Histopathological Biopsy Image Classification Based On Convolutional Neural Network

Vaibhav K. Kakade[1], Neelam Kumari[2]

Assistant Professor, Dept. of ECE, Maratha Mandal Engineering College, Belgaum, Karnataka, India[1]

PG Student [DECS], Dept. of ECE, Maratha Mandal Engineering College, Belgaum, Karnataka, India[2]

ABSTRACT: Convolutional Neural Network (CNN) is a kind of machine learning classifier used for classifying images based on training and testing networks. In this paper CNN classifier is employed to classify the histopathological images either as benign or malignant tumors. For this paper, lung tissues are taken as an example to demonstrate the computerized tumor detection which replaces the traditional method of observing the histopathological images under microscope. CNN is able to make the decision based on the observed features and stored features. Methodology implemented in this paper consists of two phases training and testing phase, which includes steps: Pre-processing, feature extraction of biopsy image using GLCM and morphological feature, PCA reduction and finally implementation of CNN classifier. In this way our project aims at giving the noteworthy accuracy.

KEYWORDS: Convolutional Neural Network (CNN), Gray level Co-occurrence Matrix (GLCM), Principal Component Analysis (PCA), Feature extraction Benign cells and Malignant cells.

I.INTRODUCTION

Lung cancer is one among the foremost common and deadly cancer within the world. There were an estimated 228,190 new cases and 159,480 deaths in 2013 [9]. Four typical histological types of respiratory organ cancers embody glandular carcinoma, squamous carcinoma, small cell malignant neoplastic disease, and large cell malignant neoplastic disease with the 5-year survival rate below two hundredth, each of that requires a completely different treatment [10]. Consequently, early diagnosing and differentiating these histological varieties is terribly necessary. Bronchial diagnostic test is one of the foremost effective diagnosis ways for respiratory organ cancer. Traditional manual exams of research area unit labor intensive, time consuming, and not too accurate [11] as well.

Therefore, computer aided diagnosis of Histopathological biopsy images is increasing at a faster rate using the image processing technique to give the better and fast result. Histology involves the examination of tiny units of plants and animals. Diagnosis of these medical components via traditional methods is time consuming and not too precise as well. Hence, the fast and accurate result of diagnosed diseases is must at its early stage.

The obvious target of histopathology in clinical medicine is in the assessment of a biopsy images by an enthusiastic physician called a pathologist. At the preparatory stage of maturity cell origination using histopathological images is characterized by the discovery of abnormal uncontrolled cells duplication. Unhealthy cells share out and aggravate to form a tumor that may be benign (non-cancerous) or malignant (cancerous).The abnormality in the remodel of a cell and the cluster of dead cells together are the mark of the presence of malignity in a body tissue.

In this agenda we have exercised on the one type of cancer cells i.e. Lung Cancer. Lung cancer is grouped accordingly its size, cell type, and spread. This staging plus the state of health of the patient are used to determine treatment. It is the most fatal and third most leading disease found in both men and women worldwide. Cancers exist in many forms in living organism. They can be found in liver cells, cerebrum cells, platelets etc. The collection of cancerous cell forms the tumors. They are of two types: Benign and Malignant.

A. Benign cancer cell

Figure 1 shows the benign cancer cell. Benign cancerous cells are less dangerous. These cancerous cells do not spread fast and multiply rapidly to give rise tumors. This is considered as the early stage of the cancer. If diagnosed at the early stage they can be cured.

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering
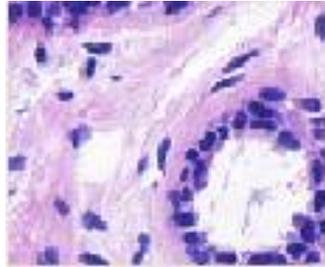
Figure 1 Benign Cancer Cell

B. Malignant Cancer cell
Figure 2 shows the malignant cancer cell. It is the most fatal cancerous cell observed in the living organism. These cells multiply at faster rate and get converted into tumor affecting the surrounding organs as well. Malignancy is the last stage of cancer which cannot be cured.
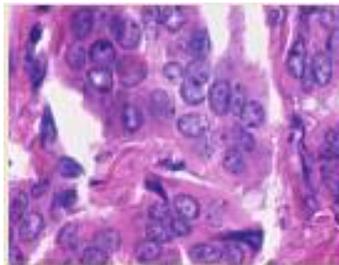

Figure 2 Malignant Cancer Cell

Thus, early detection of these diseases is clinically very important. Hence, the endeavor of this paper is to provide a noteworthy exactness in the diagnosis process so as to classify the biopsy images either as Benign effected cells or Malignant effected cells.

