polyps can be easily removed during colonoscopy and are not life-threatening, Health line (2011). If benign polyps are not removed from the large intestine, they can become malignant (cancerous) over time. Most of the cancers of the large intestine are believed to have developed from polyps [2].

Cancer of the colon and rectum (Figure 1); also referred to as colorectal cancer, Better Medicine (2011), Health line (2011) and Medicine Net (2011) can invade and damage adjacent tissues and organs. Cancer cells can also break away and spread to other parts of the body (such as liver and lung) where new tumors form. The spread of colon cancer to distant organs is called metastasis of the colon cancer. Once metastasis has occurred in colorectal cancer, a complete cure of the cancer is unlikely [3].

Figure 1: Pictorial Representation of Colon Cancer and Polyp [3]



The colon has 4 sections:

- The first section is called the *ascending colon*. It starts with a small pouch (the *cecum*) where the small bowel attaches to the colon and extends upward on the right side of the abdomen. The cecum is also where the appendix attaches to the colon.
- The second section is called the *transverse colon* since it goes across the body from the right to the left side in the upper abdomen.
- The third section, called the *descending colon*, continues downward on the left side.
- The fourth and last section is known as the *sigmoid colon* because of its "S" or "sigmoid" shape.

The waste matter that is left after going through the colon is called *feces* or *stool*. It goes into the *rectum*, the final 6 inches of the digestive system, where it is stored until it passes out of the body through the *anus*.[4]

II. Techniques Used For Colon Cancer

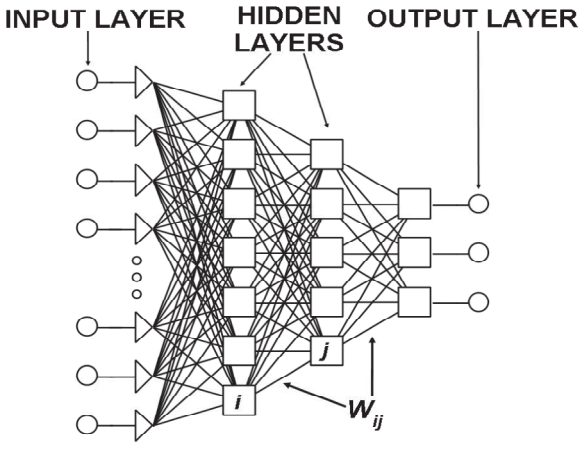
a. Artificial Neural Network

An ANN is a mathematical representation of the human neural architecture, reflecting its "learning" and "generalization" abilities. For this reason, ANNs belong to the field of artificial intelligence [5]. ANNs are widely applied in research because they can model highly non-linear systems in which the relationship among the variables is unknown or very complex. A review of various classes of neural networks can be found in (Aleksander and Morton 1995, Zupan and Gasteiger 1999).

- Mathematical background

A neural network is formed by a series of "neurons" (or "nodes") that are organized in layers. Each neuron in a layer is connected with each neuron in the next layer through a weighted connection. The value of the weight wij indicates the strength of the connection between the i-th neuron in a layer and the j-th neuron in the next one. The structure of a neural network is formed by an "input" layer, one or more "hidden" layers, and the "output" layer. The number of neurons in a layer and the number of layers depends strongly on the complexity of the system studied. Therefore, the optimal network architecture must be determined [5]. The general scheme of a typical three-layered ANN architecture is given in Figure 3. The neurons in the input layer receive the data and transfer them to neurons in the first hidden layer through the weighted links. Here, the data are mathematically processed and the result is transferred to the neurons in the next layer. Ultimately, the neurons in the last layer provide the network's output [5].

Figure 2: General Structure Of A Neural Network With Two Hidden Layers.



