

An exhaustive analysis of Feature extraction techniques for identifying diseased lungs from CT Images

¹Dr. M. MUTHURAMAN

¹Assistant Professor, P.G. and Research Department of Computer Science, H.H. The Rajah's College (A), Pudukkottai, muthuraman70@gmail.com

Abstract

LDs (Lung Diseases) when considered cumulatively, are a major cause of morbidity and mortality. Many times, CTSIs (Computed Tomography Scan Images) are obtained by doctors for evaluation of LDs and condition of patients including pneumonia, COVID-19, cancer, blood clots or other damages caused in the lungs. CTSIs of internal organs, bones, soft tissue, and blood vessels detail about these parts to clinicians and specifically their details on soft tissues and blood vessels are of great use. Hence, assessments of LDs can be done using CTSIs. These images can also be processed using IPTs (Image Processing Techniques) which are non-invasive ways of examinations. The most important part of IPTs in CTSIs are FEs (Feature Extractions) which are central to diagnosis or classifications or detections of LDs. FEs in the case of LDs from CTSIs narrows down to identification of diseased areas precisely where multitude of techniques are used. This paper presents a thorough analysis of the existing techniques for FEs with comparative performance charts.

Keywords: LDs (Lung Diseases), CTSIs (Computed Tomography Scan Images), COVID-19.

I. INTRODUCTION

LDs have high causality rates where more than 1.3 million people die LDs throughout the world. The projections by American Cancer Society on 1.74 million deaths by 2018 [1] was also surpassed. Two main reasons contribute towards high mortality rates due to LDs: Delayed diagnosis and the poor prognosis [2] as more than 70% When LDs are detected at an advanced stage, prognosis is no longer helpful. As a result, early detection of LDs is critical for improving patient's chances of survival. CTSIs provide a greater number of pictures, including 3D images that can be resized in several planes. They depict interior organs, bones, soft tissue, and blood arteries which can be examined by physicians on computers. Further, quantity of radiations in CTSIs can be considerably lowered using strategies including modifications in radiation

dosages. Low dosed chest CTSs generate pictures of adequate quality where lung illnesses and abnormalities are identified. They emit far less radiations than traditional CTSs scans. These CTSIs are commonly used to assess congenital lung abnormalities including pneumonia, interstitial LDs, and tumours. However, these procedures need highly trained radiologists, an issue for distant and impoverished areas. Furthermore, manual examinations are prone to human mistake, necessitating the use of CADs to aid [2] radiologists in diagnosis and lower bogus reports. CTSIs can identify anomalies, their kind, size, and other characteristics using DIPTs (Digital image processing methods). Medical image processing is being used more and more to create expert assistance systems for the diagnosis of a variety of disorders, including LDs[3]. A lot of work is being done in the field of early LD diagnosis utilizing

CAD systems [4]. The nature of the data produced in CTSIs necessitates the usage of automated systems. Lung CTSIs often yield more than 250 pictures in a single scan, making it difficult, time-consuming, and tiresome to examine these large datasets. Furthermore, the type of anomalies that determine a patient's fate is confounded by the fact that their form and size vary from slice to slice. They are sometimes connected to other lung structures like arteries or bronchioles. It's also possible that the hue that appear in CT scans are different. These variables contribute to the difficulty of locating abnormalities. However, on discovery, the same characteristics aid in determining approach's directions. Accurate segmentation of diseased sections from CTSIs is a crucial step in detection of LDs where several strategies have been presented. Some methods need seed pixels in lungs region of pictures before using region growing approaches to segment the lungs [5]. Lungs have been segmented using region-growing strategies in studies using CTSIs [6] [7]. The center pixel of slices was as used as seeds. Method have also segmented the abnormalities with the assistance of the radiologists [8]. CADs have also been used to divide CTSIs into four classes namely lung wall, parenchyma, bronchioles, and abnormalities [9] where active contours segmented lungs from CTSIs images.



Fig. 1 – *Diseased and Normal Human Lungs*

Several strategies for detecting anomalies in CTSIs in the literature involve intensity or colour thresholding. CTSIs are often pre-processed using IMPTs to increase sharpness, and then the picture is thresholded to extract

related components. These segmented areas' area and pixel values are passed as features for further analysis of LDs anomalies. Rules [10], sizes [11], thresholding [12] [13], template matching [14], 3D templates [15], binarization [16], multi-scale filtering [17], shape features for each segmented region [18] have all used for detecting LDs in CTSIs. DLTs (Deep learning techniques) have also been explored lately for lung nodule detections [19]. Thus, the major contribution of this paper is in discussing challenges and proposals for automated extraction of features from CTSIs which can be inferred by other researches for preparing a clean line of action related to them.

Following this introductory section, a detailed review of literature of feature extraction techniques is section two. Section three discusses issues and challenges while section four concludes this paper.

