Classification of Lung Cancer using Deep Learning Algorithm

Dr. M. Sangeetha
Professor, Department of Information Technology,
V.S.B Engineering college, Karur,
Tamil Nadu, India.

Ms. P. Sangeetha, B.Tech.,
Student, Department of Information Technology,
V.S.B Engineering college, Karur,
Tamil Nadu, India.

Ms. P. Pavithra, B.Tech.,
Student, Department of Information Technology,
V.S.B Engineering college, Karur,
Tamil Nadu, India.

Ms. G. Divya Bharathi, B.Tech.,
Student, Department of Information Technology,
V.S.B Engineering college, Karur,
Tami Nadu, India

Abstract— **Lung cancer is one of the most killer diseases in the developing countries and the detection of the cancer at the early stage is a difficult task. Analysis and cure of lung malignancy have been faced by humans over the most recent couple of decades. In this project, we use Support Vector Machine (SVM) based on deep learning algorithm to classify the tumors found in lung as malignant (cancerous tumor) or benign (non-cancerous tumor). By using deep SVM technique, the lung nodule can be detected by three dimensional image processing technique and modules are based on data mining technique. The accuracy obtained by means of SVM is 96%, which is more efficient when compared to accuracy obtained by the traditional neural network system.**

Keywords — Support Vector Machine (SVM), image processing, computed tomography, feature extraction.

I) INTRODUCTION

Among different type of cancer, lung cancer is most dreadful diseases. Begin of lung cancer starts with abnormal cells grows out of the control in lungs.Two major types of lung cancer are non-small cell lung cancer and small cell lung cancer. Lung cancer forms tumor in any part of the lung that reduce person's ability to breathe. The main symptoms of lung cancer include shortness of breath, weight loss, wheezing, changes to a person's voice, such as hoarseness.The cancer cells in the lungs can spread, or metastasize, to the lymph nodes and other parts of the body. The detection of lung tumor is done by using many imaging techniques such as Computed Tomography (CT) scans. The CT scan provides an multiple slices in the lung structure to identifies an tumors in the lung. After talking the CT scans, the radiologists must compare the current

CT scans with the previous ones. If the lung nodule on earlier CT scans has not changed in size, shape or appearance, it is probably non-cancerous, otherwise it determines to cancerous.

II) EXISTING SYSTEM

The classification performance of a baseline pattern recognition system, designed to discriminate between benign and malignant tumors with state-of-the art feature extractors and classifiers, with the aim of showing the difficult of the problem. The classification system is based on 4 machine learning models, trained with different textural representations and key point detectors. To give an insight about the discriminative power of the textural representations we have used, we also present the performance of the oracle. CNN process is implemented using ANN technique in existing system. Lung cancer is detected by using the techniques of Image processing. The system formed can take any type of medical image within the three choices consisting of CT, MRI and Ultrasound images. The developed using PSO, Genetic Optimization and CNN algorithm used for feature selection and classification. Extension of image processing using lung cancer detection and produces the result of feature extraction and feature selection after segmentation. The system formed accepts any one of medical image within the three choices consisting of MRI, CT and Ultrasound image as input. After preprocessing of image, canny filter is used for Edge detection to detect the cancerous cells effectively from the CT, MRI scan and Ultrasound images. Super pixel Segmentation has been used for segmentation and Gabor filter is used for De-noising the medical images. The major image modalities have been studied in this survey of cancer detection through image processing used on CT, MRI and Ultrasound images. We proposed a method for segmentation of MRI, CT and Ultrasound images. Correct identification of cancer cell is done by studying the necessary features extracted for the two images. Ultrasound images as well to detect the validity of this system. We used feature selection as well by the use of PSO, Genetic Optimization and CNN algorithm giving an accuracy of about 89.5% with reduction in false positive. The disadvantages of existing system are low image quality, low accuracy in classification and recognizing the cancer image.

III) PROPOSED SYSTEM

This Project proposed two approaches to classifying lung cancer histology images into benign and malignant, as well as their sub-classes. Support vector machine (SVM) based technique to classify the lung tumors as malignant or benign stage. A SVM is type of a DNN consists of multiple hidden layers such as convolution layer, RELU layer. Pooling layer and fully connected a normalized layer. The input to a

convolution layer is an image of size m x m x r, where r is the number of channels. Different results are generated from these methods based on type of convolution used. Perform two tasks.

