Journal of Critical Reviews

ISSN- 2394-5125 Vol 7, Issue 4, 2020

Review Article

MORPHOLOGY BASED LUNG NODULE DETECTION USING MACHINE LEARNING

Dr.R. Yamini[1], Muskan Lawania[2], Aashna Bansal[3]

[1]Assistant Professor, SRM Institute of Science and Technology, Chennai, India. yaminir@srmist.edu.in
[2]Computer Science and Technology, SRM Institute of Science and Technology, Chennai, India. lawaniamuskan@gmail.com
[3]Computer Science and Technology, SRM Institute of Science and Technology, Chennai, India. bansal.aashna49@gmail.com

Received: 05.12.2019 *Revised: 21.01.2020* *Accepted: 22.02.2020*

Abstract

Early recognition of lung malignant growth from registered tomography (CT) examines is basic in helping radiologists with the assistance of Computer-supported detection (CADe) of pneumonic nodules. One of the huge reasons of death in India is Lung Cancer. Different information investigation and order procedures have been utilized for the analysis and discovery of lung malignancy. Since the reason for lung malignant growth isn't found, the counteraction is inconceivable. Early recognition is the best way to fix it. Henceforth, lung nodule identification framework utilizing picture pre-processing and AI is utilized to group the nearness of lung nudules in a CT-picture. Tolerant CT check pictures are characterized as typical and irregular. The strange pictures are exposed to division to concentrate on tumor partition. Order is done on highlights separated from the pictures. This paper proposes a productive strategy to identify the lung malignancy and its stages effectively and furthermore means to have progressively precise outcomes by utilizing KNN and Image Processing systems.

Keywords: Early Cancer Detection, Pulmonary Nodules, Image Processing, Segmentation, Feature Extraction, KNN Technique.

 DOI: http://dx.doi.org/10.31838/jcr.07.04.44

INTRODUCTION

This paper features the significance of information examination and AI in forecast in wellbeing sciences, explicitly in identifying hazardous illnesses like malignant growth. Here right now, consider lung malignancy for our examination. For this reason, prior lung disease patients' information are gathered to get the outcomes. Informational collection which is as analytic pictures is run past Matlab for investigation and anticipating. Picture handling is utilized. Clinical picture division and arrangement are finished. The patients' Computed Tomography (CT) lung pictures are sorted as would be expected or strange. The irregular pictures are exposed to division to concentrate on the tumor parcel. Characterization relies upon highlights extricated from the pictures. The fundamental part is the component extraction stage to yield better characterization execution. This data is then taken care of to AI calculations to perceive an example that can give some great bits of knowledge into what mix of highlights are well on the way to bring about an anomaly. A definitive objective is to distinguish powerful and regular strategies for grouping utilizing some entrenched AI calculations like KNN and Image Processing procedures. The primary goal of our venture is to expand the discovery execution and it should have less computational expense, or practically identical cost when contrasted and the current methodologies. The proposed various leveled identification approach is quick, versatile, and completely programmed. The introduced CADe framework should yield tantamount recognition exactness and more computational proficient than existing frameworks, which ought to be use for clinical utility. Performance and it should have less computational cost, or comparable cost when compared with the existing approaches. The proposed hierarchical detection approach is fast, adaptive, and fully automatic. The presented CADe system should yield comparable detection accuracy and more computational efficiency than existing systems, which should be use for clinical utility.

EXISTING SYSTEM

The fundamental goal is to build up a system which can recognize lung nodules utilizing processed tomography imaging technique. The reason for existing system is to build up the CADe model by utilization of fuzzy c-means grouping procedures.

Fuzzy clustering otherwise called soft clustering or soft k-means is a type of bunching where every datum point can have a place with more than one group.

Clustering includes task of information focuses to groups. The things in a similar group are as comparative as could be allowed, while things having a place with various bunches are as disparate as could be expected under the circumstances. Bunches are recognized by means of likeness measures. These likeness measures incorporate separation, availability, and power. Distinctive closeness measures might be picked dependent on the information or the application.

Limitations

The existing methods are not computationally fast and adaptive.