II. LITERATURE REVIEW

FEs are the base for diagnostics, classifications, clustering and detections. IMPTs, irrespective of image types namely binary, colored or gray scale, can extract features. Study of FEs evokes interest in researches for choosing suitable features in applications. This section provides details of studies in references to FE proposals. Figure 2 depicts a generic classification using FEs for input CTSIs.

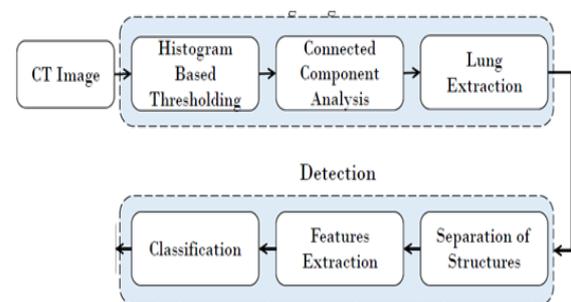


Fig. 2 – *FE based Classifications*

FEs can be extracted from images using their Geometrical, Statistical, Textural and Color features and are utilized to obtain as much information as possible where selection and effectiveness of features chosen and extracted is a major challenge [20]. These base

characteristics can be divided further into sub-types as depicted in Figure 3.

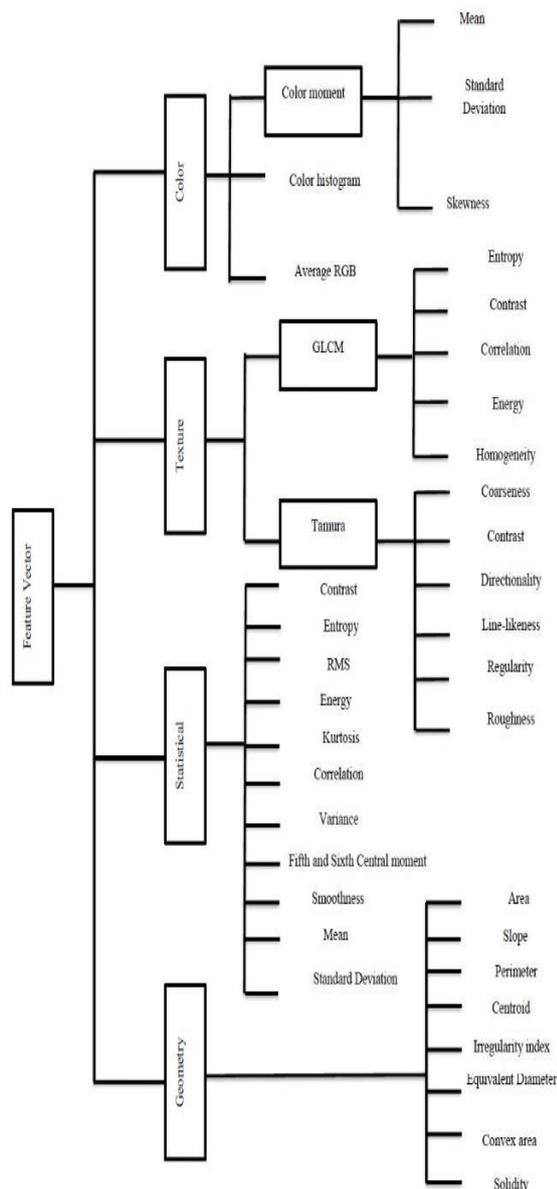


Fig. 3 - Image qualities used in FEs

- **Color Features:** Moments, histograms, and average RGBs [21] are colour attributes and are detailed below:

- o **Color moments:** They are colour scales that distinguish pictures based on colour properties and are probability distributions of means, SDs, and Skews. The average colour values of pictures are their means. SDs (Standard Deviations) are the square root of distribution variances, whereas skew values are

a measure of the degree of asymmetries in distributions.

- o **Color histograms:** Colors are commonly utilised features due to their intuitive natures in comparisons to other characteristics and have more significant information and simplicity in extractions from images. Histogram distributes these colours using a set of boxes.

- o **Average RGBs:** The purpose of employing this feature is to do picture filtering when using multiple features, and another reason for doing so is because representing features as vectors requires just a tiny amount of data [22].

