INTERNATIONAL DEVELOPMENT PLANNING REVIEW
ISSN:1474-6743 | E-ISSN:1478-3401

EXPLORING MACHINE LEARNING TECHNIQUES IN LUNG CANCER DETECTION VIA ABSOLUTE INTEGRAL-BASED ANALYSIS

Dr.S.Sumathi
Associate Professor, Department of ECE, Sri Sairam Engineering College, Chennai.

Dr.K.Sumathi
Associate Professor, Department of ECE, Sri Sairam Engineering College, Chennai

Dr.T.Mangayarkarasi
Associate Professor, Department of ICE, Sri Sairam Engineering College, Chennai.

Dr.B.Panjavarnam
Associate Professor, Department of ECE, Sri Sairam Engineering College, Chennai.

R. Tamezheneal
Assistant professor, Department of ECE, Sri Sairam engineering college, Chennai

A.E Prabhu
Assistant Professor, Department of ECE, Rajalakshmi Engineering College, Chennai.

Abstract

This paper aims to provide accurate lung cancer detection using a machine-learning technique has been performed. Here, the proposed machine learning technique, Absolute Integral Based Analysis has been implemented to detect accurate cancer affected parts in the lungs. Matlab coding has been used to perform AIA for the detection of cancer-affected parts or cells in the lungs. This paper has taken a dataset from the appropriate link where research data are available. Performance parameters like accuracy, sensitivity, specificity, precision, recall, f1 score, and gmean and validation accuracy are evaluated accordingly. The Above proposed method Absolute Integral Based Analysis (AIA) machine learning technique detects accurate cancer affected parts in the lungs efficiently and it provides improved accuracy and gmean.

Keywords: DIP (digital image processing), ML (machine learning), AIA (Absolute Integral based Analysis), AI (Artificial Intelligence), CAD (Computer Aided Diagnosis).

Introduction

Image Processing

Digital image processing is the process of developing an algorithm that helps to gather some useful information from the images. Many methods deal with image processing and can be used mainly used in medical or telehealth applications. According to the algorithm to be performed, images are taken in the form of pixels and the calculations are made with respect to computerized handling, and signal handling is divided into two aspects, which are advanced picture handling and multiphase frameworks.

The development of image processing methods is categorized into three variables, first, improvement in computer vision, second, using advanced continuous and discrete signals, third, implementation of image processing methods in numerous applications like remote sensing, agribusiness, telehealth, military, etc.

Imaging in CAD systems to diagnose lung cancer

CAD (Computer Aided Diagnosis) is one of the best methods to detect lung cancer. CAD can also be called computer-aided detection or Image based automated diagnosis. It provides better clarification for doctors while detecting the cancer-affected part, especially in the lungs in lesser time. CAD systems provide effective output in the form of images or videos and have probable applications in fields like digital pathology, to detect breast cancer, colon cancer, Alzheimer's disease, and the strategy of the diabetic. Machine learning algorithms are mainly used to detect lung cancer easily and to examine H and E stains. [1]

For the detection of cancer in the lungs, CT with 3D CAD systems is designed and 3000 individual images can be visualized and considered to perform the appropriate algorithm. By using CAD, the initial stage of lung cancer can be detected with a size of 1mm, and up to 5 – 10 mm size can be identified easily with the help of lung nodules. Virtual dual-energy image processing increases the function of CAD systems in chest radiography. [2]

Machine Learning

Machine Learning is a field of request to understand and build strategies that 'realize', that is, techniques that influence information to further develop execution on some arrangement of tasks. Machine learning can also be called man-made reasoning. Machine learning techniques will proceed to do calculations and build a model by using all the extracted information. Customized expectations and test pieces of information are pursued as per the blocks derived. The above techniques are used in a wide variety of applications in the medical field, agriculture, remote sensing, etc, and places where the appropriate tasks or methods to be performed in a faster manner. Methods of AI techniques are related to computational activities, where computer vision techniques are widely used. AI techniques can also be called fast learning techniques. [5]

