LUNG DISEASE DETECTION

Nikita Ashok Pingale
Computer Engineering
Pimpri Chinchwad College of Engineering
Pune,India

Prerna Lahanu Rokade
Computer Engineering
Pimpri Chinchwad College of Engineering
Pune,India

Sayali Bapurao Rane
Computer Engineering
Pimpri Chinchwad College of Engineering
Pune,India

Sonali Uttam Rupnar
Computer Engineering
Pimpri Chinchwad College of Engineering
Pune,India

***Abstract*-Aim of this paper is to build "Predictive Diagnostic System" of infectious lung by using the concept of image processing in conjunction with machine learning. Proposed system will detect human breathe in a real-time fashion to evaluate the ideal state of the aspiratory phase of a breath. So as, to de ne a proper timing of CT scan trigger and to use preprocessing technique that will remove the noise. Post this, feature extraction process is applied to extract the useful features of underlying image, and feature selection technique will further optimize the topranking features. SVM algorithm is then applied to classify the images for detection of lung disease. This diagnostic**

system will detect diseases like: asthma, bronchiectasis, lung cancer, hanta virus, influenza and pneumonia, chronic obstructive pulmonary (COPD) disease by using SVM. After detection of disease, report will be generated and submitted to patient.

Keywords : SVM (Support Vector Machine); Computed Tomography (CT),Feature Extraction; Preprocessing; Prediction.

I. INTRODUCTION

Lung disease is any problem in the lungs that prevents the lungs from working properly. There are three main types of lung disease:
Airway diseases -These diseases affect the tubes (airways) that carry oxygen and other gases into and out of the lungs. They usually cause a narrowing or blockage of the airways. Airway diseases include asthma, COPD and bronchiectasis. People with airway diseases often say they feel as if they're "trying to breathe out through a straw."

Lung tissue diseases -- These diseases affect the structure of the lung tissue. Scarring or inflammation of the tissue makes the lungs unable to expand fully (restrictive lung disease). This makes it hard for the lungs to take in oxygen and release carbon dioxide. People with this type of lung disorder often say they feel as if they are "wearing a too-tight sweater or vest." As a result, they can't breathe deeply Lung infection (pneumonia) are examples of lung tissue disease.

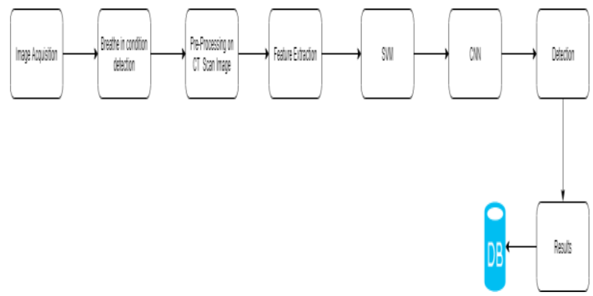
Lung circulation diseases -- These diseases affect the blood vessels in the lungs. They are caused by clotting, scarring, or inflammation of the blood vessels. They affect the ability of the lungs to take up oxygen and release carbon dioxide. These diseases may also affect heart function. An example of a lung circulation disease is pulmonary hypertension. People with these conditions often feel very short of breath when they exert themselves. The term lung disease refers to many disorders affecting the lungs, such as asthma, COPD, infections like influenza, pneumonia and tuberculosis, lung cancer, and many other breathing problems. Some lung diseases can lead to respiratory failure.Respiratory failure is a condition in which your blood doesn't have enough oxygen or has too much carbon dioxide. Sometimes you can have both problems. When you breathe, your lungs take in oxygen. The oxygen passes into your blood, which carries it to your organs. Your organs, such as your heart and brain, need this oxygen-rich blood to work well. Another part of breathing is removing the carbon dioxide from the blood and breathing it out. Having too much carbon dioxide in your blood can harm your organs. So it is necessary to take proper precautions of lung from any diseases which can cause severe health issues to the patient.In existing system, There was manual techniques used in hospitals previously patients CT

