ISSN 2278-3091
Volume 8, No.5, September - October 2019
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse52852019.pdf
https://doi.org/10.30534/ijatcse/2019/52852019

Classification of Chronic Obstructive Pulmonary Disease (COPD) Using Regression with Gabor Filtration and Random Forest Classification

V.Porkodi[1], Dr. S. Anbu Karuppusamy[2]
[1]Research Scholar, Department of Computer Science and Engineering, Shri Venkateshwara University, Gajraula, Amroha, Uttar Pradesh, India. Email Id: porkodisivaram@gmail.com
[2]Professor, Department of Electronics and Communication Engineering, Excel Engineering College, Namakkal, Tamilnadu, India. Email Id: anbuksamy@gmail.com

ABSTRACT

In this paper, a lung tissue classification using Regression with Gabor Filtration and Random Forest Classification (RGRFC) method was created. For classification of the lung dataset, the random forest model has been used. The assessment of Lung Tissue shows promising outcomes in classification. The study offers assessment over COPD datasets to classify between moderate, normal and abnormal smokers. The technique has been tested for its precision, sensitivity and specificity for COPD Datasets. The result shows that the proposed method achieves higher precision, sensitivity and classifier than other methods.

Key words: COPD, Gabor Filtration, Lung Tissue Classification, Random Forest, Regression.

1. INTRODUCTION

A significant cause of worldwide morbidity and mortality is Chronic Obstructive Pulmonary Disease (COPD) with a worldwide incidence of 11.7% [1], with an annual death rate of around 3 million. Currently COPD is the world's 4th leading cause of death, however it might be the 3rd by 2030 [2]. In the 30 years to come, the smoking prevalence is projected to increase, given the growing incidence of the developing and the populations of aging in high-income countries [3].

The natural history of COPD is characterized by exacerbations, which seems to accelerate the reduction of lung function [4] leading to decreased physical exercise, poorer standard of living [5], and enhanced risk of death. COPD policies are primarily based on reducing symptoms and future risks, i.e. aggravation, mortality and a decrease in lung function. Since 2011, COPD patients have been categorized in four different classifications (class A, B, C and D) under the Global Initiative for Chronic Obstructive Lung Disease (GOLD). In 2017, a GOLD update with resulting therapeutic consequences took place.

In prior GOLD 2011 patients were categorized according to symptoms and the risk of further occurrence and evaluated according to COPD evaluation test or modified Medical Research Council Dyspnea Scale (mMRC) [6].

In GOLD 2017 patients will not be affected by category-allocation forcible expiratory quantity in one second according to aggravated history and respiratory symptoms. The argued reason for the shift was: first to explain what is being assessed; secondly, while for patients with obstruction GOLD 3 or 4, the risk of exacerbation is considerably greater. FEV1 lacks enough accuracy to be used as a predictor, clinically; and thirdly, a history of past occurrences is the best predictor of future exacerbations [7]. However, the consequences of shift in COPD classification are missing from the present research.

A thorough examination of various literature [8] – [13] showed that current works focus on the algorithm of machine learning. Existing study method does not provide an efficient classification for unmarked parameters for the classification of lung tissues. Unlabeled Lung tissue spectral clustering needs efficient information training that takes time. Further current methods have the disadvantage of categorizing lung function tissues based on certain guidelines lagging behind in efficient classification. Therefore, a supervised classifier was built in this research article which is valid for efficient classification with unscheduled Lung Tissue information.

The research aimed at classification of evaluating COPD patients into different categories. In this study, lung tissue classification using RGRFC method is created. For the segmentation of Lung Tissue process widely used for Lung tissue segmenting and classification, the suggested RGRFC uses a stochastic threshold regression model. For classification of the lung picture dataset, the random forest model has been used. This operation is used primarily to disconnect the abnormal portion from the pulmonary picture. The assessment of Lung Tissue shows promising outcomes in classification.

2. RESEARCH METHODOLGY

The aim of this research to effectively perform the medical image processing for Lung Tissue classification. As per the review of existing report it is identified that cancer disease is severe threat to human begins life. Hence in this research concentrate on examining Lung Tissue and analyzing classification of image processing especially lung cancer. In cancer detection dataset are occurred from COPD dataset and then it undergoes further processing. The data for the Lung Tissue classification are evaluated using proposed RGRFC approach (figure 1).