Machine learning techniques have well exposure to mathematical or numerical calculations and in the field of space, where it is used to predict future values. In data mining, ML plays a vital role to examine the data. As discussed earlier, AI techniques are used to predict future values, mainly used in the medical field to execute networks that impersonate the operation of the organic brain. So compared to all fields, AI can also be called, a predictive investigation method. [4]

Machine learning in the Medical field

Machine learning is a particular kind of computerized reasoning that permits frameworks to gain from information and distinguish designs absent a lot of human mediation. Rather than being determined what to do, PCs that utilize AI are shown examples and information which then permits them to arrive at their own decisions. AI calculations have different capabilities, such as assisting with separating email, distinguishing objects in pictures, and breaking down huge volumes of progressively complex informational indexes. PCs use AI frameworks to naturally go through messages and track down spam, as well as perceive things in pictures and cycle huge information. AI strategies are a developing field of exploration with numerous likely applications. As understanding information turns out to be all the more promptly accessible, AI innovation will turn out to mean a lot to medical care experts and well-being frameworks for separating significance from clinical data. [9]

For the medical services industry, AI is especially important on the grounds that it can assist us with getting a handle on the monstrous measures of medical care information that is created consistently inside electronic well-being records. Utilizing AI in medical services like AI calculations can assist us with tracking down examples and experiences that would be difficult to physically find. As AI in medical services acquires far and wide reception, medical services suppliers have a potential chance to adopt a more prescient strategy that makes a more bound together framework with further developed care conveyance and patient-based processes. [6]

The most accepted medical services used in AI techniques are clinical charging, scientific predictions, and improvement in analytic suggestions based on well-known frameworks. Effective and eminent undeniable calculation level of AI has plenty of methods that have been implemented mainly in telehealth applications and in the field of science and medication. MD Anderson, an information researcher, identified and exposed the vital role of deep learning in medical services. Many algorithms have been developed to forecast different levels in patients involved in the medication as well as in radiation for head and neck diseases. [11]

In clinical work processes, the information created by profound learning in medical services can recognize complex examples consequently and offer an essential consideration supplier clinical choice help at the place of care inside the electronic well-being record. Enormous volumes of services regarding patient information can be viewed with the help of AI techniques but practically proved that 80% of the data are "locked" such that without permission of the appropriate person those data cannot be viewed.

In previous days, patient records are viewed physically by the appropriate doctors or technicians. Human language, or "normal language," is exceptionally intricate, lacking consistency, and consolidates a colossal measure of equivocalness, language, and unclearness. [12]

To change over these records into additional helpful and analyzable information, AI in healthcare often relies on computerized reasoning like normal language processing (NLP) programs. Most profound learning in medical care applications that utilization of NLP requires some type of medical care information for AI.

Literature Review

Chapala Venkatesh et al, 2022, this paper represents one of the most widely recognized reasons for death from disease for all kinds of people is a cellular breakdown in the lungs. Lung knobs are basic for the screening of malignant growth and early acknowledgment grants treatment and upgrade the pace of restoration in patients. Albeit a great deal of work is being finished around here, an exactness expansion is expected to enlarge the patient constancy rate.

In any case, conventional frameworks don't section malignant growth cells of various structures precisely and no system attained more prominent dependability. A powerful screening methodology is proposed in this work to recognize cellular breakdown in the lung sores quickly as well as to increment exactness.

In this system, the Otsu thresholding segmentation is used to achieve the ideal separation of the chosen region. The cuckoo search calculation is used to characterize the best qualities for dividing disease knobs. By utilizing a nearby double example, the important highlights of the injury are recovered. The Convolutional Neural Network classifier is intended to detect whether a lung injury is pernicious or non-malevolent given the recovered highlights. The proposed structure accomplishes an exactness of 96.97% percent. This article uncovers the exactness that precision is improved, and the outcomes are ordered. Utilized Molecule mob improvement and hereditary calculations are done. [1]

Timor Kadir et al, 2018, in this paper, AI-based techniques are used to predict the cellular breakdown in lungs and predictive methods have been proposed which will be useful in the medical field to diagnose pneumonic knobs.