The j-th neuron in a hidden layer processes the incoming data (xi) by: (i) calculating the weighted sum and adding a "bias" term (θj) according to Eq. 1:

$net_j = \sum_{i=1}^{m} x_i \times w_{ij} + \theta_j \ (j = 1,2, \dots \dots, n(1)$

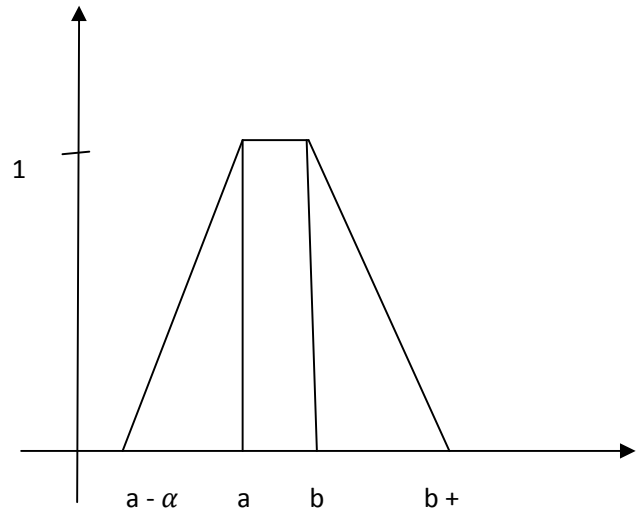
(ii) transforming the netj through a suitable mathematical "transfer function", and (iii) transferring the result to neurons in the next layer. Various transfer functions are available, however, the most commonly used is the sigmoid one:

$f(x) = \frac{1}{1+e^{-x}}$ (2)

b. Fuzzy Logic

Fuzzy Logic (FL) helps computers paint vivid pictures of the uncertain world. Fuzzy sets were introduced by Zadeh [6] as a means of representing and manipulating data that are not precise, but rather fuzzy. Fuzzy logic provides an inference morphology that helps appropriate human reasoning capabilities to be applied to knowledge-based systems. The theory of fuzzy logic provides a mathematical strength to capture the uncertainties associated with human cognitive processes, such as thinking and reasoning. A fuzzy set A is called trapezoidal fuzzy number (Figure 3) with tolerance interval [a, b], left width α and right width β if its membership function has the following:

Figure 3: Trapezoidal Fuzzy Number



Expert systems are knowledge-based systems that contain expert knowledge. An expert system is a program that can provide expertise for solving problems in a defined application area in the way the experts do. They use human knowledge to solve problems that normally would require human intelligence. These expert systems represent the expertise knowledge as data or rules within the computer. These rules and data can be called upon when needed solve problems, PCAI [7]; NIJ [8] and Steffen [9].

III. Proposed Work

The proposed work improves the classification accuracy by using the neural network. the pattern recognition neural network is used to recognize the pattern of the attributes that will decide whether a gene is normal or not. The data set used is arranged with the help of UCI repository. The data set is partitioned in two parts one is testing dataset and the other is training dataset. The 40% of the total dataset is used for training while remaining 60% for the testing purpose. The neural network gets trained with the help of training dataset and then recognizes the testing dataset appropriately. The process can be understood by following steps:

1. Read Dataset
2. Preprocess the dataset
3. Partition dataset in two parts i.e. training dataset, testing dataset.
4. Train the NN using the training dataset.
5. Input the testing dataset to NN
6. Get the output of NN
7. Compare the output of NN with the original output.

The preprocessing of the dataset contains the steps to convert dataset in to appropriate form, so that it can be used. The comparison of actual output and NN output gives the classification accuracy.

IV. Simulation Results

The proposed process is implemented using the MATLAB.

Table 1: Result Analysis

Number Of Runs	Performance	Accuracy
1	0.0526	100
2	0.0238	100
3	0.0279	100
4	0.000004	95.65
5	0.0259	100
6	0.0520	100
Average	0.0303	99.275

The above table shows the results of the various runs. The accuracy is almost 100% but sometime it gets reduced to 95%. So the average accuracy is 99.2%. The average performance is 0.0303. These results can be compared graphically as shown in following figures.

