International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-8, June 2020

Phase Recognition of Lung Cancer via Steerable Riesz Wavelets with Rf Algorithm

S. N. V. Nishanth, S. Suryadev, Ch. Charan Teja Reddy, S. Kalaivani

Abstract: ***Lung cancer is one of the diseases which has a high mortality. If the condition is detected earlier, then it is easier to reduce the mortality rate. This lung cancer has caused more deaths in the world than any other cancer. The main objective is to predict lung cancer using a machine learning algorithm. Several computer-aided systems have been designed to reduce the mortality rate due to lung cancer. Machine learning is a promising tool to predict lung cancer in its early phase or stage, where the features of images are trained using a classification model. Generally, machine learning is used to have a good prediction, but in some models, due to lack of efficient feature extraction value, the training has not been done more effectively; hence the predictions are poor. In order to overcome this limitation, the proposed covariant texture model utilizing the steerable Riesz wavelets feature extraction technique to increase the effectiveness of training via the Random Forest algorithm. In this proposed model, the RF algorithm is employed to predict whether the nodule in the image is benign or malignant ii) to find the level of severity (1 to 5), if it is a malignant nodule. Our experiment result can be used as a tool to support the diagnosis and to analyze at an earlier stage of cancer to cure it.***

Keywords: ***Benign nodule, Malignant nodule, Random Forest, Random Walker, Steerable Riesz wavelets.***

I. INTRODUCTION

Lung cancer is one of the diseases which will lead to death and ill-health condition. Radiology and CAD system are generally preferred in diagnosis. Even days and month of delay has become uncommon to the people [1]. Late treatment largely causes a lot of death in lung cancer, as early detection is complex [2]. Late diagnosis causes severe problems for the person if it is not treated on time and also spread over the body. Due to the lack of a medical facility, the diseased person needs to wait for the treatment, which will increase the risk of spreading cancer [3].

Revised Manuscript Received on June 30, 2020.
***** Correspondence Author
S. N. V. Nishanth , School of Electronics Engineering, Vellore institute of technology, Vellore, India.
Email: sn.venkatanishanth2016@vitstudent.ac.in
S.Suryadev, School of Electronics Engineering, Vellore institute of technology, Vellore, India.
Email: sangeetham.suryadev2016@ vitstudent.ac.in
Ch.Charan Teja Reddy, School of Electronics Engineering, Vellore institute of technology, Vellore, India.
Email: cvs.charantejareddy2016@ vitstudent.ac.in
S. Kalaivani*, School of Electronics Engineering, Vellore institute of technology, Vellore, India.
Email: kalaivani.s@vit.ac.in

To overcome this without sacrificing the quality of treatment, a computer-aided system is introduced and encouraged. The survey says that only 15% of the people from the world population are cured by lung cancer. Lung cancer can be predicted using a segmented tissue density region of the affected person [4]. Initially, the analysis is carried out for suspiciousness, and then the Classification is done to analyze the stages of malignancy level [5]. The analysis has been done using a feature extraction technique. Texture feature analysis has to be carried out to predict whether the person has lung cancer or not [6]. A two-step process is followed to do so. Training is the first step where a large number of datasets for malignant and benign images are used to train the machine. Testing is the second step, where the current image is analyzed and predicted based on the training data [7].

II. EXISTING SYSTEM

In many of the existing models, preprocessing has been done through smoothing, enhancement, and segmentation. The segmentation is done using morphological operators. Then the feature extraction is performed using the GLCM(Gray level co-occurrence matrix), as shown in Fig. 1. Support Vector Machine classifier is utilized to predict whether the nodule is benign or malignant. But the process is still lacking in accuracy because of limited measures. The proposed system will overcome this issue.

III. DESIGN ASPECTS

In the proposed system, an automatic segmentation and classification process are done using improved Random Walker and Random Forest algorithm.

A. Improved Random Walker algorithm [8] is implemented to perform the automatic segmentation process. In this feature value of the texture, intensity, and shape of an object are taken, and then the weight is calculated to estimate the segmented ROI.

