JOURNAL OF NORTHEASTERN UNIVERSITY

Volume 25 Issue 04, 2022 ISSN: 1005-3026 https://dbdxxb.cn/ Original Research Paper

COLON CANCER CLASSIFICATION USING CONVOLUTIONAL NEURAL NETWORK

V T Ram Pavan Kumar
Research Scholar, Department of CSE, Annamalai University
Mrpphd2018@gmail.com

M Arulselvi
Associate Professor, Department of CSE, Annamalai University, marulcse.au@gmail.com

K B S Sastry
Associate Professor, Department of Computers, Andhra Loyola College,
sastrykbs@gmail.com

Abstract:
CNN represents a particular type of neural networks works extraordinally for image processing applications. CNN is mainly comprised with multi-layers interconnected neurons especially trained effectively for classification and feature extraction. CNN provides faster classification because feature extraction is done by itself. Colon cancer is one of the harmful diseases in worldwide and it is commonly found in older adults. The rapid and uncontrollable development of cells inside the colon within a short period causes serious impacts on patient's body such that it leads to death. In order to encounter such problems, this work presents a Convolutional Neural Network (CNN) for classifying the colon cancer and effectively detect whether the image data is Cancerous or Non-Cancerous.

Keywords: Colon cancer, Convolutional Neural Network (CNN).

1. INTRODUCTION

Cancer is a disease, which is mainly originated due to the rapid and uneven growth of cells in the tissue. Cancer cells can grow at any region of the body. Cancer is similar in some characteristics but it varies in the way of its growing and spreading[12]. At its initial stage, the cells grow in abnormal way and it vanish the healthy cell. The body will find very complex to proper functioning of the body parts. Cancer can occur anywhere of the body parts, like breast, brain, lungs, cervix. Different types of cancer results mashed cells to split at a lower rate, whereas others produce rapid cell growth [2] [8]. Most of the [5]colon cancers emerge with the production of small cell clots known as adenoma polyps but they are not cancerous cells. Such clots then gradually spread over the body in an uncontrollable manner.

In fact, there is no symptom at an early stage of colon cancer [3] and people even do not realize they are really suffering from colon cancer. So, early detection of colon cancer saves the patient's life and reduces the mortality rate. There are numerous techniques related to the classification of colon cancer, but they really failed to provide better classification accuracy.

Submitted: 06/09/2022 **Accepted**: 04/11/2022

However [1] automatic classification of colon cancer is very complex because of various factors, such as same colors in various biological components of [13] histopathological image data. Colon cancer happens when there is [4] genetic mutation and DNA cells grows specific regions of the body will develop abnormally by destroying the parts slowly. The major benefit of CNN is that the same technique can be utilized for the extraction of strong features as well as image classification. The main objective of this work is to perform colon cancer classification automatically. So, CNN is used to classify the image into normal (non-cancerous) and abnormal (cancerous). Finally the performance of the classifier is evaluated using the metrics such as accuracy, sensitivity and specificity.

2. RELATED WORK

[7] Presented a Grey Wolf Optimization (GWO) algorithm for the classification of colon cancer in an effective manner. The essential features from input patterns are selected utilizing Information Gain (IG). After that, the selected features were minimized by considering the GWO algorithm. Finally, Support Vector Machine (SVM) was exploited to categorize the type of cancer. The developed technique provided higher stability as well as better classification accuracy.

The major disadvantage exists in this developed method was imbalanced data and the problem of unlabeled data can be effectively handled by introducing the unsupervised learning in future investigation.[6].Naive Bayes Classifier is a prediction model, which was solely dependent on simple probabilistic and also utilized the Bayes theorem with a powerful consideration. The presented model achieved high classification accuracy while classifying the type of colon cancer with low complexity. However, it failed to enhance the accuracy at some part of the information where attributes are associated with each other. An amalgamated approach [10] of exploiting bi-level feature selection mechanisms was developed for colon cancer classification. Here, the limitation existed in colon cancer classification problem was encountered by employing MRMR(Maximum Relevance and Minimum Redundancy) and various optimization algorithms. Finally, it was further classified with five suitable classifiers. However, the method failed to determine the microarray-based colon cancer classification.