Scan images or MRI were taken by the doctors and technician for detection of lung cancer or any disease in internal organs. Generated medical reports and images are sent to super specialist and super specialist analyze the images and if any diseases are detected from the images then specialist will give the further consultation and will suggest medical treatment. In real this entire process takes around one to two days so patient is not able to do further process to precure the further consequences of defected body part. For breathe in condition detection we are using image subtraction operation from image processing. As we get breathe in condition in system takes CT scan images of lung. The lung CT image is engaged asthe input. The original image is transformed to gray scale image. After that, removal of the noises and contrast enhancement is done for obtaining the enhanced images. After that, removal of the noises and contrast enhancement is done for obtaining the enhanced images. According to image, system will classifying diseases like asthma, bronchiectasis, lung cancer, hanta virus, influenza, pneumonia, chronic obstructive pulmonary (COPD) disease by using SVM. SVM is effective in high dimensional spaces and in cases where number of dimension is greater than the number of samples. It uses a subset of training points in the decision function (called support vectors), so it is also memory efficient. Therefore in this paper we are using SVM for classification of diseases. According to image, system will classifying diseases like asthma, bronchiectasis, lung cancer, hanta virus, influenza, pneumonia, chronic obstructive pulmonary (COPD) disease by using SVM. SVM is effective in high dimensional spaces and in cases where number of dimension is greater than the number of samples. It uses a subset of training points in the decision function (called support vectors), so it is also memory efficient. Therefore in this paper we are using SVM for classification of diseases.

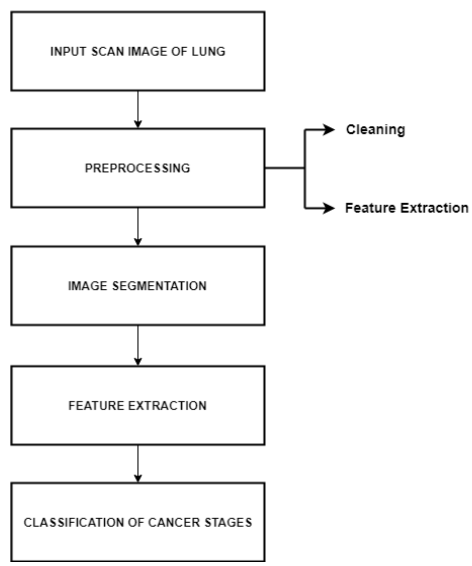
II. PROPOSED SYSTEM

Objective of the proposed system is to introduce a unique "Predictive Diagnostic System" system which first finds out breathe in or breathe out condition of a patient and then takes the images when person is in breathe in condition. For breathe in condition detection we are using image subtraction operation from image processing. As we get breathe in condition in system takes CT scan images of lung. The lung CT image is engaged as the input. The original image is transformed to gray scale image. After that, removal of the noises and contrast enhancement is done for obtaining the enhanced images. After that, removal of the noises and contrast enhancement is done for obtaining the enhanced images. The system first finds out breath motion and according to that the system will capture image on breathe in position is detected. so the resource will be utilized. After image acquisition the system perform pre-processing on image. Find out affected regions and their characteristics in form of data. This data is classified using SVM and CNN. SVM and CNN classify it as normal or diseases lung and identify lung diseases.

III. ARCHITECTURE DIAGRAM



IV. BLOCK DIAGRAM



V. COMPARATIVE STUDY

[1] Title: Classification of Lung Diseases Using a Combination of Texture, Shape and Pixel Value by K-NN Classifier.
Author: Latika A. Thamke, Madhav V. Vaidya.