II. LITERATURE SURVEY

M. Gurcan et al. [1] has given attention on the digitized tissue histopathology in an attempt which has currently modernized it into an important use of electronic image investigation and to boot machine learning procedures.
J. C. Caicedo [2] has shown (CAD) for the motive of type and retrieval of picture .In spite of the diversity of interests in solving issues and variations within the microscopic anatomy slide pictures, a trend to measure some main characteristic options in pictures, such as tissue structures like the shading, composition and morphology has additionally been displayed.
C. Demir et al. [3] has presented that the machine steps involved in automated CAD identification is primarily based entirely on histopathology steps like: pre-processing, feature extraction and cantered the malignancy area. The information sets have conjointly demonstrated correlations of the different alternatives and styles to prevent misleading estimation and also these frameworks have permitted the subsets to distinguish one from the other the choices for development of machine-controlled malignancy ID.
In the reference by O. Sertel et al. [4], the procedure proposed has assembled the image into an occasional or excessive grade on the premise of the degree of cytological segments present in it. For division reason the lower assessments are named low and mid assessments an approach of shading surface examination has been used which has adjusted the occasion of dull level in an image slide. The lattice philosophy has been utilized as non-direct shading quantization technique with the vocation of self sorting out element maps.
P.A.Hall et al., [5] has indicated evaluation is done for identifying the three essential useful sorts of a cell, i.e., factual, restrictive restoration, and persistently recharging. The act of histopathology includes immediate or backhanded appraisal of the cell multiplication, the points of interest and drawbacks of the better-known techniques that are utilized for evaluating cell expansion as a part of histopathological slide pictures.
In this paper, X. Zhang et al [6] proposed a capable structure for cell-level investigation of histopathological pictures, by directing CBIR amid a broad sum for Mobile pictures. This huge scale recovery depends on weighted hashing with

different segments, which can separate various signs of cells and model them in hash entries. He connected this system on the separation of two sorts of respiratory organ growths, the threatening neoplastic ailment carcinoma and squamous carcinoma, and accomplished promising precision and strength.

III. PROPOSED WORK

Our proposed undertaking aims at sorting out the two types of cancers i.e. benign and malignant cancer cells. Framework is divided into two segments: Training phase and testing phase. Initially, idea starts with the collection of tissue samples of lung tissues from an unhealthy person and creates the database for the collected samples. The collected samples are stored for the succeeding processing purpose. The different steps involved in classification of histopathological image are as follow: pre-processing, conversions of images into Gray and HSV form, Extraction of features like texture and shape, and finally applying the CNN training to classify the histopathological images. The proposed prototype is shown in below figure 3.
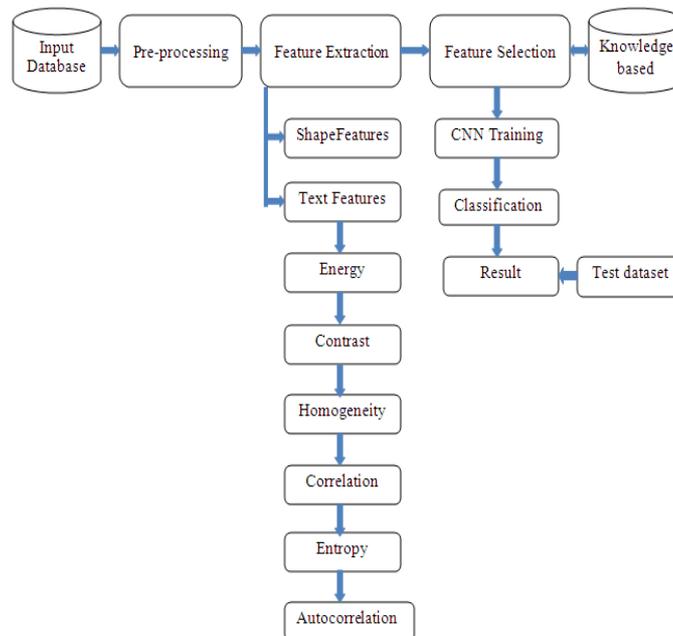

Figure3 Overview of proposed work

1. IMAGE ACQUISITION
At the outset of the framework, the image samples for the two types of cancerous cells namely benign cells and malignant cells are collected and have created the input database. All the acquired images are stored in tiff format on the disk, which creates the database. The best possible database creation is the vital part as it is in charge of the exactness of classification.