Figure 4: Comparison of Performance

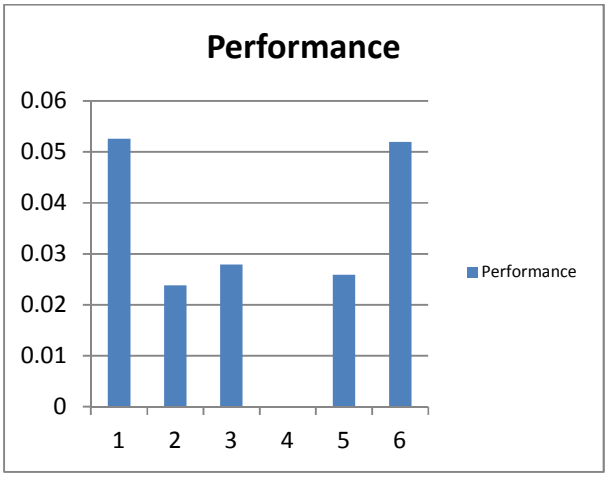
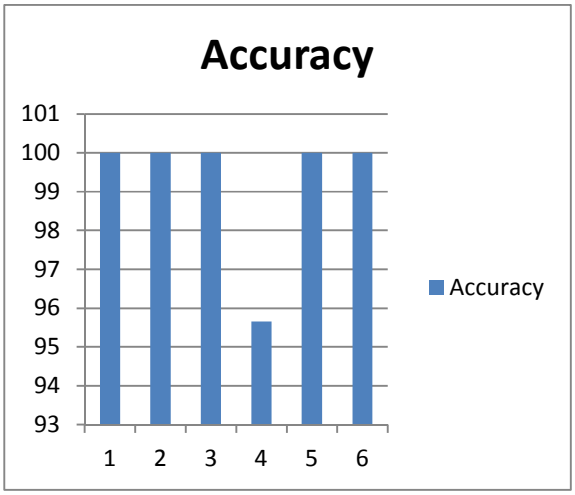


Figure 5: Comparison of Accuracy



The comparison of accuracy shows the proposed system has accuracy around 99%. It remains above than 95% in all cases, while in the existing system the accuracy is around 90%. The improved accuracy of classification confirms the better performance of the proposed system.

V. Conclusion

This research improves the classification accuracy to detect the colon cancer by using the neural network. The pattern recognition neural network is used to recognize the pattern of the attributes that will decide whether a gene is normal or not. The data set used is arranged with the help of UCI repository. The data set is partitioned in two parts one is testing dataset and the other is training dataset. The 40% of the total dataset is used for training while remaining 60% for the testing purpose. The neural network gets trained with the help of training dataset and then recognizes the testing dataset appropriately. The implementation is done using the MATLAB. The neural network toolbox is used while implementing the proposed system. The comparison of accuracy shows the proposed system has accuracy around 99%. It remains above than 95% in all cases, while in the existing system the accuracy is around 90%. The improved accuracy of classification confirms the better performance of the proposed system. In future work it will perform the proposed system can be used to predict the other diseases.

References

[1] Nithya, N. S., and K. Duraiswamy. "An Information Gain based Fuzzy Classifier for Predictive Analysis in Colon Cancer Data." *International Journal of Computer Applications* 31 (2011).

[2] Durai, Rajaraman, et al. "Principles and applications of gene therapy in colon cancer." Journal of Gastrointestinal and Liver Diseases 17.1 (2008): 59.

[3] J. C. Obi and A. A. Imianvan, "Fuzzy Neural Approach For Colon Cancer Prediction", Scientia Africana, Vol. 11 (No.1), June 2012. pp 65-76© Faculty of Science, University of Port Harcourt, Printed in Nigeria.

[4] Hiraoka K, Kimura T, Logg CR, et al. "Therapeutic efficacy of replication-competent retrovirus vector-mediated suicide gene therapy in a multifocal colorectal cancer metastasis model." Cancer Res 2007; 67: 5345-5353.

[5] Filippo Amato, Alberto López, Eladia María Peña-Méndez, Petr Vaňhara, Aleš Hampl, Josef Havel, "Artificial neural networks in medical diagnosis",ISSN 1214-0287.

[6] Zadeh L.A. (1965), "Fuzzy sets. Information and control, Vol.8, pp.338-353.

[7] PCAI (2000), "Expert System: Introduction, retrieved from http://PCAI.com/web/ai_info/expert.systems.html

[8] NIJ: National Institute of Justice (2011), "Expert system technologies for criminal justice application" retrieved from https://www.ncjrs.gov/pdffiles1/nij/sl000959.pdf.

[9] Steffen L. (2011), "Expert system and local Computation", University of Oxford, Graduate Lectures Hilary Term 2011.