This proposed method helps to reduce changeability in knob order and it produces a path to minimize the numerous knobs which are unnecessary. This article has given a path for the forecast approach of methods involved in the treatment of lungs and it helps to figure out overall aspects and disadvantages in the methods to proceed. [16]

Suren Makaju et al, 2018, this paper indicate Cellular breakdown in the lungs is one of the most hazardous and life-taking sicknesses on the planet. In any case, early finding and treatment can save a life.

According to the survey which explains that a Computer Tomography scan is the most effective imaging method in the clinical field, it is very difficult for specialists to decide and identify cancer-affected parts from CT examination pictures. Hence PC supported analyses are helpful for technicians to distinguish malignant cells easily. Various PC-helped strategies that apply picture handling and AI has been investigated and executed.

The fundamental way of this technique is to analyze the different PC-oriented strategies and investigate the growing procedure to figure out their bounds and disadvantages to propose an effective method that results in better visualization compared with the current model.

The strategy utilized was to arrange and record cellular breakdown in the lung recognition methods based on their identification precision. The methods were dissected on each step and by large constraint, demerits were also taken into an account. It is seen that some have low accuracy and some have huge precision yet not nearly 100 percent. Hence, our expectations focus to build precision nearly 100 percent. [18]

Meraj Begum et al, 2021, the fundamental goal of this paper is to figure out the beginning phase of the cancer stage in the lungs and investigate the precision levels of different AI calculations. After a methodical writing review, we figured out that a few classifiers have low precision and some are higher exactness yet challenging to come to closer 100 percent.

Low exactness and high execution cost because of ill-advised management with DICOM pictures. For clinical picture handling various sorts of pictures are utilized however PC Tomography (CT) examines are for the most part favored due to less commotion.

Profound learning is shown to be the best strategy for clinical picture handling, lung knob location and arrangement, highlight extraction, and cellular breakdown in the lungs stage forecast. The primary phase of this framework utilized picture-handling strategies to separate lung districts. The division is finished utilizing K Means. The elements are separated from the sectioned pictures and the grouping are finished utilizing different AI calculation. The exhibitions of the proposed approaches are assessed in light of their exactness, awareness, particularity, and characterization time. [20]

Manickavasagam et al, 2019, this paper explain one of the maximum well-established reasons for dying from advanced cancerous stages in all kinds of people. Lung knobs are the basic process for covering malignant growth and initial acknowledgment of licensed medication and it improves the effect of recovery in patients.

A successful screening is proposed in this work to distinguish cellular breakdown in the lungs sores quickly as well as to increment precision. In this technique, the Otsu thresholding division is used to achieve the ideal detachment of the chosen region, and the cuckoo search calculation is used to characterize the efficient qualities for apportioning malignant growth knobs. By utilizing a nearby paired design, the significant elements of the injury are recovered. [23]

Proposed Absolute Integral Based Analysis (AIA) Method

The proposed method Absolute Integral Based Analysis (AIA), a machine learning technique has been adopted to detect accurate cancer affected parts or cells in the lungs. Performance metrics like sensitivity, algorithm accuracy, specificity, precision, recall, f1 score, and gmean are evaluated accordingly.

The proposed system has been implemented in the software Matlab version 2021a. Matlab inbuilt functions like the image inline, image location, image data store, and image index, vision cascade and vision blob analysis are used. According to the inbuilt parameters, a formula has been derived to perform AIA.

Formula to perform AIA can be derived as,

$$y = x\,(0{:}4\pi / 8*2\pi/4)$$

In this method, initially, a data store has been created and samples of 100 images are considered for the process which has been stored in the particular data directory.