2. PRE-PROCESSING
The natural image taken for image analysis will not be well defined in its morphological structure. Therefore, it is mandatory to remove the background noise and unwanted parts to enhance the quality of the image and formulate a better image for preceding image processing. Pre-processing of the image also involves the resizing of all the images included into the data base.
A) RGB to Gray scale: RGB is a three plane model with Red, Green and Blue as its primary colors. As the process of feature extraction individually from these three different planes is time taking. As a result, RGB image is converted into Gray scale image where all the pixels in the image have the same intensity which makes the extraction easier and faster.
B) RGB to HSV: Images from RGB are translated into the HSV. Where, HSV stands for the Hue, saturation and the Value planes in an image. HSV images are more perception as compared to RGB color images. Hue element portray

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

the shades in the image, Saturation element gives the idea of white light added to the hue element and value element gives the brightness in the image.

3. FEATURE EXTRACTION

It is the second most important step in the image processing. Since, Features obtained from the images helps in classifying and identifying the images. Feature extraction gives the detailed view of the image to be analysed. For our frame work we have extracted the texture feature and shape feature.

A) Texture Feature: To calculate the Texture feature Gray level co-occurrence matrix (GLCM) is used. The Haralick et al gave the GLCM algorithm. [6]. It is a second order spatial technique which is accounted for to have the capacity to portray compositions as a by and large or normal spatial relationship between dim tones in a picture. Its advancement was motivated by the guessed from Julesz that second request probabilities were adequate for human segregation of composition.

In GLCM, number of rows by number of columns gives the gray level present in an image. Using the two parameters i.e. the distance d between the pair of pixels from an image and the relative orientations in the four directions (e.g. 0º, 45 º, 90 º and 135 º) concurrence matrix is calculated. The formulas used in calculations are given below:

- Entropy

$$\text{Entropy (entro)} = \sum_M \sum_N g_{MN} \, log_2 g_{MN} \qquad (1)$$

Where, MxN is the rows and columns of a GLCM matrix g.
This attribute calculates the dissorderness present in an image.

- Contrast

$$\text{Contrast (contr)} = \sum_M \sum_N (M - N)^2 \qquad (2)$$

Where, MxN is the rows and columns of a GLCM matrix g.
This attribute calculates the local changes happening in an image. It determines the difference between the highest and the lowest values present in a two continuous pixel pairs.

- Energy

$$\text{Energy (energ)} = \sum_M \sum_N g_{MN}^2 \qquad (3)$$

Where, MxN is the rows and columns of a GLCM matrix g.
It gives the information about recurrence of the pixel pairs in an image.

- Homogeneity

$$\text{Homogeneity (homom)} = \sum_M \sum_N \frac{1}{1+(M-N)^2} g_{MN}^2 \qquad (4)$$

Where, MxN is the rows and columns of a GLCM matrix g.
This attribute gives back a quality that measures the closeness of the circulation of components in the GLCM to the GLCM diagonal.

- Autocorrelation: We get the data about the extent of the tonal primitives

B) Morphological Feature: To remove the artifacts such as shadow, non-uniformity and illumination present in the real image binarization of the image is done. It removes the foreground and the background objects present in an image making it easy to perform the pattern recognition. Binarised image is presented in the form of 0's and 1's. Hence, it is easy to extract the shape of the binarised image.

To calculate the shape feature the geometrical structural features of the cells are considered. The geometrical features calculated are area, perimeter, major axis length, minor axis length, roundness and compactness. The morphological operations that it performs are dilation and erosion. In dilation wherever the image pixel is represented by 1 leads to addition of pixels to the object boundaries while in the erosion object boundaries are removed wherever the image pixel is represented by zero.

4. PRINCIPAL COMPONENT ANALYSIS

 Principal component analysis (PCA) [7] is a linear transform technique applied in signal and image processing field to obtain the compressed data set and reduced dimensionality. There are two different types of PCA techniques namely lossless PCA and lossy PCA. Lossless PCA implementation leads to the data compression without the loss of the data while the lossy PCA implementation causes the data compression with the loss of data. Therefore, for proposed framework lossless PCA technique is applied. Here, data is reduced based on the variables having highest correlation with the original variables. The variables with the high correlation are considered and that with the lower value are reduced.

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

PCA first calculates the covariance matrix for the variables. Next, it finds the Eigen vector and the Eigen value for the covariance matrix. Eigen vector is a vector quantity whose direction does not change when a linear transformation is performed in it. Eigen value gives the amount of variance present in the data in that direction.