- **Statistical Features:** Approaches do not attempt to comprehend texture's hierarchical structures and use non-deterministic features for determining distributions/connections between grey levels to express textures indirectly. Contrast, Entropy, RMS, Energy, Kurtosis, Correlation, Variance, Fifth and sixth central moments, and Smoothness are all statistical texture properties of pictures [27]. The statistical contrast between the reference pixel and its neighbour is a measure of intensity or gray-level fluctuations. The brightness of the item colour, as well as the brightness of other objects in the same display region, determines contrast. There is a little contrast when two pixels differ by one, and the weight is one. When I and j differ by two, the contrast increases and the weight increases to four. In thermodynamics, entropy is a measurement of system disruption. Entropy measurement is an excellent approach to determine the quantity of information contained in an event as well as the level of unstable signal disruption. With the progression of mistake, RMSEs (Root Mean Square Errors) steadily grow in value. They do not, however, give information on any specific incipient fault stage, although the value continuously rises as the fault progresses. Energy in statistics can be describe as a measurement of information while assessing probabilities (maximum a priori) coupled with Markov Random values where positive measures maximize while negative measures minimize. Kurtosis can assess stability of

distributions, which relates to normal distributions. Correlations are basic processes used to extract the information from the images while Variance defines the mean of a signal's square that can be computed. Fifth and sixth central moments are used to compute deviations from averages. Smoothness is used to assess comparative smoothness from gray level disparities and used to create relative smoothness recipes.

- **Geometry Features:** These features refer to shapes and sizes of shapes. There are eight types of geometry features as following :

- o **Area:** It Is the difference between perimeters and extensions of forms. There are several known formulae for basic shapes like triangles, rectangles, and circles. Any polygon area may be determined using these formulae and by splitting the polygon into triangles or circles to generate curved forms with boundaries, which can then be gathered once the areas have been calculated, and when polygons are irregular, the areas can be calculated using Trapezoidal Gauss equation [28]

- o **Slope:** A straight line is a collection of points with a constant slope between any two locations. The value of the ratio of vertical change to horizontal variation is commonly used to estimate the slope of a straight line. The slope generally refers to the two-point line's slope. The horizontal line, Zero, is defined as the parallel line of the x-axis. The vertical line is a parallel line on the y-axis with an unknown slope. The slope of two parallel lines is always equal [29].

- o **Perimeter:** It refers to the length of the line that encircles two-dimensional forms like circles, squares, rectangles, and irregular shapes. Equilateral and non-equilateral perimeters can both be determined [30].

- o **Centroid:** The centroid is a fixed location in the object through which the lines that indicate the object's weight travel. The centroid differs from one another in terms of shape or acclimatisation, and this difference determines the status of a centroid [31].

- o **Irregularity Index:** The irregularity index (L) is equal to one only for circles and is less than one for all other forms when calculating the boundaries of irregular shapes [32].

- o **Convex Area:** In the Euclidean level, the closed convex or convex is the set X of points in which the smallest convex set includes X. When you're X, for example, you're only a little part of the plane. The number of pixels in the convex picture is the convex area. The dimensions of the square that surrounds the region. A convex hull[34] is used as the bounding box.

- o **Solidity:** Determines the pixel ratio of a convex hull in the area[35].

- **Texture Features:** The most essential aspect of medical/sensor pictures is texture, which is described as a surface expression of human visual systems of natural things. These characteristics are simple to perceive but difficult to quantify in terms of mathematical matrices, generally used in quantitative/qualitative analysis. They can be categorized into GLCMs (Gray Level Co-occurrence Matrices) and Tamura [23].

- o **GLCMs:** Histograms can recover textures of damaged tissues in photos by measuring grey values that occur at each particular image offset [24]. Entropy, Contrast, Correlation, Energy, and Homogeneity are GLCM-specific texture characteristics [25]. The texture of input pictures is distinguished using entropy, a statistical measure of unpredictability. In pictures, contrast is used to calculate densities between pixels and neighbouring pixels. Scales that assess the likelihood of certain pixel pairings are known as correlations. In GLCMs, energy is defined as the sum of squared components, also known as angular second moments or uniformities. The measure of distribution approximations in the elements of GLCMs is called homogeneity.

- o **Tamura:** These qualities give descriptions for quantitative analysis, and Tamura's attributes include Contrast, Directionality, Coarseness, Roughness, Line-Likeness, and Regularity [26], which are

standard descriptions for all sorts of picture textures. Coarseness is a term used to describe how coarse something is. Essentially tied to the size of primal elements that compose textures and distances between grey levels in changes. Scales, duplications, and principal image textures can all be directly linked to it. Even in the event of a smaller tissue, coarseness seeks to discover the maximum size in which the tissue is present, and differences between pairs of averages belonging to non-overlapping regions are computed. Tamura's Contrast is a measurement of grey level distribution that changes depending on whether it's black or white. In its computations, the central moments of the fourth and second orders of grey levels are utilised. The frequency of local edges directed against directional angles in a distribution is referred to as directionality. Within an area, it is a global value. It can estimate overall degrees of picture directivities by discriminating between an image's consistencies within areas, albeit it can't discern between trends or patterns. Line-Likeness is a term used to describe the forms of texture primitives. Straight or wave-like primitives in a line-like texture can't be fixed in their orientation. Frequently, the line-like texture is also directed. In pictures, regularity is defined as a consistent pattern or equivalent components. The total of coarseness and contrast measurements is roughness.