In AIA, the number of iterations is performed and the corresponding images are displayed with a pixel size of 300 x 300. As a result, 100 images are displayed in a single Matlab window. For the second iteration, the proposed formula will set a reference image for the comparison of normal and cancer-affected images. As a result, two images that have been taken for comparison are displayed in a single Matlab window. The Block diagram of the proposed method can be pictured as,

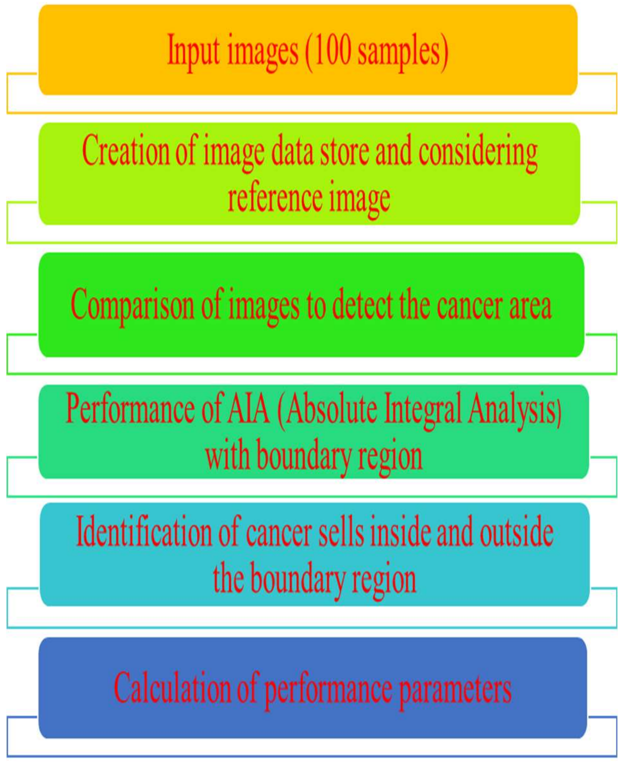

Figure 1 Block diagram of AIA

After comparison, cancer affected part has been indicated with the boundary box. Using region extraction cancer cells are detected inside and outside the boundary box. Then cancer affected images and normal images are considered for mapping where the cancer cells are indicated through straight lines.

Here the cells are indicated in red color and the lines are indicated in yellow color so that the mapping can be viewed clearly. In this proposed method, mostly damaged cells by cancer as well as cells that are going to get damaged are indicated.

The AIA method performs the process and indicates the stage of the cell. Training progress has been performed and achieved 98.76% efficiently along with the algorithm accuracy 90%. The proposed method can withstand 1000 samples to perform a number of iteration levels. If the number of samples is increased by more than 1000 then the value of algorithm accuracy and validation accuracy has been reduced. The Value of performance metrics is evaluated and indicated.

Formulas used for calculation of performance parameters:

idx = (ACTUAL value ==1);

p = length (ACTUAL value of (idx));

n = length (ACTUAL (~idx));

N = positive + negative;

True positive = sum (ACTUAL(idx)==PREDICTED(idx))

True negative = sum(ACTUAL(~idx)==PREDICTED(~idx));

False positive = n-tn;
False negative = p-tp;
tp_rate = tp/p;
tn_rate = tn/n;

algorithm accuracy = (true positive +true negative)/N;
sensitivity = true positive_rate;
Specificity = true negative_rate;
Precision = true positive/(tp+fp);
Recall = sensitivity;

f1 score = 2*((precision*recall)/(precision + recall));

Gmean = sqrt (tp_rate*tn_rate);

These are the formulas indicated to evaluate the performance metrics which has been mentioned above.

