International Journal of Engineering &Technology, 7 (4) (2018) 6122-6130

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET
doi: 10.14419/ijet.v7i4.23127
Research paper

Steerable wavelet transforms with modified extreme learning machine for lung segmentation and cancer detection in lung CT images

J. Vijayaraj [1]*, Dr. D. Loganathan [2]

[1] Ph.D. Scholar, Department of Computer Science and Engineering, Pondicherry Engineering College, Puducherry India
[2] Affiliation Professor, Department of Computer Science and Engineering, Pondicherry Engineering College, Puducherry India
**Corresponding author E-mail: Vijayaraj.j@pec.edu*

Abstract

Lung Segmentation is an imperative measure of both Computer Aided Detection (CAD) and Diagnosis frameworks for lung cancer, since the execution of those frameworks are relied upon the execution of the Lung Segmentation in Computed Tomography (CT) pictures of lungs. In this paper, a Modified Extreme Learning Machine (MELM) based classification algorithm is proposed to segment the tumor region of lung pictures. The fundamental guide of this work is to exhibit an algorithm for distinguishing the tumor from segmented lung pictures. In this way, this segmentation procedure, at first, the input CT lung pictures are pre- processed to evacuate Poisson noise by utilizing improved kaun filter. At that point, the lope of lung has been segmented by utilizing Steerable Wavelet Transform (SWT) for efficacious diagnosis. After lung segmentation, the Gray Level Co-Occurrence Matrices (GLCM) based features are utilized for tumor identification. In this process, the features are changed as an input to Modified Extreme Learning Machine (MELM), and it classifies the benevolent and harmful area of lung pictures. At long last, the threatening caner locales are segmented for diagnosis process. This exhibited scheme is assessed by utilizing open database. The simulation results comes about demonstrate the execution of presented scheme (i.e. MELM) has achieved better outcomes as far as high accuracy, purity, rand index, jaccard coefficient, F-measure, G-mean, precision and recall thought about than existing modal based classification algorithms.

Keywords: *Lung Cancer; CT Images; Kaun Filter; SWT; GLCM; MELM.*

1. Introduction

As indicated by the American Cancer Society, around 1,688,780 new cancer cases would have been analyzed and around 589,430 disease passings. Around 13% of all recently analyzed malignancies were lung tumor and lung tumor passings represent around 27% of all cancer deaths in 2015 in the USA [1]. Lung disease has turned into the main source of danger related passings around the world. One of the real reasons for the high death rate of lung tumor is that lung malignancy is probably not going to create the obvious indications until the point when the illness is progressed. Around 85% of lung tumors are analyzed after metastatic disease movement, and the subsequent treatment isn't accessible. Lung disease is curable in the event that it is timely diagnosed and suitably treated. In this manner, early finding is vital. It can enhance multiyear survival rate from just 15 to 18% [1]. Lung disease possibly shows itself as pneumonic nodules at a beginning time [2]. Computed tomography (CT) is a standout amongst the most conspicuous imaging modalities for examining and diagnosing lung disease because of its non-obtrusiveness, high spatial determination and low cost. Even though most pulmonary nodules are benevolent, the results of creating harmful nodules are extremely severe. In this way, the requests for early-stage classifications of benevolent and dangerous pulmonary nodules have expanded also In clinical practice, the characterization of generous and threatening pulmonary nodules is still judged by radiologists; however, it is to a great degree tedious and subjective. Consequently, automatic classification of benevolent and harmful pulmonary nodules has a critical requirement for checking the disease progression, arranging the medicines and foreseeing the patient results. Tragically, it has gotten extensively less consideration. The classification of benevolent and threatening pulmonary nodules is an extremely difficult issue because of low difference, variable sizes, unpredictable shapes and irregular positions. In spite of some unsupervised or supervised classification techniques for classifying benevolent and dangerous pulmonary nodules have been broadly examined over the previous decade, there is as yet an absence of legitimate order strategy with the elite, which is the focal point of this paper.
Volume estimations and perceptions of tumor in 2D is essential for legitimate conclusion and treatment. The first and principal advance for lung picture examination is the segmentation of lung. segmentation is the procedure by which region of interests of a computerized picture are partitioned and marked in various fragments utilizing properties, for example intensity, texture and so forth., which is considerably simpler to investigate. Noises in lung segmentation would produce false data, so precise segmentation is important [3]. The segmentation of lung tumor is extremely heuristic in nature as the CT pictures incorporates bones, bronchioles, trachea, delicate tissues, background artifacts which shares properties force, network with malignancy sores. Despite the fact that CT filter gives better imaging of lung, the programmed identification of the tumor is troublesome. Consequently, computer aided detection (CAD) is performed for tumor nodule segmentation. Scientists proposed distinctive techniques for the tumor nodule segmentation. Frequently clinician hones manual segmentation [4] which is administrator needy and has a tendency to be flawed. Machine

learning based calculation gives better outcomes, however they require training data. On the off chance that legitimate and adequate preparing information isn't accessible, at that point machine learning based calculation can't be connected. However, Support Vector Machine (SVM) based segmentation [4,5] just characterizes lung tumor into benevolent and harmful writes from two dimensional (2D) pictures.

In this work, a modified ELM with SWT is proposed for automatic cancer segmentation and classification in CT lung pictures. In his procedure, to start with, the input pictures are handled. At that point, the disease module is portioned to enhance the cancer classification accuracy. From that point onward, the GLCM features are extricated from segmented pictures. At last, the cancer is named threatening and starts in CT picture. The experimental results demonstrate that the proposed design achieved superior thought about than existing schemes. The rest of this paper is sorted out as takes after: The current lung tumor segmentation and classification plans are talked about in Section 2. The proposed technique is exhibited in Section 4. Simulated and analytical outcomes for the assessment and approval are given in Section 4. At last, helpful comments and conclusions are incorporated into Section 5.

2. Related work

In this area, some current lung tumour segmentation and classification schemes have been talked about and their benefits and drawbacks additionally introduced.

Li et al., [6] proposed an enhanced Random Forest (RF) calculation for characterization of benevolent and threatening pneumonic nodules in thoracic processed tomography pictures. Initially, an enhanced random walk calculation is proposed to consequently segment pulmonary nodules. At that point, intensity, geometric and texture features based on the grey-level co-occurrence matrix, pivot invariant uniform local binary pattern and Gabor channel strategies are consolidated to create a successful and discriminative feature vector. Common data is utilized to diminish the dimensionality. At long last, an enhanced RF classifier is prepared to order benevolent and threatening nodules. The proposed characterization technique on the lung pictures dataset consortium dataset accomplishes an sensitivity of 0.92 and the territory under the receiver-operating-characteristic curve of 0.95. Yet, the computational cost of this calculation was high.

Rodrigues et al., [7] displayed a novel Structural Co-Occurrence Matrix (SCM)- based way to deal with characterize nodules into threatening or benevolent nodules and furthermore into their harm levels. The SCM method was connected to remove features from pictures of nodules and classifying them into threatening or benign nodules and furthermore into their harm levels. The computed tomography exams from the lung picture database consortium and picture database asset activity datasets give data concerning nodule positions and their harm levels. The SCM was connected on both grayscale and Hounsfield unit pictures with four channels, to wit, mean, Laplace, Gaussian, and Sobel channels making eight distinct setups. This approach beat alternate strategies in the two assignments; it accomplished 96.7% for both accuracy and F-Score measurements in the primary errand, and 74.5% accuracy and 53.2% F-Score in the second. Be that as it may, the preparation time of this plan was high.

