Automatic Classification of Ovarian Cancer Types Using Cytological Images With the help of DCNN

Raubins Raj [1], S.S.Bhosale [2], Arati Role [3], PoojaPrajapat [4]
[1,2,3,4] *Dept. of E & TC Engg., Smt. Kashibai Navale College of Engineering, Savitribai Phule Pune University, Pune*

[1]*raubinssinghh@gmail.com*
[2]*sonali.bhosale_skncoe@sinhgad.edu*
[3]*aratirole02@gmail.com*
[4]*prajapatpooja111@gmail.com*

Abstract

Ovarian cancer is one of the most common gynaecologic tumours and using manual methods it is difficult to detect it, so to ease the hectic process and reduce time required to detect we have developed a process. In this project computer aided diagnosis (CADx) is used which helps the pathologists to determine and to diagnose the tumour correctly. There are basically four types of ovarian cancers.They are serous carcinoma, mucous carcinoma, endometrial carcinoma, clear cell carcinoma. Here cancer types will be detected by using Alex net architecture based on DCNN. In this project, input images areof cytological type. Theimage manipulation is done using different filters to remove noise and obtainsharpened image. There are two types of inputs to DCNN which are training set and testing set. In DCNN there are five convolution layers, three max pooling layers and two full reconnect layers. Input image is passed through DCNN module and respectively type of cancer will be detected.

Keywords*—CADx, DCNN, Alex net, Clinical Diagnosis, Gynaecologictumour, Cytological Images.*

I. INTRODUCTION

Ovarian cancer is the most aggressive and frequent gynaecologic cancer. Primary epithelial ovarian carcinoma are sub classified into subtypes as - serous, mucinous, endometrial, and clear cell.It is usually gets difficult to precisely differentiate the four subtypes from cytological images with the help of only pathologists' eyes and capability; especially when the images to be analysed and diagnosed are largein number, errors can occur.

In order to improve the accuracy of diagnosis and reduce pathologists' workload, we have used computer technology which follows certain methods in doing the pathologic diagnosis.Computer-aided diagnosis (CADx) schemes can potentially make a differential diagnosis more accurate and independent on the skill of the pathologists.

A large amount of training data is required in a deep neural network model. Overfitting and other mistakes can be observed due to insufficient size of training. In our study, wehave increased the sample size manipulating images so as to improve the accuracy of classification. Image manipulation includes image enhancement and image rotation. To improve the image clarity and edge sharpness a Gaussian High Pass-filter with kernel size = 3*3 and La pass filter were applied to the image. When the image was acquired by the microscope and camera the directions of H&E stained tissue sections was invariable. So,to increase the sample sizeswe rotated the original images (size: 227*227) from0◦ to 270◦ in 90 steps around their centrepoint .Image enhancement and rotation process is shown in Figures 2 and 3. Two independent recognition models were made by our two-group data, one group used original image dataset as training data without image augmentation, and the other one used image dataset augmented as training data, whose sample size was 11 times bigger than original image sets.

ISSN: 2233-7857 IJFGCN

II. LITERATURE SURVEY

[1] AUTHOR: Rinki Singh; AnupSom
TITLE: “Profiling of Ovarian Cancer Reveals Common Features Shared by Sub-types of Ovarian Cancer”

Ovarian cancer is the most lethal gynecological cancer affecting women and hence an important public health issue. Identifying the underlying biology and neoplastic progression of ovarian cancer is very important to understand and advancing the treatment of the disease. The aim of this study is to identify the crucial genes which are common in all the cases of ovarian cancer irrespective of their subtypes, suggesting essential neoplastic progression and the common molecular mechanism through which cancer cells progress. These genes might prove to be good therapeutic drug targets. Our analysis revealed 320 genes, which functionally interacted among different subtypes of epithelial ovarian cancer. JUN, FOS, MYC, STAT5B, NR3C1, KRAS, PNN, SFN, SMARCA4, GSK3B, JUND, CAV1, and MAPK13 are genes played crucial role in ovarian cancer progression and thus can act as good therapeutic targets.

