CANCER RECOGNITION OF PROSTATE GLAND USING DEEP LEARNING, KERAS & TENSORFLOW

Vedant Sharma[1], Prithvee Raj Singh[2]
[1,2]Scholar, Department EC, SSEC, Bhavnagar, Gujarat, India
E-MAIL ID: anilved2530@gmail.com[1], prithvee11901@gmail.com[2]

Abstract-Cancer is characterized as the second catalyst of all-embracing demise and selectively that of the prostate; it is the second variety with an enormous figure of nascent cases disclosed each year in Spain and the most perilous between the male population. A way that can help to diagnose cancer is using Digital Pathology: A biopsy (sample of tissue collected from an organ.) is promptly analyzed under a microscope with the company of a specialist. But, the advancement in digital pathology has led to a different method such as scanning the biopsy under a powerful scanner that collects data from the collected biopsy which can be later analyzed by software. These algorithms are used to detect cancer in the images, using conventional neural networks and computer vision.
Keywords: Cancer, Machine learning, Image processing, Feature Extraction, Deep neural networks, Keras, Tensor flow, Support vector machine, Naïve bayes, k-nearest neighbors.

1. Introduction

As an organism is made up of cells only visible through microscopes. When these cells damage occur that cannot be repaired. When these control mechanisms have an uncontrolled division of descendant's cells, then it leads to a nodule tumor. Now, a nodule tumor is like a lump which can be malignant (cancer).So, because of not 100% effective treatment of cancer it causes death.

Currently, one of the diagnostic programs for detection of cancer is pathological anatomy, which is the branch dealing with the study of morphological techniques and in this the oncologist pathologist (who specialized in analysis of cancerous tissue) will examine the tissue. So, with significant progress in technology there we get an innovative solution provided by the field of digital image processing which we are going to use in this project. In this we will analyze, marking, and labeling the high resolution images known as whole slide imaging through a computer program. Through this method we can find the image is cancerous one or normal.

2. Analysis and Approach

In Early Works, we ordinarily opt for the rudimentary microscopic examination of tissue segments. We have practiced Computer-Aided syndrome Diagnosis (CAD) for exposure. However, in the recent actions, we acquired the technique of developing the models for detecting the parameters and features, and these models are known as the Automated Cancer Detection Models.

Using essential detection techniques for machine vision. Normally, this consists of appropriate characteristics from the images. We do classification using SVM (Support Vector Machine).

Then, applying deep learning techniques, more specifically, a CNN (Convolution Neural Network) is utilized. It's a type of neural network which is concentrated to be used in data analysis and processing. Once the technique has been chosen.

The Analysis and Approach consist of the ML model that is trained to get the best accuracy and get the correct results of cancer. So, to get more exact results. Magnetic Resonance Images (MRI) are utilized in the character of an example model. The tracking down is accomplished

utilizing K-Nearest Neighbor (KNN) with Linear Discriminant Analysis (LDA).

Thermographs and mammograms are also utilized as an example that makes use of the technology support machine vectors (SVM). Automatic cancer discovery models are used that uses different parameters like the domain of interest, the variance of information (VOI), and the false error rate.[1,2]
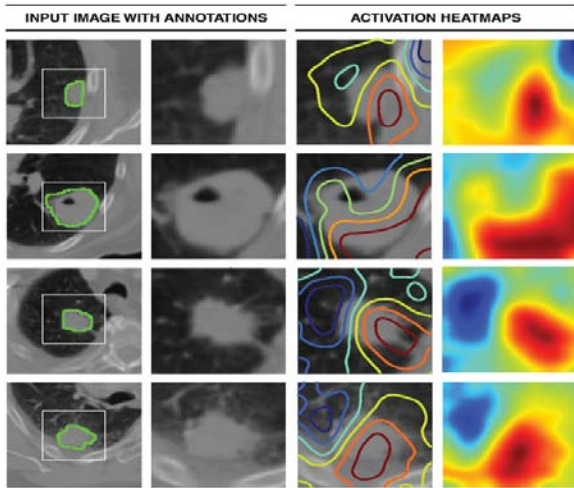

Figure 1: Input Images With Annotation.

