Colon Cancer Biopsy Image Analysis using Deep Learning

Anuja A. Ubale, S. G. Shikalpure

Abstract: ***According to WHO (world health organization) colon cancer is third most commonly identified cancer in male as well as female. It is cancer of colon or rectum, located in inner side of colon wall or rectum wall in large intestine. There are different method used to diagnoses and staging of colon cancer like, colonoscopy, blood testing, biopsy method .Clear analysis of these method used for diagnose the disease. In these method some methods are time consuming if approach is manual analysis and perform on medical images. That's the reason it required some automated, efficient and accrete machine or technique. To design such machine biopsy image analysis by deep learning play an important role to design and creation of machine that used for biopsy method diagnoses. We work on three model, CNN, VGG16 and InceptionV3 .it provide different accuracy. CNN gives 72%, VGG16 gives 71% and InceptionV3 gives 89% accuracy .Hence proposed method used to significant accuracy improvement automated diagnoses system.***

Index Terms**: *Deep learning, CNN, VGG16, IncepctionV3, Colon Cancer.***

I. INTRODUCTION

Cancer is Collection of related diseases in which some of the body cells begin to Divide without stopping & spread into surrounding tissues.[1]Basic unit of human body is Cells as per body need, there is development of cell like it grows, divide and may create new cells. And when cell is too old it die, then new cell taken their place. Cancer begins when genetic interface with this cell order process when cell start grow uncontrollably and spread into surrounding that may create mass, called tumor. Tumor can be benign or cancerous. Cancerous tumor called as malignant, that can grow and spread to other parts of body. Following are type of cancers, like Carcinomas, sarcomas, leukemia's, lymphomas. In Carcinomas begins in skin or tissues that cover organs &gland these is basically solid tumor, this is most common type of cancer. Examples of Cancer in these is, lungs cancer, prostate cancer, colorectal cancer.[1]In these paper we working on colorectal cancer Colon Cancer happen when tumorous growth expand in large intestine. In US it is third most frequent type of Cancer. It may causes death. In 2017, 95,520 new diagnoses are excepted to occurs [2]. There are different type of methods used to screening colon cancers. Like high sensitivity focal occult, blood test, stool-DNA test, colonoscopy and sigmoidoscopy. Biopsy method are applied under colonoscopy and sigmoidoscopy. Microscopic inspection study of that biopsy samples may create result but it is laborious task and time consuming for histopathogists lead to significant inter observer variation in grading. In these heterogeneity of feature in some region also effects diagnosis. Image classification is techniques used for it. The main focus to finding malignant and normal tissues by classification method apply on colon biopsy images. Different methods used for detection and classification of cancer images. In Korusk Sirinukunwattna et al. used data set of tissue images, with cell nuclei in histopathology images. They used Spatially Constrained convolution neural network (SC-CNN) for nuclei detection and for classification of nuclei novel neighboring ensemble predicator (NEP) it more accurately predict class labels of detected cell nuclei [4] In Erdem Ozdemior and Cenk Sokmensuer work with markovian module for automated colon cancer on tissue dataset they mainly work on robust images variation. They propose new resembling framework to simulate variation in tissue images .So performance of generalization capacity of learn by size variation and improve performance .But there may future work different feature extraction and incorporate different features in these framework.[5] .In Aparna Ratheesh et al. Proposed advanced algorithm for polyp detection using depth segmentation in colon endoscopy. They use two type of segmentation methods, linear thresholding is used to find sutured region by HSV images and use markovian random field for depth wise segmentation. Then use SVM classifier to predict disease condition using color corrlogram vectors and texture vector. [6]. By Omeria Bradhi in paper of Automatic colon polyp detection by CNN encoder and decoder module. They work on their different data set with CNN auto encoder with 3 different dataset. They are not work with images processing at starting using on dataset like augmentation it may increase the performance.[7] Akshay M Godkhindi et al. proposed to apply convolution neural network and some machine learning algorithm like k-nearest neighbor (KNN) and random forest. For extraction of texture features use LBP and for shape feature use the HOG .For analysis they use dataset (TCIA) from cancer imaging archiver .On the result they told that deep learning accuracy will better than machine learning algorithm they use.[3] For Cancer prevention when applying existing knowledge, need to increase the use of new theology, screening tests and recommended accurate diagnosis by different methods, to check that all the patient receive timely, standard treatment and prevent death.