[2] AUTHOR: Hye-Jeong Song; Eun-Suk Yang; Jong-Dae Kim; Chan-Young Park ; Yu-Seop Kim ; Min-Sun Kyung
TITLE: “Improving performance for classifying ovarian cancer with menopause information”
Tumor biomarker testing is a relatively simple test using blood, which allows low-cost cancer screening. In previous studies, a diagnostic model of ovarian cancer was made with a combination of two or three optimal biomarkers. This study proposes a classification model of ovarian cancer screening by adding menopausal information, which is important clinical information for diagnosis of ovarian cancer, to biomarker combination. The classification model was evaluated by the area under the ROC curve (AUC). The performance of the classification model including menopausal clinical information showed better performance than the classification model which did not include clinical information.

[3]AUTHOR: Andrew Janowczyk ;SharatChandran ; Rajendra Singh ; DimitraSasaroli ; George Coukos ; Michael D.Feldman ; AnantMadabhushi
TITLE: “High-Throughput Biomarker Segmentation on Ovarian Cancer Tissue Microarrays via Hierarchical Normalized Cuts”
We present a system for accurately quantifying the presence and extent of stain on account of a vascular biomarker on tissue microarrays. We demonstrate our flexible, robust, accurate, and high-throughput minimally supervised segmentation algorithm, termed hierarchical normalized cuts (HNCuts) for the specific problem of quantifying extent of vascular staining on ovarian cancer tissue microarrays. The high-throughput aspect of HNCut is driven by the use of a hierarchically represented data structure that allows us to merge two powerful image segmentation algorithms-a frequency weighted mean shift and the normalized cuts algorithm. HNCuts rapidly traverses a hierarchical pyramid, generated from the input image at various color resolutions, enabling the rapid analysis of large images (e.g., a 1500 × 1500 sized image under 6 s on a standard 2.8-GHz desktop PC). HNCut is easily generalizable to other problem domains and only requires specification of a few representative pixels (swatch) from the object of interest in order to segment the target class. Across ten runs, the HNCut algorithm was found to have average true positive, false positive, and false negative rates (on a per pixel basis) of 82%, 34%, and 18%, in terms of overlap, when evaluated with respect to a pathologist annotated ground truth of the target region of interest. By comparison, a popular supervised classifier (probabilistic boosting trees) was only able to marginally improve on the true positive and false negative rates (84% and 14%) at the expense of a higher false positive rate (73%), with an additional computation time of 62\% compared to HNCut. We also compared our scheme against a k-means clustering approach, which both the HNCut and PBT schemes were able to outperform. Our success in accurately quantifying the extent of vascular stain on ovarian cancer

TMAs suggests that HNCut could be a very powerful tool in digital pathology and bioinformatics applications where it could be used to facilitate.

[4] AUTHOR: Beant Kaur; Kulvinder Singh Mann; Manpreet Kaur Grewal.
TITLE: "Ovarian cancer stage based detection on convolutional neural network"
Ovarian cancer is the fifth most common cancer affecting women today. Ovarian cancer is a cancer that begins in the ovaries. The ovaries are female generative organs situated in thepelvis, approximately the size of an almond. The ovaries produce eggs (ova) for reproduction. In this research paper, detect the ovarian cancer and found the stage of the cancer in the malignant cancer image. The proposed algorithm is used to feature extraction technique using SIFT algorithm. Any object there are many features, interesting points on the object, that can be extracted to provide and feature, a description of the object. In genetic algorithm used to optimize the extracted feature with the help of the fitness function. In fitness function depends upon three parameters i.e, each feature, total features and classification error rate. The detection of the ovarian cancer and stages found using a convolutional neural network. The accuracy is achieved with CNN classifier is 98.8% and with SVM is 85.01%. The performance parameters used are Sensitivity Specificity and accuracy.

[5]AUTHOR: ShubhamNegi; Poornima Mitta; BrijeshKumar ;Anamika Bhatia
TITLE: "Detection of Ovarian Cancer using Organic Light Emitting Diodes"
The modern world has been affected by a lot of diseases, and day by day new ones keeps on emerging. But no disease has threatened the mankind as much as cancer has. The reason behind it is that it has taken more human lives than any other disease. According to data of American Cancer Society, in USA alone half a million people will die because of cancer in 2018. Further this disease has a lot of different types which affects a whole lot of people varying in age and gender. Cancer even after all this is still a curable disease, but the problem with it that makes it so dangerous is late detection. If detected at early stages it can very easily be treated. Ovarian cancer is one of the categorization of cancer which has a very low survival rate often because of late detection and diagnosis. In this article a non-invasive sensor mechanism has been proposed for detection of ovarian cancer. This method will result in a portable device that can be used for giving an early diagnosis of ovarian cancer and can easily be used at home.