5. CONVOLUTIONAL NEURAL NETWORK

Convolutional Neural Network (CNN) [8] is a kind of machine learning classifier which has the ability to make the self decision by learning the relationship between the features collected from our data and the observed data responses.
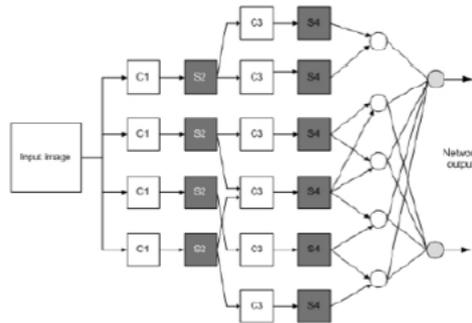

Figure 4 Convolutional Neural Network layer; C=Convolutional layer; S=subsampling layer

CNN works on the 2-D images. The architecture of CNN is designed in the three layers. The three types of layers are: convolutional layer, subsampling layer and output layer. Each layer in the CNN acts as a filter and holds the specific feature patterns existing in the original data. First layer of the CNN analysis some features from the original data. Subsequent layers detect few more data i.e. increasingly more abstracted features. The last layer of the CNN is proficient to make an ultra particular grouping by joining all the particular components recognized by the past layer within the input file. Above figure 4 depicts the three layers of CNN classifier.

IV.EXPERIMENTAL RESULT

The experimental results carried out for benign and malignant biopsy images are shown in the below figures 5(a) and 5(b) respectively. The experiment is carried out for the created dataset. For the testing purpose a query image is selected. This is converted into the Gray scale, HSV and binary forms. From the converted images features are extracted using the GLCM algorithm and shape feature. Finally, the CNN classifier is applied to classify the histopathology image either as benign or malignant cells based on the features obtained and features stored.
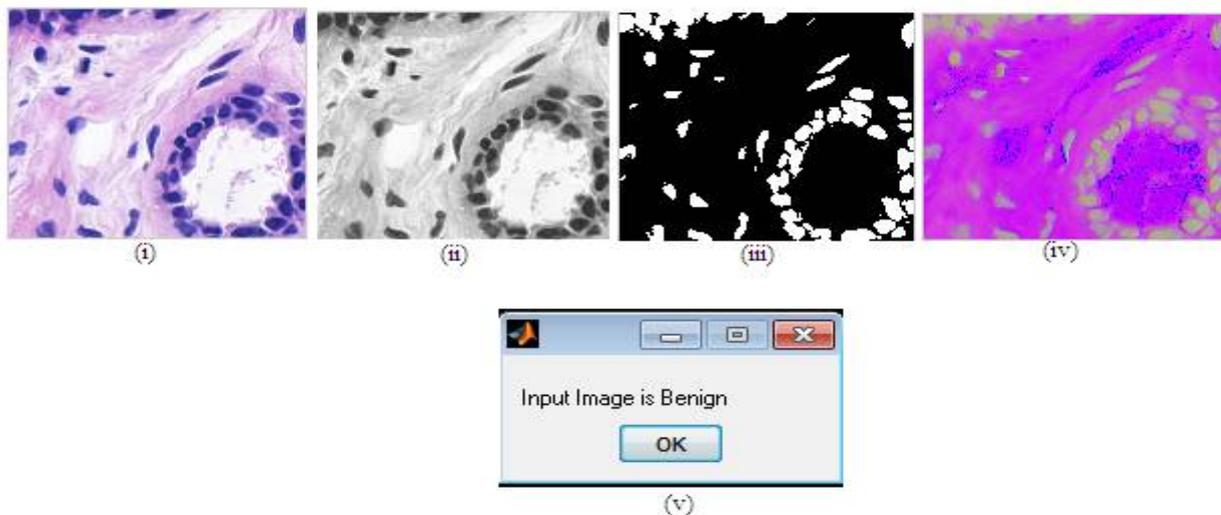