Abdelrahman and Abdelwahab [8] introduced a novel Accumulated Gray-Level Image (AGLI) strategy for gene depiction, where each base in gene string is spoken to by aggregated number in light of its request in gene sequence and after that reflected into picture area. AGLI is consolidated with 2D principle component analysis to assemble exact and low-dimensional calculation for grouping the hereditary changes. Proposed calculation was connected on the best 10 successful genes in lung disease, where a accuracy of 99.27% was accomplished.

Chen et al. [9] proposed a Neural Network (NN) ensemble based Computer-Aided Diagnosis (CAD) conspire for characterizing amiable and threatening pneumonic nodules, which accomplished regions under the Receiver-Operating-Characteristic (ROC) bend of 0.79. Chen et al. [10] additionally played out an examination between artificial NNs (ANNs) and multivariable logistic regression (LR) for separating harmful nodules from benevolent knobs. ANNs had a higher classification execution than LR, which got a area under the ROC curve (AUROC) of 0.955, a accuracy rate of 90.0%.

Lin et al. [11] proposed a fractal-based feature set got from the fragmentary Brownian movement model for separating harmful nodules from amiable single pulmonary nodules. The Support Vector Machine (SVM) classifier was utilized to recognize threatening nodules from benevolent nodules. They acquired an AUROC of 0.8437, a accuracy of 83.11%, an sensitivity of 90.92%, a specificity of 71.70%, a positive prescient estimation of 80.05% and a negative prescient estimation of 87.52%. Gabor and LBP-based texture features were utilized to develop three dimensional (3D) Haralick includes by Han et al. [12] for classifying amiable and threatening nodules. Thinking about pulmonary nodules with a composite rank of threat '1' and '2' as considerate and '4' and '5' as dangerous, this paper accomplished an AUROC of 0.94 by utilizing a SVM classifier.

Cheng et al. [13] proposed a deep learning based computer-aided diagnosis (CADx) for classifying considerate and threatening knobs. The stacked denoising auto-encoder was utilized to naturally remove features. Dhara et al. [14] connected a SVM for characterization of considerate and harmful pulmonary nodules. The shape-based, texture based and edge based features were extricated from semi- automated segmented nodules. They accomplished an AUROC of 0.9505 when design 1 (composite rank of malignancies '1' and '2' as favourable and '4' and '5' as threatening) was utilized.

Liu et al. [15] proposed a machine-learning strategy to foresee malignancy status, which scored 24 radiological characteristics and fabricated a linear classifier to characterize pulmonary nodules. They had gotten an AUROC of 0.88, an exactness of 81%, and sensitivity of 76.2% and a specificity of 91.7%. Tajbakhsh and Suzuki [16] played out an examination between massive training ANNs (MTANNs) and convolutional NNs (CNNs) for characterization of benevolent and dangerous knobs in CT pictures. MTANNs got an AUROC of 0.8806, which was more prominent than the CNNs with an AUC of 0.7755.

Ahmed Soliman et al., [17] segmented neurotic and well lungs for dependable computer-aided disease diagnostics, a pile of chest CT filters is demonstrated as an example of a spatially inhomogeneous joint 3D Markov-Gibbs random field (MGRF) of voxel-wise lung and chest CT picture signals (powers). The proposed learnable MGRF coordinates two visual appearance sub models with a versatile lung shape submodel. The first-order appearance submodel represents both the first CT picture and its Gaussian scale space (GSS) separated adaptation to determine local and global flag properties, individually. Each observational peripheral likelihood dissemination of signals is nearly approximated with a linear combination of discrete Gaussians (LCDG), containing two positive overwhelming and various sign-interchange sub14 ordinate DGs. The estimation is isolated into two LCDGs to portray exclusively the lungs and their experience, i.e., all other chest tissues. The second-arrange appearance submodel evaluates restrictive pairwise force conditions in the closest voxel 26-neighborhood in both the first and GSS19 separated pictures. The shape submodel is worked for an set of training data and is adjusted amid segmentation utilizing both the lung and chest appearances. The accuracy of the proposed segmentation structure is quantitatively surveyed utilizing two open databases (ISBI VESSEL12 test and MICCAI LOLA11 challenge). From the above writing, the classification accuracy of lung cancer detection is still tested. Thus, to enhance the recognition exactness, here, the machine learning calculation of MELM with SWT is proposed.

3. Research methodology

In this section, the proposed MELM based lung segmentation and classification has been discussed.

3.1. System overview

The general design of proposed MELM is outlined in figure 1. It demonstrates the four stages of proposed conspire like pre-preparing, segmentation, feature extraction and classification. In this procedure, to start with, the input images are handled to lessen the t poison noise. At that point, the tumor is segmented by using SWT. From that point onward, the GLCM features are separated for classification. At last, the MELM classifier is anticipated the growth arrange like generous and harmful.

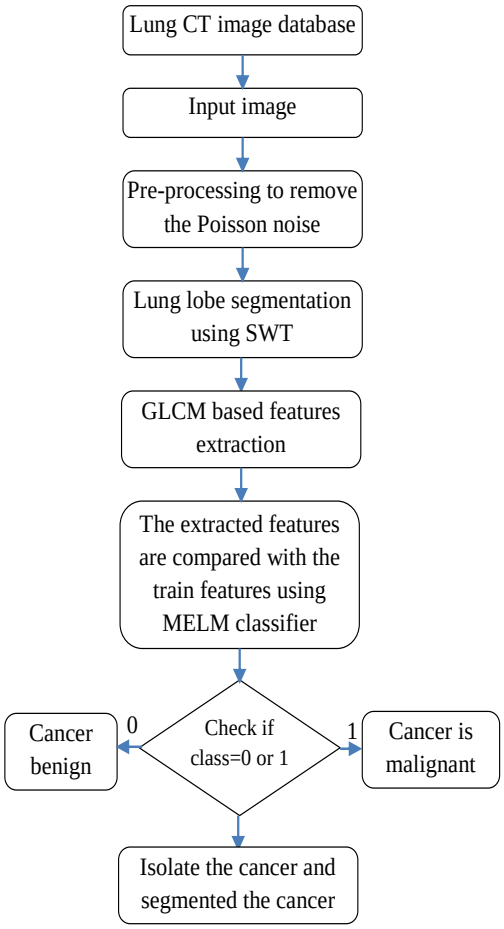

Fig. 1: Overall Architecture of MELM Based Cancer Classification.

3.2. Pre-processing stage

This strategy has redesigned the idea of picture and obliterates noise. To evacuate above poison noise and to have exactness of the lung CT picture, the enhanced Kaun filter has been proposed in this framework, which works without expelling edges or features in the lung CT pictures. In any case, the noise is changed from multiplicative noise display into a picture dependent additive noise model. From that point onward, the minimum Mean Square Error (MSE) condition is connected to the model for evaluating the first type of picture. Pre- processing coming about of grey-level value GL is given as takes after.

$$GL_{i,j} = \sum_{i=1}^{m=3,n=3} \sum_{j=1}^{m=3,n=3} Cp_{i,j} * W_{i,j} + M_{i,j} * (1 - W_{i,j}) \quad (1)$$

Where, (i, j) speaks to the picture pixels, (m, n) speaks to a rows and columns of the pixel information. Cp_ (i,j) is characterized as a middle pixel in filter window M_(i,j), which is alluded as a mean estimation of force inside window and W_(i,j) is said as weighting factor. Rely upon this weight, the noise decrease execution has been anticipated. Along these lines, to enhance the picture quality and noise lessening, the weight factor has been streamlined by Random Search Algorithm (RSA). Along these lines, this capacity is with KF named as enhanced KF.