FEs are also important to AIS (artificial intelligence system) and MLTs (Machine learning techniques) as techniques can extract relevant image features and label them for classifications which are based on labels. Moreover, performance accuracies of classification models depend on FEs. Figure 4 depicts initial segmentation of CTSIs.

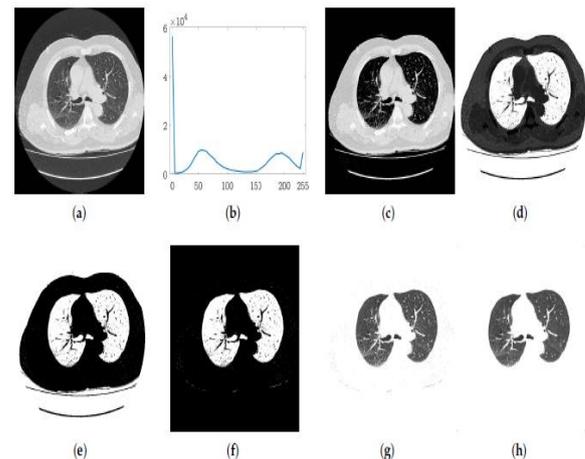


Fig. 4 –initial segmentation of CTSIs: (a) Input Image; (b) Input image's histogram; (c) thresholded image; (d) complemented image; (e) binary segmentation map; (f) binary map after connecting components (g) Identification of lungs; (h) noise removed lungs

This demonstrative research looked into data from 160 CTSIs of patients with sick lungs from 3.2 Iowa University's Hospital. Patients with (a) lung abnormalities (4–29 mm) and (b) malignant abnormalities confirmed on histology or benign abnormalities verified on histopathology or by size stability for at least 24 months were included in the study. Table 1 contains demographic statistics..

Table 1 – Dataset's Demographic features

	Malignant	Benign
Number of patients	100	60
Female	55 (55%)	36 (60.0%)
Male	45 (45%)	24 (40.0%)
Age, yrs (mean \pm SD)	67.5 \pm 12.1	52.8 \pm 12.9
Pack-years (mean \pm SD)	34.8 \pm 32.1	12.2 \pm 15.8

To establish ROIs (regions of interest) surrounding each anomaly, a graduate student skilled in medical image processing did manual segmentations. With the obtained feature information, each ROIs were designated abnormality or normal. The intensity, shape, and texture of ROIs were assessed using these biomarkers. This was a secondary examination of de-identified data that had been obtained with the University's permission. The performances based on FEs used are listed as Table 2.

Table 2 – *FEs and their comparative performances*

Database Contents	Samples Count	Features	Accuracy Percentage	
			Training	Testing
Lung CTSIs	Training:100 CTSIs	Geometry	99	96
		Statistic	98.5	95
	Testing :60 CTSIs	Color	97	94
		Texture	96.5	93
		Geometry	99	94

III. Discussion

Medical imaging is often evaluated visually or subjectively, leaving a significant amount of latent information in the pictures unused. Extraction of quantitative data from clinical images is one technique to gain access to this concealed information. Visual evaluation, in particular, can not appropriately detect variability in imaging data. Tumor aggressiveness and poor patient outcomes are linked to intratumoral heterogeneity. In studies evaluating the diagnosis, prognosis, and treatment response of LDs, metrics, notably texture analysis metrics, have been used to quantify intra-tumor heterogeneity. A typical

radiomic evaluation involves a texture analysis, shape, and size [37] [38], [39]. The technique's basic concept is that the gray-scale values that create the picture of the anomalies, as well as their spatial and temporal interrelationships, represent the tumor's phenotypic changes, which are suggestive of genetic and other molecular alterations [40]. Although there is a lot of interest in employing features for noninvasive LD evaluation, the lack of consistency and generalization of results makes clinical procedures difficult to translate [41] [42] [43] [44]. Texture is described in material science as a measure of a surface's variation; a rough-textured material would have a high rate of change in the high and low points of a surface compared with smooth textures [45]. Image texture in radiology refers to changes in the grayscales that constitute a ROI. When compared to a smooth-textured material, the picture of a rough-textured material would show a high rate of change in the high and low points of a surface (the gray-scale value). A typical radiomics workflow consists of four modules: picture capture, image segmentation, feature extraction, and statistical analysis at the most basic level. The initial step in the radiomics procedure is image acquisition. Figure 5 shows a diagram of an image-based process for clinical evaluations.