In this work, Author has proposed Features Extraction Techniques for arrangement of Lung Computed Tomography Images. A Mix of Texture, Shape and Pixel Coefficient Feature are created for Classifying the CT pictures of lung sickness. The proposed framework can arrange lung pictures naturally as

Typical Lung, Pleural Effusion, Emphysema and Bronchitis. The proposed System contains four stages. In the underlying advance, the pictures are pre-handled. In the subsequent advance, the pictures are fragmented by Thresholding and Edge Detection. In the third step, the Texture, Shape and Pixel Coefficient Feature are determined utilizing the GLCM (Gray Level Co-event Matrix), Minute Invariant and WHT (Walsh Hadamard Transform) and joined to frame the single descriptor. In the last advance, the KNN, Multiclass-SVM and Decision Tree classifiers are utilized for characterization of Lung pictures. The pictures are the CT check pictures. The complete datasets contain 400 pictures, 100 pictures of each infection like the Normal, Pleural Effusion, Emphysema and Bronchitis. The 280 pictures are utilized for Training and 120 pictures are utilized for Testing. The order precision of collapsing strategy achieved by the K-NN classifier with Global Thresholding is 97.50% for WHT +GLCM, 97.50% for WHT + MI, 94.45% for GLCM + MI, 97.50% for WHT +GLCM+MI. The K-NN classifier with Global Thresholding decreases the time and furthermore gives better outcomes when contrasted with different strategies and classifiers.

[2] Title: Evaluate the Malignancy of Pulmonary Nodules Using the 3-D Deep Leaky Noisy-OR Network.
Author: Fangzhou Liao , Ming Liang, Zhe Li, Xiaolin Hu

Fangzhou Liao , Ming Liang, Zhe Li, Xiaolin Hu has proposed a 3-D deep neural network to take care of this issue. The model comprises of two modules. The first is a 3-D locale proposition organize for knob recognition, which yields every suspicious knob for a subject. The subsequent one chooses the main five knobs based on the identification certainty, assesses their malignant growth probabilities, what's more, joins them with a flawed boisterous OR door to get the likelihood of lung malignancy for the subject. The two modules share a similar spine organize, a changed U-net. The overfitting brought about by the lack of the preparation information is lightened by preparing the two modules on the other hand.

[3] Title: Multiple Resolution Residually Connected Feature Streams For Automatic Lung Tumor Segmentation From CT Images
Author: Jue Jiang, Yu-chi Hu, Chia-Ju Liu, Darragh Halpenny, Matthew D. Hellmann, Joseph O. Deasy.Jue Jiang, Yu-chi Hu, Chia-Ju Liu, Darragh Halpenny, Matthew D. Hellmann, Joseph O. Deasy proposed two numerous goals excessively associated system (MRRN) definitions called steady MRRN and thick MRRN. Author has assessed their technique on an aggregate of 1210 non-little cell (NSCLC) lung tumors and knobs from three datasets comprising of 377 tumors from the open-source Malignancy Imaging Archive (TCIA), 304 propelled organize NSCLC treated with against PD-1 checkpoint immune therapy from inward foundation MSKCC dataset, and 529 lung knobs from the Lung Picture Database Consortium (LIDC). The calculation was prepared utilizing the 377 tumors from the TCIA dataset and approved on the MSKCC and tried on LIDC datasets. The division exactness contrasted with master depictions was assessed by processing the Dice Similarity Coefficient (DSC), Hausdorff separations, affectability and accuracy measurements. Best performing steady MRRN technique created the most astounding DSC of 0.74±0.13 for TCIA, 0.75±0.12 for MSKCC and 0.68±0.23 for the LIDCdatasets. There was no huge distinction in the estimations of volumetric tumor changes processed utilizing the steady MRRN technique contrasted and master division. In rundown, author have built up a multi-scale CNN approach for volumetrically fragmenting lung tumors which empowers precise, computerized ID of and sequential estimation of tumor volumes in the lung.