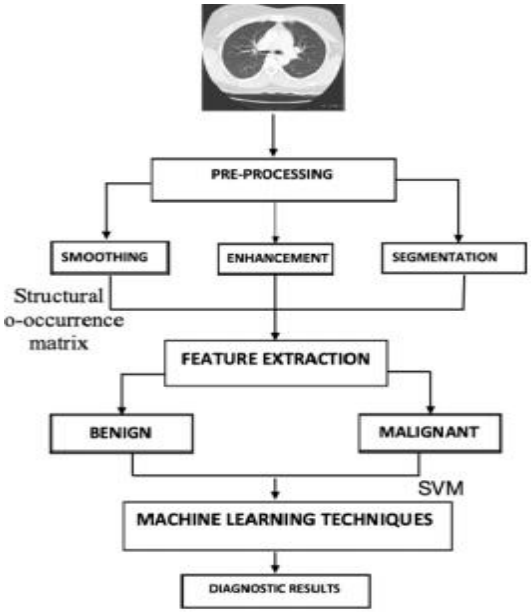


Fig. 1 Existing system block diagram

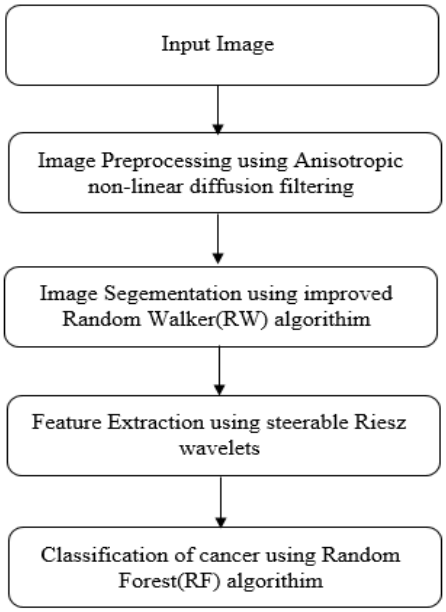


Fig. 2 Proposed system Block diagram

B. Steerable Riesz wavelets algorithm[9] is used to extract the feature value of an object in an image.

C. The Random forest algorithm is implemented to perform the classification part to define whether the nodule is benign or malignant. Along with the prediction of cancer, the stage of malignancy is also determined.

IV. PROPOSED SYSTEM

Lung cancer detection using a machine learning algorithm is proposed in this work. The block diagram of the proposed system is shown in Fig. 2.

Initially, to improve the quality of an image, the anisotropic non-linear diffusion filter is used as preprocessing. Then to extract the ROI the random walker algorithm is implemented by using the features values such as intensity, texture, and shape structure values to calculate the weight value of the segmented region.Once the segmentation is done using RW, then the feature extraction is carried out using steerable Riesz wavelets. The RF is implemented based upon the feature value to train the machine, and then the testing is done for the current image depend upon the trained data.

V. THE ARCHITECTURE OF THE SYSTEM

The architectural diagram of the proposed method for real-tie implementation is depicted in Fig.3 [10]. In this, User A and User B lung cancer data are taken and stored in the database. The features of both the users are subsequently and compared using RF architecture to predict the matching.

Initially, a large number of datasets can be trained based upon the feature values extracted from the system. Then the testing sample for the specific user is taken and tested for whether the person has benign or malignant.

