Lung Cancer Detection Using Multi-Class SVM And CNN

Akhila Mohan[1], Neevan R[2]
[1]Dept of Master Of Computer Applications
[2]Asst.Professor, Dept of Master Of Computer Applications
[1, 2] College Of Engineering, Trivandrum, Kerala, India

Abstract- *Several new approaches has been used to remove the snap blur caused due to the camera shake, with one or more multiple images, by providing solutions for an inverse and inherently illposed deconvolution problem. Currently, in all new digital cameras a mode is available to combine a burst of snaps to produce a sharp clean snapshot This solution doesn't take care of the inverse problem and the blur estimation. This paper proposes a method which uses weights that depends on the Fourier spectrum magnitude and then calculates the weighted average in the Fourier domain.*

Keywords- Convolutional Neural Network(CNN),Lung Cancer, Multi-class Support Vector Machine(SVM).

I. INTRODUCTION

One of the major causes of cancer death in the world is lung cancer. The reason behind this is that the symptoms of lung cancer come into light at the final stage. As a result, diagnosing the disease at the beginning stage is difficult.

Some researches regarding lung cancer have been reported in recent years. Some of the researches are Murphy, Messay, Hardie, Rogers, etc. The problem with these works is, some techniques were cramped to a solitary nodule, detects only diseases and doesn't predict the stages.

In this proposed paper, a few problems were solved with improved accuracy. The developed system can be used to diagnose lung cancer and would predict its stages. At first multi- class SVM was used and for further improvement, CNN was also used.

II. METHODOLOGY

Lung cancer detection makes use of lung CT scans provided by the kaggle dataset. The system uses python and Open CV based implementation to predict whether the input lung image is cancerous or not and if it is cancerous then the system can detect the stage of cancer.

This system consists of four main modules. There is a module which deals with the image acquisition and pre-processing where the binarization, morphological operations, and segmentation are performed. The second module is a feature extraction module in which the features such as area, perimeter, diameter and eccentricity of the lung nodule are extracted. The extracted features are added to a csv file. The third module is for training the data using Support Vector Machine. Training and testing are performed to predict whether an input lung CT image is malignant or not. The fourth module deals with training and testing the images of different classes using a convolutional neural

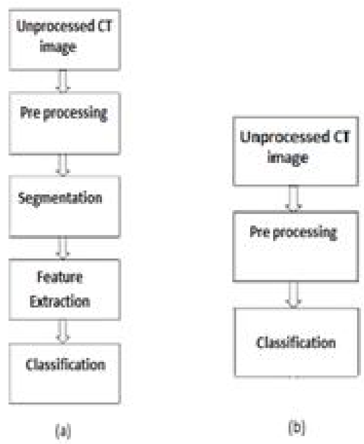

Fig -1: a) Overall process in multi-class SVM and (b) shows the process over CNN.

III. IMPLEMENTATION

Now it is the time to articulate the research work with ideas gathered in above steps by adopting any of below suitable approaches:

A. Data Collection and Preprocessing

The first step is to acquire normal and abnormal CT images that are collected from the available database. The given below image shows the obtained data.

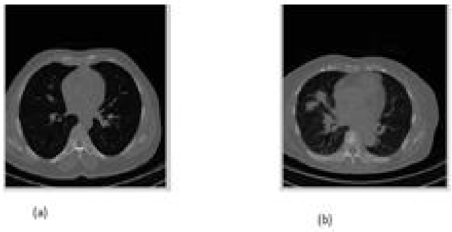

Fig -2: (a) Normal CT image of lungs
(b) Cancerous image(Stage 1).

On these dataset images, pre-processing is applied. For the SVM module, the pre-processing steps involve cropping, binarizing, border clearing, binary erosion, etc. For the CNN module, the binarized images are resized.

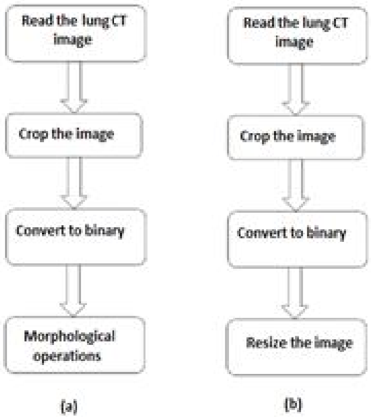

Fig -3: Preprocessing steps
(a) multiclass SVM (b) CNN

For training with SVM, the pre-processed image should be further segmented. This uses the binarization approach. The pre-processed binary image is filtered according to the maximum area and a contour is plot in order to make it ready for feature extraction.