Task 1 : SVM using 2d convolution.

2D Convolution layers take a three-dimensional input, typically an image with three color channels. They pass a filter, also called a convolution kernel, over the image, inspecting a small window of pixels at a time, for example 3×3 or 5×5 pixels in size, and moving the window until they have scanned the entire image. The convolution operation calculates the dot product of the pixel values in the current filter window with the weights defined in the filter.

Task 2: SVM using depth wise separable convolution process broken into 2 operations-
  ➢ Depth wise convolution
  ➢ Point wise convolutions

1. Depth wise convolution

Convolution is a very important mathematical operation in artificial neural networks (ANN's). Support vector machines (SVM's) can be used to learn features as well as classify data with the help of image frames. There are many types of SVM's. One class of SVM's are depth wise separable support vector machines.

These type of SVM's are widely used because of the following two reasons –

They have lesser number of parameters to adjust as compared to the standard SVM's, which reduces over fitting. They are computationally cheaper because of fewer computations which makes them suitable for mobile vision applications

2) Point wise convolutions

The qualifier point wise is used to indicate that a certain property is defined by considering each value. The point wise operations, that is operations defined on functions by applying the operations to function values separately for each point in the domain of important relations. It can also be defined point wise. The advantages of proposed system are high accuracy, easy to deal with unbalance data.
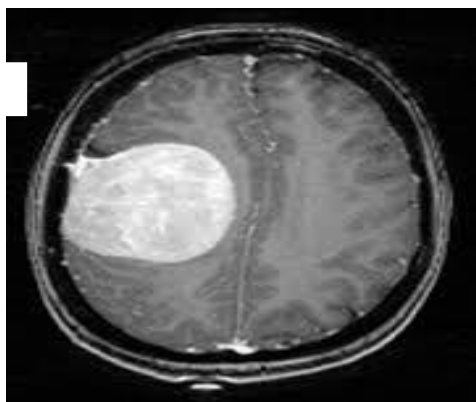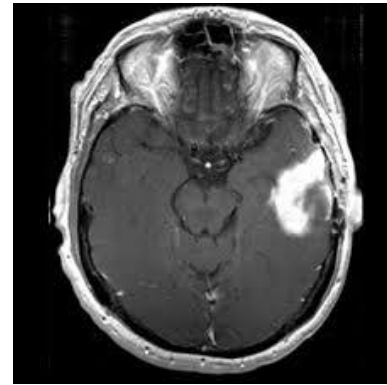

Figure 1: Normal Lung Image

Figure 2: Abnormal lung image.

IV) IMAGE PROCESSING

In image processing, Segmentation is important process. In image segmentation process, enhanced image is segmented into one or more sub-segments that will be easier to the images for further extracting the datasets.

A. Image Acquisition

The first stage of methodology starts with collecting the datasets (CT images). The datasets contain normal and abnormal images. The images are in raw data format. So, it needs to pre- processed the images to improve contrast transparency.

B. Image Preprocessing

The image preprocessing technique commences with image enhancement. The preprocessing technique is mainly used for obtain the better image compare to other computerized image. Image enhancement technique is split into two parts:
  ➢ Frequency domain techniques- which operate in Fourier transform image.
  ➢ Spatial domain techniques- which operate directly on pixels.

  Gabor filter is used during preprocessing technique.
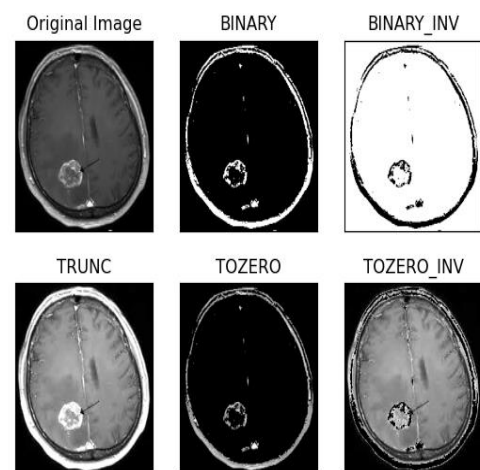
C. Image Segmentation

Image Segmentation helps to analyze the image easily. In this method, the images are segmented in different ways. Image segmentation is typically used to locate objects and

Special Issue - 2020

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
RTICCT - 2020 Conference Proceedings

boundaries (lines, curves, etc.) in images. More precisely, image segmentation is the process of assigning a label to every pixel in an image such that pixels with the same label share certain visual characteristics.