A general random search algorithm is depicted by a succession of iterates$\{W_k\}, k = 0, 1, \ldots, k$, which may rely upon past pixel focuses and algorithmic parameter θ and it comprises of the mean vector and the covariance matrix. The current emphasize W_k represented an accumulation of focuses. Likewise, the cycles are promoted to indicate that they are random variables, mirroring the probabilistic idea of the RSA.

The generic random search algorithm is relies upon two essential procedures. They are candidate points, and the update procedure.

3.3. Segmentation using SWT

The accurate segmentation of pneumonic nodules is a basic advance for consequent feature extraction step. Albeit monstrous pulmonary nodule segmentation strategies [18-24] have been created over the previous decade, how to accurately section pneumonic nodules still remains an open test. Here, the wavelet based steerable transformation scheme is proposed for strong segmentation.

The steerable pyramid is a specific acknowledgment of recursive multi-scale changes. The disintegration can be characterized in the Fourier area, where the subbands are polar detachable. For the instance of n=1. {B_k ($\omega^{\rightarrow}$ | k=0, 1)} are band pass oriented filters, H_0 ($\omega^{\rightarrow}$) is a non-arranged high-pass channel and L_1 ($\omega^{\rightarrow}$) is a narrowband low-pass channel [25]. The SWT conspire incorporates two sections, for example, decomposition and synthesis. In this, at first, the picture is deteriorated by left half of the plan. The CT pictures are disintegrated into low pass and high pass groups with steerable filter L_0 andH_0. The low pass band keeps on separating into an arrangement of band pass subbands B_0,…,B_Nand lower low pass subband L_1.The bring down low pass subband is subsampled by a factor of 2 along the x and y bearings. The recursive structure is acknowledged by rehashing the shaded region to L_i (i means the present deterioration scale). At a similar time, the right side of the graph gives the combination part of the steerable wavelet plot. The integrated picture is reproduced by up sampling the lower low pass subband by the factor of 2 and including with the arrangement of band pass subbands and the high pass subbands. For i=1, the recreation in the frequency domain is:

$$\hat{X}(\vec{\omega}) = \left\{ |H_0(\vec{\omega})|^2 + |L_0(\vec{\omega})|^2 \left(|L_1(\vec{\omega})|^2 + \sum_{k=0}^{n} |B_k(\vec{\omega})|^2 \right) \right\} X(\vec{\omega}) + a.t. \quad (2)$$

Where a.t. demonstrates the associating terms. To guarantee end of the associating terms, the L_1 filter ought to be obliged to have a zero reaction for frequencies higher than π/2 in both ω_x and ω_y and the transfer function of the framework ought to be equivalent to one keeping in mind the end goal to evading amplitude distortion. In this proposed scheme, the preprocessed pictures is splitted as low pass and high pass channel based parts and after that the disease affected region is portioned as synthesis picture. SWT gives more exact segmentation.

3.4. Feature extraction using GLCM

Image feature extraction is vital phase of PC based framework. Feature extraction gives certain parameters, based on which PC framework takes choice. After the segmentation is performed on lung area, the features can be acquired from it and the conclusion

control can be intended to distinguish nodules in the lung. The whole component which are ascertained from the picture, pass on some data in regards to lung nodule. This data is exceptionally useful in identifying lung nodule as threatening or non-harmful. In this manner the features separated from the CT picture can be utilized as demonstrative pointers [26]. The features that are utilized as a part of this investigation are Texture features utilizing Co-occurrence matrix representation. The GLCM is a second-arrange factual technique for the extraction of texture f eatures. To start with, the picture is converted into a l-grey-level image and GLCM is produced by counting occurrences of intensity pairs between the current and neighbor pixels for each scale and orientation. We develop a feature vector by utilizing the average of matrices for each scale and all orientations. The standardized GLCM is computed in the equation beneath:

$$G(i,j) = \frac{N(i,j)}{\sum_{m=0}^{l-1}\sum_{n=0}^{l-1} N(m,n)} \tag{3}$$

Where i and j are grey values in the l- grey level picture. N (i,j) is the co- occurrence relative recurrence frequency matrix by the equation underneath:

$$N(i,j) = num\left(\left((x_1,y_1),(x_2,y_2)\right)\middle| x_2 - x_1 = d\cos\theta, y_2 - y_1 = d\sin\theta, I(x_1,y_1) = i, I(x_2,y_2) = j\right) \tag{4}$$

Where $(x_1, y_1)$and (x_2, y_2) are pixel positions, and I(·) is the grey level of the pixel. num · means the quantity of the pixel matches that fulfill the comparing conditions. In this paper, 15 texture features are ascertained from the comparing GLCM.

Auto correlation	$\sum_{i=0}^{l-1}\sum_{j=0}^{l-1}(i,j)P(i,j)$
contrast	$\sum_{i=0}^{l-1}\sum_{j=0}^{l-1}\lvert i-j\rvert^2 P(i,j)$
cluster prominence	$\sum_{i=0}^{l-1}\sum_{j=0}^{l-1}\left(i+j-u_x-u_y\right)^4 P(i,j)$
dissimilarity	$\sum_{i=0}^{l-1}\sum_{j=0}^{l-1}\lvert i-j\rvert P(i,j)$
energy	$\sum_{i=0}^{l-1}\sum_{j=0}^{l-1}P(i,j)^2$
entropy	$\sum_{i=0}^{l-1}\sum_{j=0}^{l-1}P(i,j)\log\left(P(i,j)\right)$
maximum probability	$\max_{i,j} P(i,j)$
sum of squares	$\sum_{i=0}^{l-1}\sum_{j=0}^{l-1}(i-j)^2 P(i,j)$
sum average	$\sum_{j=0}^{l-1} i^2 P_{x+y}(i)$
sum variance	$\sum_{j=0}^{l-1}\left(i-\sum_{j=0}^{l-1} i^2 P_{x+y}(i)\right)^2 P_{x+y}(i)$
sum entropy	$-\sum_{i=2}^{2l} P_{x+y}(i)\log\left[P_{x+y}(i)\right]$
difference variance	$\sum_{i=0}^{l-1} i^2 P_{x-y}(i)$
difference entropy	$-\sum_{i=0}^{l-1} P_{x-y}(i)\log\left[P_{x-y}(i)\right]$
information measure of correlation	$\frac{HXY - HXY1}{\max(HX,HY)}$
inverse difference normalized (INN)	$\sum_{i=0}^{l-1}\sum_{j=0}^{l-1}\frac{P(i,j)}{1+\lvert i-j\rvert^2/l}$

The extracted features are changed to MELM training process for identification of tumor.

3.5. MELM for classification

Extreme learning machine (ELM) is another class of single-concealed layer feedforward neural system (SLFN), which is straightforward in principle and quick in implementation. However, the current revealed ELM and its enhanced rendition are just in light of the experimental hazard minimization standard, which may experience the ill effects of overfitting. In this paper, to bring down the overfitting phenomena of the extreme learning machine algorithms, we allude to the Least-Squares Support Vector Machine (LS-SVM) calculation, draw the basic hazard minimization rule into the ELM calculation, and propose a Modified ELM named as MELM.