III. BLOCK DIAGRAM

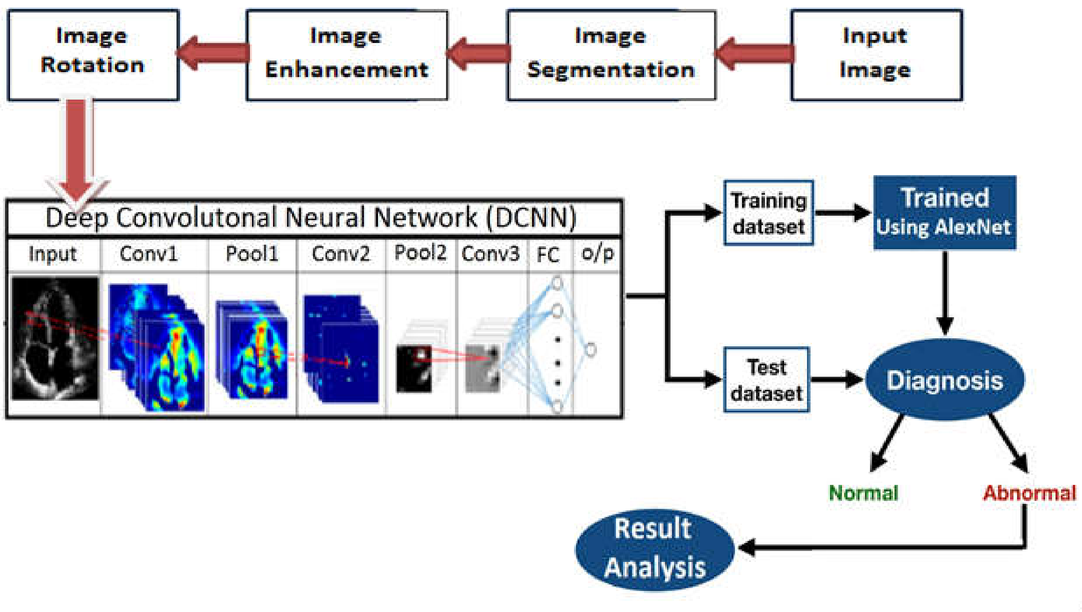

Fig. 1 Ovarian Cancer Detection Using DCNN

ISSN: 2233-7857 IJFGCN

In this project we are using Cytological image as a input. After that Segmentation is performed on the input cytological image. The method of applying different filters to the image is image Enhancement. Segmented image is passed through Gaussian high pass filter to obtain a sharpened image and then we apply Laplace high pass filter to sharpen the edges of the image. To increase the sample size, the next step is torotate image from 0 to 270 degree. Hence image modification is done.

Then we will apply the modified image on DCNN based on Alex Net to automatic classify the ovarian cancer cytological images. The DCNN for our study had five convolutional layers, three max pooling layers, and two full reconnect layers. Each of the layers was followed by a Rectified Linear Unit (ReLU) as the activation function. Three max pooling layers whose size was 3*3 pixels and two fully reconnect layers were applied to reduce the size of image, which was the input of next convolutional layer.

Two full connected layers consisting of a large number of the neurons were applied at the end part of the DCNN. As a layer which is fully connected occupies most of the parameters, it is prone to the overfitting. One method to reduce overfitting was dropout which was employed in our networks. An efficient method for reducing the overfitting is Dropout , and it is usually used to improve the performance of neural networks on supervised learning tasks in vision, computational biology, document classification, and obtaining state-of-the-art results on many benchmark data sets. The dropout rate applied is 50%. The output is the probability, for four ovarian cancer types, which were calculated by the SoftMax function.