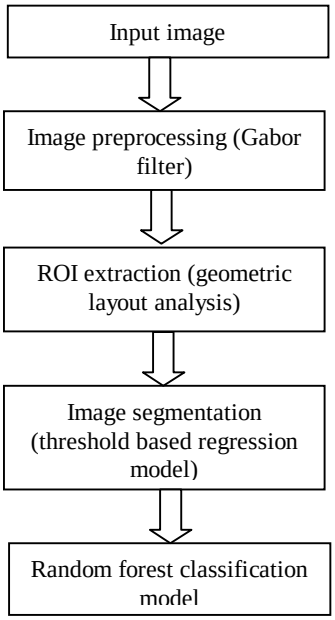

Figure 1: Proposed Architecture

2.1 Feature Extraction

For a pulmonary tissue, the feature extraction model G is given as follows

$$G=\left\{G_1^M,G_1^C,....,G_i^M,G_i^C,....,G_{14}^M,G_{14}^C,i\in N_+,i[1,14]\right\} \quad (1)$$

G_i^M is the average i[th] feature with computation is given as follows:

$$G_i^M=\frac{\sum_{d\in D} g_i^d}{n}, D=\left\{0^0,45^0,90^0,135^0\right\} \quad (2)$$

G_i^C is the variance of i[th] feature with formulation is estimated as below:

$$G_i^C=\frac{\sum_{d\in D}\left(g_i^d-G_i^M\right)^2}{n}, D=\left\{0^0,45^0,90^0,135^0\right\} \quad (3)$$

where,
n – total selected angles (0◦, 45◦, 90◦, 135◦) and n = 4.
d - direction angle.
g_i^d - texture of ith feature from a specific angle $g^d=\left\{g_1,.....,g_i,....,g_{14}\right\}$ is the parameteric set with Gabor features.

The Gabor feature parameteric computation is estimated based on statistical image analysis with pixel intensity elements obtained through GLCM [14]. In this situation, Pixel strength represent the pixel gray space dependency and the gray level correlation rule not only quantitatively but it defines a two-dimensional distribution of a gray level on a specific local structure. Consequently, a measurement of texture function parameters includes acquisition of likelihood parameters of pixel frequency. It defines the estimation of the likelihood parameters of pixel size.

The neighbor gray level pixel are considered to be m pixels away from the central pixel direction and its incidence number Set P(i,j) to calculate in every direction its pixel intensity.

The gray level probability of (i,j) pixel is estimated as p(i,j) = P(i,j)/R, where total number of occurrences of gray level on a specific direction. Further, p(i,j) represents the (i,j)th pixel intensity input element.

$$p_x(i)=\sum_{j=1}^{N_g} p(i,j) \quad (4)$$

Similarly, j[th] input value py(j) of the marginal-probability matrix about Y is considered as a total number of element available in p(i,j)

$$p_y(j)=\sum_{i=1}^{N_g} p(i,j) \quad (5)$$

The two joint probability density functions are as below:

$$p_{x+y}(k)=\sum_{i=1,i+j=k}^{N_g}\sum_{j=1}^{N_g} p(i,j), k=2,3,...,2N_g \quad (6)$$

$$p_{x-y}(k) = \sum_{i=1,|i-j|=k}^{N_g} \sum_{j=1}^{N_g} p(i,j), k = 0,1,..., N_g - 1 \quad (7)$$

Algorithm 2: Lung Tissue size

Input: N denoted as nodule volume

Maximum Lung Tissue radius is stated as rmax, maximum nodule radius.

Specified parameters are P, Q, V.

Output: M is output vector with sample matrix P, Q, V for N.

$D \leftarrow$ Generate lung inner centers P.

$r \leftarrow r_{\max} / Q$

$n = 1 \rightarrow p, q = 1 \rightarrow Q \; do$

$r_{cur} \leftarrow r \times q$

$circle \leftarrow build\_circle(C_n, r_{cur})$

$k = 1 \rightarrow k \; do$

$circle \leftarrow build\_circle(C_n, r_{cur})$

$M_{pqv} \leftarrow linear\_\text{int}(N, circle);$

Return M;

2.2 Data Classification using RGRFC

The proposed method are validated for data testing and validation performance. Tissues in COPD tissues are optimized based on pixel intensity based on nodule size and classification. Lung Tissue classification performed via random selection with nodule radius and sorting.

