FOREX Publication
Open Access | Rapid and quality publishing

International Journal of
Electrical and Electronics Research (IJEER)
Research Article | Volume 11, Issue 2 | Pages 495-499 | e-ISSN: 2347-470X

A Diagnostic Study of Content-Based Image Retrieval Technique for Studying the CT Images of Lung Nodules and Prediction of Lung Cancer as a Biometric Tool

Rajeev Dixit[1*], Dr. Pankaj Kumar[2] and Dr. Shashank Ojha[3]

[1]*Department of Computer Science, United College of Engineering & Research, Prayagraj Dr. A.P.J. Abdul Kalam Technical University, Lucknow India,*
[2]*Department of Computer Science, Sri Ramswaroop Memorial Group of Professional Colleges Lucknow, University: Dr. A.P.J. Abdul Kalam Technical University, Uttar Pradesh, Lucknow, India*
[3]*Chest Department, Guru Shri Gorakshnath Chikitsalaya, Gorakhpur, India*

*__Correspondence:__ Rajeev Dixit; dixitrajeev19@gmail

ABSTRACT- Content Based Medical Image Retrieval (CBMIR) can be defined as a digital image search using the contents of the images. CBMIR plays a very important part in medical applications such as retrieving CT images and more accurately diagnosing aberrant lung tissues in CT images. The Content-Based Medical Image Retrieval (CBMIR) method might aid radiotherapists in examining a patient's CT image in order to retrieve comparable pulmonary nodes more precisely by utilizing query nodes. Intending a particular query node, the CBMIR system searches a large chest CT image database for comparable nodes. The prime aim of this research is to evaluate an end-to-end method for developing a CBIR system for lung cancer diagnosis.

Keywords: Medical image retrieval, Content-based image retrieval, Lung CT images.

ARTICLE INFORMATION
Author(s): Rajeev Dixit, Dr. Pankaj Kumar, Dr. Shashank Ojha;
Received: 15/03/2023; **Accepted**: 30/05/2023; **Published**: 30/06/2023;
e-ISSN: 2347-470X;
Paper Id: IJEER230215;
Citation: 10.37391/IJEER.110234
Webpage-link:
https://ijeer.forexjournal.co.in/archive/volume-11/ijeer-110234.html

Crossref member
CROSSREF.ORG
THE CITATION LINKING BACKBONE

Publisher's Note: FOREX Publication stays neutral with regard to Jurisdictional claims in Published maps and institutional affiliations.

1. INTRODUCTION

Lung cancer endures plaguing mankind as the prominent root of cancer-related fatalities. Timely recognition of pulmonary nodes, which can be an indicator of lung cancer, is the unique approach to recuperate survival rates. On a CT scan, a node manifests as a spherical mass with a diameter of up to 3cm. There are hundreds of slices (images) in CT scan for a patient, and manually processing such a large image set is time-consuming and inefficient. CAD systems are meant to help radiotherapists in the same way, but they have certain limitations because they functioning is somewhat similar to a black box. In other words, a highly effective CAD system might determine whether a patient is malignant or not based on the results of a CT scan. A CBIR system assists radiotherapists with diagnosis by allowing them to view a number of scans from the database of images that are optically comparable to the image of input query.

The CBIR system has two modes of operation and also has a high diagnostic potential. An image repository and a character base are available in offline mode. A vast number of images of mild and malicious nodes may be found inside the image library. Numerous characteristics are retrieved and saved in the character base using the character extraction procedure. The CBIR require an input query image of a mild or malignant node using an HCI in online mode. This image is subjected to a character extraction procedure, and the extracted characteristics are compared to those in the character database. One of the several distance measurements aids this comparison. Visually most comparable pictures are recovered and shown for the domain expert using an HCI based on the threshold value. This study offers a step-by-step approach for developing a CBIR system for lung cancer diagnosis. [1]

Radiotherapists utilize CT imaging as a base to detect lung illnesses since it may offer the essential info for exhibiting whether the tissues of lung are ordinary or irregular [2].