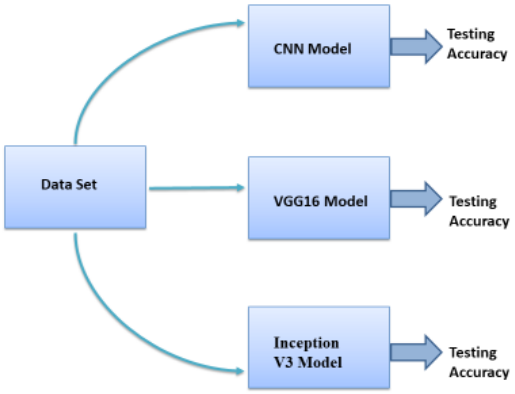
Revised Manuscript Received on July 22, 2019.

Anuja Ubale, Department of Computer Science and Engineering Government College of Engineering, Aurangabad, Maharashtra, India. E-mail:Ubaleanuja@gmail.com

Assistant Prof. S. G. Shikalpure, Department of Computer Science and Engineering ,Government College of Engineering, Aurangabad, Maharashtra, India. E-mail:Shikalpure@geca.ac.in

II. PROPOSED ARCHITECTURE

For image analysis or processing using deep learning CNN play an important role.VGG16 ,IncepctionV3,ResNet ,AlexNet this all model based on neural network. For biopsy image analysis we us following architecture as show in fig(a) . It show that how data set is use for different model analysis.



Fig(a). Flow for colon cancer Biopsy image analysis

Detail explanation of above model use in proposed architecture is explained in methodology. For Implementing above model we used kaggle kernel ,because it take less time compared to run on common specification computers .Kaggle proved us workspace in python ,GPU services and Cloud storage on which we can work easily.

A. *Dataset*

For prediction of cancer use the dataset from MNIST dataset of colorectal and cancer category from kaggle web at CIFAR10.Get dataset *from(https://www.kaggle.com/kmader/colorectal-histology-mnist*).Which contain data with different size of images ,with different type of biopsy tissues like Tumor, stroma ,Complex, lympho, debris, mucosa, adipose, empty. These 8 class play important role for applying supervised classification. It contain 5000 images .each class contain 625 image with 64*64 size .Data set contain 2400, nuclei which are being and malignant. [8]

B. *Methodology*

CNN

Convolution Neural network (CNN) has been successfully applied in carious medical image diagnosis such as, breast cancer prediction etc. CNN module is based on local respective field, weight sharing, subsampling in the spatial domain [9]. In CNN contain mainly three layers. Convolution layer, subsampling layer, output layer. They are arranged in feedforward network. In that neuron receive sum input, performs a dot product and optionally follows it with a non-linearity [10].Subsampling are hidden pooling layer, here CNN work on 2D image. Each plane firstly computes the convolution between 2D input and convolution mask. That output submersed and together add with adjustable scalar known as bias team. Then apply activation function which is different for layers. Convolution layer gives output with visual feature and given pixel location pass to subsampling layer. Following fig (b) shows the concept of CNN .Input layer (224*224*3) which hold image in pixel format of width, height and color of RGB. Convolution layer will compute output in neuron, that computation contain dot product between weight and small region connected in input column. They get result in [224*224*64] with 16 filters. RELU apply on activation function element wise max (O, X) and mean (o, x). POOL layer dividing image into non overlapping block. For each block sum of pixel Value is calculated with spatial dimensions (width and height) they may result [112*112*128].Fully connected layer will compute the class score ,so get volume as [1*1*1000] A ordinary neural network and as name impels each neuron in this layer will be connected to all previous volume.[10]