[4] Title: Investigation of Sub-Centimeter Lung Nodule Quantification for Low-Dose PET
Author: Yihuan Lu , Kathryn Fontaine, Mary Germino, Tim Mulnix, Michael E. Casey, Richard E. Carson, and Chi Liu

In this paper, author played out a complete reenactment and smaller than expected Derenzo apparition concentrate to investigate the quantitative precision of sub-centimeter knobs utilizing the Siemens Biograph mCT scanner. Author reenacted knobs running from 4 to 10 mm in breadth, with 2:1 to 8:1 difference level at 1% to 100% (70 million) tally level, and with reasonable respiratory movement amplitudes (unrivaled sub-par/foremost back bearings) of 5/3, 10/6, and 20/12 mm. Pictures were recreated utilizing movement pay requested subset desire amplification list-mode calculation for goals recuperation reproduction. Author additionally examined diverse reproduction voxel sizes of 0.5, 1.0, and 2.0 mm for both reenactment and apparition contemplates. Recreations with 1.0- and 2.0-mm voxel size were upsampled to those with 0.5 mm preceding assessment and diverse upsampling strategies were analyzed. The outcomes from recreation and ghost studies were steady. Author found that knobs estimated 6 mm or more noteworthy brought about an inclination of mean institutionalized take-up worth (SUVmean) littler than 20%, notwithstanding when the tally levels dropped to 4%. SUVmean was diminished with 5/3 mm movement contrasted with static sweeps by 18 ± 4% with the 100% check level, with little extra decreases found for bigger movement amplitudes. Pictures remade with voxel sizes of 1.0 and 0.5 mm brought about increasingly precise evaluation and decreased twisting contrasted with those with 2-mm voxels. The outcomes demonstrated that it is practical to accomplish exact measurement for knobs ≥ 6 mm utilizing low-portion PET, with respiratory movement remedy and fine remaking voxel size.

[5] Title: Monitoring Tumor Lung Irradiation with Megavoltage Patient-Scattered Radiation: A Full System Simulation Study. Author: Joana Lencart, Paulo J. B. M. Rachinhas, Joao A. M. Santos

The framework comprises in recognizing mega voltage patient scattered radiation that is transmitted at right edges regarding the shaft hub. Since photon dispersing in the patient happens with higher power in tissues of higher thickness, a multi-cut photon identification framework situated oppositely to the bar hub yields a sign connected with patient morphology, including the tumor. Author in this manner report on GEANT4 recreations completed with a human so as to dissect the ability of the framework to recognize appropriate and clinically-important situations for example lung tumor deviation and tumor relapse movement. The sign conveyance got with practical full framework (counting the multi-cut collimator, the scintillator precious stones, and the charge electronic readout mode) demonstrate a high visual understanding both with the mimicked, recommended portion, and with the tumor area/size, just as with the ghost structures. The ability of this framework to get morphological pictures without X-beam source pivot possibly permits to exceedingly diminish portion in solid tissues and organs in danger in regard to other existing picture guided radiation treatment systems, in this manner supplementing them.

[6] Title: Knowledge-based Collaborative Deep Learning for Benign-Malignant Lung Nodule Classification on Chest CT. Author: Yutong Xie, Yong Xia, Jianpeng Zhang, Yang Song, Dagan Feng, Fellow, Michael Fulham, Weidong Cai.