Results and Discussions

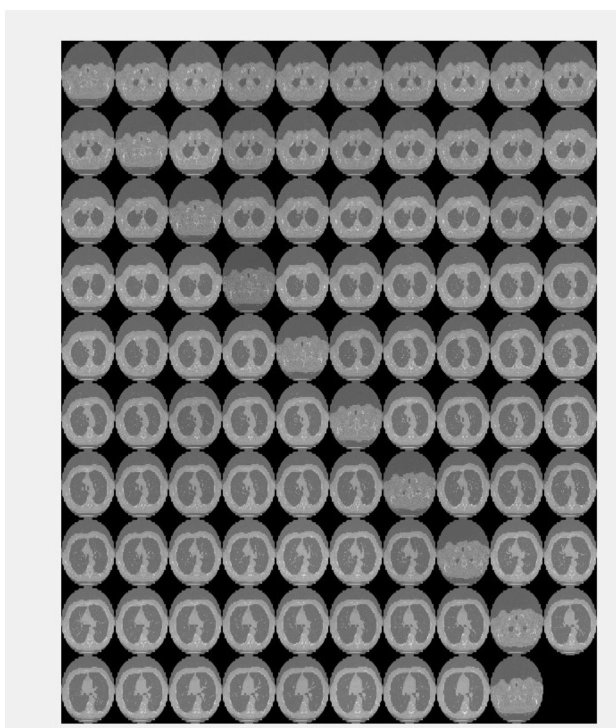

Figure 1 Input Sample Images

Figure 1 shows that the dataset of 100 lung images are considered to perform comparison and detection of cancer-affected part

Figure 2 Reference Image

Figure 2 indicates the reference image taken from sample input images for detection

Figure 3 Two images from the dataset are taken for comparison

Figure 3 indicates a comparison of two images taken with reference to the query image

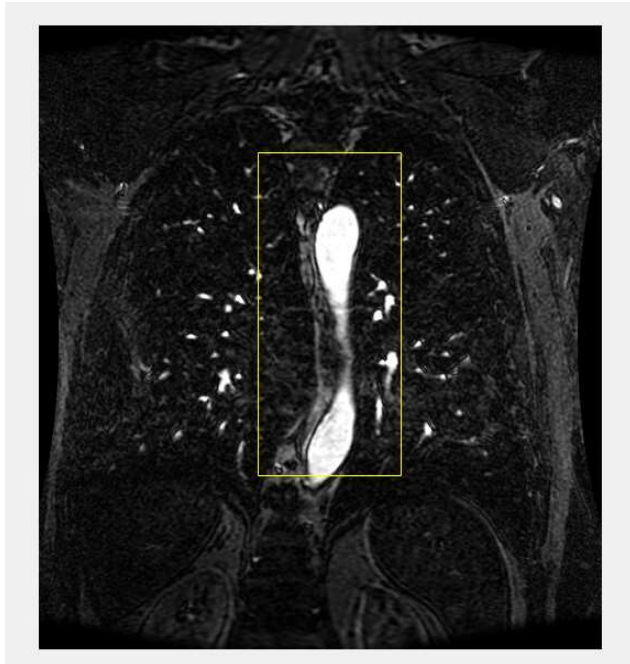

Figure 4 Indication of the most affected cancer part

Figure 4 shows that the part affected by cancer with the indication of the boundary box

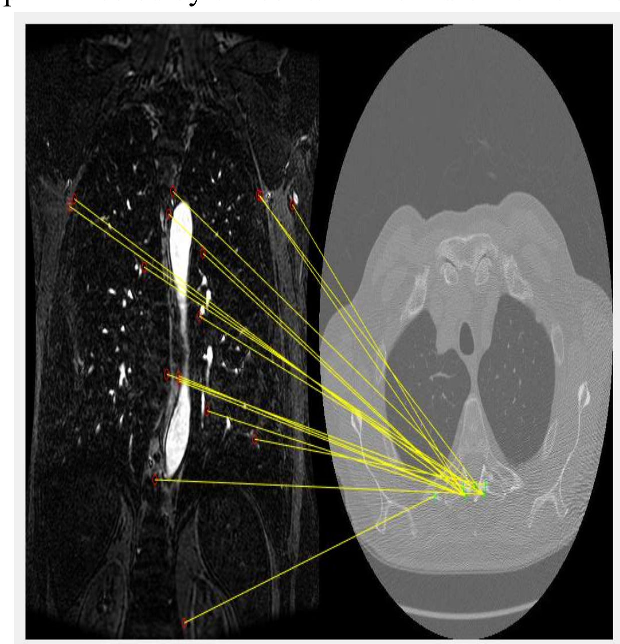

Figure 5 Indication of partially affected cells near most and outer of the region

