A Study of Methods Researched for Lung Cancer Detection

Aishwarya Nalawade, Shrinivas A. Patil

Textile & Engineering Institute Ichalkaranji, India

Corresponding author: Suraj Pillai, Email: aishwarya.t.nalawade@gmail.com

Advancement in medical science has sharply increased the overall average lifespan of humans since the last century. But this in turn has led to an increase in the incidence or rather possibility of cancer. It has been found that the risk of cancer steadily increases as age increases and more than 1,000 people out of 1,00,000 in the age group of 60 years and older suffer from some type of cancer. This, coupled with the mounting pollution and lifestyle changes has resulted in cancer becoming one of the top five non communicable diseases together responsible for 70% global deaths as per WHO data. Air pollution and global warming due to various causes has made lung cancer a major concern. It becomes essential to detect presence of cancer as early as possible for to minimize the risk of fatality and recurrence. Development of intelligent systems is a ray of hope for automation of the entire process of anomaly detection, identification of a disease, localization of the disease as well as analysis of the severity of the disease. Systems are being developed for early detection of lung cancer through results of various diagnostic systems. This paper studies the numerous preprocessing techniques implemented and different algorithms used for the detection of lung cancer through CT scans.

Keywords:Lung cancer, Classification, Biomedical, Image Processing, Tumor Detection, Artificial Intelligence.

2023. In Saroj Hiranwal & Garima Mathur (eds.), *Artificial Intelligence and Communication Technologies*, 493–499. Computing & Intelligent Systems, SCRS, India. https://doi.org/10.52458/978-81-955020-5-9-47

1 Introduction

According to the Global burden of Disease data of 2018, although India has 18% of the world's population, it has 32% of the cases of respiratory diseases in the world.[1] In India, respiratory disease burden comes next to that of ischemic heart disease.
The Gbest-guided Artificial Bee Colony (GABC) algorithm is a latest swarm intelligence-based approach to solve optimization problem. In GABC, the individuals update their respective positions by drawing inspiration from the global best solution available in the current swarm. The GABC is a popular variant of Artificial Bee Colony (ABC) algorithm and is proved to be an efficient algorithm in terms of convergence speed. But, in this strategy, each individual is simply influenced by the global best solution, which may lead to trap in local optima. Therefore, in this paper, a new search strategy, namely "Fully Informed Learning" is incorporated in the onlooker bee phase of ABC algorithm. The developed algorithm is named as Fully Informed Artificial Bee Colony (FABC) algorithm. To validate the performance of FABC, it is tested on 20 well known benchmark optimization problems of different complexities. The results are compared with GABC and some more recent variants of ABC. The results are very promising and show that the proposed algorithm is a competitive algorithm in the field of swarm intelligence-based algorithms.
The largest contributor behind this huge load of respiratory diseases is air pollution, both indoor and outdoor. [2] Some other major contributors for these diseases are smoking, infection, and genetics. Of the total deaths in India in 2017, 1.24 million deaths, equivalent to 12·5 percent of total mortalities, could be attributed to air pollution, as per a paper by India State Level Disease Burden Initiative, published in The Lancet. [3]
Lung cancer like any other cancer is basically the uncontrolled growth of abnormal cells in the lungs leading to a tumor that does not actually perform the working of normal lung cells. In most countries but especially in India, lung cancer is most prominent type of cancer and cause of cancer related mortality in men. [4,5] This is made worse by the fact that India is the country with 21 out of 30 most air polluted cities according to IQAir AirVisual's 2019 World Air Quality Report. [6]
Early diagnosis can prevent many complications and raise the life expectancy of patients after suffering from lung cancer. But manually going through hundreds of scans from different angles and searching for minute details in those monochromatic scans is both tedious and a task bound to give erroneous results. Also, it is very difficult to differentiate between actual data and noise
Several methodologies have been researched and implemented for efficient and speedy diagnosis of lung cancer which involve the use of ultrasound scans, X-rays, CT scans, biomarkers, exhaled breath analysis for detection purpose. Further numerous data pre-processing and feature extraction techniques have been studied to diagnose cancer using different types of input data format.