For a perception data set (i.e. number of features) with N nodes in the concealed layer and the excitation work G, the extreme learning machine model can be communicated as

$$f(x) = \sum_{i=1}^{N}\beta_i G(a_i, b_i, x_i) = \beta.h(x) \tag{5}$$

Where β_ithe yield weight of the ith is concealed layer node and the yield neuron, a_i is the information weight of the input neuron and the ith concealed layer node, and b_iis the counterbalanced of the ith concealed layer node. Consider h (x) = [G (a_1,b_1,x_1) . . . G(a_N,b_N,x_N)] signifies the output matrix of concealed layer. ai and bi are arbitrarily chosen before training and continue as before in the training system. The yield weights β_ican be gotten by comprehending the least-squares solutions of the accompanying linear equation:

$$\min\sum_{i=1}^{N}\lVert\beta_i.h(x_i - y_i)\rVert \tag{6}$$

The least-squares solution to the equation is

$$\beta = H^{+}Y \tag{7}$$

Where H^+ is called the Moore-Penrose generalized inverse of the hidden layer output matrix H.

LS-SVM Regression. Consider that an input and output test informational index for regression analysis is T= {(x_1 ,y_1),...,(x_l ,y_l)}, where x_i∈R^nandy ∈ R,i = 1,...,l. LS-SVM regression algorithm maps the information X into a high-dimensional element space F through a nonlinear mapping ϕ and does linear regression in the space F. The regression estimation for the perception data set given above can be figured as underneath, where ω and b^*are the regression factors,

$$f(x) = \left(\omega.\phi(x)\right) + b^{*} \tag{8}$$

LS-SVM regression method is utilized to unravel the weight vector Φ and deviation $b*$. In view of the auxiliary hazard minimization, the advancement model of the optimal regression function [27– 29] can be built up as

$$\min\frac{1}{2}\lVert\omega\rVert^2 + \frac{1}{2}\gamma\sum_{i+1}^{l}\xi_i^2 \tag{9}$$

Subject to y_i-f(x_i)=ξ_i,ξ_i≥0,i=1,2,…,l

Where is the penalty constant, which is a tradeoff amongst multi-faceted nature and fitting exactness of regression model. Higher esteem γ implies higher fitting degree. ξ_i is the slack variable. LS-SVM changes the imbalance imperatives into uniformity requirements by characterizing loss functions not quite the same as those in the standard SVM. It develops the accompanying Lagrange function:

$$L_{LS-SVM} = \frac{1}{2}\|\omega\|^2 + \frac{1}{2}\gamma\sum_{i+1}^{l}\xi_i^2 - \sum_{i+1}^{l}\alpha_i[y_i - f(x_i) - \xi_i] \tag{10}$$

Where α_iis the Lagrange multiplier. As indicated by Karush-Kuhn-Tucker (KKT) ideal conditions, the linear equations can be acquired as takes after:

$$\begin{bmatrix} 0 & 1_v^T \\ 1_v & \Omega + \frac{I}{\gamma} \end{bmatrix}\begin{bmatrix} b^* \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix} \tag{11}$$

Where y =[y_1,y_2,...,y_l]^T,1_V= [1,1,...,1]^T,I is the $l \times l$ identity matrix, Ω is a square matrix, and Ω_(i,j)=K(x_i,x_j) is the ith row and the jth column data element, where Ω_(i,j)=K(x_i,x_j) is the kernel function that fulfills the Mercer condition.

Weighted ELM: In [30], the authors proposed a weighted extreme learning machine for irregularity realizing, which characterized a N×N diagonal matrix W related with each training samplex_i. Generally if training data x_i originates from a minority class (thought to be positive class), the related weight W_iiwill be set moderately bigger than others.

To expand the marginal distance and to limit the weighted cumulative error as for each example, an improvement issue scientifically is composed as

$$minimize: \|H\beta - T\|^2, \|\beta\| \tag{12}$$

Where T= [t_1,...,t_N]. More precisely,

$$minimize: L_{F_{ELM}} = \frac{1}{2}\|\beta\|^2 + \frac{1}{2}CW\sum_{i+1}^{N}\|\xi_i\|^2 \tag{13}$$

Subject to: h (x_i) β=t_i^T-ξ_i^T,i=1,…,N

Where h (x_i) is the element mapping vector in the concealed layer as x_i, β represents the yielded weight vector interfacing the concealed layer and output layer, and C is the regularization parameter to speak to the exchange off between the minimization of training errors and the boost of the marginal distance. ξ_i, the training error of testx_i, is caused by the distinction of the desired output t_iand the genuine yieldh (x_i) β.

MELM algorithm: The conventional extreme learning machines depend on the experimental hazard minimization rule and the training error minimization rule, whose downside is that it is probably going to experience the ill effects of overfitting, which diminishes the generalization capability thusly.

As per the statistical theory, the genuine dangers incorporate the experimental and auxiliary dangers, and a model with great speculation execution ought to have the capacity to adjust observational and basic dangers to acquire the best trade off. So we lead the auxiliary hazard minimization guideline into the ELM calculation and propose a modified weighted ELM.

Accept that an input and output test informational index for regression analysis isT = {(x_1, y_1)... (x_n,y_n)}, where $xi \in Rn$ and $y \in R$, $i = 1 \ldots l$. We draw into the state of the auxiliary hazard and alter the extent of the observational and basic dangers by ζ rather than the C in formula (13), and the optimization model of the ideal regression function can be built up as takes after:

$$\min\frac{1}{2}\|\beta\|^2 + \frac{1}{2}\zeta\sum_{i+1}^{N}\|\delta_i\|^2 \tag{14}$$

Subject to $\begin{cases} y_i - \beta_i h(x_i) = \delta_i \\ \delta_i \geq 0, i = 1,..,N \end{cases}$

Where δ_i^2, the addition of the square errors, speaks to the observational hazard and ‖β‖^2 speaks to the auxiliary hazard, as indicated by the maximal margin principle in statistical theory [31]. As indicated by formula (10), the formula above is the contingent extraordinary issue and can be changed into the Lagrange condition as takes after:

$$L_{ELM} = \frac{1}{2}\|\beta\|^2 + \frac{1}{2}\zeta W\sum_{i+1}^{N}\|\delta_i\|^2 - \sum_{i+1}^{l}\alpha_i[y_i - \beta h(x_i) - \delta_i] \tag{15}$$

Where the Lagrange multiplier α_iis the consistent factor of test x_iin the linear combination to form the ultimate decision function. Further, by making the halfway subsidiaries concerning factors (β,δ_i,α_i)all equivalent to zero, the KKT optimality conditions are gotten:

$$\begin{bmatrix} 0 & 1_v^T \\ 1_v & \Omega + \frac{I}{\zeta} \end{bmatrix}\begin{bmatrix} 0 \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix} \tag{16}$$

The$i$th row and the jth column data element is

$$\Omega_{i,j} = K_{ELM}(x_i, x_j) = h(x_i).h(x_j) = [G(a_1, b_1, x_i), \ldots, G(a_N, b_N, x_i)]. \tag{17}$$
$$[G(a_1, b_1, x_i), \ldots, G(a_N, b_N, x_i)]^T$$

Where G is the excitation function. The sigmoid function is utilized in this paper as follows:

$$G(a, b, x) = \frac{1}{(1 + \exp(-(a.x) + b))} \tag{18}$$

Solve the linear equations and after that get the accompanying nonlinear mapping condition beneath that is inferred:

$$f(x) = \sum_{i=1}^{N}\alpha_i K_{ELM}(x_i, x) \tag{19}$$

In view of the above MELM, the input features are prepared and the expectation comes about are put away in the framework.