V) PROJECT DESCRIPTION

MODULE

- Pre - Processing.
- Data set training.
- Classification
- Clustering
- Feature Extraction

A. Pre-Processing

Pre-processing refers to the transformations applied to our data before feeding it to the algorithm.Data preprocessing is a technique that is used to convert the raw data into a clean data set. In other words, whenever the data is gathered from different sources it is collected in raw format which is not feasible for the analysis.

B. Data Set Training

The training data is used to make sure the machine recognizes patterns in the data, the cross-validation data is used to ensure better accuracy and efficiency of the algorithm used to train the machine, and the test data is used to see how well the machine can predict new answers based on its training.

C. Classification

The classification process is used to identify the category of the data's. The classification is used to identify impossible data combinations, missing data's, out of range value, etc. The classification is used to remove the damaged data's, and the empty data's in the overall dataset.

D. Clustering

Clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense) to each other than to those in other groups (clusters).

It is a main task of exploratory data mining, and a common technique for statistical data analysis, used in many fields, including machine learning, pattern recognition, image analysis, information retrieval, bioinformatics, data compression, and computer graphics.

E. Feature Extraction

Feature extraction is used to reduce the amount of resources required to describe a large set of data. It starts from an initial set of measured data and builds derived values. When the input data to an algorithm is too large to be processed and it is suspected to be redundant, then it can be transformed into a reduced set of features. Determining a subset of the initial features is said to be feature selection / Extraction.
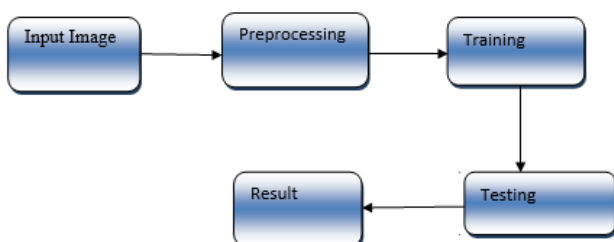
VI) CONCLUSION

The performance comparisons it is evident that our proposed model 3DDSVM attained the highest results against other state-of-the-art systems for sensitivity and FPs per scan. Although the current tested performance metric of 3DDSVM is relatively high, it could be further improved. The performance was relatively less accurate in detecting micro nodules, therefore future work will investigate the detection of micro nodules whose diameter is less than 3 mm. To ensure our solution is scalable, future work will consider extending the training stage to include data from hospitals worldwide. Integrating more data-augmentation methods to increase the training sample in order to achieve more robustness and reduce the problem brought by local optimal. Another future direction for lung cancer CAD system is to propose CAD system that performs well on all nodule types maintaining good performance in terms of sensitivity and FPs/Scan, even if the dataset contains relatively less amount of such nodule type's samples.

VII) REFERENCE

[1] R. L. Siegel, K. D. Miller, and A. Jemal, ``Cancer statistics, 2018,'' CA, Cancer J. Clin., vol. 68, no. 1, pp. 730, Jan. 2018.

[2] S. Park, S. J. Lee, E. Weiss, and Y. Motai, ``Intra- and inter-fractional variation prediction of lung tumors using fuzzy deep learning,'' IEEE J. Transl. Eng. Health Med., vol. 4, 2016, Art. no. 4300112.

[3] T. Tan et al., ``Optimize transfer learning for lung diseases in bronchoscopy using a new concept: Sequential ne-tuning,'' IEEE J. Transl. Eng. Health Med., vol. 6, 2018, Art. no. 1800808.

[4] H. Jiang, H. Ma, W. Qian, M. Gao, and Y. Li, ``An automatic detection system of lung nodule based on multigroup patch-based deep learning network,'' IEEE J. Biomed. Health Inform., vol. 22, no. 4, pp. 12271237, Jul. 2018.

[5] X. Du, S.Yin, R. Tang,Y. Zhang, and S. Li, ``Cardiac-DeepIED: Automatic pixel-level deep segmentation for cardiac bi-ventricle using improved endto- end encoder-decoder network,'' IEEE J. Transl. Eng. Health Med., vol. 7, 2019, Art. no. 1900110.