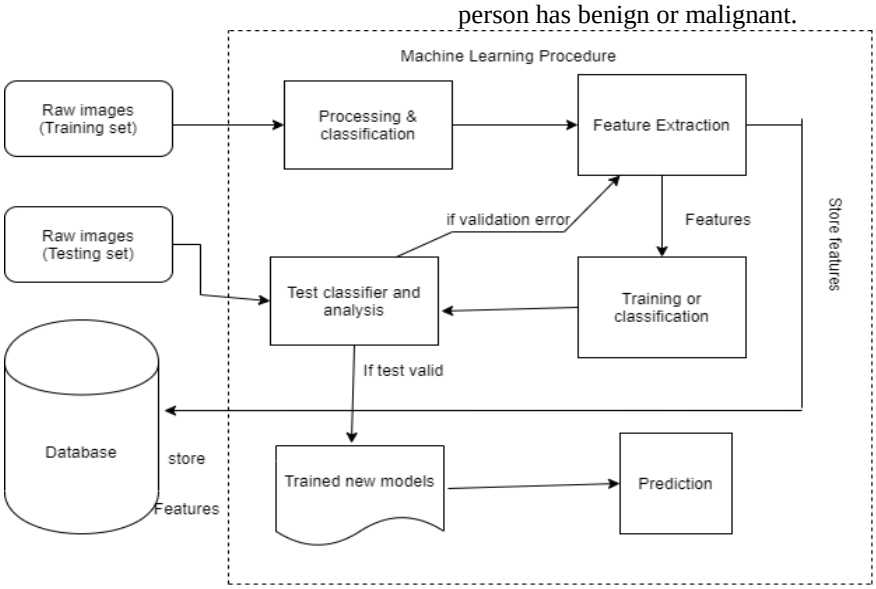


Fig. 3 Architecture diagram

VI. METHODOLOGY

The proposed method is composed of four major steps, as shown in Fig. 2, they are the preprocessing, image segmentation using improved random walker, feature extraction using steerable Riesz wavelets, classification of cancer using RF algorithm. These methods are elaborated in the following sections.

A. Preprocesing

The preprocessing technique is a crucial step in doing the segmentation of pulmonary nodules similarly. In this paper, "Anisotropic non-linear diffusion filter" is hired to suppress the image noise while preserving nodule barriers very well, as shown in Fig. 4.

Based on partial differential equations (PDE), several algorithms are used to carry out the elimination of noise from a photo and also enhancement of a photo in clinical processing.

Formally, let $\Omega \subset \mathbb{R}^2 \Omega \subset \mathbb{R}^2$ denote a subset of the plane and $I(\cdot,t) : \Omega \longrightarrow \mathbb{R}$ be a circle of relatives of gray scale photographs, then anisotropic diffusion is defined as

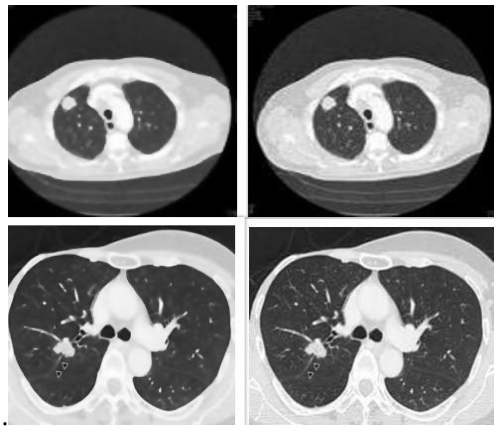
$$\frac{\partial I}{\partial t} = div(c(x,y,t)\nabla c \cdot \nabla I + c(x,y,t)\Delta I \quad (1)$$

Pietro Perona and Jitendra malik spearheaded the idea of anisotropic diffusion and anticipated abilities for the diffusion coefficient:

$$c(\|\nabla I\|) = e^{-(\|\nabla I\|/k)^2} \quad (2)$$

And

$$c(\|\nabla I\|) = \frac{1}{1+\left(\frac{\|\nabla I\|}{k}\right)^2} \quad (3)$$

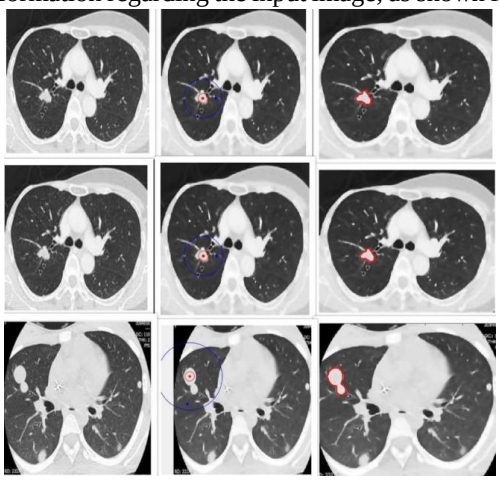


(a) (b)

Fig. 4 Examples of images after preprocessing (a)input images (b) images after preprocessing

B. Image segmentation(Improved random walker)

Accurate segmentation of the pulmonary nodules is an essential step for the subsequent extraction of the feature. In this paper, we use the improved method of random walkers from [8]. Improved RW method is automatic in the pulmonary nodule segmentation step automated seed acquisition system is implemented. All training images are derived from radiologist notes to get binary images of pulmonary nodules. Geodesic distance is then used to locate centers of the nodules. Lastly, nodule seeds are sampled from a circle centered onto a radius R. The background seeds are sampled from a circle centered on it, with a 4R radius. Nodule and background seeds are sampled according to the information regarding the input image, as shown in Fig.5.