4. Result and discussion

In this section, the proposed MELM based lung disease characterization has been assessed. The performance of proposed scheme is compared with existing RF and SCM networks. The segmentation accuracy and robustness were evaluated by using publicly available SIMBA dataset. The performance is evaluated by using given below matrices.

$$Accuracy = (TP + TN)/(TP + FN + TN + FP \tag{20}$$

$$Precision = TP/(TP + FP) \tag{21}$$

$$Recall = TP/(TP + FN) \tag{22}$$

$$\text{F-measure} = 2.\frac{Pr.Re}{Pr + Re} \tag{23}$$

$$\text{G-mean}=\sqrt{TP\_rate * TN\_rate} \quad (24)$$

Where at the whole image, let TP indicates the value of true positive rate of images, FP mentions the value of false positive rate of images, TNindicates the value of true negative rate of images and FN mentions the value of false negative rate of images.

$$Purity = \sum_i p_i (\max j \, p_{ij} / p_i) \quad (25)$$

$$Rand\ index = \frac{\left[\binom{n}{2} - \Sigma_i \binom{n_{i.}}{2} - \Sigma_j \binom{n_{.j}}{2} + 2\Sigma_{ij} \binom{n_{ij}}{2}\right]}{\binom{n}{2}} \quad (26)$$

$$Jaccard\ coefficient = \frac{\Sigma_{ij} \binom{n_{ij}}{2}}{\left[\Sigma_i \binom{n_{i.}}{2} + \Sigma_j \binom{n_{.j}}{2} - \Sigma_{ij} \binom{n_{ij}}{2}\right]} \quad (27)$$

Where p_{ij} is represented as segmentation point among i and j pixels, n_{ij} is the number of points among i and j. The step by step simulation results are shown in figure 2.

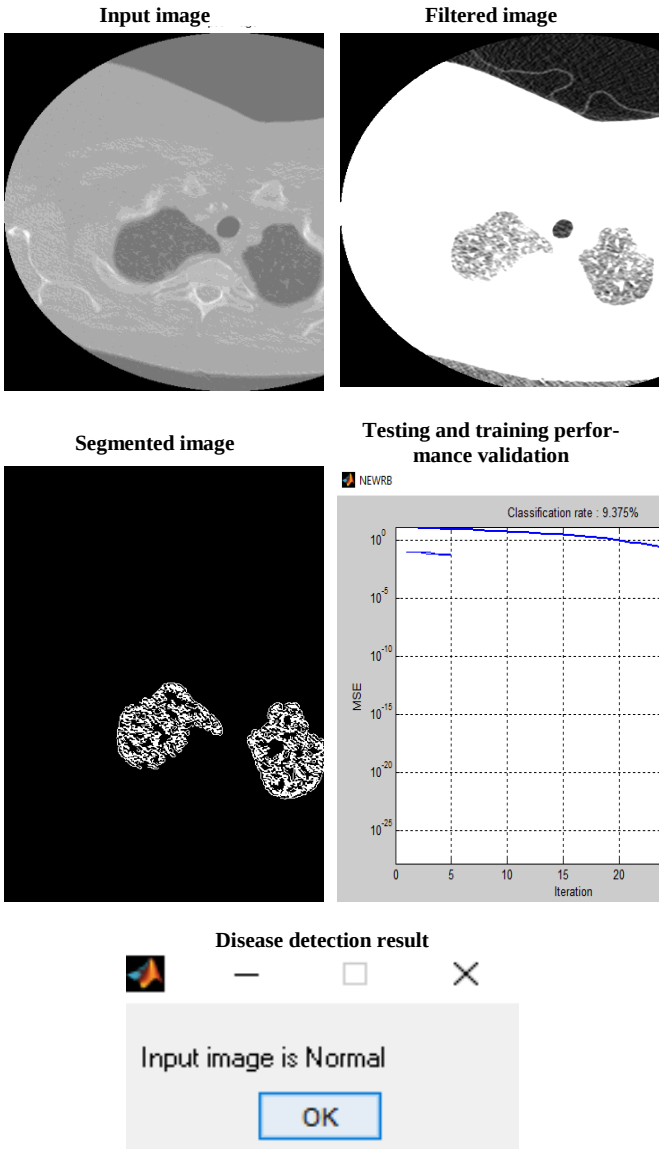

Fig. 2: Proposed MELM-SWT Based Results for Input Lung Image.

Table 1: The Performance Measure among Pre-Processing Schemes

Performance measures	Gaussian scale space (GSS) filter	Kaun filter	Proposed IKF
MSE	31.2232	31.1939	31.1717
PSNR	33.2034	33.2241	33.2272

The GSS, KF and IKF examination results has been appeared in table 1. It represents the IKF accomplished great outcomes contrasted than other two plans due with the effective streamlining. The execution has been estimated as far as PSNR and MSE esteems. It demonstrates the IKF accomplished great PSNR looked at than GSS and KF and in addition less MSE with great streamlined outcome.

The general exhibitions of all segmentation schemes graphical depiction is showed up in Figure 3. It demonstrates that the proposed SWT has achieved superior as far as accuracy rate of 92.13%, specificity rate of 91.05% looked at than existing WT and MGRF schemes. Because of the enhanced procedure of SWT, the segmentation accuracy has been expanded.

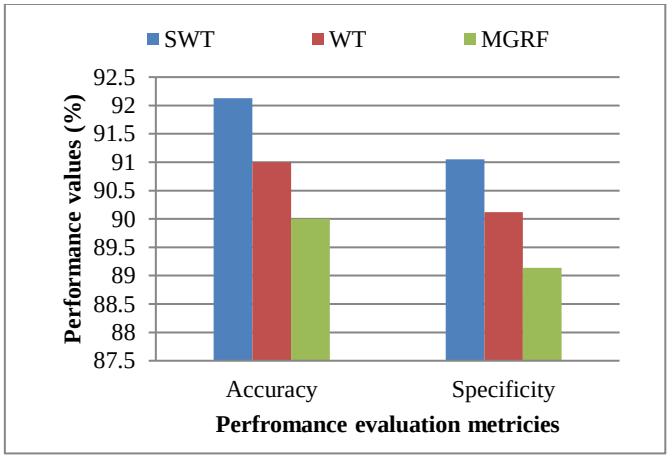

Fig. 3: Prediction of Overall Performances for All Lung Segmentation Schemes.

The accuracy execution of all lung illness acknowledgment designs graphical depiction is showed up in figure 4. From Figure 4, the precision of MELM is 2% and 2.2% extended than RF and SCM. Inferable from the valuable request diminishment and viable teeth region with advanced division with morphological post getting ready has been upgraded the accuracy result 87.2% of the MELM accomplished better results.