Instead of measuring whole node, the content-based image retrieval technique of pulmonary nodes assess an illustrative share of the node [3]. Furthermore, the previous analysis depends on radiotherapists to manually segment nodes. For clinical usage, the retrieval system with the least amount of user interaction is recommended. The applicability of a retrieval system is also determined by its performance. The current system's objective is to decrease user interaction while improving retrieval accuracy. The diagnostic efficiency and accuracy can be enhanced by radiotherapists are provided deep knowledge of history. The CBMIR approach is useful to solve this challenge. Considering the outcomes of the survey, a novel method for lung cancer detection and prediction will be proposed, which will aid in the prognosis of lung cancer and early recognition. As a result, the patient's chances of survival will improve.

2. CONTENT BASED MEDICAL IMAGE RETREIVAL SYSTEM (CBMIR)

Content-Based Image Retrieval are utmost commonly utilized method in Health Image Retrieval. Shape, Texture, Color, and additional information in an image are retrieved using the Content-Based Image Retrieval technique. CBMIR is a leading research field in supercomputer vision and image processing. The primary goal of the Content-Based Medical Image Retrieval System (CBMIR) technique is to extract images quickly. As a result, the suggested CBMIR approach not only combines visual and semantic info to overwhelm their limitations but also makes use of the database's basic structure for improved retrieval efficiency.

2.1 Computed Tomography (CT) Image

Although there are other ways of diagnosing lung cancer [4], such as PET, MRI, and X-ray, the CT scan image is regarded to be the distinctive method for providing a thorough description of lung nodes. It has less distortion and noise, as well as improved clarity, than additional images, making it easier to compute and predict the texture characteristics of the image.

2.2 Dataset

The dataset of lung images was obtained from the NIH/NCI dataset consortium of lung images, which offers computed tomography images of the lung on the network-based cancer [5] imaging archive location. It is open to the public and simple to use.

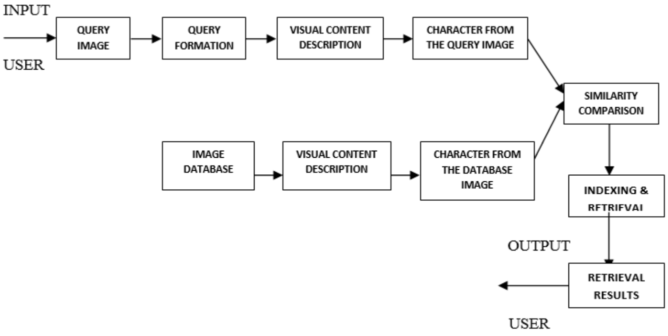

Figure 1: Content Based on Image Retrieval (CBMIR)

The worker provides the retrieval system with input in the form of an image, as illustrated in *figure 1*. Then, together the pictures in the database and the query image, image characteristics are extracted. The retrieval system uses similarity measurement to compute the displacement amid the images in the database and the query image.

3. NODE CLASSIFICATION STATEMENTS

There are primarily two research approaches for the organization of nodes in lung CT imaging. *(i)* 2-type cataloging of node, and *(ii)* four-type classification of node (built on appearance, shape, and location).

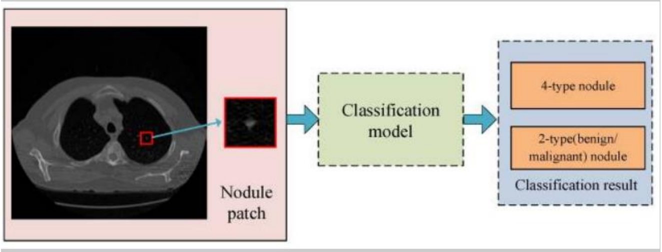

Figure 2: Statements of problem classification of lung node

3.1 Four-Type Node Arrangement

Lung nodes are tiny, spherical lumps that develop inside the lungs. Adjacent structural assemblies, for instance, arteries and pleura, might deform nodes. As a result, the look and surrounding regions of the lung node are generally used to describe it. The most commonly used method classifies nodes into four kinds, as defined by Reference. As shown in *figure 4*, Well-circumscribed (*W*) is positioned at center in the lung deprived of somewhat connections to supplementary muscles, Vascularized (*V*) is alike to W but connects to Vascular Structures, Juxta-Pleural (*J*) is completely associated to the Pleural exterior, Pleural-Tail (*P*) is near to Pleural but individual associated by a tinny tail.