The precise distinguishing proof of harmful lung knobs on chest CT is basic for the early identification of lung malignant growth, which likewise offers patients the most obvious opportunity with regards to fix. Profound learning techniques have as of late been effectively acquainted with PC vision issues, albeit generous difficulties stay in the location of threatening knobs because of the absence of enormous preparing datasets. In this paper, we propose a multi-see information based collective (MV-KBC) profound model to isolate dangerous from favorable knobs utilizing constrained chest CT information. Our model learns 3D lung knob attributes by deteriorating a 3D knob into nine fixed perspectives. For each view, we develop an information based collective (KBC) submodel, where three kinds of picture patches are intended to tweak three pre-prepared ResNet-50 arranges that describe the knobs' general appearance, voxel and shape heterogeneity, separately. We mutually utilize the nine KBC submodels to group lung knobs with a versatile weighting plan got the hang of during the blunder back proliferation, which empowers the MV-KBC model to be prepared in a start to finish way. The punishment misfortune capacity is utilized for better decrease of the bogus negative rate with an insignificant impact on the general execution of the MV-KBCmodel. We tried our technique on the benchmark LIDC-IDRI dataset and contrasted it with five best in class grouping draws near. Our outcomes demonstrate that the MV-KBC model accomplished a precision of 91.60% for lung knob arrangement with an AUC of 95.70%. These outcomes are notably better than the best in class draws near.

[7] Title: Bounded Fuzzy Possibilistic Method reveals information about lung cancer through analysis of metabolomics. Author: Hossein Yazdani, Leo Cheng, David C. Christiani, Azam Yazdani The precise ID of dangerous lung knobs on chest CT is basic for the early discovery of lung malignancy, which additionally offers patients the most obvious opportunity with regards to fix. Profound learning techniques have as of late been effectively acquainted with PC vision issues, albeit generous difficulties stay in the discovery of threatening knobs because of the absence of enormous preparing datasets. In this paper, we propose a multi-see learning based communitarian (MV-KBC) profound model to isolate threatening from kind knobs utilizing constrained chest CT information. Our model learns 3D lung knob qualities by decaying a 3D knob into nine fixed perspectives. For each view, we develop an information based collective (KBC) submodel, where three kinds of picture patches are intended to adjust three pre-prepared ResNet-50 arranges that describe the knobs' general appearance, voxel and shape heterogeneity, separately. We together utilize the nine KBC submodels to characterize lung knobs with a versatile weighting plan got the hang of during the blunder back spread, which empowers the MV-KBC model to be prepared in a start to finish way. The punishment misfortune capacity is utilized for better decrease of the bogus negative rate with a negligible impact on the general execution of the MV-KBC model. We tried our strategy on the benchmark LIDC-IDRI dataset and contrasted it with five best in class arrangement draws near. Our outcomes demonstrate that the MV-KBC model accomplished an exactness of 91.60% for lung knob order with an AUC of 95.70%. These outcomes are extraordinarily better than the best in class draws near.

VI. DATASET

VIA/I-ELCAP SIMBA Public Access Database.

This Database provides a demonstration of the SIMBA framework for image documentation for chest health. In addition to the ELCAP database, described below, it provides access to the image segmentations outcomes and custom visualizations designed for the SIMBA framework for very large image dataset documentation. The automated image documentation for this database is discussed in:

A. P. Reeves, Y. Xie, and S. Liu, "Large-scale image region documentation for fully automated image biomarker algorithm development and evaluation," *Journal of Medical Imaging*,4(2):

024505, Jun. 2017. Please refer to this paper when using information from this database.

This image database was made possible by a collaboration between the ELCAP and VIA research groups. It was created to make available a common dataset that may be used for the performance evaluation of different computer aided detection systems. This database was first released in December 2003 and is a prototype for web-based image data archives. The SIAN version of this database contains image segmentation analysis.Database Contents**:**The database currently consists of an image set of 50 low-dose documented whole-lung CT scans for detection. The CT scans were obtained in a single breath hold with a 1.25 mm slice thickness. The locations of nodules detected by the radiologist are also provided.The website provides a set of interactive image viewing tools for both the CT images and their annotations and their analysis. All images and their annotations may be downloaded from the website.

VII. CONCLUSION

The objective of proposed system is to build "Predictive Diagnostic System" of infectious lung. Proposed system works on the concept of image processing techniques such as feature subtraction, extraction, selection along with SVM classifier algorithms (machine learning) so as to precisely predict & detect various diseases of infectious lung. As an integrated approach this system will not only predict and detect various lung diseases, but it will also generate test reports that can be used for a preventative treatment.