(a) (b) (c)

Fig.5 (a) Examples of input images, (b) Nodule, and background seeds. The red and blue points indicate nodule and background seeds, respectively(c) Close-ups of the improved RW segmentation results

The texture features of LBPs are invariant in scale, and also invariant to distortions in intensity. Therefore the descriptor for LBP texture is calculated in this paper.

To improve the discriminative power between a pair of adjacent nodes, the features of strength, texture, and shape index are integrated to create a new weight function s in 4.

$$w_{ij} = \begin{cases} \exp\left(-\|I(i)-I(j)\|_2 - \|T(i)-T(j)\|_2 - \|SI(i)-SI(j)\|_2, & if\ j \in N(i) \\ 0, & otherwise \end{cases} \quad (4)$$

C. Feature extraction(steerable riesz wavelets)

AA. Texture descriptors

Due to simplicity and efficiency, LBPs have brought a revolution as one of the most prominent texture descriptors, which has been successfully applied to the texture extraction

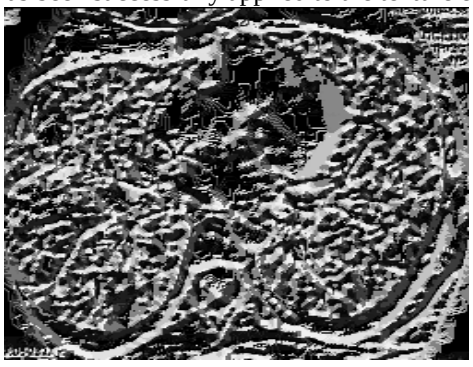


Fig. 6 Texture feature vector

of focal lesions such as breast cancer, sub-solid pulmonary nodules, and renal kidney in some of the references.

LBP features have been derived, as shown in Fig. 6, using circular neighborhood pixels uniformly at the R radius to the center pixel, referred to as LBPP, R. The formula is defined in the following equation:

$$LBP(c) = \sum_{i=0}^{p-1} s(gi - gc)2^i \qquad s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{5}$$

BB. Feature extraction

A good technique of extraction of features should be used to extract all the necessary features from an image in order to predict the value more precisely. In this paper, we used a texture-learning approach called "Steerable Riesz wavelets." Initially, by using kernel support vector machines, the linear combinations of Steerable Riesz wavelets are obtained. The signatures of the texture which lead Class-based simulation of optimal discriminating properties. Visualization of the signatures obtained enables verification of the visual relevance of the concepts learned. After that, local orientations of the signatures are modified to optimize their responses, which are analytically implemented and yet can be interpreted as a linear combination of the initial Riesz steerable models. To get the final texture signatures that covariate with rotation, the global cycle is repeated iteratively.

In this section, we describe our approach to iterative rotation – covariant texture learning using steerable Riesz wavelets. The Riesz transform and associated filterbanks are explained in subparts of the Chapter. The learning framework for the iterative texture and the validation scheme used to evaluate it are described respectively in Sections subparts.

1. Notations A generic d – dimensional signal f is considered to be indexed by the x = (x1, ..., xd) continuous – domain space variable. The Fourier transformation of d – dimensional f is noted as:

$$f(x) \overset{F}{\leftrightarrow} \hat{f}(\omega) = \int_{R^d} f(x) e^{-j\langle \omega, x \rangle} dx_1 \dots dx_d,$$
$$with\ \omega = (\omega_1, \dots, \omega_d) \in R^d \tag{9}$$