2 Organization Of The Study

The study discusses details of various methods used in lung cancer detection. First, the type of data that can be taken as input from the human body to ascertain presence of cancer is considered. Next, preprocessing techniques used for a given type of data are seen. Further, the algorithms that have been developed by researchers for cancer detection and localization are studied.

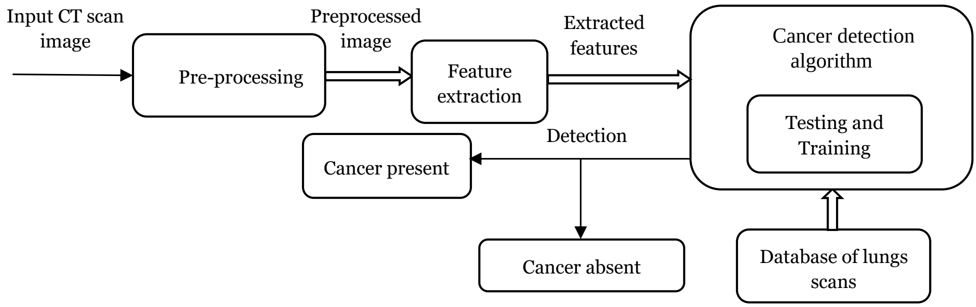

Figure 1. Block diagram of automated cancer detection system

3 Input Data

Input to any computing system is the information that is fed or taken by the system in order to process and produce results using that information. For detecting lung cancer, the most widely used data is CT scan results of the patient. Yet, other forms of information or markers have also been tested.

A. Exhaled Breath Analysis

Thiago Jos´e Barbosa Lima et. al. [7] in their review elaborate the different methods being used for early diagnosis of lung cancer. It describes the usability of biomarkers, exhaled breath analysis and CT scans for the purpose. Human body produces biomarkers in the form of proteins which can be present in hormones, tissues, antigens, oncogenes, etc. as a response to development of cancer cells in the body. The concentration and variety of markers present in the body is being widely used as a method to diagnose easy cancer stage which usually goes unnoticed in scans. Exhaled human breath contains various contents along with carbon dioxide. It consists of variety of volatile organic compounds (VOCs) which are supposed to be a result of tissue level metabolic activities in the body. Around 3000 types of VOCs are identified in human breath. The quantitative and qualitative analysis of different types of VOCs present in the sample can be used to possible tumor growth.

B. X-Rays

X-Rays are a type of electromagnetic radiation that pass variably through the human body as per the density of the body parts and create an image of the internal body structure. This diagnostic technique is mainly useful for scanning and studying bones and thick tissues as the passing of rays reduces with increasing density of the object or body. Sufficiently grown tumors can be detected in X-ray scans [8-10].

C. CT Scans

Computerized Tomography scanning is widely being relied upon for lung cancer nodule detection. A computerized tomography (CT) scan is a combination of number of X-ray images taken from different angles around an individual's body and uses computer processing to create cross-sectional images (slices) of the bones, blood vessels and soft tissues inside the body. Such scan images provide more-detailed information than plain X-rays do. This technique requires different stages like pre-processing with filtering, noise reduction, and segmentation for nodule detection. The scans are being used to design systems for benign and malignant nodule differentiation [7, 11-15]. CT scans are faster and more accurate than most diagnostic methods.

D. Ultrasound Images
Sonography or ultra sound is a medical diagnosis technique which uses high frequency sound waves to produce images of organs and tissues in the body. The sound waves are reflected back or bounced back from the organs or structures in the body which are detected to create an image of what is inside. Ultra sounds are less expensive and comparatively safer than CT scans as they do not consist of radiation. Reference [16] has developed an algorithm in order to detect contours of the lung lesions and estimate if the tumor is benign.

4 Pre-processing

Medical scanning processes like sonography, X-rays, CT scans give images as data. These images are further used as input for new age processing techniques for analysis. But there is a possibility that images coming from different machines or institutions have difference in size, colour, and other properties. To remove any inefficiency in the system due to difference of input images, preprocessing is performed on the input data.