3. Proposed Work

Scrutiny of cancer research requires a strong holographic between methodological technique and cancer control initiatives in prevention, early detection, and treatment. [3] The process for detecting cancer is alienated in different stages. The first stage for detecting cancer is very hard because of common symptoms. So, in this method all the images used in this work have been stained with hematoxylin and eosin. In this we will build up algorithm which is capable for detecting the cancer in high resolution images of needle biopsies of the gland, those region are also called as glandular cells.[4]
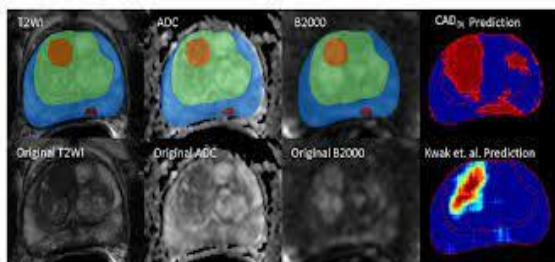
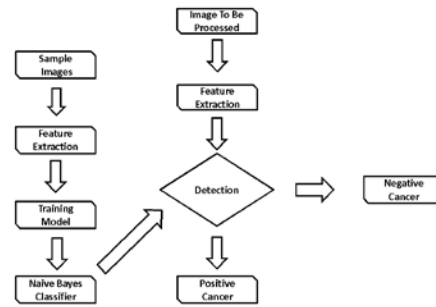

Figure2: detected cancerous cell

Figure3: The above diagram portrays the steps in cancer detection:

3.1 Sample Images: The sample images are used for the implementation of the deep net by processing in a data set or in an algorithm. In sample images we take different types of images which may be cancerous or non cancerous. The different parameters of images are achieved with the help of OpenCv library in python and there are various other libraries which are used for image processing. [5]

3.2 Feature Extraction: Now, it is the most difficult part for feature learning because the human mind is thoughtful and sympathetic to shapes whereas the computer is more delicate to pattern and texture.

The feature extraction is divided in two phases: feature selection and classification. Feature selection is crucial to the entire series of action whereas the classification is not the same. Thus the selection of a feature extraction is depending on which type of input you are allotting.

3.3 The two main objective of dataset is Training and Testing:
1) Training
2) Testing

1) Training: Image training refers to the process based on image in a training database. In this we will drive an algorithm to automatically determine the cancerous cell from the image of the cell.

2) Testing: It is a digital image file used across different institutions to test image processing and images compression algorithms. The testing types of data set which we will use after the Training and in this no outcomes are there for predictions. Now, the new images are compared and classified on the basis of parameters from which we get to know

whether the image of a cell is cancerous or not.

4. Methodology

4.1 Input Image: Input images are microscopic tested images which undergo after the biopsy. All the images used for cancer detection have been blemished using hematoxylin and eosin. First, we will take the images of benign cases in which the glands have customary shapes.

4.2 Image Pre-Processing: In this all the images will endure through many pre-processing processes such as noise removal and enhancement of image.

4.3 Noise Removal: For noise confiscation of an image we mostly use the demonizing algorithm in image processing. In input image we will normally

give the RGB (red green blue) image and it is transformed into a grayscale image. The most common noise in images is white noise, salt and pepper noise etc which is separated by using filters from the extracted image.

4.4 Image Enhancement: Image enhancement is used to improve image quality, so the final image is better than the standard image. The primary purpose of image enhancement is to enhance the appearance of the image or to provide a better transition to default image usage.

4.5 Image Segmentation: In image sharing digital photo separation is done in many stages. In the classification of the cancer image we will classify the image as: nuclei, cytoplasm and other factors separated by the region, group, or limit and a specific algorithm used.
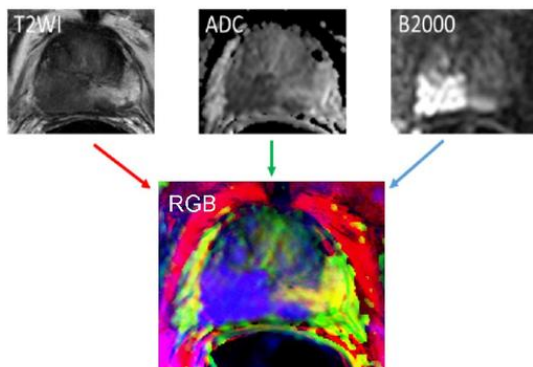