3. PROPOSED WORK

Figure.1 shows the block diagram of the proposed work. At first, the input image is acquired from the dataset and then selected input image is subjected to colon cancer classification using CNN. Finally, the CNN provides classified results. The main concept of
[9] CNN is by converting the values which are receiving from predecessor layer to more complex values and forward to successor layer for generalization. CNN [11] is mainly comprised with multi-layers interconnected neurons especially trained effectively for classification and feature extraction. When compared with the existing classification algorithms, [5] CNN provides better classification results with minimum cost within a short time. CNN consists of 3 layers named Convolutional layer, pooling layer and

fully Connected layer. Convolution Layer is the first layer of CNN network that decides the whole operation of the network. The performance of CNN is usually based on the utilization of learnable filters. The output of this layer is obtained by convolving each filter on the given image and the culmination will be a series of images where the number of images is similar to the amount of filters. Each filter in the convolution layer is a grid of discrete numbers and the procedure involves in initialization of weights randomly. For every convolution layer multiple kernels are defined and at each point underlying pixel values are multiplied and add them which give raise to corresponding output. In order to bring out the nonlinear property from CNN, activation function is used. The Rectified linear unit (ReLu) is used as activation functions in this work. ReLu removes the problem of over fitting and makes the model more adaptable to real world cases. Pooling Layer is also called as sub-sampling layer. The main function of Pooling Layer is to generate down sampled feature maps and minimizes the dimensions without losing the useful information. Consequently, it diminishes the number of parameters to learn and the amount of computation acted in the network. The common idea is max pooling which is nothing but sliding a patch over the network and for each patch the maximum value is taken in the patch. Fully connected layer is feed forward neural network. In fully connected layer all the inputs form one layer are connected to every activation unit of the following layer.
The main objective of utilizing Fully Connected layer is that it transforms the 2D feature map into a 1D feature map.

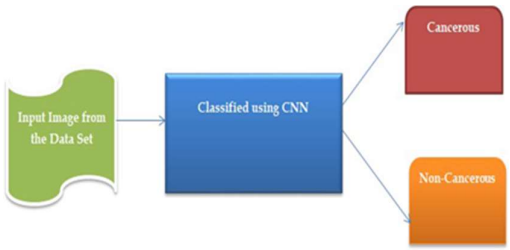

Figure 1. Block diagram of the proposed work

3.1 NETWORK ARCHITECURE

Figure 2. Gives the CNN architecture of the prospective work. CNN [15] has input layer, hidden layers, and a output layer. However, a hidden layer is comprised with convolutional layer, a pooling Layer, and a fully Connected layer. The high-level features are extracted at convolution layer. The input image of dimension 128×128 are considered as 2D vector is passed to the first convolution layer of size 128×128 with 32 filters, after applying batch normalization the second convolution layer feature vectors are determined and the size of the image is 128×128 with 64 filters. At the convolutional layer 3, the image is reduced to a dimension of 64×64 with 128 filters and at layer 4 the image is reduced to a dimension of 32×32 with 256 filters. The function that is used to activate the neurons is the ReLU activation function. The pooling layer employs max pooling with stride 2×2 and generates the feature maps with the minimized dimensions. In order to increase the accuracy the rate of the dropout layer is 0.2. Activation dropout provides boost in performance on unseen data in the test phase. In the dense layer, 100 neurons are created and the inputs are flattened so that the 2D feature

map is converted into a 1D feature map wherein the inputs can be mapped with the output by a onto function

.