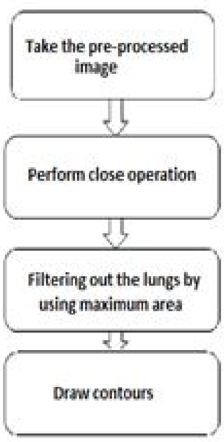

Fig -4: Segmentation steps

The thresholding approach is used for segmentation. The input image is converted to a binary image. Then binary closing is done with a radius of 10 to keep nodules attached to the lung walls. Only the portions with the largest area are kept which are the lungs. Then the image is superimposed with the original image to get the separated nodule.

B. Feature Extraction

The features such as area, perimeter and eccentricity are extracted from the lung nodule. For extracting these features a contour is drawn on each extracted nodule. Then the area of the region within the contour is found for each nodule. The nodule with maximum area is found and the perimeter, diameter, and eccentricity of the nodule are calculated.

The area, perimeter, and diameter of a cancerous image is greater than that of a non-cancerous image and also there will be a significant difference in these features for the images in different cancer stages. The image with eccentricity 1 denotes a circular nodule and is not considered as cancerous.

C. Training With Multi Class SVM

After extracting the features, they are added to a csv file and are then given as input to the training module. The aim of classification is to group items that have similar feature values.

Training an SVM involves feeding known data along with the previously known decision values, thus forming a finite training set. SVM then gets the intelligence to classify unknown data from the training set.

D. Training with CNN

The input to the CNN training module are the images after some pre-processing. The images are cropped first. Then they are binarized and resized in order to increase training speed. Images from the four classes such as non- ancerous, cancer stage 1, cancer stage 2 and cancer stage 3 are given. An input (28x28)is passed to the convolutional filters of size 20x5x5. Next, the max pooling layer with a kernel size of 2 × 2 plays its role. Next, the output from the max pool layer moves to the next convolutional filters and the then max pool layer respectively. Finally, the result is passed to the dense layer. Then softmax is applied to obtain probabilities for each class. After training with the given dataset, a trained model is created and then using the trained model, predictions are made for unknown data.

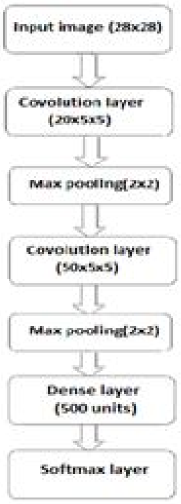

Fig -5: Training using CNN.

IV. RESULTS

The results for lung cancer detection using Multi-class SVM and CNN techniques are shown in the table.

Table 1: Accuracy table

Algorithm Used	*Input data*	*Training*	*Testing*
SVM	Lung CT image	0.869	0.779
CNN	Lung CT image	0.851	0.851

The inference from the table is that the model predicts whether a person has lung cancer or not with 77 percent accuracy using SVM and with 87 percent accuracy using CNN. CNN seems to be better than multi-class SVM classifier.

V. CONCLUSIONS

The goal of this paper is to put forth a system that would detect lung cancer and predict the stages. The features like area, eccentricity, diameter, etc seem to have the most predicting power with multi-class SVM. Some unexpected prediction with multi-class SVM has become a reason for using CNN, which is well known for its capablility to identify and quickly learn local features within an image. The flaws with the multi-class SVM was overcome by Convolutional Neural Network.

Various segmentation techniques such as p-tile thresholding, watershed segmentation can be applied to increase the accuracy of the cancer detection system. Some feature set along with morphological features can play a positive impact on accuracy. A new class can be added to the system for prediction of the possibility of lung cancer if more dataset is available.

REFERENCES

[1] JaneeAlam, Sabrina Alam, AlamgirHossan, "Multi-Stage Lung Cancer Detection and Prediction Using Multi-class SVM Classifier ", International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2), 2018.

[2] E. Cengil, A. nar, E. zbay ., " Image classification with caffe deep learning framework ", International Conference on (pp. 440-444). IEEE. , 2017.

[3] Emine CENGL, AhmetINAR .," A Deep Learning Based Approach to Lung Cancer Identification ", International Conference on Artificial Intelligence and Data Processing (IDAP), 2018.

[4] https://www.tensorflow.org/tutorials/