Figure 4: Getting RGB image from input
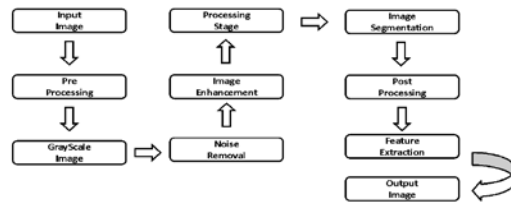
Architectural Diagram

Fig 5: Architectural Diagram

4.6 Feature Extraction: In feature extraction the dimensionality process for reduction is done in which the raw data of the initial set is divided and reduced to more manageable groups. Feature extraction method is also used to capture the visual content of images for indexing and retrieval. There are two types of texture feature measure:

1) Initial order: In the first order, the texture steps are mathematical, calculated from each pixel and do not take into account the pixel neighbor relationships. The strength of histogram and the strength of the first order statistics [1].

2) Second order: In the second system, the steps look at the relationship between the neighbors GLCM second order calculation. In this work, the texture features of Intensity Histogram, GLCM and GLRLM are extracted from the input image provided [1].
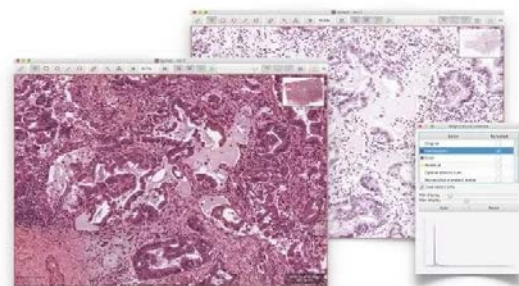
Finally images are classified by Qupath software. [6]

Fig. 6 Example of output image using in Qupath

5. Implementation

The application or implementation has two aspects:

    5.1 Image Processing
    5.2 Machine Learning

5.1 In Image Processing, the module analyses the images as input, and then it is loaded into the program as it is then separated into 12 parts and CNNs (which is Convolutional Neural Networks – Class of Deep Neural Networks) which are applied on each part.

According to the implementation four options are there which are:

1. *Benign Cancer*: It is a harmless tumor that forms around the cells. And it does not spread.

2. *In Situ Cancer:* Stage 0 cancer is considered In Situ Cancer.

3. *Invasive Cancer*: The cancer that spreads around the layer of tissue.

4. Normal Cancer: The normal type of cancer.

These are these four options that will be spotted by using our program or code. In this process, CNN clips the ratio of each option of cancer. Then afterward to conclude the type of cancer the CNN will take the average sum of the 12 parts and stores them in a different file that will be now sent to the machine learning algorithm.[8,9]
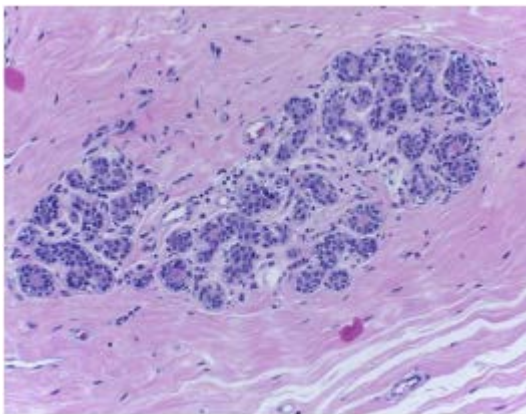

Figure 7: cell image

Now, after we process all the information in the above step. The next step will be the Machine Learning Part.

5.2 The Machine Learning algorithm will kick in and it consists of training and testing aspects.
The previous results were accessed from a file that we stored during the image processing part and the Naive Bayes theorem is applied. Naive Bayes Algorithm will be trained in such a way with data so it shows the final results as positive or negative.

What is Naïve Bayes algorithm?
This algorithm is a technique used for classification with assumption of independence among predictors.
It is a type of classifier that predicts the existence of some particular features. [9]
We'll be using the Naive Bayes Model to train in the testing phase to classify as positive or negative. To decide the result to be positive the percentage of benign cancer needs to be low and the other three aspects need to be higher that will result in the positive result and vice versa. [10]

6. **Result And Experiment**
6.1 Initial learning rate: In this method the learning rate is adjusted by the optimal method and it is also interesting to find the initial value. The procedure for this is as followed:

7 of the 11 available images for this experiment were used. Of these, 25% were present from time to time samples to be used as a participant in validation. This is moving us with 2,508 certified levels and 7,524 training courses.