2. Steerable Riesz Filterbanks: The Riesz transform is a multidimensional extension of the Hilbert transform, which maps every f(x) function to its harmonic conjugate and is a very effective method to manipulate periodic signals mathematically. The N+1 components of the Nth – order Riesz transform RN are defined as: for a 2–D signal f(x),

$$R^N\{f\}(x) = \begin{bmatrix} R^{(0,N)}\{f\}(x) \\ R^{(n,N-n)}\{f\}(x) \\ R^{(N,0)}\{f\}(x) \end{bmatrix} \tag{10}$$

with n = 0,1,...,N. A singular kernel R(n,N−n){f}(x) is defined in the Fourier domain as:

$$R^{(n,N-n)}\{f\}(x) \overset{F}{\leftrightarrow} R^{\widehat{(n,N-n)}}\{f\}(\omega), \tag{11}$$

Where

$$R^{\widehat{(n,N-n)}}\{f\}(\omega) = \sqrt{\frac{N}{n!\,(N-n)!}} \frac{(-j\omega_1)^n (-j\omega_2)^{N-n}}{\|\omega\|^N} \hat{f}(\omega), \tag{12}$$

3. An appropriate weighting scheme of the energies of the responses of the Riesz components as

$$\Gamma_c^N = \omega^T \mathcal{R} = \omega_1 \mathcal{R}^{(0,N)} + \omega_2 \mathcal{R}^{(1,N-1)} + \cdots + \omega_{N+1} \mathcal{R}^{(N,0)} \tag{13}$$

4. A multiscale texture signature is obtained as an extension of Eq. 12 with multiscale Riesz filterbanks

$$\Gamma_c^N = \omega_1 \left(\mathcal{R}^{(0,N)}\right)_{s_1} + \omega_2 \left(\mathcal{R}^{(1,N-1)}\right)_{s_2} + \cdots + \omega_{J(N+1)} \left(\mathcal{R}^{(N,0)}\right)_{s_j} \tag{14}$$

where s_j, j = 1,..., J is the index of the scale. The scale-wise signatures Γ_c^N can be obtained using only weights and corresponding Riesz models at scale j.

D. RF Classification

For the Classification of cancer stages, a machine-learning algorithm is required. We employ the RF (random forest) algorithm from [4] in this paper. The RF classifier is trained by data sets of the lung image dataset (LIDC) consortium. To predict the class label, a rational and effective method of composition of all decision trees is presented, as shown in Fig.8. RF is a good classifier for separating the pulmonary nodules from benign and malignant. Within this portion, the class labels are trained to predict an RF. At each node of the tree, a subset of features is selected to find the best split, which is generated using the bootstrap method. To send a node to the left or right child node, each split node of the tree is associated with a split function f ·.

f · is defined in the equation below:

$$f(v, \theta_j): R^d \times \chi \longrightarrow \{0,1\} \tag{15}$$

where v is a vector input function, and θj is a feature parameter. χ denotes the space of all split parameters, and Rd is the d dimensional feature space. A threshold τj is selected to determine the node split, then all training images are split into left and right

$$\Re_L(j, \tau_j) = \{[v_i; \theta_j] | \forall i, f(v_i; \theta_j) \leq \tau_j\} \tag{16}$$

$$\Re_R(j, \tau_j) = \{[v_i; \theta_j] | \forall i, f(v_i; \theta_j) > \tau_j\} \tag{17}$$

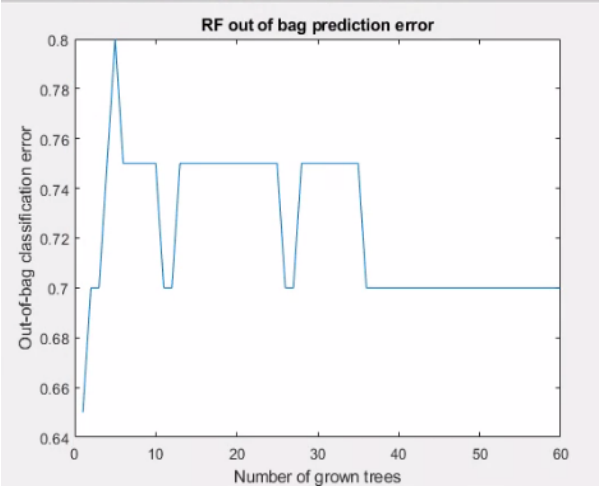


Fig. 7 RF out of the bag prediction error

In this paper, the probability density distribution for each tth tree feature vector as shown in Fig.7 is modeled as a multivariate in the following equation:

$$p_t(v|\mathcal{L}) = \frac{1}{(2\pi)^{d/2}\Lambda} e^{-(\frac{1}{2})(v-\hat{v})^T \Lambda^{-1}(v-\hat{v})} \quad (18)$$

where $\mathcal{L}$ is a binary class label indicating a malignant nodule $\mathcal{L}$=1 or a benign nodule $\mathcal{L}$=0. vi and vj are input feature vectors at ith and jth pixels, respectively.