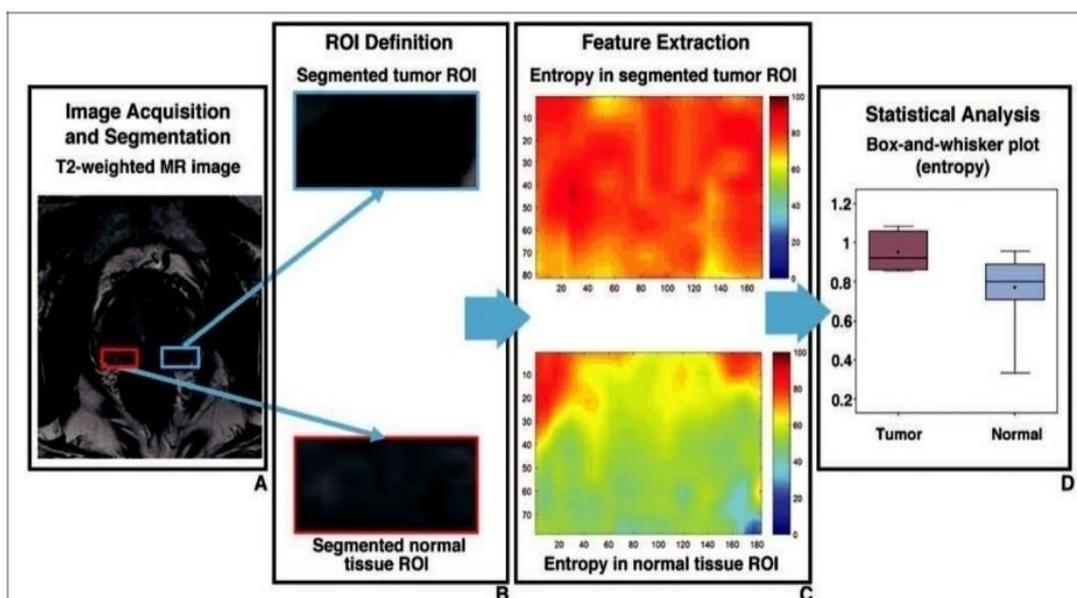


Fig. 5—Schematic display of image based Clinical assessments

Clinical imaging technologies now accessible allow for a wide range of acquisition and reconstruction techniques. This is not a barrier in terms of visual or qualitative imaging evaluation. When pictures are quantitatively evaluated to extract useful data, however, differences in acquisition and image reconstruction parameters lead to inconsistent results across datasets, especially in multicenter investigations [46]. Image Segmentation follows, which entails determining a ROIs, which can be done automatically, semi-automatically, or manually. Manual segmentations are accurate, but time-consuming and subjective [47]. Automatic segmentations are objective, but prone to errors, especially when there are image distortions and noises. Active contours [48], level sets [49], and regions/graphs based approaches [50] are some of the most extensively used automated segmentation algorithms.

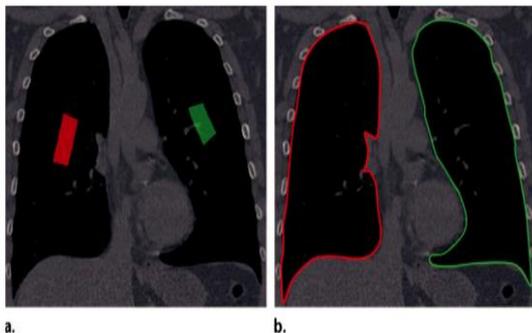


Fig. 6 - Example of object recognitions

(a) object demarcation (b) for the left lung (green) and right lung (red) on a CTSI

There is presently no approved segmentation standard. As seen in Figure 7, poor boundary identification might lead to erroneous findings.

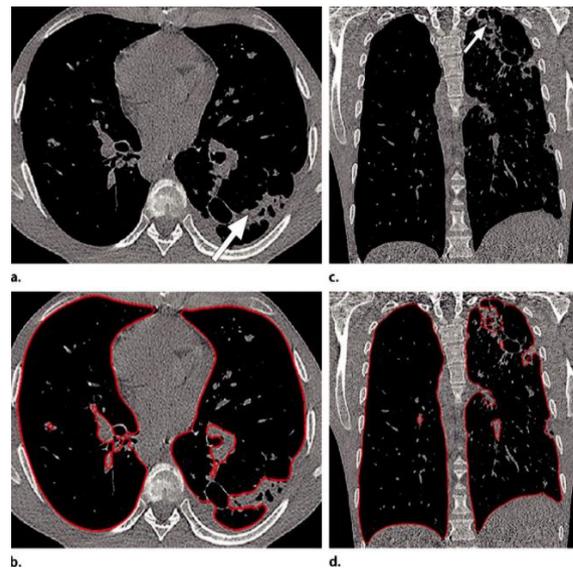


Fig. 7. Inaccurate boundary identifications

Axial (a, b) and coronal (c, d) CT images show that cavities and consolidation (arrow in a, c) can lead to inaccurate segmentation (red contours in b, d).