A. IMAGE ACQUISITION

To investigate diseases involving a wide range of body sites cytopathology is commonly used, to aid in the diagnosis of cancer but also in the diagnosis of inflammatory conditions some infectious diseases and some infectious diseases. For example, a common application of cytopathology is the Pap smear, a screening tool used to detect ovarian tumour that may lead to ovarian cancer. Cytological images have been used in our project

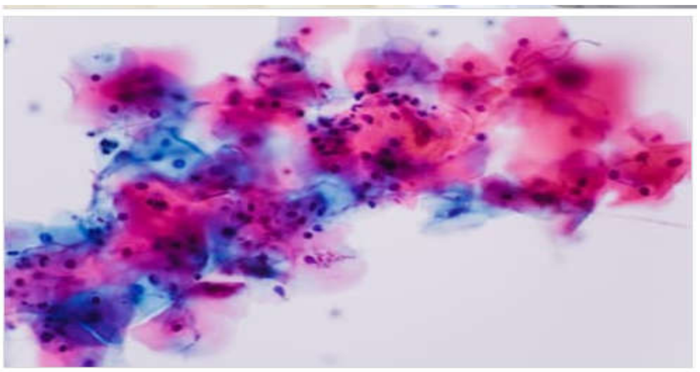

Fig. 2 Cytological image of ovarian cancer

B. IMAGE PROCESSING

To perform image processing on digital images computer algorithms are used. Digital image processing has many advantages over analog image processing in a field of digital signal processing,. It allows a much wider range of algorithms to be applied to the input data and can avoid many problems such as the build-up of noise and signal distortion during image processing. Since images are defined over two dimensions (perhaps more) digital image processing may be modelled in themultidimensionalform. Digital image processing are mainly depends on generation and development and which is affected by three factors: first, the development of first,computers; second, mathematicsdevelopment (especially the creation and improvement of discrete mathematics theory); third, Environmental demandsfor a wide range of applications, agriculture, military, industry and medical science has increased.

C. IMAGE RESIZING AND CROPPING

The most basic image editing functions which is highly used are Resizing and cropping. As they can affect image quality ,henceBoth require careful consideration. The dimensions of the image, which usually affects the file size can be change using resizing(and, thereby, image quality). Cutting away the part of the original image is done by cropping and results in some of the pixels being discarded.

D. IMAGE AUGMENTATION

To achieve good performance deep networks need large amount of training data. Image augmentation is usually required to boost the performance of deep networks and To build a powerful image classifier using very little training data. Artificially creation of training images through different ways of processing or combination of multiple processing is done by image augmentation, such as random rotation, shifts, shear and flips, etc.

E. NOISE PRESENT IN IMAGE

Random variation of brightness or colour information in images are the types of noise in an image , and is usually an aspect of electronic noise. It is generated by using image sensor and circuitry of a scanner or digital camera. Image noise can also come fromfilm grain and also from the unavoidable shot noise of an ideal photon detector. Image noise is an undesirable part of image capture which deviates the process of acquiring the desired information.

F. DCNN

For automatic classify the ovarian cancer we will apply the modified image on DCNN based on Alex Net on the cytological images. The DCNN had five convolutional layers, three max pooling layers and two full reconnect layers. Each of the layers includes Rectified Linear Unit (ReLU) as the activation function. Three max pooling layers whose size was 3*3 pixels and two fully reconnect layers were applied to reduce the size of image, which was the input of next convolutional layer.