Trees grow the depth of the tree progressively by adding new nodes. The number of points is less than a threshold or a maximum depth of the tree D, and the growth stops. The output of random forests is predicted by weighting all of the individual tree predictions, as defined in the following equation:

$$p(\mathcal{L}|v) = \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{NP} w'_i p_t(\mathcal{L}|v) \quad (19)$$

where $\mathcal{L}$ is a binary class label indicating a malignant nodule $\mathcal{L}$=1 or a benign nodule $\mathcal{L}$=0. T is the number of trees, and

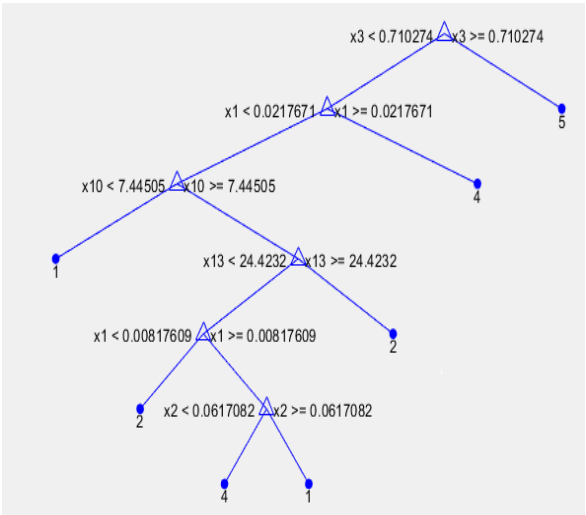


Fig. 8 Random forest decision trees

NP is the number of image pixels. $p_t(\mathcal{L}|v_i)$ is the probability density function given a multivariate feature vector of the pixel i in the tth tree.

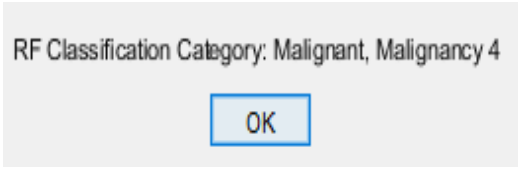
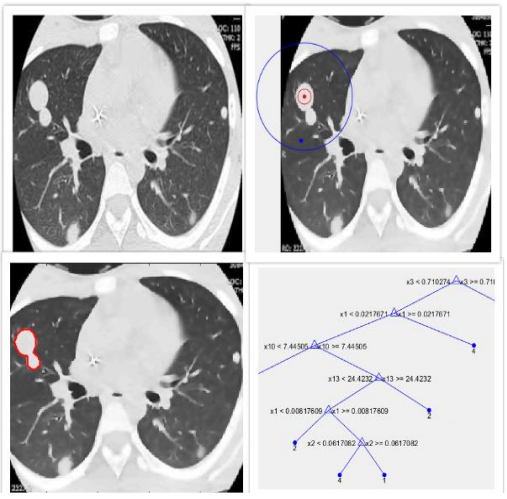


Fig.no.9 Classification of cancer by the decision trees

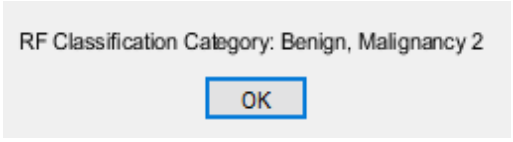
Provided an unlabeled pulmonary nodule, subsets of features are previously tested in the same way from the training collection. Each subset of test nodules features is pushed into each tree, starting at the root and adding the respective test series.

VII. RESULT AND DISCUSSION

The experimental results of the improved random walker segmentation and random forest classification obtained are shown in Fig. 10. The feature texture values are compared with standard values from the feature extraction method from [8]. The texture feature parameters obtained from GLCM and steerable riesz wavelets are compared as in the Table. 1.