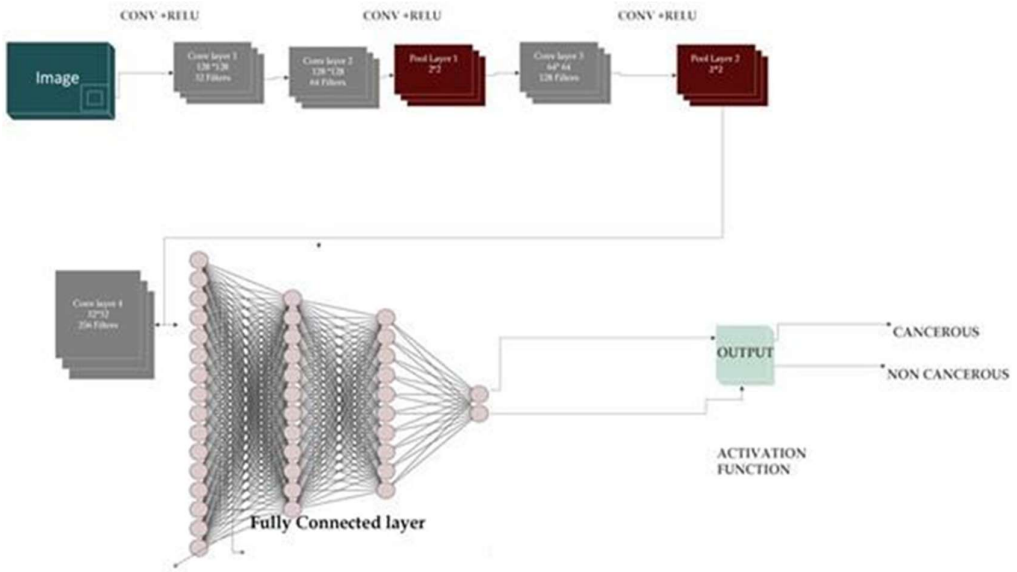

Figure 2. Architecture of CNN

3.2 Dataset

The dataset used for colon cancer classification is CT (Computed Tomography) colonography [12]. In 2004, the NCI (National Cancer Institute) Executive committee conducts the National CT Colonography trial that shares the public access of image data of colon cancer, which would provide a significant research resource to a wide range of research community. The overall count of images considered is 1000 from which 700 images are adopted for training and 300 images are employed for testing phase and the modalities which have been used for this data are CT. The 825 cases provide the polyp description and their locations. In 825 cases, 582 are positive cases and 243 are negative cases. The descriptions of the polyp and the location of the polyp in the colon segments are provided in the XLS sheet. The supine and prone DICOM (Digital Imaging and Communication in Medicine) images can be downloaded from the CT Colonography collection. The dataset contains number of image samples and it is expressed as, $D = \{I, I, ...I, ...I\}$ Here, I_n indicates the total count of samples and I j indicates selected input sample with a dimension[u × v] .

3.3 PERFORMANCE CRTERIA

The performance of developed CNN is analyzed with respect to evaluation metrics, such as Confusion Matrix, accuracy, sensitivity, and specificity.

(i) Confusion Matrix: It is employed to find the Prediction of the classifier model. True Positive is the forecasting of positive classes and True Negative is the indicator of negative classes. False Positive is the conclusion when model faulty predicts the positive class where the original is negative and false negative is the indicator when model wrongly predicts negative where original is positive. Where, T_P and T_N denote the true positive and true

negatives, respectively. However, false positive is represented F_P and F_N denotes the false negative

		Actual Values	
Predicted Values		**Positive**	**Negative**
	Positive	**T_P**	**F_P**
	Negative	**F_N**	**T_N**

(ii)Accuracy: Accuracy is defined as the proportion of corrected predictions of cancer to the total number of colon cancers examined.

$$\text{Accuracy} = \frac{(T_P+T_N)}{(T_P+F_P+T_N+F_N)} \quad \text{---------- (1)}$$

(iii) Sensitivity: Sensitivity refers the correct identification of colon cancer affected patients.

$$\text{Sensitivity} = \frac{T_P}{(T_P+F_N)} \quad \text{---------- (2)}$$

(iv) Specificity: Specificity is defined as the correct identification of unaffected colon cancer patients.

$$\text{Specificity} = \frac{T_N}{(T_N+F_P)} \quad \text{---------- (3)}$$

3.3 EXPERIMENTAL RESULTS

Figure 4 a), and 4 b), shows the cancer image and non-cancerous image. The input image of size 512*512 is taken from the CT colonography. The images of this Size have been resized to 128*128 before passing it to the CNN and the further processes are carried out with these images. Table.1 shows the results in the form of confusion matrix. Table.2 Portrays performance metrics, like accuracy, sensitivity, and specificity. In order to configure the network and reduce the over fitting of data SGD (Stochastic Gradient descent) [14] optimizer is used. Then accuracy, loss curve is drawn during the training process which is shown in Figure 3.

Table1

	T_P	T_N
F_P	500	22
F_N	0	478

Table2

Metrics/Methods	**Proposed CNN**
Accuracy	**89.1%**
Sensitivity	**93.2%**
Specificity	**79.45%**

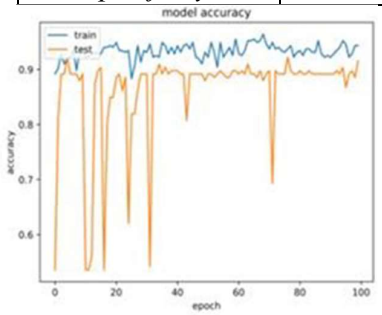

Figure 3. Training and Validation for accuracy and loss for 100 epochs.

Loss Curve is a snapshot of training phase which is used to debug a neural network and accuracy gives the progress of neural network The performance analysis of developed CNN is evaluated by considering the training and test data is split in the ratio of 70% and 30%.