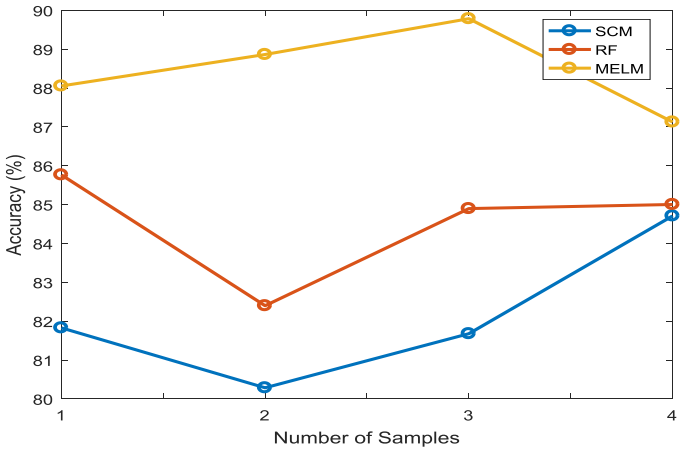

Fig. 4: Accuracy Performance Comparison among Different Lung Disease Classification Schemes.

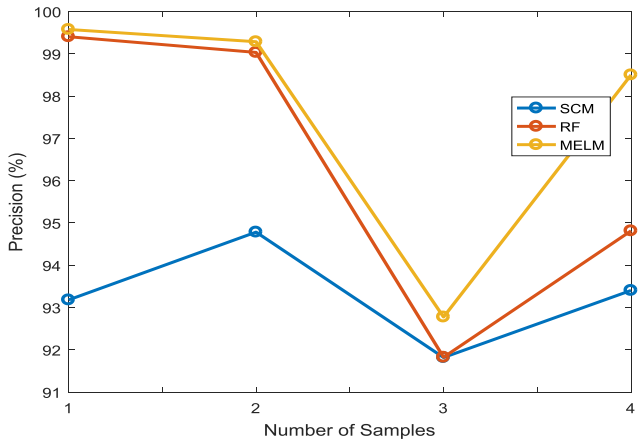

Fig. 5: Performance Comparison of Precision among Different Lung Disease Classification Schemes.

The precision execution of all lung disease recognition graphical depiction is showed up in figure 5. It shows the precision of MELM is 3.7% and 5.1% extended than RF and SCM. Inferable from high specificity and authentic positive rate, the precision of proposed MELM is extended as 98.5%. The MELM has been diminished the computational multifaceted nature thought about than machine learning calculations, in light of that the precision has been extended.

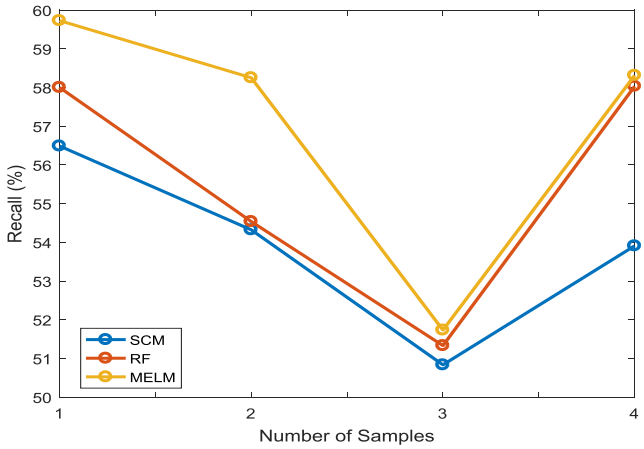

Fig. 6: Performance Comparison of Recall among Different Lung Disease Classification Schemes.

The recall execution of all lung disease recognition plans graphical depiction is showed up in figure 6. It shows the survey of MELM is 0.23% and 4.2% extended than RF and SCM. In light of the less error rate and high specificity regards, the review of proposed MELM is extended broke down than various procedures.

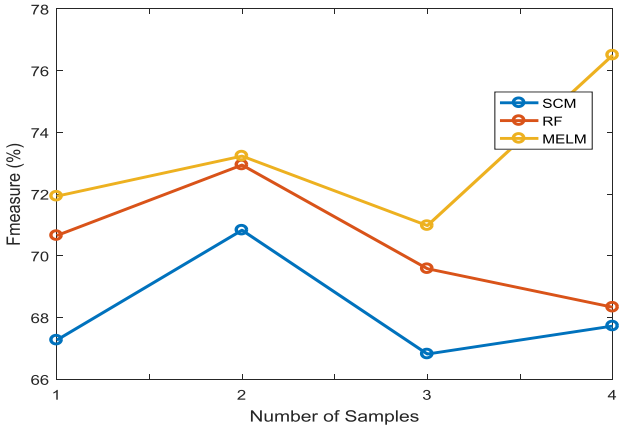

Fig. 7: Performance Comparison of F-Measure among Different Lung Disease Classification Schemes.

The F-measure execution of all lung disease recognition plans graphical portrayal is appeared in figure 7. It demonstrates the review of MELM is 8.4% and 9.2% expanded than RF and SCM. Because of the high exactness and review rate, the f-measure of proposed MELM is accomplished high esteem looked at than existing plans. The high positive rate and more positive rate of grouping has been enhanced the exhibitions of proposed conspire.

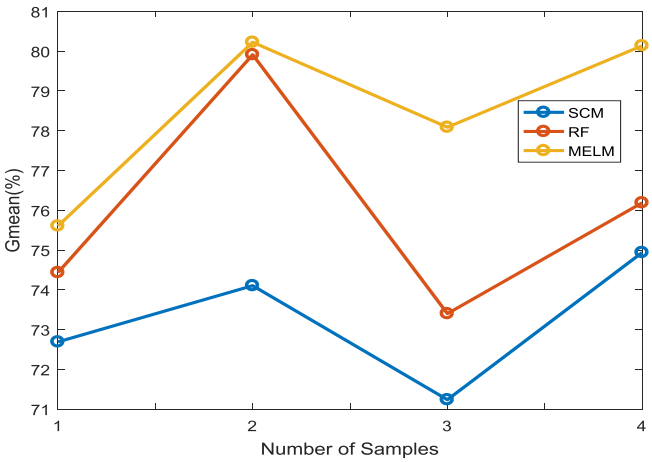

Fig. 8: Performance Comparison of G-Mean among Different Lung Disease Classification Schemes.

The G-mean execution of all lung disease recognition plans graphical portrayal is appeared in figure 8. It demonstrates the review of MELM is 4% and 5.2% expanded than RF and SCM. Due to the effectual segmentation and pre-processing schemes, the proposed scheme attained high g-mean compared than other algorithms.

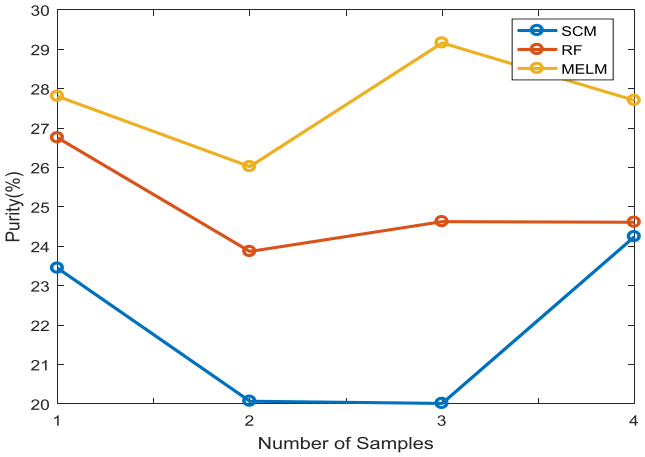

Fig. 9: Performance Comparison of Purity among Different Lung Disease Classification Schemes.