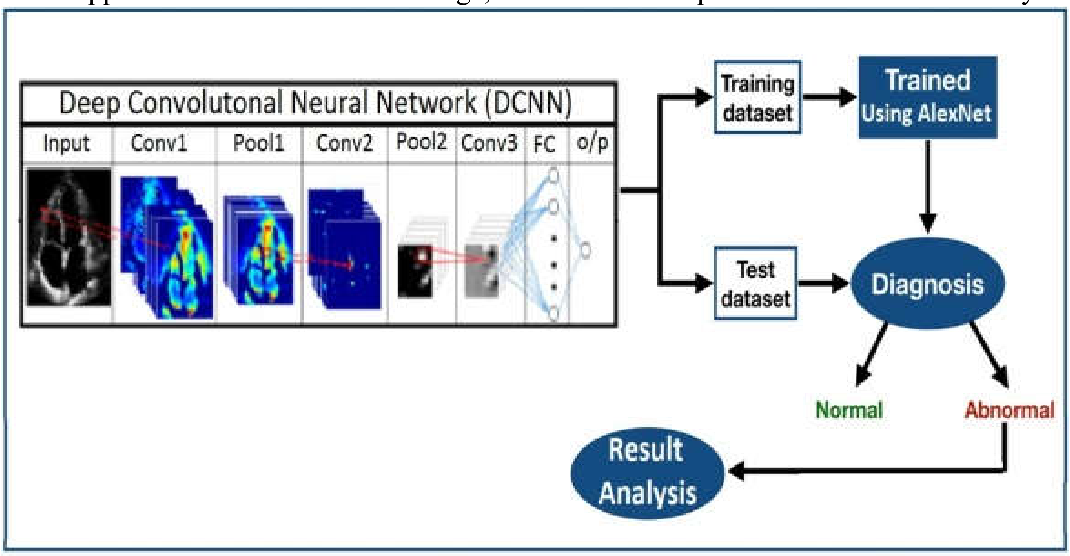

Fig. 3 DCNN

At the end part of the DCNN two full connected layers consisting of a large numbers of the neurons were applied. Because a fully connected layer occupies most of the parameters, it is prone to the overfitting. One method to reduce overfitting was dropout which was employed in our networks. Dropout is an efficient method for reducing the overfitting, and It is usually used to improve the performance of neural networks on supervised learning tasks in vision, computational biology, document classification, and obtaining state-of-the-art results on many benchmark data sets. The dropout rate we applied was 50%. The output was the probabilities for four ovarian cancer types, which were calculated by the softmax function

.

IV. CONCLUSION

In this project computer aided diagnosis (CADx) will be used which will provide the useful advice for pathologists to determine and to do diagnosis correctly. Here we will automatically classify four cancer types by using Alex net architecture based on DCNN. In this project our input image are of cytological type. Then image manipulation will be done using different filters to remove noise and do image sharpening. There are two types of inputs to DCNN which are training set and testing set. Input image is passed through DCNN module and respectively type of cancer will be detected.

REFERENCES

1. Rinki Singh; AnupSom(2015) "Profiling of Ovarian Cancer Reveals Common Features Shared by Sub-types of Ovarian Cancer" .
2. Hye-Jeong Song; Eun-Suk Yang; Jong-Dae Kim; Chan- Young Park; Yu-Seop Kim; Min-Sun Kyung (2018) "Improving performance for classifying ovarian cancer with menopause information" .
3. Andrew Janowczyk ; SharatChandran ; Rajendra Singh ; DimitraSasaroli ; George Coukos ; Michael D.Feldman ; AnantMadabhushi(2016) "High-Throughput Biomarker Segmentation on Ovarian Cancer Tissue Microarrays via Hierarchical Normalized Cuts" .
4. Beant Kaur; Kulvinder Singh Mann; Manpreet Kaur Grewal (ICCES2017) "Ovarian cancer stage based detection on convolutional neural network" .
5. Shubham Negi; Poornima Mitta; Brijesh Kumar; Anamika Bhatia (2018) "Detection of Ovarian Cancer using Organic Light Emitting Diodes".
6. Miao Wu,chuanbo Yan, Huiqiang Liu and Qian Liu (2018) "Automatic classification of ovarian cancer types from cytological images using deep convolutional neural networks".
7. P.Yashodha and N.R.Anathnarayanan (July 2015) "Analysing big data to build knowledge based system for early detection of ovarian cancer",
8. Gurjeet kaur Seerha, Rajneet kaur, "Review on Recent Image Segmentation Techniques", International Journal on Computer Science and Engineering (IJCSE), Vol. 5 No. 02 Feb 2013.
9. WJ Schempp, Magnetic Resonance Imaging: Mathematical Foundations and Applications, John Wiley and Sons, New York, 1998.
10. M NessAiver, All You Really Need to Know about MRI Physics, Simply Physics,5816 Narcissus Ave, Baltimore, MD, 1997

ISSN: 2233-7857 IJFGCN