VIII. REFERENCES

[1] Thamke, Madhav V. Vaidya. “Classification of Lung Diseases Using a Combination of Texture, Shape and Pixel Value by K-NN Classifier” IEEE Xplore Part Number:CFP18OZV-ART; ISBN:978-1-5386-1442-6 (2018)

[2] Fangzhou Liao , Ming Liang, Zhe Li, Xiaolin Hu, “Evaluate the Malignancy of Pulmonary Nodules Using the 3-D Deep Leaky Noisy-OR Network. ” IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS (2019)

[3] Jue Jiang, Yu-chi Hu, Chia-Ju Liu, Darragh Halpenny, Matthew D. Hellmann, Joseph O. Deasy. “Multiple Resolution Residually Connected Feature Streams For Automatic Lung Tumor Segmentation From CT Images” IEEE Transactions on Medical Imaging (2018)

[4] Yihuan Lu , Kathryn Fontaine, Mary Germino, Tim Mulnix, Michael E. Casey, Richard E. Carson, and Chi Liu. “Investigation of Sub-Centimeter Lung Nodule Quantification for Low-Dose PET” IEEE TRANSACTIONS ON RADIATION AND PLASMA MEDICAL SCIENCES, VOL. 2, NO. 1, JANUARY 2018

[5] Joana Lencart, Paulo J. B. M. Rachinhas, Joao A. M. Santos. “Monitoring Tumor Lung Irradiation with Megavoltage Patient-Scattered Radiation: A Full System Simulation Study.” IEEE Transactions on Radiation and Plasma Medical Sciences(2017).

[6] Yutong Xie, Yong Xia, Jianpeng Zhang, Yang Song, Dagan Feng, Fellow, Michael Fulham, Weidong Cai. “ Knowledge-based Collaborative Deep Learning for Benign-Malignant Lung Nodule Classification on Chest CT.” IEEE Transactions on Medical Imaging (2017) .

[7] Hossein Yazdani, Leo Cheng, David C. Christiani, Azam Yazdani. “Bounded Fuzzy Possibilistic Method reveals information about lung cancer through analysis of metabolomics” IEEE TRANSACTIONS ON COMPUTATIONAL BIOLOGY AND BIOINFORMATICS, 2018.

IX.LITERATURE SURVEY

TITLE	TECHNOLOGY	ACCURACY	LIMITATIONS
Classification of Lung Diseases Using a Combination of Texture, Shape and Pixel Value by K-NN Classifier.	1) GLCM (Gray Level Co-event Matrix), Minute Invariant and WHT (Walsh Hadamard Transform) 2) KNN, Multiclass-SVM and Decision Tree	90.50%	It ignores the spatial relationship between the texture patterns and sensitive towards image noise.
Evaluate the Malignancy of Pulmonary Nodules Using the 3-D Deep Leaky Noisy-OR Network.	3-D Convolution Neural Network	85.96%	It is more expensive
Multiple Resolution Residually Connected Feature Streams For Automatic Lung Tumor Segmentation From CT Images	Multiple resolution residual network-based deep convolutional neural network	85%	The method used does not gives the clear output
Investigation of Sub-Centimeter Lung Nodule Quantification for Low-Dose PET	Siemens Biograph mCT scanner	90%	These procedure is time consuming.
Monitoring Tumor Lung Irradiation with Megavoltage Patient-Scattered Radiation: A Full System Simulation Study.	GEANT4 simulations	80%	Geant simulations require more time for execution.
Knowledge-based Collaborative Deep Learning for Benign-Malignant Lung Nodule Classification on Chest CT.	MV-KBC model	91.60%	It takes long procedure to view from multiple sides of image.
Bounded Fuzzy Possibilistic Method reveals information about lung cancer through analysis of metabolomics.	Bounded Fuzzy Possibilistic Method BFPM algorithm	91%	