More recently, MLTs can also be used for FEs. They also select suitable features (Dimensionality Reduction) when used for FEs. These algorithms can be categorized as filters, wrappers and embedded methods. Filters function without taking the classifier into account. As a result, they are extremely computationally efficient. Multivariate and univariate approaches are the two types. Multivariate approaches can uncover correlations between features, whereas univariate methods look at each feature individually. Wrappers are better at picking features because they train and test in the feature space, taking into consideration the model hypothesis. As a result, wrappers have a significant disadvantage: computational inefficiency, which becomes increasingly obvious as the feature space develops. They can discover feature dependencies, unlike filters. Wrappers are divided into two types: random and deterministic. Embedded Techniques are faster than wrappers in terms of computing, but they make classifier-specific choices that may not work with any other classifier. This is because the best collection of features is produced when the classifier is built, and the selection is influenced by the

classifier's assumptions. Random forests are a well-known embedding approach. Ensembles are a collection of classifiers. They create decisions iteratively by discarding small fractions of features with lowest importance. Table 3 lists comparative performances of MLTs where ensembles perform the maximum in terms of classification accuracy with extracted feature subsets..

Table 3 - comparative performances of MLTs

Classifier	Sensitivity	Specificity	Precision	Accuracy
KNN	0.784245	0.98994	0.97482	0.887355
DT	0.900165	0.9282	0.92421	0.914235
SVM	0.857955	0.970935	0.960855	0.91455
RF	0.952035	0.9534	0.952665	0.952665

IV. Conclusion

Detecting LDs in early stages highly increases the treatment probability. AIs and MLTs are used to help detect these abnormalities and based on extracted features. This paper provides a critical review of existing lung segmentation methodologies on CTSIs to help doctors make better judgments when choosing tools for lung field segmentations. The advantages and disadvantages of various FEs were highlighted in this study. The factors that affect speed gains based on picture texture metrics are discussed in this article. The major purpose is to make recommendations that may be used to assist plan and assess future investigations. The study also acquired 160 images and proposed FEs were evaluated for their extraction capabilities. Classifiers were also evaluated on the extracted feature sets. It can be concluded that increasing the accuracy can be achieved by using ensembles and generated FEs can be based more geometrical features.

Reference

S[1] Siegel, R.L.; Miller, K.D.; Jemal, A. Cancer statistics, 2018. *CA-Cancer J. Clin.* 2018, 68, 7–30.
 [2] Capocaccia, R.; Gatta, G.; Dal Maso, L. Life expectancy of colon, breast, and

testicular cancer patients: An analysis of US-SEER population-based data. *Ann. Oncol.* 2015, 26, 1263–1268.
 [3] Nasrullah, N.; Sang, J.; Alam, M.S.; Mateen, M.; Cai, B.; Hu, H. Automated Lung Nodule Detection and Classification Using Deep Learning Combined with Multiple Strategies. *Sensors* 2019, 19, 3722.
 [4] Rajan, J.R.; Chelvan, A.C.; Duela, J.S. Multi-Class Neural Networks to Predict Lung Cancer. *J. Med. Syst.* 2019, 43, 211.
 [5] Akram, S.; Javed, M.Y.; Akram, M.U.; Qamar, U.; Hassan, A. Pulmonary Nodules Detection and Classification Using Hybrid Features from Computerized Tomographic Images. *J. Med. Imaging Health Inform.* 2016, 6, 252–259.
 [6] Netto, S.M.B.; Silva, A.C.; Nunes, R.A.; Gattass, M. Automatic segmentation of lung abnormalities with growing neural gas and support vector machine. *Comput. Biol. Med.* 2012, 42, 1110–1121.
 [7] Cascio, D.; Magro, R.; Fauci, F.; Iacomì, M.; Raso, G. Automatic detection of lung abnormalities in CT datasets based on stable 3D mass-spring models. *Comput. Biol. Med.* 2012, 42, 1098–1109.
 [8] Lee, S.L.A.; Kouzani, A.Z.; Hu, E.J. Automated identification of lung z. In *Proceedings of the 10th Workshop on Multimedia Signal Processing (MMSP)*, Cairns, Australia, 8–10 October 2008; pp. 497–502.
 [9] Keshani, M.; Azimifar, Z.; Tajeripour, F.; Boostani, R. Lung nodule segmentation and recognition using SVM classifier and active contour modeling: A complete intelligent system. *Comput. Biol. Med.* 2013, 43, 287–300.
 [10] Özokes, S. Rule-Based Lung Region Segmentation and Nodule Detection via Genetic Algorithm Trained Template Matching. *Istanbul Ticaret Üniversitesi Fen Bilim. Derg.* 2007, 6, 17–30.
 [11] Pu, J.; Zheng, B.; Leader, J.; Wang, X.; Gur, D. An automated CT based lung nodule detection scheme using geometric analysis of signed distance field. *Med. Phys.* 2008, 35, 3453–3461.
 [12] Messay, T.; Hardie, R.C.; Rogers, S.K. A new computationally efficient CAD system for pulmonary nodule detection in CT imagery. *Med. Image Anal.* 2010, 14, 390–406.