Figure 5 indicates the partially affected cells which are closer and outer to the boundary box

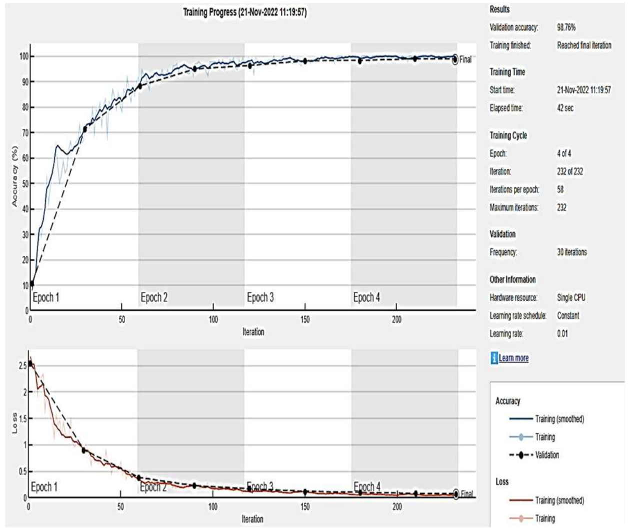

Figure 6 Training Progress of the Proposed AIA Algorithm

Figure 6 indicates the performance of the proposed AIA algorithm and the progress achieved with a validation accuracy of 98.76%

The performance parameters are calculated accordingly and the proposed method achieved algorithm accuracy of 90%, a sensitivity of 100, a specificity of 80, a precision of 66.6667, an f1 score of 80, a recall of 100, and a gmean of 89.447. Step by step the AIA algorithm has been explained with the result of images. The bar chart has been represented according to the values obtained through AIA.

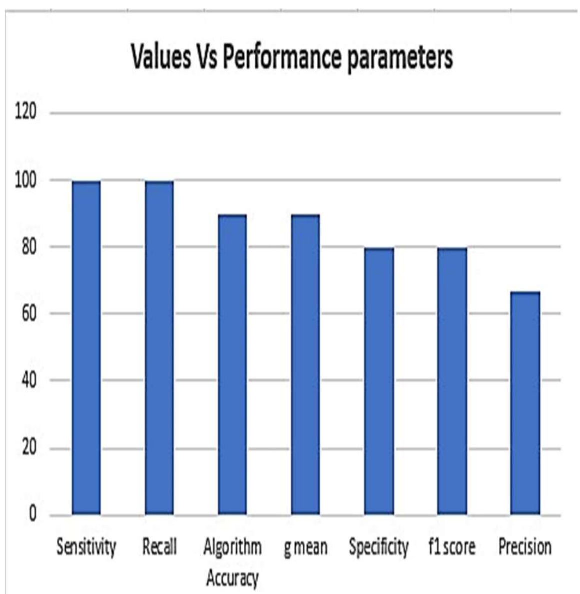

Table 1 Comparative result based on Accuracy value

For comparison regarding previous work, the algorithm accuracy parameter has been considered. According to the proposed method, AIA has obtained better algorithm accuracy of 90% compared to previous work. The table has been indicated below

Previous Work	Comparative values of Algorithm Accuracy (%)
Lung Cancer Detection using CT scan images	88.4
Neural Network and optimization based Lung Cancer Detection	89.90
Lung Cancer Prediction using Machine Learning and Advanced Imaging Techniques	80
Lung Cancer Detection and Classification using Machine Learning	84.4
Lung Cancer Detection using Lung Knobs by Cuckoo Search Algorithm	88
Proposed Absolute Integral Based Analysis for Lung Cancer Detection	90

Conclusion

The proposed method Absolute Integral Based Analysis (AIA) is adopted and obtained efficiently improved algorithm accuracy of 90% and validation accuracy of 98.76%. Using Region extraction damaged cells are identified inside and outside of the boundary box with an improved quality level. Performance parameters like Sensitivity, Specificity, Recall, Precision, Algorithm accuracy, Validation accuracy, f1 score, and gmean are evaluated accordingly. The proposed method can be used in hospitals in real time where a larger number of patient records are taken for cancer diagnosis. The above work can be extended for future research work by considering a greater number of image samples for a larger number of patients.