[6] H. Chung, H. Ko, S. J. Jeon, K.-H. Yoon, and J. Lee, ``Automatic lung segmentation with juxta-pleural nodule identication using active contour model and Bayesian approach,'' IEEE J. Transl. Eng. Health Med., vol. 6, 2018, Art. no. 1800513.

[7] S. Shen, S. X. Han, D. R. Aberle, A. A. Bui, and W. Hsu, ``An interpretable deep hierarchical semantic support vector machine for lung nodule malignancy classication,'' Expert Syst. Appl., vol. 128, pp. 8495, Aug. 2018.

[8] A. Teramoto and H. Fujita, ``Fast lung nodule detection in chest CT images using cylindrical nodule-enhancement lter,'' Int. J. Comput. Assist. Radiol. Surg., vol. 8, no. 2, pp. 193205, 2013.

[9] K. He, X. Zhang, S. Ren, and J. Sun, ``Deep residual learning for image recognition,'' in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2016, pp. 770778.

[10] L. Lu, Y. Tan, L. H. Schwartz, and B. Zhao, `Hybrid detection of lung nodules on CT scan images,'' Med. Phys., vol. 42, no. 9, pp. 50425054, 2015.

[11] A. Masood et al., ``Computer-assisted decision support system in pulmonary cancer detection and stage classication on CT images,'' J. Biomed. Inf., vol. 79, pp. 117128,Mar. 2018.

[12] M. Tan, R. Deklerck, B. Jansen, M. Bister, and J. Cornelis, ``A novel computer-aided lung nodule detection system for CT images,'' Med. Phys., vol. 38, no. 10, pp. 56305645, Oct. 2011.

[13] B. van Ginneken, A. A. A. Setio, C. Jacobs, and F. Ciompi, ``Off-the-shelf support vector machine features for pulmonary nodule detection in computed tomography scans,'' in Proc. IEEE Int. Symp. Biomed. Imag., Apr. 2015, pp. 286289.

[14] W. Shen et al., ``Multi-crop support vector machines for lung nodule malignancy suspiciousness classication,'' Pattern Recognition., vol. 61, pp. 663673, Jan. 2017.

[15] A. A. A. Setio et al., ``Pulmonary nodule detection in CT images: False positive reduction using multi-view convolutional networks,'' IEEE Trans. Med. Imag., vol. 35, no. 5, pp. 11601169, May 2016.

[16] S. Hamidian, B. Sahiner, N. Petrick, and A. Pezeshk, ``3D support vector machine for automatic detection of lung nodules in chest CT,'' in Proc. SPIE, vol. 10134, Mar. 2017, Art. no. 1013409.

[17] C. Zhao, J. Han, Y. Jia, and F. Gou, ``Lung nodule detection via 3D U-Net and contextual support vector machine,'' in Proc. IEEE Int. Conf. Netw. Netw. Appl. (NaNA), Oct. 2018, pp. 356361.

[18] J. Zhang, Z. Huang, T. Huang, Y. Xia, and Y. Zhang, ``Lung nodule detection using combined traditional and deep models and chest CT,'' in Proc. Int. Conf. Intel. Sci. Big Data Eng. Cham, Switzerland: Springer, 2018, pp. 655662.

[19] B. Wang, G. Qi, S. Tang, L. Zhang, L. Deng, and Y. Zhang, ``Automated pulmonary nodule detection: High sensitivity with few candidates,'' in Proc. Int. Conf. Med. Image Comput. Comput.- Assisted Intervent. Cham, Switzerland: Springer, 2018, pp. 759767.

[20] A. A. A. Setio et al., ``Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: The LUNA16 challenge,'' Med. Image Anal., vol. 42, pp. 113, Dec. 2017.

[21] S. Wang et al., ``Central focused support vector machines: Developing a data-driven model for lung nodule segmentation,'' Med. Image Anal., vol. 40, pp. 172183, Aug. 2017.

[22] S. Ren, K. He, R. Girshick, and J. Sun, ``Faster R-SVM: Towards realtime object detection with region proposal networks,'' IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 6, pp. 11371149, Jun. 2017.

[23] J. Dai, Y. Li, K. He, and J. Sun, ``R-FCN: Object detection via regionbased fully convolutional networks,'' in Proc. Neural Inf. Process. Syst., 2016, pp. 379387.