- [13] Tan, M.; Deklerck, R.; Jansen, B.; Bister, M.; Cornelis, J. A novel computer-aided lung nodule detection system for CT images. *Med. Phys.* 2011, 38, 5630–5645. *J. Imaging* 2020, 6, 6
- [14] Jo, H.H.; Hong, H.; Goo, J.M. Pulmonary nodule registration in serial CT scans using global rib matching and nodule template matching. *Comput. Biol. Med.* 2014, 45, 87–97.
- [15] Ozekes, S.; Osman, O.; Ucan, O.N. Nodule detection in a lung region that's segmented with using genetic cellular neural networks and 3D template matching with fuzzy rule based thresholding. *Korean J. Radiol.* 2008, 9, 1–9.
- [16] Dolejsi, M.; Kybic, J.; Polovincak, M.; Tuma, S. The Lung TIME: Annotated lung nodule dataset and nodule detection framework. In *Proceedings of the Medical Imaging 2009: Computer-Aided Diagnosis*, Lake Buena Vista (Orlando Area), FL, USA, 7–12 February 2009; p. 72601U.
- [17] Narayanan, B.N.; Hardie, R.C.; Kebede, T.M.; Sprague, M.J. Optimized Feature Selection-Based Clustering Approach for Computer-Aided Detection of Lung Nodules in Different Modalities. *Pattern Anal. Appl.* 2019, 22, 559–571.
- [18] da Silva Sousa, J.R.F.; Silva, A.C.; de Paiva, A.C.; Nunes, R.A. Methodology for automatic detection of lung nodules in computerized tomography images. *Comput. Methods Programs Biomed.* 2010, 98, 1–14.
- [19] Negahdar, M.; Beymer, D.; Syeda-Mahmood, T. Automated volumetric lung segmentation of thoracic CT images using fully convolutional neural network. In *Proceedings of the Medical Imaging 2018: Computer-Aided Diagnosis*, Houston, TX, USA, 10–15 February 2018; Petrick, N., Mori, K., Eds.; Volume 10575, pp. 356–361.
- [20] Ethem Alpaydin 2014 *Introduction to Machine Learning* (MIT Press).
- [21] S.R. Kodituwakku and S.Selvarajah 2011 Comparison of Color Features for Image Retrieval (*Indian Journal of Computer Science and Engineering* Vol 1, No 3) pp 207-211.
- [22] Amara H.M alzoubi 2015 Comparative Analysis of Image Search Algorithm using Average RGB, Local Color Histogram and Global Color Histogram and Color Moment HSV(thesis, Faculty of Computer Science and Information Technology Universiti Tun Hussein Onn Malaysia).
- [23] C. Umamaheswari, Dr. R. Bhavani and Dr. K. Thirunadana Sikamani 2018 Texture and Color Feature Extraction from Ceramic Tiles for Various Flaws Detection Classification (*International Journal on Future Revolution in Computer Science & Communication Engineering* Vol 4 ,Issue. 1)pp 169 – 179.
- [24] M. S. Ahmad, M. S. Naweed, and M. Nisa 2009 Application of texture analysis in the assessment of chest radiograph (*International Journal of Video & Image Processing and Network Security (IJVIPNS)* Vol 9, No 9) pp 291-297.
- [25] Fritz Albrechtsen 2008 Statistical Texture Measures Computed from Gray Level Co-occurrence Matrices (*Image Processing Laboratory Department of Informatics University of Oslo*).
- [26] Peter Howarth and Stefan Ruger 2004 Evaluation of Texture Features for Content-Based Image Retrieval (Department of Computing, Imperial College London South Kensington Campus London)
- [27] Jaspinder Kaur, Nidhi Garg and Daljeet Kaur 2014 Segmentation and Feature Extraction of LungRegion for the Early Detection of Lung Tumor(*International Journal of Science and Research(IJSR)* Vol 3, Issue 6).
- [28] A. H. Stroud 1971 *Approximate Calculation of Multiple Integrals*(Prentice-Hall Inc., Englewood Cliffs, N. J.
- [29] Christopher Clapham, James Nicholson 2009 *Oxford Concise Dictionary of Mathematics*(OUP oxford).
- [30] Dr Yeap Ban Har, Dr Joseph Yeo, Teh Keng Seng, Loh Cheng Yee, Ivy Chow, Neo Chai Meng and Jacinth Liew 2018 *New Syllabus Mathematics Teacher's Resource Book 1*(OXFORD UNIVERSITY Press) 7th edition.
- [31] Dan B. Marghitu and Mihai Dupac 2012 *Advanced Dynamics* (Springer) chapter 2 pp 73-141.
- [32] S. A. Patil and V. R. Udipi 2010 Chest x-ray features extraction for lung cancer classification(*Journal of Scientific and Industrial Research* Vol 69) pp 271-277.