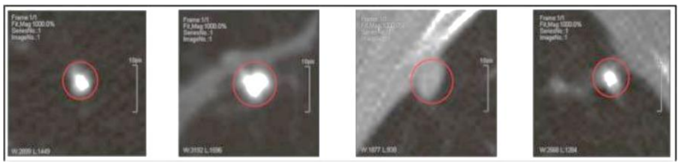

Figure 3: Example of 4 sorts of CT image of lung node, made known from left to right, W, V, J, and P, respectively [6]. Rings in red signify the positions of the node

This problem recognized in a number of articles. The type is the output, and the input is an image patch of lung node. Let *x* be an image patch of lung node by a size of *l*w*c*, as indicated in *equation (1)*. (*l*, *w*, and *c* signify length, width, and channel respectively). The node type is denoted by letter y. The sorting technique is f ().

$$y=f(x)\ (x\in R^{l*w*c},\ y\in \{W, V, J, P\}) \qquad (1)$$

3.2 Two-Type Node Classification

Many research has been conducted on the issue 2-type node categorization (mild/malicious ranking).

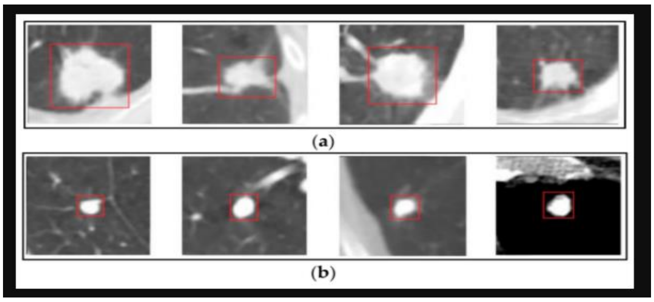

Figure 4: (a) More menace apprehensive cases, (b) suspicious cases with little malice are assumed. Red border boxes mean the positions of the node

It is too regarded a complication in the categorization of lung node malice shiftiness (possibility of node enmity). The formulation is identical to *equation (1)*, with *y* denoting the malice of the node (mild and enmity are signified by 0 and 1). *Figure 5* depicts node instances by great malice suspected and little malice.

4. MAIN DATASETS

For training and testing the ground truth dataset is required developed algorithms for lung node image categorization. It is difficult to get datasets as easily as in other disciplines due to rules and patient privacy. As a result, some academics rely on local hospitals or commercial databases for their data. Few datasets available publically, such as ANODE09, ELCAP, DLCST, LIDC, JSRT, LIDC-IDRI, NELSON and are currently broadly applied to assess algorithm presentation [7]. In this part, details about these datasets are discussed.

4.1 LIDC-IDRI and LIDC

The image database resource initiative and lung image database consortium (IDRI- LIDC) used thoracic CT images with marked-up annotated scratches for diagnosis and lung cancer screening. The national cancer institute (NCI) was considered as the driving force behind the creation of this database. The majority of the information for the pre-stage lung cancer investigation came from six academic institutes and eight medical imaging firms [8].

IDRI-LIDC has 1018 scans or a total of 1010 patients. There were three types of nodes in the lesion's region: *(i)* nodes >= 3.2 mm, *(ii)* nodes 3.2 mm, and *(iii)* non-nodes >= 3.2 mm.