With this data, Adam optimizes with different initial values of the reading level, several models are trained repeatedly which brings the best selected F1 Score.Study values range from 0.0001 to 100 multiplied by 10 eachstep (to create a line space). Table 4.1 shows the results obtained by each of them. [11]

From 0.01, the steps taken by the optimizer are very broad, skipping close minimum. On the other hand, the winner is obviously 1e-3, this is the first amount of learning scale to be used in another study. [12]

Learning Rate	F1 Score	Loss	Accuracy
1e-4	0.8515	0.3447	0.8698
1e-3	0.8706	0.2951	0.8826
1e-2	0.6248	6.0484	0.6247
0.1	0.6248	6.0484	0.6247
10	0.6241	6.0484	0.6247
100	0.5414	6.0484	0.6247

Table1   Results of the initial LR selection

6.2 Penalization L2:
It is a form of habit that, as its name implies, sets the penalty (L1, L2 ...) for work expenses. With these fines we enforce that you choose the simplest model available one. In particular, we will add a L2 () penalty of 0.01 to the fully connected layer and the final partition of SoftMax.

Two tests were performed: one with this type of habit and one without. As a result, we will be able to better discuss the impact of fines on the model. They were both trained for 25epocastimes, using Adam with the pre-selected Learning Rate as well as the same training and validation data as in the previous trial.[13]

The error in verification begins to grow rapidly while the train continues to slow down. The obvious gap that appears between the two lines is the Iquique sign that it is over productive.[14] This happens when the model reads very well-read samples en (low train error), but it cannot happen before unprecedented examples(high validation error)
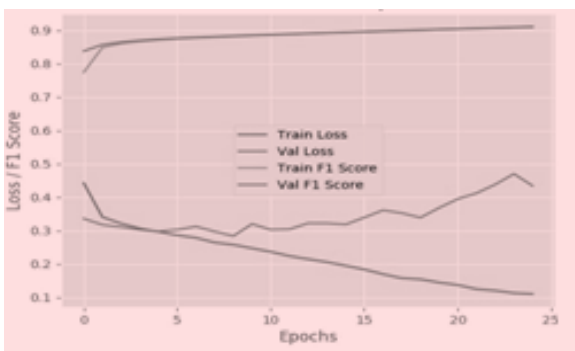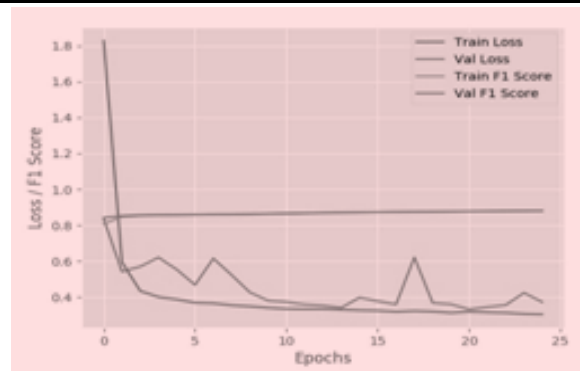

Figure 8: Learning curve L2.

Figure 9: Learning Curve with L2.

	Train	Validation
Loss	0.1100	0.4337
F1 Score	0.9091	0.9098
Accuracy	0.9573	0.8848

Table 2: Results without penalty L2.

	Train	Validation
Loss	0.3045	0.3727
F1 Score	0.8808	0.8812
Accuracy	0.9049	0.8829

Table 3: Results with L2 penalty.

7. Conclusion

As we have seen so far, we understood a way of detecting cancer using image processing and digital pathology.

First, the image is pre-processed which means it will first be converted to grayscale and extra noise will be removed from it and then makes the image usable for further processing.

The image processing is usually done by using the Open library which is widely used for basic image processing.

Then the image is treated with image processing and processed to be then used for feature extraction and training and testing of our machine learning deep networks model that is based on the Naive Bayes algorithm that we trained in such a way so it would first take the average of the four types of cancer that we give as input to the Naive Bayes Classifier and then it

gave the results in positive or negative. And we'll be using Qu path software for showing the results

The results we get may get improved or degraded if we use better data or bad data. It depends a lot on input. Good input equals good output, bad input equals bad output.

The machine learning part is done with the help of TensorFlow libraries with some other supporting libraries for small tasks.