(a)



(b)

Fig.no. 10 (a)Final results obtained in each subsequent module(preprocessing, image segmentation, RF classification) (b) Final output obtained after the RF classification(the class of cancer)

TABLE. 1 Comparison between Mean values of Texture Feature parameters from 20 samples from the dataset.

Texture feature parameter	GLCM	Steerable Riesz
Energy	0.9270	0.99720
Entropy	0.0145	0.07276
sum of square	1.0566	1.3339
sum average	2.0195	4.5237
sum variance	4.2948	7.500
difference variance	0.0101	0.3564

The comparison shows an improvement in the texture vectors in the proposed method.

VIII. CONCLUSION

To identify the benign and malignant model of an image, the improved Random Walker and Random Forest Algorithm are implemented. The principal benefit of the improved RW is, it would segment the nodule automatically using the method of seed acquisition. Because of the proper process of texture and shape feature value, the improved model is more accurate than the previous traditional model. The system is improved with the use of steerable Riesz wavelets as feature extraction functionality. After the extraction of texture, shape and, intensity feature vectors, a rational and effective method of composition of all decision trees is presented by the RF algorithm for the prediction of the class label.

REFERENCES

1. N.Camarlinghi, ``Automatic detection of lung nodules in computed tomography images: Training and validation of algorithms using public research databases," Eur. Phys. J. Plus, vol. 128, no. 9, p. 110, Sep. 2013.
2. R. L. Siegel, K. D. Miller, and A. Jemal, ``Cancer statistics, 2016," CA, Cancer J. Clin., vol. 66, no. 1, pp. 730, Jan. 2016
3. D. Kumar, A. Wong, and D. A. Clausi, ``Lung nodule classification using deep features in CT images," in Proc. 12th Conf. Comput. Robot Vis. (CRV) Jun. 2015, pp. 133138.
4. Zakarie Hashi, Kalamazoo, MI.Rabia Almamlook "Lung Cancer Survival Prediction Using Random Forest-Based Decision Tree Algorithms,".2017
5. W. Shen et al., ``Multi-crop convolutional neural networks for lung nodule malignancy suspiciousness classification," Pattern Recognit., vol. 61, pp. 663673, Jan. 2017.
6. Han, F., Wang, H., Zhang, G., *et al.*: 'Texture feature analysis for computer-aided diagnosis of pulmonary nodules', *J. Digit. Imaging*, 2015, 28, pp. 99– 115
7. Hyo Kyung Lee, FengJu, , Raymond U. Osarogiagbon, Nicholas Faris, Xinhua Yu, FedoriaRugless, Shan Jiang, and JingshanLi " A System-Theoretic Method for Modeling, Analysis, and Improvement of Lung Cancer Diagnosis-to-Surgery Process,", 2017
8. Xiang-Xia Li1, Bin Li1, Lian-Fang Tian1, Li Zhang "Automatic benign and malignant classification of pulmonary nodules in thoracic computed tomography based on RF algorithm," 1.
9. Adrien Depeursinge, Antonio Foncubierta-Rodriguez "Rotation–Covariant Texture Learning Using Steerable Riesz Wavelets,"
10. "Skin Disease Diagnosis System using Image Processing and Data Mining." R. S. Gound ,Priyanka S. Jyoti B., India.

AUTHORS PROFILE

S. N. V. Nishanth is pursuing B. Tech in Electronics and Communication Engineering from VIT Vellore, Tamil Nadu.

S. Suryadev is pursuing B. Tech in Electronics and Communication Engineering from VIT Vellore, Tamil Nadu.

Ch. Charan Teja Reddy is pursuing B. Tech in Electronics and Communication Engineering from VIT Vellore, Tamil Nadu.

Dr.S.Kalaivani obtained her Ph.D in Medical Image processing from ANNA University, Chennai, Post graduate in Communication Systems and Undergraduate in Electronics and Communication Engineering. She has published her work in more than 25 reputed international journals, 20 international conferences and more than 20 national conferences and journals. Her research interests are neonatal brain image analysis, Hyperspectral Image processing for mineral identification and crop area detection, SAR (Synthetic Aperture Radar) image processing for change detection in urban areas, coastal area monitoring, deforestation and vegetation monitoring, Ultrasound image processing, ECG signal analysis etc.