Preprocessing is the treatment of input data to make it more accurate before the actual algorithm works on it for increasing the validity of the system results. It involves enhancement of the images using various operators in the initial stages of a system especially to make it easier to extract the required information in the form of features. It also improves the quality of required data. It includes cropping, scaling, resizing, noise reduction, contrast enhancement, sharpening, edge detection, segmentation, colour correction, angular correction, etc. Each type of preprocessing category has number of algorithms that can be applied as per need. Some of the commonly used preprocessing algorithms for medical images are Otsu thresholding, adaptive thresholding (for thresholding); Sobel operator, Canny operator (for edge detection); resampling and interpolation (for resolution adjustment), etc. [17]

5 Feature Extraction

The parts or properties of an image that contain the information desired for a specific process are known as features. It could be the color, edges, shapes, corners, etc. in the image. First, the system needs to identify the required features and then extract or collect them for further processing. For an efficient and speedy system, minimum number of features with maximum information should be chosen. Some features indicating development of lung cancer are presence of tumors, their size, shape, position, and presence of enlarged lymph nodes. Researchers have designed systems by utilizing distinct variety of features.

A. Geometric Features
Geometric features of convexity, solidity, circularity, and compactness are studied in [14]. Convexity and circularity use convex hull to get the shape of tumor. Solidity identifies the amount and concavities in an object boundary. Compactness compares the area of a shape to area of a circle with same perimeter as the given shape to measure how compact the shape is.

B. Texture Features
In digital image processing, texture of an image can be described as function of spatial variation of pixel intensities. It denotes the roughness or the smoothness of the image. Wavelet transformed or compressed image can be treated with Gray Level Co-occurrence Matrix (GLCM) to extract texture of the image. GLCM calculates the frequency of occurrence of a combination of pixels in four directions (0, 45, 90, 135 degrees) at a given distance. This gives 4 GLCMs. Further, from each GLCM distinct features can be extracted. Haralick features have been extracted from the GLCM to determine the region of interest [15].

C. Filtering
Filtering in digital image processing is usually done to suppress particular frequencies. Images can be filtered in the frequency or spatial domain. Gabor filter analyzes whether there is any specific frequency content in the image in specific directions in a localized region around the point or region of analysis and entropy filter has been used to extract features to locate a tumor by analyzing local texture differences [18].

D. Histogram
Histogram is a graphical representation of the number of pixels of an image having a particular intensity value. It gives an overall view of the intensity variation or concentration in an image. Histogram of Oriented Gradients HOG has been used to extract features in [18, 19]. HOG extracts the gradient (that is the magnitude) as well as orientation (that is the direction of edges) of an object in the area of interest. Gabor filter analyzes whether there is any specific frequency content in the image in specific directions in a localized region around the point or region of analysis whereas, the entropy filter has been used to extract features to locate a tumor by analyzing local texture differences. Scanning processes like sonography, X-rays, CT scans give images as data. These images are further used

6 Classification

Post extraction of features, a classification model is usually utilized to detect whether the extracted features denote presence of cancer or not. Various classifier models have been developed for the purpose and studied. Some of the researched classifiers are Convolutional Neural Networks, Support Vector Machines (SVMs), U-Net, Multi-Layer Perceptron, etc.

A. Convolutional Neural Network

A Convolutional Neural Network is a Deep Learning multi-layer network model. The five parts of CNN are input layer, pooling layer, convolutional layer, fully connected layer, and output layer. In comparison to conventional classification models, CNN requires lesser pre-processing. By increasing number of convolutional layers, a CNN model can be designed to extract more complex features and minute details from an image. The fully connected layer converts the output into a column vector. A CNN uses feed forward neural network and applies back propagation during each iteration to update weights. CNN algorithm with two convolutional layers consisting of 32 and 16 filters respectively has been utilized [20] to distinguish between malignant tumor and benign tumor.