The fuzzy c-means technique separates the image into different clusters, but each time the cluster change its place while running the program which means the system can't call the particular image at all-time that leads to give the wrong output.

This segmentation is going to segment not only the cancer parts alone but segment the background also leads to wrong calculations of cancer area.

The deep learning techniques needs more advanced hardware requirements leads to more cost.

PROPOSED SYSTEM

The paper proposes lung nodule discovery framework by utilizing image processing procedures. The Pulmonary nodules (lung nodules) are a mass of delicate tissue situated in the lungs which can be analyzed utilizing any radiography methods. Lung nodules doesn't bring about any indications until become harmful. Threatening nodules are frequently brought about by lung malignant growth, yet can likewise be brought about by disease elsewhere in the body, for example, bosom malignant growth and colon disease regularly spread to the lungs. The proposed lung malignancy location framework by utilizing image processing strategies and AI. The framework is utilized to

arrange the nearness of lung malignancy in a CT-picture. CT filter reports are more successful than Mammography; in this manner tolerant CT check pictures are classified in typical and strange.

Architecture

Fig 1.1 shows the engineering of the proposed framework. Every component is briefly depicted in the paper.

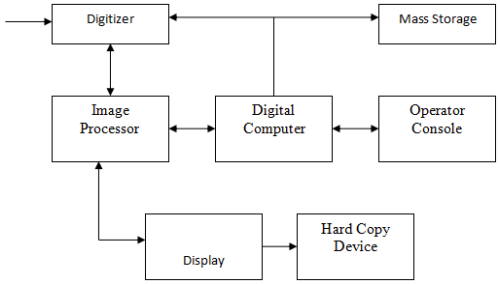


Fig. 1.1: Block Diagram for Image Processing System

Digitizer:

A Digitizer changes over a picture into a numerical portrayal appropriate for contribution to an advanced PC. Some normal digitizers are

i. Microdensitometer
ii. Flying spot scanner
iii. Image dissector
iv. Videocon camera
v. Photosensitive solid- state arrays.
vi. Image Processor

A image processor does the elements of image acquisition, storage, preprocessing, segmentation, representation, recognition and interpretation lastly shows or records the subsequent picture. The fig1.2 gives the crucial grouping associated with a picture preparing framework.

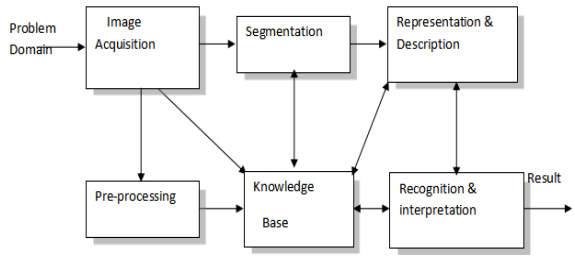


Fig. 1.2: Block Diagram of Fundamental Sequence Involved in an Image Processing System

As definite in the fig1.1, the initial phase in the process is picture securing by an imaging sensor related to a digitizer to digitize the picture. The subsequent stage is the preprocessing step where the picture is improved being taken care of as a contribution to the next processer. Preprocessing regularly manages improving, expelling clamor, detaching areas, and so on. Division parcels a picture into its constituent parts or articles. The yield of division is typically crude pixel information, which comprises of either the limit of the area or the pixels in the locale themselves. Portrayal is the way toward changing the crude pixel information into a structure valuable for resulting preparing by the PC. Depiction manages separating highlights that are essential in separating one class of items from another. Acknowledgment appoints a name to an item dependent on the data gave by its descriptors. Translation includes doling out importance to an outfit of perceived articles. The information about an issue area is fused into the information base. The information base aides the activity of each handling module and furthermore controls the association between the modules. Not all modules need be fundamentally present for a particular capacity. The arrangement of the picture handling framework relies upon its application. The frame rate of the picture processor is typically around 25 edges for every second.

Digital Computer*:*

Scientific handling of the digitized picture, for example, convolution, averaging, expansion, subtraction, and so on are finished by the PC.

Mass Storage:

The secondary storage devices regularly utilized are floppy disks, CD ROMs and so forth.