The proposed RGRFC uses RF classifier for Lung Tissue classification. RF is a kind of algorithm that makes ensemble learning. RF puts together several of poor learners and uses decision trees for the creation of strong learners. The word random in RF is regarded as the random feature. Next, RF extracts the original data collection by averaging and bagging by way of bootstrap into random samples. Bagging utilizes 2/3 of the initial training data set and 1/3 for screening. A standard definition of instances of substitution is used to retrieve random samples from the learning dataset.

Untapped decision trees were established on each collected dataset after the compilation of the random samples. Random features are used to describe the optimal break in each tree node rather than using all the features. Trees work separately to achieve their own effects in the woods. The final outcome of the woods emerges from the big vote on the outcome of all decision trees.

The lower OOB oblivion rate indicates the RF's improved performance. In particular, RF mistakes are normally related in accordance with couple predominant facts: (i) The kindred about anybody couple forest selection trees then (ii) the choice plant strength. The greater inter-treatment the more complicated the arrangement concerning the RF leads, who raises the OOB carelessness rate. The better thriving decision chain, concerning the mean side, the lower the OOB confusion rate. Thus the RF is quintessential agreement optimum depth and minimal interaction are in imitation of be achieved. The RF outturn may additionally stay increased by using optimizing the variety of decision trees because of the woodland and the number regarding lamely applications ancient in imitation of decide the superior cut up over each node. The initial RF as much a law establishes the number on around traits so the foundation concerning the quantity number about characteristics. Depending concerning the OOB carelessness rate, the variety regarding timber into the woodland might also keep picked.

Using the Bootstrap resampling process, several samples are derived from original samples and sub data sets are created, then the sub dataset forms and trains the core decision tree. RF was implemented into Decision tree learning in random attribute collection for each node of the decision chain, initially from the node attribute attributes in a random K attribute subset choice and from the subset, an ideal node splitting attribute was chosen which can discriminate between the decision trees and maximize process heterogeneity. Eventually, the effects of the rating by voting method are shown in the flow chart in Figure 2, in order to improve classification efficiency.

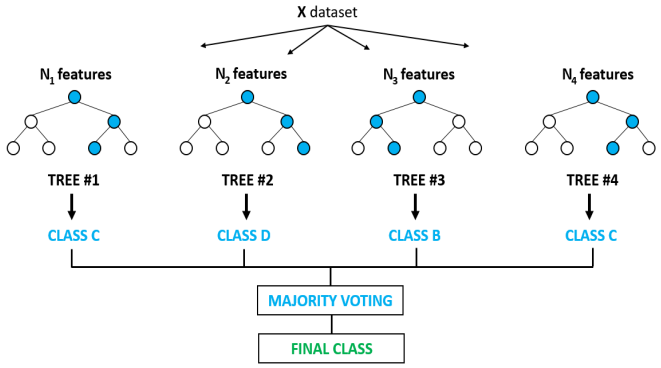

Figure 2: Random forest algorithm flowchart

The central stage of the RF algorithm is node separation. The nodes can be separated into a whole decision tree. The branch formation of each tree is based on certain laws of branching. The guidelines comprise specifically the total data profit, the highest gain frequency and the minimum gain factor. A separated value is then selected. Based on its section, the branch of the decision tree rises. The consistency of the node decreases as a consequence of the partitioning process,

ensuring that the node test is in the same class as much as possible.

2.3 Data Classification

First randomly appear upon with L pairs over black parameter (wi,bi); then according according to the input then the randomly got here on along parameters, using footsie IV to count the individual stolen bed output matrix H. Finally, the usage of footsie 6 in conformity with tell the same output poise vector β.

Algorithm 3: Lung Tissue Classification
1. Input: Number of hidden layers: L
Class data: N = {(xj,tj)|xj ∈Rn,tj∈Rm, j = 1, 2, ...,N};
2. Output: Three parameters of ELM: w,b,β;
3. for i = 1 to L do
4. Randomly select hidden layer parameter (wi,bi);
5. Calculate the single layer output matrix H;
6. Calculate the output weight vector β = H*T;
7. Return (w,b,β);

When the usage of enter some then certain group regarding pulmonary nodule CT dataset, such begins the feature extraction in accordance after Algorithm 1. According in conformity with the price about w and b arrived from the education process, reckon the odd stolen seam casting H. Finally, in accordance after f(x) = h(x)β = h(x)H*T gain the prognosis result beyond unknown result image.