B. U-Net
U-Net is type of Convolutional Neural Network designed specifically for biomedical image processing. It caters the need of not simply classifying into whether diseased or not, but further localizes the abnormality in the image[21] has implemented noisy U-Net (NU-Net) with a better sensitivity towards small nodules by adding specific noise to the training layers for earlier detection of lung cancer[22] uses 3D U-Net model to diagnose lung cancer from CT scans. Unlike most conventional CNNs, U-Net has the same output size as the input image. The network model looks like the letter 'U' and has a symmetric structure. The initial half of U-Net network is known as Contracting Path and consists of convolutional layers while second half is known as the expansive path in which transposed convolution is applied post which the output is expanded to the size of original image. The CNN is trained similar to a neural network.

C. Support Vector Machines
Support Vector Machine (SVM) is a supervised machine learning algorithm usually used for performing classification. In an SVM, an n-dimensional plane represents the n features to be considered and every item is plotted on this plane. The value allotted to each feature of an item is taken to get the corresponding coordinates. A hyperplane is found in the space, such that it distinguished the classes clearly [23] utilizes an SVM for pulmonary nodule detection.

D. Multi Layer Perceptron

A Multi Layer Perceptron (MLP) is supervised learning model that uses back propagation. Except for the input layer, nonlinear functions are used in this model. It is basically a feed forward ANN. MLP has been used for classification after extracting geometric features like solidity and convexity from lung CT scans for the purpose of diagnosis of lung cancer [14].

7 Conclusion

This paper studies different techniques that have been researched and implemented for lung cancer diagnosis. First and foremost a reliable input data format is selected according to convenience of data availability. The input data is preprocessed using various methods. Useful features are then extracted from the preprocessed data. The extracted features are used for classification of data using an algorithm. Various types of neural networks and classifiers have been studied which can implement a model to distinguish between data of patients suffering from lung cancer and those not suffering from lung cancer.

References

[1] A. Ghosh, "India bears 32 per cent global burden of respiratory diseases: Global Burden of Disease study", 2018. [Online]. Available: https://indianexpress.com/article/india/india-respiratory-diseases-global-burden-of-disease-study-5353493/. [Accessed: 22- Oct- 2020].

[2] R. Parveen, "More Indians are being diagnosed with lung cancer, and it" s not because they are smoking", 2018. [Online]. Available: https://theprint.in/india/governance/has-pollution-caused-a-spike-in-lung-cancer-cases-in-india/39878/. [Accessed: 22- Oct- 2020].

[3] S. Ghosh, "Air pollution major, underappreciated contributor to ill health in India", 2019. [Online]. Available: https://india.mongabay.com/2019/01/air-pollution-major-underappreciated-contributor-to-ill-health-in-india/. [Accessed: 22- Oct- 2020].

[4] John Stoddard Cancer Center, "Top Five Most Dangerous Cancers in Men and Women", 2020. [Online]. Available: https://www.unitypoint.org/desmoines/services-cancer-article.aspx?id=c9f17977-9947-4b66-9c0f-15076e987a5d. [Accessed: 22- Oct- 2020].

[5] Prabhat Singh malik and Vinod Raina, "Lung cancer: Prevalent trends & emerging concepts". [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4405940/#:~:text=In%20India%2C%20lung%20cancer%20constitutes,rate%2028.3%20and%2028.7%20per. [Accessed: 22- Oct- 2020].

[6] H. Regan, "21 of the world" s 30 cities with the worst air pollution are in India", 2020 [Online]. Available: https://edition.cnn.com/2020/02/25/health/most-polluted-cities-india-pakistan-intl-hnk/index.html. [Accessed: 22- Oct- 2020].