Hard Copy Device:

The hardcopy device is utilized to create a perpetual duplicate of the picture and for the storage of the product in question.

Operator Console:

The operator console comprises of gear and game plans for confirmation of transitional outcomes and for adjustments in the product as and when require. The administrator is additionally fit for checking for any subsequent blunders and for the section of imperative information.

Algorithms

GLCM (Gray Level Co-Occurance Matrix) algorithm is used to for feature extraction of the processed image.

It is a factual technique for analyzing surface that thinks about the spatial relationship of pixels. The GLCM capacities describe the highlights by checking how frequently matches of pixels are orchestrated in a spatial relationship. Contrast, Correlation, Energy and Homogeneity are the statistics which are being measured by the algorithm. Inverse Difference Moment is calculated using these statistics and it is added to the features array.

K-Nearest Neighbours algorithm is used for classification of healthy and unhealthy lung image.

It is a straightforward and simple to actualize regulated machine learning calculation utilized for classification and regression issues. KNN utilizes the possibility that comparable things exist in closeness. In this paper we are using the IDM values for healthy and unhealthy images as the classification factor. The features of the training images are laoded and according to that the classification of the dataset images is done. The use of KNN algorithm is better than the Fuzzy c-means algorithm because of the lower computational time.

Values Used

For the Segmentation of the image, parameters like area and density are considered. For the density, values from 0-1 are tested and an accurate result is being found for the values greater than 0.8. For area of the tumor region, values from 100-600 pixels are tested and an accurate result is being found for the area greater than 360 pixels.

The values calculated by the GLCM algorithm for the healthy and unhealthy training set images are specified in the table.

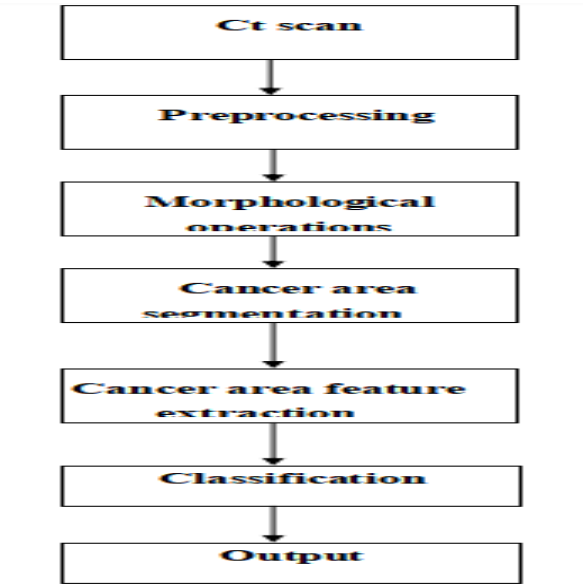
Features	Healthy	Unhealthy
Contrast	0.0111	3.142e-04
Correlation	0.9333	0.9688
Energy	0.8234	0.9896
Homogeneity	0.9945	0.9996
IDM	172.4749	0.7373
max_area	88	795

Proposed System Advantages

i. This paper shows quick and versatile recognition of aspiratory nodules in chest CT checks.
ii. Compared with existing CADe frameworks assessed on a similar lung picture database, our methodology indicated a practically identical discovery ability however a lower computational expense.
iii. The proposed morphology based nodule recognition approach is quick, versatile, and easy to comprehend and furthermore section the malignancy zone alone which prompts give right yield.