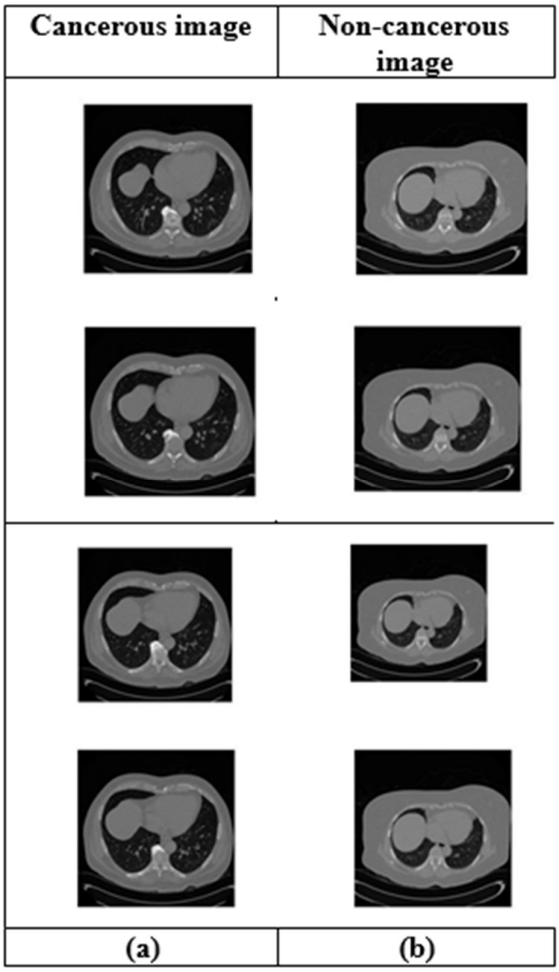

Figure 4 (a). Cancerous and (b). Non- Cancerous Images

4. CONCLUSION

In this work CNN is used for the classification of Colon Cancer. It is a binary classification problem. The Proposed CNN gives an accuracy of 89.1%. The accuracy can be achieved more by fine tuning parameters.

A major setback of CNN is the Max Pool Operation which is Slow if the number of layers increases. If the input images comprises of rotation or tilt then using ConvNet it is difficult to classify the image. Max Pooling layer is used as a messenger for transferring the information from one layer to next layer but it doesn't transfer the spatial information between the parts of an image. The Drawbacks of CNN will fueled further research by using the Transfer Learning techniques using Alex Net, Google Net and by using recurrent neural networks.

5. REFERENCES

[1] Classification tree analysis: a statistical tool to investigate risk factor interactions with an example for colon cancer (United States) accepted in revised form 1st June 2002.

[2] Kim, K.J. and Cho, S.B., "Prediction of colon cancer using an evolutionary neural network", Neurocomputing, vol.61, pp.361- 379, 19th Nov 2003.
[3] Identification of candidate colon cancer biomarkers by applying a random forest approach on microarray data on 8th June 2012.
[4] Cancer classification based on gene expression using neural networks Published December 21, 2015
[5] RCCNet: An Efficient Convolutional Neural Network for Histological Routine Colon Cancer Nuclei Classification arXiv:1810.02797v3 [cs.CV] 8th June 2019
[6] Salmi, N. and Rustam, Z., "Naïve Bayes classifier models for predicting the colon cancer", In IOP Conference Series: Materials Science and Engineering, IOP Publishing, Vol. 546, no. 5, p. 052068, 12th June 2019.
[7] Dual CNN Models for Unsupervised Monocular Depth Estimation accessed on 6th Nov 2019.
[8] Loey, M., Jasim, M.W., El-Bakry, H.M., Taha, M.H.N. and Khalifa, N.E.M., "Breast and colon cancer classification from gene expression profiles using data mining techniques", Symmetry, vol.12, no.3, p.408, 4th March2020
[9] Ganeshan R, "Skin Cancer Detection with Optimized Neural Network via Hybrid Algorithm", Multimedia Research, vol.3, no.2, April 2020.
[10]Khare, S.K. and Bajaj, V., "Time- frequency representation and convolutional neural network-based emotion recognition", IEEE transactions on neural networks and learning systems, 31st July 2020.
[11] Prabhakar, S.K., Rajaguru, H. and Kim, S.H., "An Amalgamated Approach to Bilevel Feature Selection Techniques Utilizing Soft Computing Methods for Classifying Colon Cancer", BioMed Research International, 13th October2020.
[12]CTColonogarphy,"https://wiki.cancerimagingarchive.net/display/Public/CT+COLONOGRAPHY#e88604ec5c654f60a897fa77906f88a6", accessed on 6th June 2021.
[13]Malignant mesothelioma of the pleura: interobserver variability accessed on 6th June 2021
[14]Combining Optimization Methods Using an Adaptive Meta Optimizer accessed on 19th June 2021.
[15]Image Net Classification with Deep Convolutional Neural Networks accessed on 13th July 2021.