3. COMPARISON OF CLASSIFICATION PERORMANCE

The final objective of our suggested clustering algorithm is to distinguish benign and malignant tissues. We contrasted the Random Forest in combination with threshold and Gabor algorithms with the current classifiers for the classification of benign and malignant tumor clusters recorded in order to validate the effectiveness of RGRFC on Lung Tissue Diagnostics against LDAs, SVM and ELM.

A few parameters must be learned in order to enhance classification efficiency before the performance comparison. Fivefold cross validation [15] is used for selecting the RGRFC parameters. You can use the highest value of c = 2 and g = 0.0313. The cluster nodes are set to 2000 in the RGRFC. For these studies, we extract 745 nodular samples. The information is on average split into five groups. For the training of classifiers, 4 folds are used and the remaining fold is used as test information, so loop 5 times.

In order to assess classification efficiency, precision is regarded. The classification performance of the various classifiers is shown in Figure 3 – Figure 5. The results of accuracy, sensitivity and specificity shows that the lung tissue dataset classification is accurate than other methods.

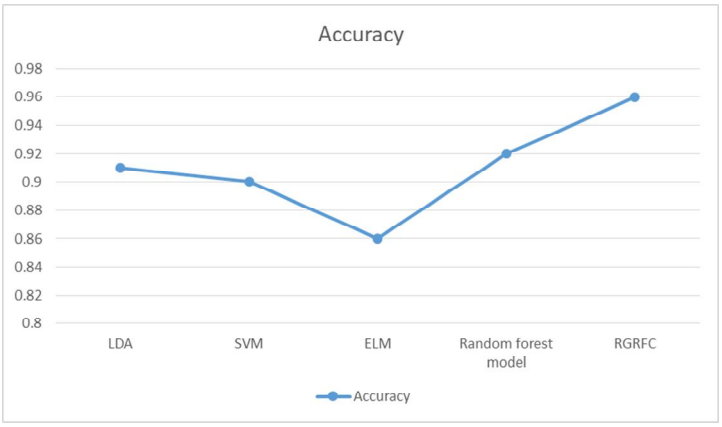

Figure 3: Results of Accuracy

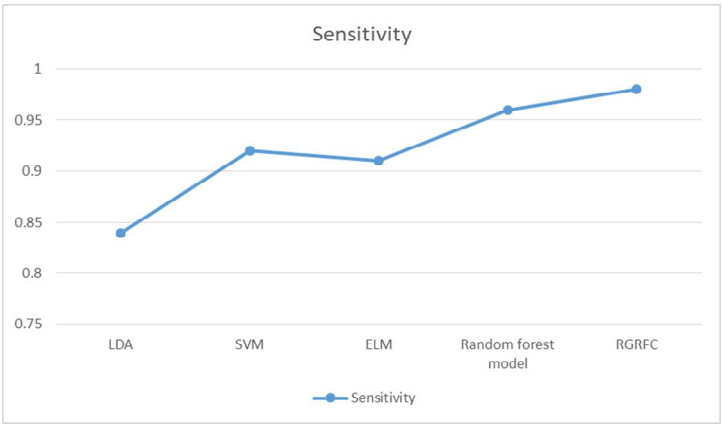

Figure 4: Results of Sensitivity

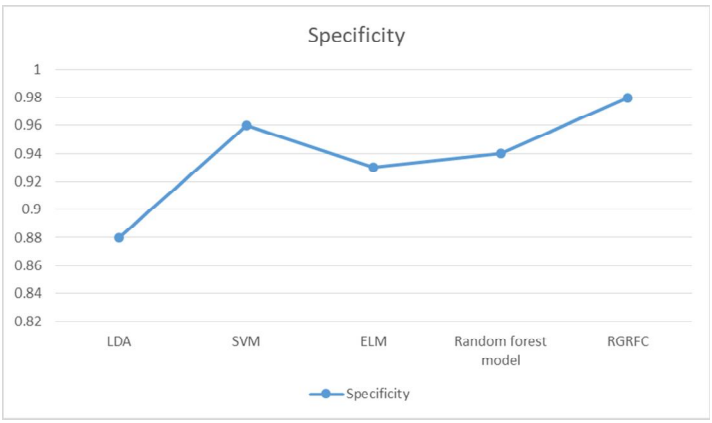

Figure 5: Results of Specificity

4. CONCLUSION

In this study, RGRFC method is designed for classification of Lung Tissue which is affected with smoking behavior. For