Description

Recognition of pulmonary nodules crucially affects the analysis of lung malignant growth, however the discovery is a nontrivial task, not just on the grounds that the presence of pneumonic nodules changes in a wide range, yet in addition since nodules densities have low complexity against neighboring vessel sections and other lung tissues. Computed tomography (CT) has been appeared as the most mainstream imaging methodology for nodule recognition, since it can give solid picture surfaces to the discovery of little nodules. The advancement of lung nodule CADe frameworks utilizing CT imaging methodology has gained great ground over the previous decade. For the most part, such CADe frameworks comprise of three phases: 1) image preprocessing, 2) initial nodule candidates (INCs) identification, and 3) false positive (FP) reduction of the INCs with preservation of the true positives (TPs). In the preprocessing stage, the framework plans to a great extent lessen the search space to the lungs, where a division of the lungs from the whole chest volume is normally required. Due to the high image contrast between lung fields and the encompassing body tissue, image intensity based straightforward thresholding is compelling, and is at present the most usually utilized system for lung division. Be that as it may, the assurance of an exact limit is extraordinarily influenced by image aquisition conventions, scanner types, just as the inhomogeneity of powers in the lung district, particularly toward the division of obsessive lungs with extreme pathologies. This paper proposes a versatile answer for relieve the trouble of thresholding-based technique in lung division. Adequate location power for nodule candidates is definitely joined by many (self-evident) FPs. The guidelines gained from the prepared dataset can be applied to the separation among nodules and nonnodules in the test dataset.



Input Image

Read and Display an input Image. Add a picture to the workspace, utilizing the imread() method or camera. In picture preparing, it is characterized as the activity of recovering a picture from some source, typically an equipment based source for handling. It is the initial phase in the work process grouping on the grounds that, without a picture, no preparing is conceivable. The picture that is gained is totally natural.

Preprocessing

Processing is a typical name for tasks with pictures at the most minimal degree of deliberation. Both input and yield are intensity pictures. The point of pre-handling is an improvement of the picture information that stifles undesirable contortions or upgrades some picture highlights significant for additional preparing. Picture pre-processing strategies utilize the extensive excess in pictures. Neighboring pixels comparing to one item in genuine pictures have basically the equivalent or comparative brightness value. In this way mutilated pixel can frequently be reestablished as an average value of neighboring pixels.

RESIZING THE INPUT IMAGE

All the input images are resized into same measurements. In the event that the predetermined size doesn't create a similar viewpoint proportion as the input picture, the yield picture will be distorted.

Segmentation

Image Segmentation is a usually utilized system in computerized picture preparing and examination to segment a picture into various parts or districts, regularly dependent on the attributes of the pixels in the picture. In PC vision, Image Segmentation is the way toward subdividing an advanced picture into various sections (sets of pixels, otherwise called super pixels). Division is a procedure of collection together pixels that have comparable properties. Picture Segmentation is the way toward parceling a picture into non-converging areas with the end goal that every locale is homogeneous and the association of no two contiguous districts is homogeneous Pixels in a locale are comparative as indicated by some homogeneity criteria, for example, color, intensity or texture in order to find and distinguish items and limits (lines, curves, etc.) in a picture. Segmentation accuracy decides the possible achievement or disappointment of electronic investigation method.

Colour Space Conversions*:*

For some uses of image processing, color information data doesn't support us. In the event that you get into the matter of endeavoring to recognize hues from each other, at that point one explanation behind changing over RGB picture to BLACK AND WHITE or GRAYSCALE organizes in picture.

Morphological Operations

Morphological image handling is an assortment of non-linear operations identified with the shape or morphology of features in a picture. Morphology is a wide arrangement of image processing tasks that procedure pictures dependent on shapes. Morphological tasks apply an organizing component to an input picture, making a yield picture of a similar size. Some morphological activities are given beneath:

a) Dilation
b) Erosion
c) Areaopen
d) Border clearing

Dilation

Dilation is one of the two fundamental administrators in the territory of numerical morphology, the other being erosion. It is ordinarily applied to binary images, however there are renditions that chip away at grayscale pictures. The essential impact of the administrator on a twofold picture is to step by step develop the limits of areas of frontal area pixels (for example white pixels, commonly). Along these lines zones of frontal area pixels develop in size while gaps inside those areas become littler.

Erosion
Erosion is one of the two essential administrators in the region of numerical morphology, the other being dilation. It is commonly applied to twofold pictures, however there are variants that chip away at grayscale pictures. The essential impact of the administrator on a parallel picture is to dissolve away the limits of districts of frontal area pixels (for example white pixels, regularly). In this manner zones of frontal area pixels shrivel, and openings inside those regions become bigger.