Acknowledgment

Datasets considered for the proposed Absolute Integral Based Analysis method are taken from the below website. https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=70224216

References

- Venkatesh C, Ramana K, Lakkisetty SY, Band SS, Agarwal S and Mosavi A (2022) A Neural Network and Optimization Based Lung Cancer Detection System in CT Images. Front. Public Health 10:769692. doi: 10.3389/fpubh.2022.769692.
- Meraj Begum Shaikh Ismail, 2021, Lung Cancer Detection and Classification using Machine Learning Algorithm, Turkish Journal of Computer and Mathematics Education Vol.12 No.13 (2021), 7048- 7054.
- Manickavasagam R, Selvan S. Automatic detection and classification of lung nodules in CT imaging neuro-fuzzy neuro-fuzzy classifier with h cuckoo search algorithm. J Med Syst. (2019) 43:1–9. doi: 10.1007/s10916-019-1177-9.
- Kadir and Gleeson et al, 2018, Lung cancer prediction using machine learning and advanced imaging techniques, Transl Lung Cancer Res 2018;7(3):304-312.

- Suren Makaju et al, 2018, Lung Cancer Detection using CT Scan Images, Procedia Computer Science 125 (2018) 107–114 6.
- Brown MS, McNitt-Gray MF. Patient-specific models for lung nodule detection and surveillance in CT images. Proc IEEE Trans Med Imag. (2001)20:1242–50. doi: 10.1109/42.974919
- Cosman PC, Tseng C, Gray RM. Tree-structured vector quantization of CT chest scans: image quality and diagnostic accuracy. Proc IEEE Trans Med Imag. (1993) 12:727–39.
- Dewes P, Frellesen C, Al-Butmeh F, Albrecht MH, Scholtz J-E, Metzger SC, et al. Comparative evaluation of non-contrast CAIPIRINHA-VIBE 3T-MRI and multidetector CT for detection of pulmonary nodules: in vivo evaluation of diagnostic accuracy and image quality. Eur J Radiol. (2016) 85:193–8. doi: 10.1016/j.ejrad.2015.11.020
- Frontiers in Public Health | www.frontiersin.org 8 June 2022 | Volume 10 | Article 769692 Venkatesh et al. Deep Learning for Lung Cancer Detection
- Georg HomannMD, Mustaf DF. Improved detection of bone metastases from lung cancer in the thoracic cage using 5- and 1-mm axial images versus a new CT software generating rib unfolding images. J Acad Radiol. (2015) 2015–12. doi: 10.1016/j.acra.2014.12.005
- De Nunzio G, Massafra A. Approaches to juxta-pleural nodule detection in CT images within the MAGIC-5 collaboration. J Nucl Instrum Methods Phys Res. (2011)8:103–6. doi: 10.1016/j.nima.2010.12.082
- Dhanamjayulu C, Nizhal UN, Maddikunta PKR, Gadekallu TR, Lwendi C, Wei C, et al. Identification of malnutrition and prediction of BMI from facial images using real-time image processing and machine learning. IET Image Process. (2021) 16:647–58. doi: 10.1049/ipr2.12222
- American Cancer Society. Key Statistics for Lung Cancer. Available online at: https://www.cancer.org/cancer/lung-cancer/about/key-statistics HTML (accessed January 5, 2022).
- Prokop M, Galanshi M. Spiral and Multislice Computed Tomography of the Body. Stuttgart: Thime Medical Publishers (2003)
- Abbas S, Jalil Z, Javed AR, Batool I, KhanMZ, Noorwali A, et al. BCD-WERT: a novel approach for breast cancer detection using whale optimization based efficient features and extremely randomized tree algorithm. Peer J Comput Sci. (2021) 7:e390. doi: 10.7717/peerj-cs.390
- Ananya C, Rajamenakshi, Subramanian R, Gaur S. A novel approach for tumor segmentation for lung cancer using multi-objective genetic algorithm and connected component analysis. In: Proceedings of the 2nd International Conference on Data Engineering and Communication Technology, Advances in Intelligent Systems and Computing Springer Nature. Singapore (2019). pp. 367–76.
- Venkatesh C, Polaiah B. A novel approach for lung lesion segmentation using optimization technique. Helix Sfic Explore. (2019) 94:4832– 7. doi: 10.29042/2019-4832-4837