The purity execution of all lung disease recognition plans graphical portrayal is appeared in figure 9. It demonstrates the review of MELM is 3.3% and 4% expanded than RF and SCM. Because of the high exactness and review rate, the f-measure of proposed MELM is accomplished high esteem looked at than existing plans. The high positive rate and more positive rate of grouping has been enhanced the exhibitions of proposed conspire.

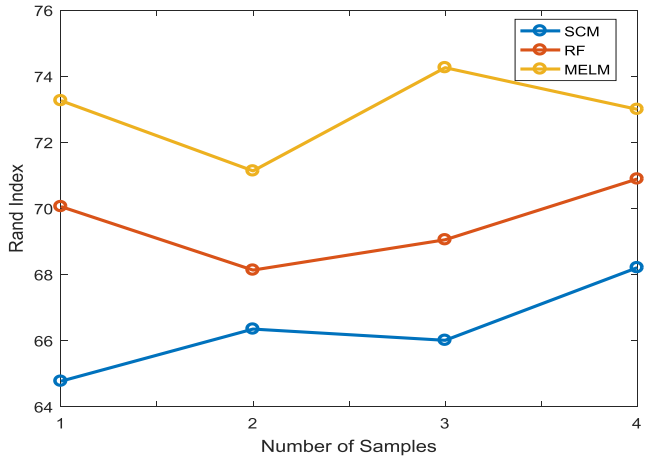

Fig. 10: Performance Comparison of Rand Index among Different Lung Disease Classification Schemes.

The rand intex execution of all lung disease recognition plans graphical portrayal is appeared in figure 10. It demonstrates the review of MELM is 2% and 5% expanded than RF and SCM. Due to the accurate segmentation point by using SWT, the proposed scheme attained high rand index.

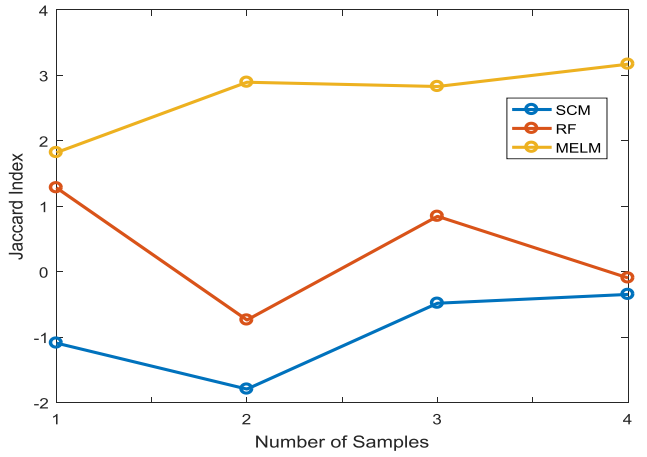

Fig. 11: Performance Comparison of Jaccard Coefficeint among Different Lung Disease Classification Schemes.

The jaccard coefficient execution of all lung disease recognition plans graphical portrayal is appeared in figure 11. It demonstrates the review of MELM is attained high jaccard coefficient compared than RF and SCM due to the effectual segmentation.

Table 2: Overall Performance Numerical Values for All DA Detection Methods

Performance matrices	Proposed MELM	RF	SCM
Accuracy	87.2	85	84.8
Precision	98.5	94.8	93.5
Recall	58.23	58	54.03
F-measure	76.8	68.4	67.8
G-mean	80	76.2	75
Purity	27.8	24.6	23.8
Rand index	73	71	68.2
Jaccard coeficient	3.2	0	-1.8

CT picture and their numerical evaluation regards are foreseen in Table 2. It exhibits that the proposed MELM is 0.23% and 31.23% extended than RF and SCM. Exactly when the amount of pictures assembles suggests the numerical evaluation results of proposed MELM moreover expands, in light of the valuable division and features.

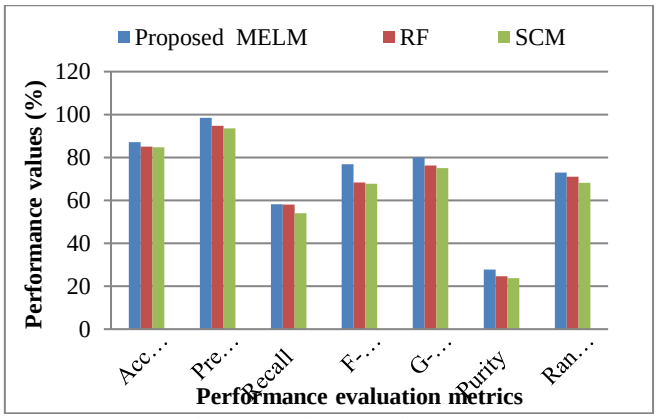

Fig. 12: Prediction of Overall Performances for All Lung Disease Detection Schemes.

The overall performances of all lung disease recognition plans graphical portrayal is appeared in figure 12. It shows that the proposed MELM has attained high performance in terms of accuracy rate of 87.2%, precision rate of 98.5%, recall rate of 58.23%, F-measure rate of 76.8%, G-mean rate of 80%, purity rate of 27.8%, rand index of 73 % compared than exiting RF and SCM schemes. Because of the effectual preprocessing and effective optimal segmentation with effectual features, the proposed scheme attained improved performance.

5. Conclusion

Accurate segmentation and classification of lung images, especially in the case of pathological lungs with, e.g., lung nodules, is essential for reliable CAD of pulmonary and lung diseases. Most segmentation errors are due to missing pathological tissues; however, these errors negligibly affect the overall segmentation accuracy because the pathological tissue volume is often very small compared to the whole lung volume. So, to solve the above problem MELM with SWT segmentation and classification algorithms are proposed in this paper. Initially, the input images are preprocessed to improve the quality of image. Then, the pathological lungs are segmented by using SWT for improving the classification accuracy. After that, the GLCM features are extracted for effectual classification. Finally, the lung disease accurately predicted. The experimental results show that the proposed MELM-SWT attained high performance compared than existing schemes. In future, some other features and segmentation schemes will focus to improve the classification accuracy.

References

[1] Siegel, R.L., Miller, K.D., Jemal, A.: 'Cancer statistics, 2015', CA Cancer J. Clin., 2015, 65, pp. 5–29 https://doi.org/10.3322/caac.21254.

[2] Girvin, F., Ko, J.P.: 'Pulmonary nodules: detection, assessment, and CAD', Am. J. Roentgenol., 2008, 191, pp. 1057–1069 https://doi.org/10.2214/AJR.07.3472.

[3] A. Mansoor et al.,"Segmentation and Image Analysis of Abnormal Lungs at CT: Current Approaches, Challenges, and Future Trends," Radio Graphics, vol. 35 no. 4 pp. 1056-1076, 2015. https://doi.org/10.1148/rg.2015140232.

[4] J. Wang et al., "Prediction of Malignant and Benign of Lung Tumor using a Quantitative Radiomic Method," Proceedings of International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 1272-1275, 2016. https://doi.org/10.1109/EMBC.2016.7590938.

[5] E. R. Gonzalez, and V. Ponomaryov, "Automatic Lung nodule segmentation and classification in CT images based on SVM," 9th International Kharkiv Symposium on Physics and Engineering of Microwaves, Millimeter and Submillimeter Waves (MSMW), pp.1-4, 2016.

[6] Li, X. X., Li, B., Tian, L. F., & Zhang, L. (2018). Automatic benign and malignant classification of pulmonary nodules in thoracic computed tomography based on RF algorithm. *IET Image Processing*, *12*(7), 1253-1264. https://doi.org/10.1049/iet-ipr.2016.1014.