All radiotherapist findings were compiled in an XML file to a given CT sequence. It included the elaboration of altogether nodes larger than 4.2mm (outlines and personal node evaluations), estimated 3d center of mass of nodes less than 4.2mm, and altogether non-nodes (no outline or personal valuations of features). The IDRI-LIDC had 7372 lesions labeled “node” by on slightest radiotherapist, with 2670 of them recognized as node >= 3mm by a minimum of one radiotherapist. *Figure 6* shows an example image with shown marks. *Table 1* shows the most commonly used node descriptions.

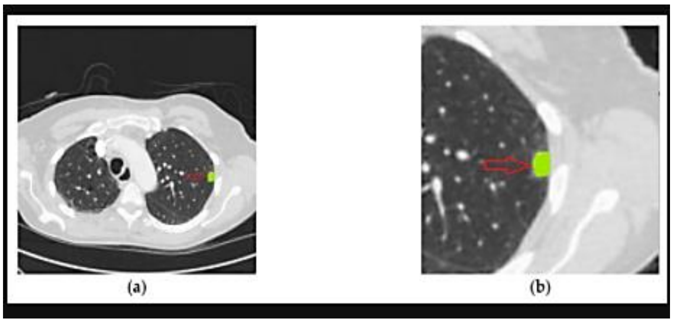

Figure 5: CT image of Lung: The green color shows the Inventive image with node, (b) the portion of CT image with node > 3.2 mm ROI (green color).

Table 1: Explanation of IDRI- LIDC

Event	Node Id	X loc.	Y loc.	Z loc.	Outline of the node	Malignancy
06	Node 001	195	291	36	((190,281),(188,281),...(191,282), (189,280)	5
06	IL057_159747	294	267	31	((290,281),(291,278),…(287,278), (288,281)	5

Each row represents a node's important information. The primary column contains a component of the affected node ID. The other column contains the node's unique identification. The 3rd and 4th columns show the node's x and y coordinates. The *Z* location specifies the portion position where the node is outlined. The 6th column denotes the assortment of (x, y) locations that form the node's border (individual a share of synchronizes are tabulated). The last column indicates the radiotherapists' assessment of the menace of this node. The first two rows give info about (position, contours, and evaluations) for nodes larger than 3mm, whereas the others provide info (location) for nodes less than 3mm.

4.2 Public Lung Image Database: ELCAP

In December 2003, the early lung cancer action program (ELACP) database was unrestricted for the primary time. This database is designed in association between the ELCAP and VIA research groups, and it is often used to evaluate the effectiveness of various CAD schemes [11].

There were 380 unrepeated lung node CT pictures and 50 reported reduced whole-lung CT scans in the database. The CT scans stayed acquired in a sole snort grasp with a slice thickness of 1.26 mm. The radiotherapist also gave the positions of the nodes found.

In ELCAP, the proportions of node types be located at W-16 percent, V-17 percent, J-29 percent, and P-40 percent [12]. ELCAP differed after IDRI- LIDC in 2 ways: Non-nodes were not included, and its node size was lowered [13]. The individual situation included an additional *.csv file providing the center location of each detected node, tabulated 2. A piece row represents a lung node. The 1st column contains the scan identification. The 3rd and 4th columns indicate the location of the lung node's center. The position of the node on this portion has been described in the last column. ELCAP lung CT scans are shown in *figure 6*.

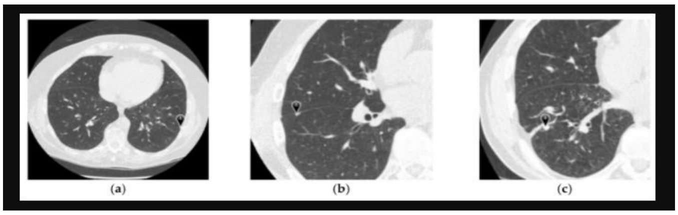

Figure 6: From ELCAP illustration of CT images of lung [9]: (a) the comprehensive CT pictures, (b, c) shows the slice of CT scans. The sign “1” is the position of node

Table 2: Arrangement of lung node location

Scan	Type	X	Y	Slice
W00001	Node	99	219	55
W00001	Node	55	225	171
W00003	Node	159	357	81

4.3 Others Datasets

Both Japanese Societies of radiological technology collaborated Japanese radiological society to create a database of the JSTR. There were 155 nodes and 94 non-nodes with labels in the database [14]. Age of patient elusiveness grade, gender, node size, and organization were some of the other pieces of data.