[7] Thiago Jos´e Barbosa Lima, Fl´avio Henrique Duarte de Ara´ujo, Antˆonio Oseas de Carvalho Filho, Ricardo de Andrade Lira Rabˆelo, Rodrigo de Melo Souza Veras, and Mano Joseph Mathew, "Evaluation of data balancing techniques in 3D CNNs for the classification of pulmonary nodules in CT images", IEEE Symposium on Computers and Communications (ISCC), July 2020

[8] Ali Afzali, Farshid Babapour Mofrad, &Majid Pouladian, "Feature Selection for Contour-based Tuberculosis Detection from Chest X-Ray Images", 26th National and 4th International Iranian Conference on Biomedical Engineering (ICBME), 27-28 November 2019, pgs. 194-198

[9] Norliza Mohd. Noor, Omar Mohd. Rijal, Ashari Yunus, Aziah A. Mahayiddin, Gan Chew Peng, Ong Ee Ling, S. A. R. Abu Bakar, "Pair-wise Discrimination of Some Lung Diseases using Chest Radiography", 2014 IEEE Region 10 Symposium, 14-16 April 2014, pgs. 151-156

[10] Anuradha D. Gunasinghe, Achala C. Aponso, Harsha Thirimanna, "Early Prediction of Lung Diseases", 2019 5th International Conference for Convergence in Technology (I2CT), Pune, India, March 29-31, 2019

[11]Yuyun Y, Miao Tian, Qiyu Liu, & Heng-Ming Tai, “Pulmonary Nodule Detection using V-Net and High-Level Descriptor Based SVM Classifier”, IEEE Access (Volume: 8), 23 September 2020

[12] Ruchita Tekade, Prof. Dr. K. Rajeswari, “Lung Cancer Detection and Classification using Deep Learning”, 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), 16-18 Aug. 2018

[13] Tiantian Fang, “A Novel Computer-Aided Lung Cancer Detection Method Based on Transfer Learning from GoogLeNet and Median Intensity Projections”, 2018 IEEE International Conference on Computer and Communication Engineering Technology (CCET), pgs. 286-290

[14] Yosefina Finsensia Riti, Hanung Adi Nugroho, Sunu Wibirama Budi Windarta, Lina Choridah, “Feature Extraction for Lesion Margin Characteristic Classification from CT Scan Lungs Image”, IEEE 1st International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE), 2016

[15] Moffy Vas, Amita Dessai, “Lung cancer detection system using lung CT image processing”, IEEE International Conference on Computing, Communication, Control and Automation (ICCUBEA), 2017

[16] Achim Cristian, Rusu-Both Roxana, Dulf Eva-Henrietta, Chira Romeo Ioan, “Lung Cancer Diagnosis based on Ultrasound image processing”, IEEE 22nd International Conference on System Theory, Control and Computing (ICSTCC),2018

[17] Fangfang Han, Jinzhu Yang, Yang Liu , Hong Zhao, “Research on Preprocessing Algorithm for PET-CT Image Registration”, IEEE Third International Conference on Digital Manufacturing & Automation, 2012

[18] Mohammad A. Alzubaidi, Mwaffaq Otoom, and Hamza Jaradat, “Comprehensive and Comparative Golbal and Local Feature Extraction Framework for Lung Cancer Detection using CT Scan Images”, IEEE Access (Volume: 9),19 November 2021

[19] Haythem AMEUR, Abdelhamid HELALI, Mohsen NASRI, Hassen MAAREF, and Anis YOUSSEF, “Improved feature extraction method based on Histogram of Oriented Gradients for Pedestrian Detection ”, IEEE Global Summit on Computer & Information Technology (GSCIT),2014

[20] Wadood Abdul, “An Automatic Lung Cancer Detection and Classification (ALCDC) System Using Convolutional Neural Network”, IEEE 13th International Conference on Developments in eSystems Engineering (DeSE), 2020

[21] Wenkai Huang, and Lingkai Hu, “Using a Noisy U-Net for Detecting Lung Nodule Candidates”, IEEE Access (Vol. 7), 2019

[22] Yinglong Wang, and Lyu Zhou, “A Lung Nodule Detector based on U-Net and 3D-CNN Model”, IEEE International Conference on Computer Science, Electronic Information Engineering and Intelligent Control Technology (CEI), 2021

[23] Yuyun Y, Miao Tian, Qiyu Liu, & Heng-Ming Tai, “Pulmonary Nodule Detection using V-Net and High-Level Descriptor Based SVM Classifier”, IEEE Access (Volume: 8), 23 September 2020