[7] Rodrigues, M. B., Da Nóbrega, R. V. M., Alves, S. S. A., Rebouças Filho, P. P., Duarte, J. B. F., Sangaiah, A. K., & De Albuquerque, V. H. C. (2018). Health of Things Algorithms for Malignancy Level Classification of Lung Nodules. *IEEE Access*, *6*, 18592-18601. https://doi.org/10.1109/ACCESS.2018.2817614.

[8] Abdelrahman, S. A., & Abdelwahab, M. M. (2018). Accumulated grey-level image representation for classification of lung cancer genetic mutations employing 2D principle component analysis. *Electronics Letters*, *54*(4), 194-196. https://doi.org/10.1049/el.2017.1890.

[9] Chen, H., Xu, Y., Ma, Y.J., et al.: 'Neural network ensemble-based computeraided diagnosis for differentiation of lung nodules on CT images clinical evaluation', Acad. Radiol., 2010, 17, pp. 595–602 https://doi.org/10.1016/j.acra.2009.12.009.

[10] Chen, H., Zhang, J., Xu, Y., et al.: 'Performance comparison of artificial neural network and logistic regression model for differentiating lung nodules on CT scans', Expert Syst. Appl., 2012, 39, pp. 11503–11509 https://doi.org/10.1016/j.eswa.2012.04.001.

[11] Lin, P.-L., Huang, P.-W., Lee, C.-H., et al.: 'Automatic classification for solitary pulmonary nodule in CT image by fractal analysis based on fractional Brownian motion model', Pattern Recognit., 2013, 46, pp. 3279–3287 https://doi.org/10.1016/j.patcog.2013.06.017.

[12] Han, F., Wang, H., Zhang, G., et al.: 'Texture feature analysis for computeraided diagnosis on pulmonary nodules', J. Digit. Imaging, 2015, 28, pp. 99– 115 https://doi.org/10.1007/s10278-014-9718-8.

[13] Cheng, J.-Z., Ni, D., Chou, Y.-H., et al.: 'Computer-aided diagnosis with deep learning architecture: applications to breast lesions in US images and pulmonary nodules in CT scans', Sci. Rep., 2016, 6, pp. 24454–24466 https://doi.org/10.1038/srep24454.

[14] Dhara, A.K., Mukhopadhyay, S., Dutta, A., et al.: 'A combination of shape and texture features for classification of pulmonary nodules in lung CT images', J. Digit. Imaging, 2016, 29, pp. 466–475 https://doi.org/10.1007/s10278-015-9857-6.

[15] Liu, Y., Balagurunathan, Y., Atwater, T., et al.: 'Radiological image traits predictive of cancer status in pulmonary nodules', Clin. Cancer Res., 2017, 23, pp. 1442–1449 https://doi.org/10.1158/1078-0432.CCR-15-3102.

[16] Tajbakhsh, N., Suzuki, K.: 'Comparing two classes of end-to-end machinelearning models in lung nodule detection and classification: MTANNs vs. CNNs', Pattern Recognit., 2017, 63, pp. 476–486 https://doi.org/10.1016/j.patcog.2016.09.029.

[17] Ahmed Soliman, Fahmi Khalifa, Ahmed Elnakib, Mohamed Abou El-Ghar, Neal Dunlap, Brian Wang, Georgy Gimel'farb, Robert Keynton, and Ayman El-Baz, Accurate Lungs Segmentation on CT Chest Images by Adaptive Appearance-Guided Shape Modeling,

[18] Kubota, T., Jerebko, A.K., Dewan, M., et al.: 'Segmentation of pulmonary nodules of various densities with morphological approaches and convexity models', Med. Image Anal., 2011, 15, pp. 133–154 https://doi.org/10.1016/j.media.2010.08.005.

[19] Farag, A.A., Abd El Munim, H.E., Graham, J.H., et al.: 'A novel approach for lung nodules segmentation in chest CT using level sets', IEEE Trans. Image Process., 2013, 22, pp. 5202–5213 https://doi.org/10.1109/TIP.2013.2282899.

[20] Netto, S.M.B., Silva, A.C., Nunes, R.A., et al.: 'Automatic segmentation of lung nodules with growing neural gas and support vector machine', Comput. Biol. Med., 2012, 42, pp. 1110–1121 https://doi.org/10.1016/j.compbiomed.2012.09.003.

[21] Chen, K., Li, B., Tian, L.F., et al.: 'Vessel attachment nodule segmentation using integrated active contour model based on fuzzy speed function and shape-intensity joint Bhattacharya distance', Signal Process., 2014, 103, pp. 273–284 https://doi.org/10.1016/j.sigpro.2013.09.009.

[22] Sun, S.S., Guo, Y., Guan, Y.B., et al.: 'Juxta-vascular nodule segmentation based on flow entropy and geodesic distance', IEEE J. Biomed. Health Inf., 2014, 18, pp. 1355–1362 https://doi.org/10.1109/JBHI.2014.2303511.

[23] Messay, T., Hardie, R.C., Tuinstra, T.R.: 'Segmentation of pulmonary nodules in computed tomography using a regression neural network approach and its application to the lung image database consortium and image database resource initiative dataset', Med. Image Anal., 2015, 22, pp. 48–62 https://doi.org/10.1016/j.media.2015.02.002.

[24] Li, B., Chen, Q.L., Peng, G.M., et al.: 'Segmentation of pulmonary nodules using adaptive local region energy with probability density function-based similarity distance and multi-features clustering', Biomed. Eng. Online, 2016, 15, pp. 49–76 https://doi.org/10.1186/s12938-016-0164-3.

[25] Unser,M., "Texture Classification and Segmentation using Wavelet Frames," IEEE Trans. Image Processing,vol.4, no.9,1995 ,pp.1549-1560. https://doi.org/10.1109/83.469936.

[26] K. P. Aarthy and U. S. Ragupathy, "Detection of lung nodule using multiscale wavelets and support vector machine," International Journal of Soft Computing and Engineering (IJSCE), vol. 2, issue 3, July 2012.

[27] J. A. K. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural Processing Letters*, vol. 9, no. 3, pp. 293–300, 1999. https://doi.org/10.1023/A:1018628609742.

[28] J. A. K. Suykens, J. De Brabanter, L. Lukas, and J. Vandewalle, "Weighted least squares support vector machines: robustness and sparce approximation," *Neurocomputing*, vol. 48, no. 1–4, pp. 85–105, 2002. https://doi.org/10.1016/S0925-2312(01)00644-0.

[29] J. A. K. Suykens and J.Vandewalle, "Trainingmultilayer perceptron classifiers based on a modified support vector method," *IEEE Transactions on Neural Networks*, vol. 10, no. 4, pp. 907–911, 1999. https://doi.org/10.1109/72.774254.

[30] W. Zong, G.-B. Huang, and Y. Chen, "Weighted extreme learningmachine for imbalance learning,"Neurocomputing, vol.101, pp. 229–242, 2013. https://doi.org/10.1016/j.neucom.2012.08.010.

[31] N.-Y. Liang, G.-B. Huang, P. Saratchandran, and N. Sundararajan, "A fast and accurate online sequential learning algorithm for feedforward networks," IEEE Transactions on Neural Networks, vol. 17, no. 6, pp. 1411–1423, 2006. https://doi.org/10.1109/TNN.2006.880583.