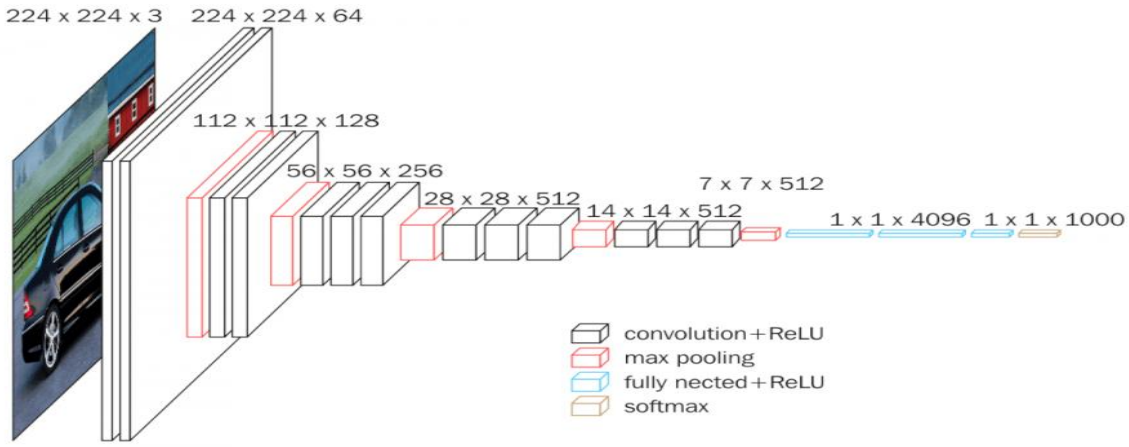


Fig (b). Neural Network Layer [10]

VGG16

One of the convolution natural network architecture is VGG16 i.e.-Visual geometry group oxford. In these module weight are pertained on ImageNet. Which is used for large scale image recognition and deep convolution network. Fig (c). Shows the flow chart of VGG16, which contain different layers as fig. (c)Indicates in different colors likes, convolution layer, Pooling layer and fc as flatter and Dence layer. Where in proposed architecture also use soft max layer which is customized layer at end of module. In these module use keras package to visualize input which contain different activation filters in different layers of VGG16.

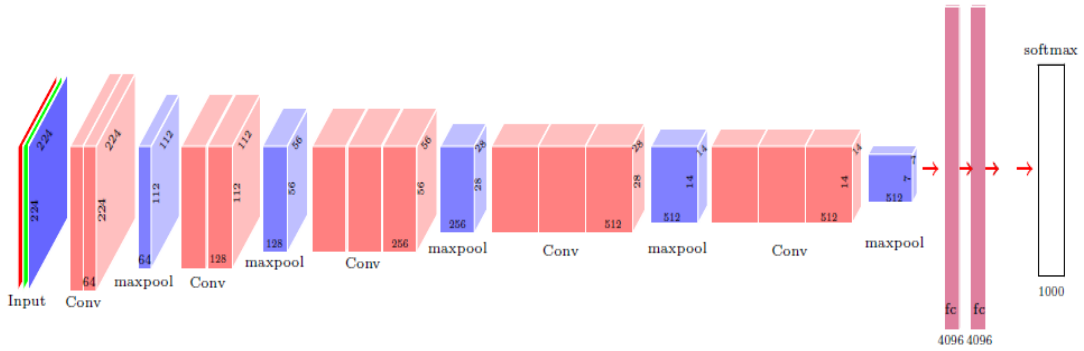


Fig (c).Structure of VGG16

Convolution 2D is convolution layer applied on different stages. Convolution layer will compute output of neuron that connected to local input .It contain dot product of weight and region they connected. Max pooling layer reduce spatial dimensionality of output volume. Where in dense layer fully connected layer where all neurons in layer are connected to next layer. The drop out drops connections of neuron from the dense layer to prevent overfitting. Flatter layer used in keras for tensor reshape that have equal to number of elements contained in tensor non including batch dimensions. Softmax Layer is applied on last i.e.- output layer where activation to net input coming from previous layer. In softmax layer gives output is equivalent to probability distribution which indicates any class are true.

InceptionV3

Transfer learning is one of the type of machine learning which utilized pre-trained neural network. In that incepctionV3 play very important role .IncpetionV3 is one of the higher version of googleNet, where use CNN Called inceptionNet. Purposed of inceptionV3 is it reduce the high computational overhead if convolution layer by breaking the large convolution into smaller convolution [11] . In contain mainly two parts feature extraction and classification done by CNN and fully connected layer respectively.

Following fig (d) shows concept of inceptionV3 module. In which firstly applied CNN and then pooling with both types, max pooling as well as average pooling, Then use dropout layer which is used to avoid over fitting of neural network, in which randomly select neurons and ignore them during training. At end applied full connected layer and softmax layer. At middle of module use auxiliary classifier for regularizer in inceptionV3.grid size reduction play important role it is less expensive and still efficient network achieved by effective gride sized reduction.