Areaopen
Opening and shutting are two significant administrators from scientific morphology. They are both gotten from the essential activities of erosion and dilation. Like those administrators they are typically applied to binary images, in spite of the fact that there are likewise graylevel variants. The fundamental impact of an opening is to some degree like disintegration in that it will in general evacuate a portion of the frontal area (splendid) pixels from the edges of districts of closer view pixels. Anyway it is less dangerous than erosion when all is said in done. Likewise with other morphological administrators, the specific activity is controlled by an organizing component. The impact of the administrator is to protect forefront regions that have a comparable shape to this organizing component, or that can totally contain the organizing component, while killing every other area of frontal area pixels.

Border Clearing
Right now undesirable outskirts are eliminated by applying the border clearing activities. Imclearborder (I) smothers structures in picture that are lighter than their environment and that are associated with the picture border. Utilize this capacity to clear the picture border. For grayscale pictures, imclearborder will in general diminish the general intensiy level notwithstanding stifling border structures. The yield picture is grayscale or double, contingent upon the info.

Watershed Segmentation
In the study of image processing, a watershed is a change characterized on a grayscale picture. The name alludes figuratively to a geographical watershed, or drainage divide, which isolates contiguous drainage basins. Watershed calculation is utilized in picture preparing principally for segmentation reason.

Feature Extraction
In machine learning, pattern recognition and in image processing, feature extraction begins from an underlying arrangement of estimated information and constructs inferred values (highlights) planned to be useful and non-repetitive, encouraging the resulting learning and speculation steps, and at times prompting better human translations. Feature extraction is identified with dimensionality decrease. At the point when the input information to a calculation is too enormous to ever be handled and it is suspected to be repetitive (for example a similar estimation in the two feet and meters, or the tedium of pictures introduced as pixels), at that point it tends to be changed into a diminished arrangement of highlights (likewise named an element vector). Deciding a subset of the underlying features is called feature selection. The chose features are relied upon to contain the significant data from the information, with the goal that the ideal assignment can be performed by utilizing this diminished portrayal rather than the total starting information. Given below are the types of faetures:

a) Shape features
b) Geometrical features
c) Texture features

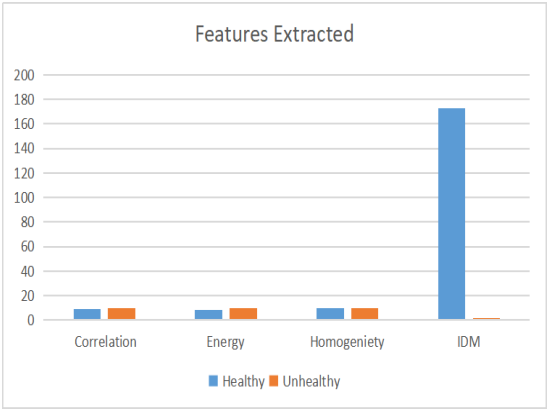
Shape Features
Visual features of objects are known as the shape characteristics or visual features. For instance, round article or triangular items or different shapes, edge limit of the article, the distance across of the fringe, etc. The visual features demonstrated naturally are all has a place with shape features.

Geometrical Features
Geometric features are highlights of items developed by a lot of geometric components like focuses, lines, bends or surfaces. These features can be corner highlights, edge highlights, Blobs, Ridges, remarkable point's picture surface, etc., which can be recognized by feature detection strategies.

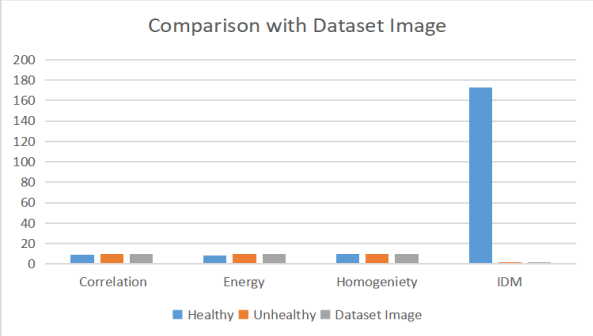
Texture Features
An image texture is a lot of measurements determined in image processing intended to evaluate the apparent surface of a picture .Image Texture gives us data about the spatial course of action of colors or intensities in a picture or chosen region of a picture.
Some feature extraction techniques are GLCM (Gray level cooccurance matrix).