Since 2003, the Netherlands-Leuvens Longkanker Screenings Onderzoek (NELSON) has gathered information from 15,524 people in four institutions. Yearly CT screenings were understood first at a residential facility and subsequently at a fundamental place. They offered resources such as semi-automated measurement of lung node volume, mild vs. malicious node classification, automated lung node identification, and dissection [15].

The Automatic Node Detection 2009 (ANODE09) database contained 50 test scans and five example scans. [16]. ANODE09 served as a testing ground for node identification algorithms. Any team can submit their findings for assessment as a benchmark [17]. Researchers used databases gathered from hospitals that are operated privately or other organizations in several studies. These databases are less well-known, and around isn't much information about them. *Table 3* provides a quick overview of various databases.

Table 3: Arrangement of lung node spot

Database	Sample Number	Classification
Shanghai Zhongshan hospital database(ZSDB)	From 360 patients' CT scans	AIS, IA, MIA, AAH,
SPIE-AAPM lung CT challenge (23,24)	80 series of 22 500 CT scans	Vile and benign
Universal hospitals of Guangzhou Military command (GHGMC) dataset	130 malignant and 178 mild lung nodes	Vile and benign
NSCLC-Radiomics database (25,26)	89 patients' 13,481 CT scans	Vile and benign
Lung node Analysis challenge 2016(LUNA 16)[27]	CT scans: 888	Subset of LIDC-IDRI
Danish lung node screening Trial(DLCST)[28]	CT images from 4104 members	Node and non-node

MIA: Minimally Invasive Adenocarcinoma Hyperplasia, AIS: Adenocarcinoma In Situ, and IA: Invasive Adenocarcinoma.

IDRI-LIDC and ELCAP were the utmost widespread databases among those mentioned above. The IDRI-LIDC was mostly utilized for mild and malignant node cataloging, whereas the ELCAP was primarily used for four-type node cataloging.

5. AIM OF THE STUDY

The main purpose of this research is to examine an endwise method for developing a CBIR scheme for lung cancer diagnosis.

6. RELATED WORK

Despite the fact that the literature on generic CBIR comprises a large number of research publications, this section gives a quick overview of the work of various researchers to CBIR systems specialized to analytical radiology. Proposal of an input query image, character extraction, computation of a resemblance amount, and a Human Computer Interface (HCI) to offer imagining to end users are all part of a typical CBIR system. Among all the processing stages, character extraction is the most important. During the character extraction phase, several characteristics such as shade, figure, and feel are mined, and the literature [18] emphasizes how these characteristics can be extracted. Another key aspect in the development of the CBIR system is the computation of resemblance measures. Academics frequently use a variety of metrics to measure similarity, including Euclidean distance, Minkowski distance, Manhattan distance, Cosine similarity, and Jaccard similarity. [19] Provides a thorough analysis of these topics. When it comes to the implication of CBIR in analytic radioscopy in over-all, and CBIR for diagnosis of lung cancer in specific, it is clear the analytic and understanding correctness is steadily improving with the introduction of CBIR devices.