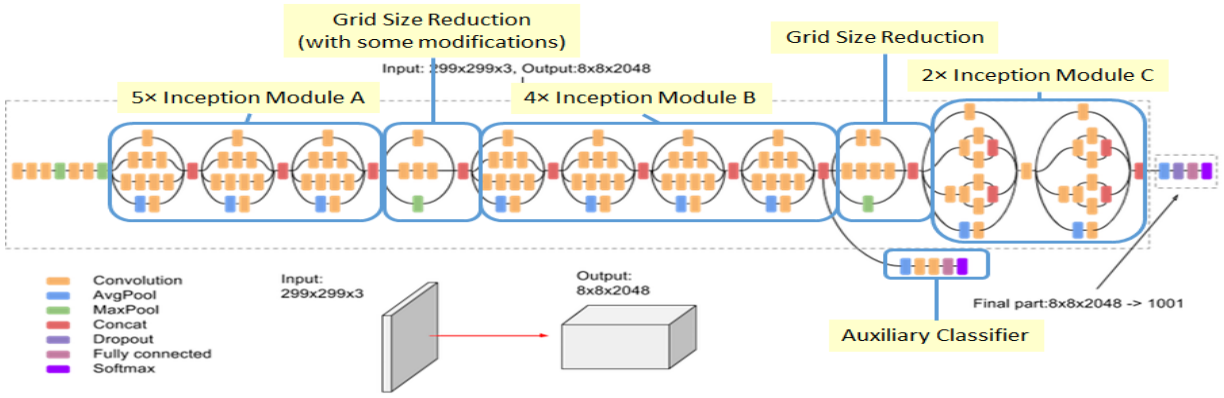


Fig (d) Structure of InceptionV3 [12]

III. RESULT

By applying the proposed method, with module as show above CNN, VGG16 and InceptionV3.All of them IncepctionV3 model shown best accuracy. Following graph shows the testing phase accuracy of all the models. In CNN model use CSV file as input, which contain pixel value of the all 5000 image in data set with (64*64)size In CNN applying some custom layers and activation functions in model creation. In these module use 20 epoch and 200 batch size. It overcome the problem of dividing data into small size and then use to pass the computer for process .Batch size is dividing the data into batch or sets .Epoch is how many time that dada set pass through network .These both the term play an important role for deep learning module. As proposed model indicate VGG16 used to detecting class of cancer. It pass images data set through module with adding custom layer relu and softmax at end .which give accuracy averagely 72% accuracy as show in following fig(e). And also show f1- score of it.f1-score is function to find accuracy based on precision and recall with formula (F1-score= 2 * Precision * Recall / Precision +Recall).

Were precision is relevant instances among the retrieved instances and recall is relevant instances that have been retrieved over the total amount of relevant instances [13].It also represent the confusion matrix which explain the classification of model. As show in fig (f).Confusion matrix is used to describe performance of classification model on test data for which true value are known.

```
[130   0   0   6   0   0   0   0]
[  0 129   1   0   8   2   4   5]
[  2   5 112   0   1   1   6   5]
[  2   0   0 119   0   0   0   0]
[  0   7   0   0 118   0   0   2]
[  0   3   1   0   1 107   0   2]
[  0  29  12   0   1   0  68   0]
[  0  13   0   0   3   1   1  93]]
```

Fig (f).Confusion Matrix values of VGG16 model

As proposed architecture indicate that use of inceptionV3 model for biopsy image analysis .It's give us best accuracy compare to above two model. It gives average 89% accuracy. Following fig (g) indicate we get best confusion matrix value and also f1-score of it. But compare to above model it take more time to run (depend on batch size and epoch).

```
[[127   0   5   4   0   0   0   0]
 [  0 108   3   0   8  12  15   3]
 [  1   2  91   0   0  10  28   0]
 [  6   0   0 114   0   0   1   0]
 [  0   1   0   0  95  14   0  17]
 [  0   3   3   0   1  94   3  10]
 [  0  20  29   0   0   5  56   0]
 [  1   2   0   0   0  12   0  96]]
```