Classification
Image classification alludes to the undertaking of separating data classes from a multiband raster picture. The subsequent raster from picture grouping can be utilized to make topical maps. The prescribed method to perform arrangement and multivariate investigation is through the Image Classification toolbar. There are many classification algorithms are available and some classification algorithm that are given below is KNN (K-NEAREST NEIGHBOUR).

RESULTS AND DISCUSSIONS
The features like Contrast, Correlation, Energy and Homogeneity are being extracted using GLCM algorithm. When the features are plotted for healthy ad unhealthy training set images, a large deviation is seen in the values of IDM calculated. Rest all the features are nearly equal for both. Same is shown in the chart given.



On the basis of IDM value, KNN algorithm is applied and the dataset images are being classified. Following chart shows the deviation of the IDM value of the dataset image with the unhealthy training image.



The input dataset image is not a healthy image, it has a cluster of nodules present in the CT Scan image as shown in the figure below.

This shows that the system proposed in this paper is accurate, more cost and time efficient.

CONCLUSION

This paper gives a review of current segmentation and classification procedures utilized in the distinguishing proof of lung malignant growth tumor from CT pictures. It might assist analysts with selecting a fitting technique. The paper likewise proposed a strategy for productive division and characterization. KNN gave great outcomes with high precision. The morphological tasks and watershed division featured the tumor in the picture and it could extricate the tumor. Later on, in future work, a PC helped finding (CAD) framework may use to identify little size pulmonary nodules from the chest computed tomography (CT) pictures through three-dimensional (3-D) technique.