- Preeti J, Bajaj SB, Aman J. “Segmentation and Detection of Lung Cancer Using Image Processing and Clustering Techniques” Springer Nature Singapore, Progress in Advanced Computing and Intelligent Engineering. Adv Intell Syst Comput. (2019) 1:3–23. doi: 10.1007/978-981-13-1708-8_2
- Senthil Kumar K, Venkatalakshmi K, Karthikeyan K. Lung cancer detection using image segmentation by means of various evolutionary algorithms. Hindawi Comput Math Methods Med. (2019) 2019:1–6. doi: 10.1155/2019/4909846
- Perumal S, Velmurugan T. Lung cancer detection and classification on CT scan images using enhanced artificial bee colony optimization. Int J Eng Technol. (2018) 7:74–9. doi: 10.14419/ijet.v7i2.26.12538.
- Ammar O, Ibrahim Al A, Abrahim B. Novel genetic algorithm for early prediction and detection of lung cancer. J Cancer Treat Res. (2017) 5:15–8. doi: 10.11648/j.jctr.20170502.13
- Kamil D, Ali H, Yoney KE. Lung lesion segmentation using Gaussian filter and discrete wavelet transform. ITM Web Conf. (2017) 11:1– 10. doi: 10.1051/itmconf/20171101018
- Mukesh CA Dr, Bhumika G. Detect mass tissue in lung images using discrete wavelet transformation. In: Proceedings of the IEEE International Conference on Information Processing (IICIP) (New Delhi), (2016):1–10.
- Santos AM, Ode A, Filho C, Silva AC, Nunes RA. Automatic detection of small lung nodules in 3D CT data using Gaussian mixture models, Tsallis entropy and SVM. Eng Appl Artif Intell. (2014)
- 36:27–39. doi: 10.1016/j.engappai.2014.07.007
- Jinsa K, Gunavathi K. Lung cancer classification using neural networks for CT images. Comput Methods Programs Biomed. (2014) 113:202–209. doi: 10.1016/j.cmpb.2013.10.011
- Anita P, Raja SK, Gandharba S. Digital image steganography using LSB substitution, PVD, and EMD. Hindawi Math Probl Eng. (2018) 2018:1–11. doi: 10.1155/2018/1804953
- Venkatesh C, Polaiah B. An investigation of diverse optimization techniques on medical imagery for detection of perilous diseases. Front J Soc Technol Environ Sci. (2017) 62:49—255. doi: 10.1108/IJPCC-10-2020-0160
- Thippa Reddy G, Khare N, Bhattacharya S, Singh S, Maddikunta PKR, In-Ho Ra, et al. Early detection of diabetic retinopathy using PCA-firefly based deep learning model. Electronics. (2020) 9:274. doi: 10.3390/electronics9020274
- Madhav BTP, Pardhasaradhi P,.Manepalli RKNR, Kishore PVV, Pisipat VGKM. Image enhancement using virtual contrast image fusion on Fe3O4 and ZnO nano-dispersed hydroxybenzoic acid. Liq Cryst. (2018) 42:1329– 36. doi: 10.1080/02678292.2015.1050704.