Fig (g).Confusion Matrix values of incapctionV3 Model

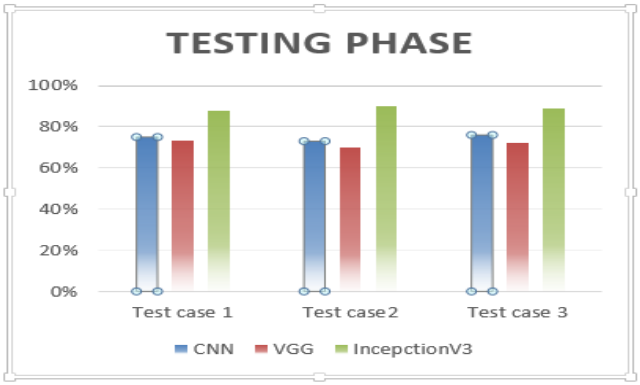


Fig (e). Testing phase accuracy

As per above analysis of all method on same data set, Testing phase indicate incepctionV3 has high accuracy compare other models show in fig (e). It may possibility that when custom layer change accuracy and complexity varies, but Inception V3 take more time compare to VGG16 and CNN.

IV. CONCLUSION

In this paper implemented the different deep learning methods for colon cancer biopsy images. We study the different techniques used before for the prediction of colon Cancer in Literature Survey. As above Implemented method indicate IncpectionV3 is gives best accuracy for colon cancer reconnection on biopsy images. In implementation we are not applying image preprocessing .In given model we directly apply deep learning on given data set. For future work of this analysis use of the image proposing on data set, may improve result. The analysis of biopsy images will provide foundation for further research in colon cancer diagnosis and detection using deep learning. The Implemented method can also use in other medical imaging diagnosis.

ACKNOWLEDGMENTS

We would like to thank Dr. P. B. Murnal. Dr. V. P. Kshirsagar, Assistant Prof. S.G. Shikalpure for their valuable support and guidance and Department of Computer Science and Engineering, GECA, Aurangabad for providing healthy environment for research work.

REFERENCES

1. S. Rathore, M. Hussain, A. Ali, A. Khan, "A recent survey on colon cancer detection techniques," IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2013.
2. History about cancer "www.medicalnewstody.com"
3. Akshay M Godkhindi et al., "Automated detection of polyps in CT colonography image using deep learning algorithm in colon cancer diagnosis" International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS-2017) :1722-1728
4. Korsuk Sirinukunwattana, et al., " Locality Sensitive Deep Learning for Detection and Classification of Nuclei in Routine Colon Cancer Histology Images" IEEE Transactions on medical imaging , Vol. 35, NO. 5, MAY 2016 .
5. Erdem Ozdemir et al ., "A Resampling-Based Markovian Model for Automated Colon Cancer Diagnosis" IEEE Transactions on Biomedical Engineering, Vol. 59, No. 1, JANUARY 2012
6. Aparna Ratheesh et al., "Advanced Algorithm for Polyp Detection Using Depth Segmentation in Colon Endoscopy" 2016 International Conference on Communication Systems and Networks (ComNet) 21-23 July 2016
7. Ornela Bardhi1 et al., "Automatic colon polyp detection using Convolutional Encoder-Decoder model "2017 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)
8. https://www.kaggle.com/kmader/colorectal-histology-mnist.
9. Havaei, Mohammad, et al. "Brain tumor segmentation with deep neural networks." Medical image analysis 35 (2017):18-31.
10. Convolutional Neural Networks for Visual Recognition :http://cs231n.github.io/convolutional-networks/
11. Kalyani wadkar et al." Breast cancer detection using ANN network and performance analysis with SVM" International Journal of Computer Engineering & Technology (IJCET) Volume 10, Issue 3, May-June, 2019, pp. 75-86,
12. Inception v3 digram.https://medium.com/@sh.tsang/review-inception-v3-1st-runner-up-image-classification-in-ilsvrc-2015-17915421f77c
13. https://en.wikipedia.org/wiki/Precision_and_recall

AUTHORS PROFILE

First Author: Received the B.E. Degree in computer science and Engineering from Deogiri Institute of engineering and management studies, Aurangabad. Under Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, Maharashtra, India in 2017, and currently perusing M.E degree in Computer Science and Engineering from Government College of Engineering, Aurangabad under.