Figure 5(a) Experimental Result for an input image after applying CNN classification (i) Query image; (ii) Gray scale image; (iii) Binary image; (iv) HSV image and (v) Result.
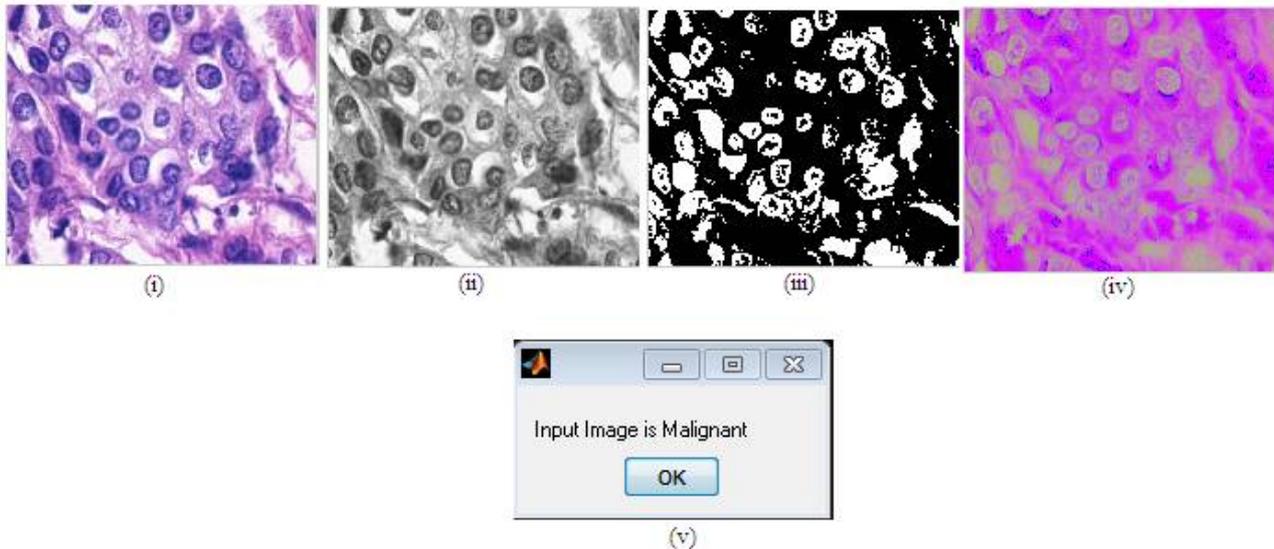
Figure 5(b) Experimental Result for an input image after applying CNN classification (i) Query image (ii) Gray scale image (iii) Binary image (iv) HSV image and (v) Result.

VI.CONCLUSION

In the framework Histopathological biopsy image classification based on CNN, We exhibited joined set up convolutional neural systems as an adaptable classifier for identifying the pathological images in diagnostic biopsy pictures.

In this work, we have achieved modernized biopsy picture examination approach for distinguishing the lung influenced by the dangerous cells. This modernized approach of classifying study demonstrates that grouping of destructive cells utilizing Convolutional Neural Network (CNN) gives better exactness.

REFERENCES

[1]   M. N. Gurcan, A. Can, L. Boucheron, A. Madabhushi, N. Rajpoot and Bulent Yener, "Histopathological Image Analysis: A Review", IEEE Reviews in Biomedical Engineering, vol. 2, (2009), pp. 147-171

[2]   C. Demir and B. Yener, "Automated cancer diagnosis based on Histopathological Images: A systematic survey", Technical Report Rensselear Polytechnic Institute Dept. Of Computer Science, TR 05-09, CiteSeerXβ, (2005), doi: 10.1.1.61.1199, pp. 1-12.

[3]   O. Sertel, J. Kong, G. Lozanski, J. Saltz, M. Gurcan, A. Shana'ah and U. Catalyurek, "Texture Classification using nonlinear color quantization: Application to Histopathological image analysis", Journal on Pattern Recognition Archive, (2008) IEEE, vol. 9, no. 1, (2008), pp. 596-600.

[4]   P. A. Hall and D. A. Levison, "Review: Assessment of cell proliferation in histological material", Journal of Clinical Pathology, vol. 43, no. 3, (1990), pp. 184-192.

[5]   X. Zhang, L. Yang, W. Liu, H. Su, and S. Zhang,"Mining histopathological images via composite hashing and online learning", In MICCAI. 2014, pp. 1-8.

[6]   R.M. Haralick, 3K. Shanmugam and I. Dinstein, "Textural Features for Image Classification", IEEE Trans. on Systems, Man, and Cybernetics SMC-3, 1973, pp. 610-621.

[7]   Pramod Kumar Pandey, Yaduvir Singh and Sweta Tripathi," Image Processing using Principle Component Analysis", International Journal of Computer Applications, Volume 15– No.4, February 2011,pp. 1-4.

[8]   Nuh Hatipoglu and Gokhan Biligin,"Classification of Histopathological Images Using Convolutional Neural Network", IEEE 2014,pp. 1-6

[9]   Siegel, R., Naishadham, D., Jemal, A.: Cancer statistics. CAJC 63(1), 11–30 (2013)

[10] Catalu˜na, J.J.S., Perpi˜n´a, M., Greses, J.V., Calvo, V., Padilla, J.D., Par´ıs, and F.: Cell type accuracy of bronchial biopsy specimens in primary lung cancer. CHEST 109(5) (1996)

[11] Freeman, D.L.: Harrison's principles of internal medicine. JAMA: The Journal of the American Medical Association 286(8), 506 (2001)