Reference

[1]N. Abbas, T. Saba, Z. Mehmood, A. Rehman, N. Islam, K.T. Ahmed An automated nuclei segmentation of leukocytes from microscopic digital images Pak J Pharm Sci, 32 (5) (2019), pp. 2123-2138

[2]N. Abbas, T. Saba, A. Rehman, Z. Mehmood, N. Javaid, M. Tahir, *et al.* Plasmodium species aware based quantification of malaria, parasitemia in light microscopy thin blood smear Microsc Res Tech, 82 (July (7)) (2019), pp. 1198-1214, 10.1002/jemt.23269

[3]N. Abbas, T. Saba, A. Rehman, Z. Mehmood, H. kolivand, M. Uddin, *et al.*Plasmodium life cycle stage classification based quantification of malaria parasitaemia in thin blood smears Microsc Res Tech (2018) doi.org/10.1002/jemt.23170

[4]N. Abbas, T. Saba, D. Mohamad, A. Rehman, A.S. Almazyad, J.S. Al-Ghamdi Machine aided malaria parasitemia detection in Giemsa-stained thin blood smearsNeural Comput Appl, 29 (3) (2018), pp. 803-818, 10.1007/s00521-016-2474-6

[5]A. Rehman, N. Abbas, T. Saba, S.I.U. Rahman, Z. Mehmood, K. Kolivand Classification of acute lymphoblastic leukemia using deep learning Microsc Res Tech, 81 (11) (2018), pp. 1310-1317, 10.1002/jemt.23139

[6]A. Rehman, N. Abbas, T. Saba, Z. Mehmood, T. Mahmood, K.T. Ahmed Microscopic malaria parasitemia diagnosis and grading on benchmark datasets Microscopic Research and Technique, 81 (9) (2018), pp. 1042-1058, 10.1002/jemt.23071

[7]A. Rehman, N. Abbas, T. Saba, T. Mahmood, H. KolivandRouleaux red blood cells splitting in microscopic thin blood smear images via local maxima, circles drawing, and mapping with original RBCs Microscopic

research and technique, 81 (7) (2018),pp. 737744, 10.1002/jemt.23030

[8]T. Saba, S.U. Khan, N. Islam, N. Abbas, A. Rehman, N. Javaid, *et al.*Cloud based decision support system for the detection and classification of malignant cells in breast cancer using breast cytology imagesMicrosc Res Tech, 82 (6) (2019), pp. 775-785CrossRefView Record in ScopusGoogle Scholar

[9]T. Saba, A. Sameh, F. Khan, S.A. Shad, M. Sharif Lung nodule detection based on ensemble of hand crafted and deep features J Med Syst, 43 (12) (2019), p. 332 View Record in ScopusGoogle Scholar

[10]S. Iqbal, M.U.G. Khan, T. Saba, A. Rehman Computer assisted brain tumor type discrimination using magnetic resonance imaging features Biomed Eng Lett, 8 (1) (2017), pp. 5-28, 10.1007/s13534-017-0050-3 View Record in ScopusGoogle Scholar

[11] S. Iqbal, M.U. Ghani, T. Saba, A. Rehman Brain tumor segmentation in multi-spectral MRI using convolutional neural networks (CNN) Microsc Res Tech, 81 (4) (2018), pp. 419-427 CrossRefView Record in ScopusGoogle Scholar

[12]J. Amin, M. Sharif, M. Yasmin, T. Saba, M. RazaUse of machine intelligence to conduct analysis of human brain data for detection of abnormalities in its cognitive functions Multimed Tools Appl, 2019 (2019), 10.1007/s11042-019-7324-yGoogle Scholar6

[13]B. Mughal, N. Muhammad, M. Sharif, A. ehan, T. Saba Removal of pectoral muscle based on topographic map and shape-shifting silhouette BMC Cancer, 18 (2018), p. 778 View Record in ScopusGoogle Scholar

[14]B. Mughal, M. Sharif, N. Muhammad, T. SabaA novel classification scheme to decline the mortality rate among women due to breast tumor Microsc Res Tech, 81 (2018), pp. 171-180 CrossRefView Record in ScopusGoogle Scholar

[15]J. Amin, M. Sharif, M. Raza, T. Saba, R. Sial, S.A. ShadBrain tumor detection: a long short-term memory (LSTM)-based learning model Neural Comput Appl (2019), pp. 1-9 View Record in ScopusGoogle Scholar