REFERENCES

1. https://www.medicalnewstoday.com/info/lung-cancer
2. Mokhled Altarawneh," Lung Cancer Detection Using Image Processing Techniques", Leonardo Electronic Journal of Practices and Technologies, 2012, Issue 20, ISSN 1583-1078.
3. Dr.T. Arumuga Maria Devi, Mr. Mebin Jose V.I, Dr.P. Kumar Parasuraman," A novel approach for automatic detection of non-small cell lung carcinoma in CT images", International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT), 2016, ISBN: 978-1-5090-5240-0, DOI: 10.1109/ICCICCT.2016.7988059.
4. P.B. Sangamithraa, S. Govindaraju," Lung tumour detection and classification using EK-mean clustering", International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), 2016, ISBN: 978-1-4673-9338-6, DOI: 10.1109/WiSPNET.2016.7566533.
5. Yosefina Finsensia Riti, Hanung Adi Nugroho, Sunu Wibirama, Budi Windarta, Lina Choridah," Feature extraction for lesion margin characteristic classification from ct scan lungs image", International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE), 2016, ISBN: 978-1-5090-1567-2, DOI: 10.1109/ICITISEE.2016.7803047.
6. Suren Makaju, P.W.C. Prasad, Abeer Alsadoon, A.K.Singh, A. Elchouemi, "Lung cancer detection using ct scan images", 6th International Conference on Smart Computing and Communications, ICSCC 2017, https://doi.org/10.1016/ j.procs. 2017.12.016.
7. Eman Magdy, Nourhan Zayed, and Mahmoud Fakhr, "Automatic classification of normal and cancer lung ct images using multiscale amfm features", International Journal of Biomedical Imaging Volume 2015, http://dx.doi.org/10.1155/2015/230830.
8. Emre Dandil, Murat Cakrolu, Ziya Eksi, Murat Ozkan, Ozlem Kar Kurt, Arzu Canan," Artificial neural network-based classification system forlung nodules on computed tomography scans", 6th International Conference of Soft Computing and Pattern Recognition (SoCPaR), 2014, ISBN: 978-1-4799-5934-1, DOI: 10.1109/SOCPAR.2014.7008037.
9. Anam Tariq, M. Usman Akram, M. Younus Javed, "Lung nodule detection in ct images using neuro fuzzy classifier", IEEE Fourth International Workshop on Computational Intelligence in Medical Imaging (CIMI), 2013, ISSN: 2326-9928, ISBN: 978-1-4673-5919-1, DOI: 10.1109/CIMI.2013.6583857.
10. Salsabil Amin El-Regaily, Mohammed Adbel Megeed Salem, Mohamed Hassan Abdel Aziz, Mohamed Ismail Roushdy," Lung nodule segmentation and detection in computed tomography", Eighth International Conference on Intelligent Computing and Information Systems (ICICIS), 2017, ISBN: 978-1-5386-0821-0, DOI: 10.1109/INTELCIS.2017.8260029.
11. Jaspinder Kaur, Nidhi Garg, Daljeet kaur," An automatic CAD system for early detection of lung tumor using back propagation network", International Conference on Medical Imaging, m-Health and Emerging Communication Systems (MedCom), 2014, ISBN: 978-1-4799-5097-3, DOI: 10.1109/MedCom.2014.7006014.
12. S. Kalaivani, Pramit Chatterjee, Shikhar Juyal, Rishi Gupta, "Lung cancer detection using digital image processing and artificial neural networks", International conference of Electronics, Communication and Aerospace Technology (ICECA), 2017, ISBN: 978-1-5090-5686-6, DOI: 10.1109/ICECA.2017.8212773.
13. Md. Badrul Alam Miah, Mohammad Abu Yousuf," Detection of lung cancer from ct image using image processing and neural network", International Conference on Electrical Engineering and Information Communication Technology (ICEEICT), 2015, ISBN: 978-1-4673- 6676-2, DOI: 10.1109/ICEEICT.2015.7307530.
14. Omnia Elsayed, Khaled Mahar, Mohamed Kholief, Hatem A. Khater, "Automatic detection of the pulmonary nodules from ct images", SAI Intelligent Systems Conference (IntelliSys), 2015, ISBN: 978-1-4673- 7606-8, DOI: 10.1109/IntelliSys.2015. 7361223.
15. Jinsa Kuruvilla, K. Gunavathi, "Lung cancer classification using neural networks for ct images", Computer Methods and Programs in Biomedicine Volume 113, Issue 1, January 2014, https://doi.org/10.1016/j.cmpb.2013.10.011.
16. Niranjan Shukla, Aakash Narayane, Aniket Nigade, Krishnakumar Yadav, Harshada Mhaske, "Lung cancer detection and classification using support vector machine", International Journal of Advanced Trends in Computer Science and Engineering, 2016, DOI: 10.18535/Ijecs/v4i11.20.
17. Manasee Kurkure, Anuradha Thakare, "Lung cancer detection using genetic approach", International Conference on Computing Communication Control and automation (ICCUBEA), 2016, ISBN: 978- 1-5090-3291-4, DOI: 10.1109/ICCUBEA.2016.7860007.
18. Allison M Rossetto, Wenjin Zhou," Deep Learning for Categorization of Lung Cancer CT Images", IEEE/ACM International Conference on Connected Health:

Applications, Systems and Engineering Technologies (CHASE), 2017, ISBN: 978-1-5090-4722-2, DOI: 10.1109/CHASE.2017.98.

19. Omar SH, Nabi Saba. "Melatonin, Receptors, Mechanism, and Uses." *Systematic Reviews in Pharmacy* 1.2 (2010), 158-171. Print. doi:10.4103/0975-8453.75069
20. Bhagyarekha U. Dhaware, Anjali C. Pise, "Lung cancer detection using bayasein classifier and FCM segmentation", International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT), 2016, ISBN: 978-1-5090-2080-5, DOI: 10.1109/ICACDOT.2016.7877572
21. Avinash. S, Dr.K. Manjunath, Dr.S. Senthil kumar," An improved image processing analysis for the detection of lung cancer using gabor filters and watershed segmentation technique", International Conference on Inventive Computation Technologies (ICICT), 2016, ISBN: 978-1-5090- 1285-5.
22. Sharma, S., Ghalaut, V.S., Dixit, R., Kumar, S., George, P.J.Microalbuminuria and C-reactive protein as a predictor of coronary artery disease in patients of acute chest pain(2013) Journal of Cardiovascular Disease Research, 4 (1), pp. 37-39.
DOI: 10.1016/j.jcdr.2013.02.014