There is a plethora of literature on CBIR systems for lung cancer. For analogous CT image retrieval, [20] employed graphic and conceptual resemblances amid the database's individual images as well as the query image. They tested their method using CT scans of the lungs. [21] Suggested a CBIR method for detecting the malignancy levels of lung nodes. Node density level and node lesion density heterogeneousness are the characteristics they use. Their research used CT scans of the lungs from the LIDC data collection [22]. Utilized 27 Haralick texture characteristics to build GLCM in 4 different alignments. The Mahalanobis distance degree was cast off to assess the correlation amid the question image of a node and the dataset's position node. Employed margin sharpness characteristics to retrieve related pulmonary nodes. Each node in the CT segment has a margin sharpness vector of size 12 retrieved for it [22]. collected and kept in the character repository for CT images from LIDC are 2D and 3D shape-, margin-, and texture-based properties. The similarity is calculated using Euclidian, Manhattan, and Chebyshev measurements. Each node had an increased scale, calcification, sphericalness, boundary, lobulation, supposition, and degree of malice. An XML parser application is built, and 760 nodes are retrieved from CT images, 375 each of mild and malignant nodes. After that, the node images are harvested to 60*60 pixels and saved in the image base. These pictures are utilized in the character abstraction and post-processing stages [23]. Proposed a model which is created on convolutional neural network (CNN) architecture with ten layers to obtain high accuracy. The model demonstrated a considerable training and validation accuracy of 94% and 92% respectively. [24] Visual Analytics for Medical

Image Retrieval is a novel procedure for medicinal CBIR proposed in this research (VAMIR). B. [25] Optimized Fuzzy Intensification parameter constants are used to minimize overexposed and underexposed areas and offers elevated contrast improvement [26]. Medical images are pre-processed using hybrid median filter to discard noise and then decomposed using integer wavelet transform and employed modified grasshopper optimization algorithm to select the optimal coefficients for efficient compression and decompression.

7. CONCLUSION

The content-based medical image retrieval technique is a quick in addition easy way to discover and repossess CT images of lungs from a huge medical image folder. CBMIR is an approach for recovering common CT imaging signs of lung diseases (CISLs) that has lately become popular. CBMIR uses a range of methodologies to identify lung diseases early. This study presented a method for developing a CBIR system for retrieving and analyzing lung nodes.

REFERENCES

[1] Dasare, Ashwini & S., Harsha. CBIR for Lung Cancer Detection. (2020) 9. 565-569. 10.35940/ijitee.B1063.1292S19.

[2] D.R. Aberle et al., A consensus statement of the Society of Thoracic Radiology: screening for lung cancer with helical computed tomography, J. Thorac. Imaging 16 (1) (2001) 65–68.

[3] Lam MO, Disney T, Raicu DS, Furst J, Channin DS: BRISC—an open source pulmonary node image retrieval framework. JDigit Imaging (2007) 20(1):63–71.

[4] Aggarwal, P., Sardana, H.K., &Vig, R., March. An Efficient Visualization and Segmentation of Lung CT -Scan Images for Early Diagnosis of Cancer. In National Conference on Computational Instrumentation (NCCI) (2010).

[5] Yadav, N. G... "Detection of lung node using content based medical image retrieval". International Journal of electrical, electronics and data communication, ISSN (P), (2013) 2320-2084.

[6] Diciotti, S.; Picozzi, G.; Falchini, M.; Mascalchi, M.; Villari, N.; Valli, G. 3-D Segmentation Algorithm of Small Lung Nodes in Spiral CT Images. Inf. Technol. Biomed. 2008, 12, 7–19.

[7] ELCAP Public Lung Image Database. Available online: http://www.via.cornell.edu/databases/lungdb.html

[8] Armato, S.G.; McLennan, G.; Bidaut, L.; McNitt-Gray, M.F.; Meyer, C.R.; Reeves, A.P.; Clarke, L.P. Data From LIDC-IDRI. The Cancer Imaging Archive.

[9] Armato, S.G., III; McLennan, G.; Bidaut, L.; McNitt-Gray, M.F.; Meyer, C.R.; Reeves, A.P.; Zhao, B.; Aberle, D.R.; Henschke, C.I.; Hoffman, E.A.; et al. The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): A completed reference database of lung nodes on CT scans. Med. Phys. 2011, 38, 915–931.

[10] Henschke, C.I.; McCauley, D.I.; Yankelevitz, D.F.; Naidich, D.P.; McGuinness, G.; Miettinen, O.S.; Libby, D.M.; Pasmantier, M.W.; Koizumi, J.; Altorki, N.K.; et al. Early Lung Cancer Action Project: Overall design and findings from baseline screening. Lancet 1999, 354, 99–105.