- [33] R. C. Gozalez and R. E. Woods 2002 Digital Image Processing Using Matlab, 2nd ed, Gatesmark (USA) chapter 12 pp 642-654.
- [34] Nitin S. Lingayat and Manoj R. Tarambale 2013 A Computer Based Feature Extraction of Lung Nodule in Chest X-Ray Image(International Journal of Bioscience, Biochemistry and Bioinformatics Vol 3, No. 6) .
- [35] K. P. Aarthy and U. S. Ragupathy 2012 Detection of lung nodule using multiscale wavelets and support vector machine(International Journal of Soft Computing and Engineering (IJSCE) Vol 2, Issue 3).
- [36] The National Lung Screening Trial Research Team. Reduced lung-cancer mortality rate with low-dose computed tomographic screening. *N Engl J Med.* (2011) 365:395–409. doi: 10.1056/NEJMoa1102873
- [37] Gillies R, Kinahan P, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology.* (2016) 278:563–77. doi: 10.1148/radiol.2015151169
- [38] Zhang B, He X, Ouyang F, Gu D, Dong Y, Zhang L, et al. Radiomic machine-learning classifiers for prognostic biomarkers of advanced nasopharyngeal carcinoma. *Cancer Lett.* (2017) 403:21–7. doi: 10.1016/j.canlet.2017.06.004
- [39] Kumar V, Gu Y, Basu S. Radiomics: the process and the challenges . *MagnReson Imaging.* (2012) 30:1234–48. doi: 10.1016/j.mri.2012.06.010
- [40] Ortiz-Ramón R, Larroza A, Ruiz-España S, Arana E, Moratal D. Classifying brain metastases by their primary site of origin using a radiomics approach based on texture analysis: a feasibility study. *Eur Radiol.* (2018) 28:4514–23. doi: 10.1007/s00330-018-5463-6
- [41] Kuruvilla J, Gunavathi K. Lung cancer classification using neural networks for CT images. *Comput Methods Prog Biomed.* (2014) 113:202–9. doi: 10.1016/j.cmpb.2013.10.011
- [42] Lambin P, Rios-Velazquez E, Leijenaar Rea. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer.* (2012) 48:441–6. doi: 10.1016/j.ejca.2011.11.036
- [43] Lin Y, Leng Q, Jiang Z, Guarnera MA, Zhou Y, Chen X, et al. A classifier integrating plasma biomarkers and radiological characteristics for distinguishing malignant from benign pulmonary nodules. *Int J Cancer.* (2017) 141:1240–8. doi: 10.1002/ijc.30822
- [44] Parmar C, Leijenaar RTH, Grossmann P, Velazquez ER, Bussink J, Rietveld D, et al. Radiomic feature clusters and Prognostic Signatures specific for Lung and Head & neck cancer. *Sci Rep.* (2015) 5:1–10. doi: 10.1038/srep11044
- [45] Zhu X, Dong D, Chen Z, Fang M, Zhang L, Song J, et al. Radiomic signature as a diagnostic factor for histologic subtype classification of non-small cell lung cancer. *Eur Radiol.* (2018) 28:2772–8. doi: 10.1007/s00330-017-5221-1
- [46] Huang Y, Liu Z, He L, Chen X, Pan D, Ma Z, et al. Radiomics signature: a potential biomarker for the prediction of disease-free survival in early-stage (I or II) non small cell lung cancer. *Radiology.* (2016) 281:947–57. doi: 10.1148/radiol.2016152234
- [47] Lambin P, Leijenaar R, Deist T. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol.* (2017) 14:749. doi: 10.1038/nrclinonc.2017.141
- [48] Krafft SP, Briere TM, Court LE, Martel MK. The utility of quantitative ct radiomics features for improved prediction of radiation pneumonitis. *Med Phys.* (2018) 45:5317–24. doi: 10.1002/mp.13150
- [49] Sun T, Wang J, Li X. Comparative evaluation of support vector machines for computer aided diagnosis of lung cancer in CT based on a multi-dimensional data set. *Comput Methods Prog Biomed.* (2013) 111:519–24. doi: 10.1016/j.cmpb.2013.04.016
- [50] Kuhn M, Johnson K. *Applied Predictive Modeling.* New York, NY: Springer (2013).