[11] Zhang, F.; Cai, W.D.; Song, Y.; Lee, M.Z.; Shan, S.; Feng, D.D. Overlapping node discovery for improving classification of lung nodes. In Proceedings of the EMBC, Osaka, Japan, 3–7 July 2013; pp. 5461–5464.

[12] Liu, X.L.; Hou, F.; Hao, A. Multi-view multi-scale CNNs for lung node type classification from CT images. Pattern Recognit. 2018, 77, 262–275.

[13] Shiraishi, J. Development of a digital image database for chest radiographs with and without a lung node: Receiver operating characteristic analysis of radiotherapists' detection of pulmonary nodes. Am. J. Roentgenol. 2000, 174, 71–74.

[14] Zhao, Y.R.; Xie, X.; de Koning, H.J.; Mali, W.P.; Vliegenthart, R.; Oudkerk, M. NELSON lung cancer screening study. Cancer Imaging. 2011, 11, 79–84.

[15] Consortium for Open Medical Image Computing, Automatic Node Detection. Available online: http://anode09.grand-challenge.org/

[16] Ginneken, B.V.; Armato, S.G.; de Hoop, B.; Amelsvoort-van, V.S.; Duindam, T.; Niemeijer, M.; Murphy, K.; Schilham, A.; Retico, A.; Fantacci, M.E.; et al. Comparing and combining algorithms for computer-aided detection of pulmonary nodes in computed tomography scans: The ANODE09 study. Med. Image Anal. 2010, 14, 707–722.

[17] G. Schaefer, An introduction to content-based image retrieval, 8th International Conference on Digital Information Management (ICDIM 2013), Islamabad, (2013) pp. 4-6.

[18] Cha, Sung-Hyuk, Comprehensive survey on distance/similarity measures between probability density functions, City 1, no. 2, 2007.

[19] Wei, Guohui, et al., A content-based image retrieval scheme for identifying lung node malignancy levels, Control and Decision Conference (CCDC), 29th Chinese, IEEE, 2017.

[20] Wei, Guohui, He Ma, Wei Qian, Hongyang Jiang, and Xinzhuo Zhao. Content-based retrieval for lung node diagnosis using learned distance metric. In 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2017, pp. 3910-3913. IEEE.

[21] Junior, José Raniery Ferreira, and Marcelo Costa Oliveira. Evaluating margin sharpness analysis on similar pulmonary node retrieval. In 2015 IEEE 28th International Symposium on Computer-Based Medical Systems, 2015, pp. 60-65. IEEE.

[22] Dhara, Ashis Kumar, Sudipta Mukhopadhyay, Anirvan Dutta, Mandeep Garg, and Niranjan Khandelwal. Content-based image retrieval system for pulmonary nodes: Assisting radiotherapists in self-learning and diagnosis of lung cancer. Journal of digital imaging 30, no. 1 (2017): 63-77.]

[23] Aluka, M., Dixit, R. and Kumar, P. 2023. Enhancing and Detecting the Lung Cancer using Deep Learning. International Journal on Recent and Innovation Trends in Computing and Communication. 11, 3s (Mar. 2023), 127–134.

[24] S.Abinaya and T.Rajasenbagam (2022), Enhanced Visual Analytics Technique for Content-Based Medical Image Retrieval. IJEER 10(2), 93-99. DOI: 10.37391/IJEER.100207.

[25] Avadhesh Kumar Dixit, Rakesh Kumar Yadav and Ramapati Mishra (2021), Contrast Enhancement of Colour Images by Optimized Fuzzy Intensification. IJEER 9(4), 143-149. DOI: 10.37391/IJEER.090408.

[26] N. Shyamala and Dr.S. Geetha (2022), Compression of Medical Images Using Wavelet Transform and Metaheuristic Algorithm for Telemedicine Applications. IJEER 10(2), 161-166. DOI: 10.37391/IJEER.100219