classification of the lung dataset, the random forest model has been used. The assessment of Lung Tissue shows promising outcomes in classification. The study offers assessment over COPD datasets to classify between moderate, normal and abnormal smokers. The technique has been tested for its precision, sensitivity and specificity for COPD Datasets. The result shows that the proposed method achieves higher precision, sensitivity and classifier than other methods. Experimental analysis of COPD shows that the proposed method achieves highest accuracy of 83%, training accuracy of 89% and testing accuracy of 87%.

REFERENCES

1. Adeloye, D., Chua, S., Lee, C., Basquill, C., Papana, A., Theodoratou, E., & Chan, K. Y. (2015). **Global and regional estimates of COPD prevalence: Systematic review and meta–analysis**, *Journal of Global Health*, 5(2). https://doi.org/10.7189/jogh.05.020415
2. Lozano, R., Naghavi, M., Foreman, K., Lim, S., Shibuya, K., Aboyans, V., & AlMazroa, M. A. (2012). **Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010,** *The Lancet*, 380(9859), 2095-2128. https://doi.org/10.1016/S0140-6736(12)61728-0
3. Lopez, A. D., Shibuya, K., Rao, C., Mathers, C. D., Hansell, A. L., Held, L. S., ... & Buist, S. (2006). **Chronic obstructive pulmonary disease: current burden and future projection**, *European Respiratory Journal*, 27(2), 397-412.
4. Donaldson, G. C., Seemungal, T. A., Bhowmik, A., & Wedzicha, J. A. (2002). **Relationship between exacerbation frequency and lung function decline in chronic obstructive pulmonary disease**, *Thorax*, 57(10), 847-852. https://doi.org/10.1136/thorax.57.10.847
5. Seemungal, T. A., Donaldson, G. C., Paul, E. A., Bestall, J. C., Jeffries, D. J., & Wedzicha, J. A. (1998). **Effect of exacerbation on quality of life in patients with chronic obstructive pulmonary disease**, *American Journal of Respiratory and Critical Care Medicine*, 157(5), 1418-1422.
6. Vestbo, J., Hurd, S. S., Agustí, A. G., Jones, P. W., Vogelmeier, C., Anzueto, A., ... & Stockley, R. A. (2013). **Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary**, *American Journal of Respiratory and Critical Care Medicine*, 187(4), 347-365.
7. Soriano, J. B., Lamprecht, B., Ramírez, A. S., Martinez-Camblor, P., Kaiser, B., Alfageme, I.,& de-Torres, J. P. (2015). **Mortality prediction in chronic obstructive pulmonary disease comparing the GOLD 2007 and 2011 staging systems: a pooled analysis of individual patient data**, *The Lancet Respiratory Medicine*, 3(6), 443-450.
8. Song, Y., Cai, W., Zhou, Y., & Feng, D. D. (2013). **Feature-based image patch approximation for lung tissue classification**, *IEEE Transactions on Medical Imaging*, 32(4), 797-808. https://doi.org/10.1109/TMI.2013.2241448
9. Depeursinge, A., Sage, D., Hidki, A., Platon, A., Poletti, P. A., Unser, M., & Muller, H. (2007, August). **Lung tissue classification using wavelet frames**, *in 2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, (pp. 6259-6262).
10. Anthimopoulos, M., Christodoulidis, S., Ebner, L., Christe, A., & Mougiakakou, S. (2016). **Lung pattern classification for interstitial lung diseases using a deep convolutional neural network**, *IEEE Transactions on Medical Imaging*, 35(5), 1207-1216. https://doi.org/10.1109/TMI.2016.2535865
11. Schlegl, T., Ofner, J., & Langs, G. (2014, September). **Unsupervised pre-training across image domains improves lung tissue classification**, *in International MICCAI Workshop on Medical Computer Vision*, (pp. 82-93).
12. Depeursinge, A., Racoceanu, D., Iavindrasana, J., Cohen, G., Platon, A., Poletti, P. A., & Müller, H. (2010). **Fusing visual and clinical information for lung tissue classification in high-resolution computed tomography**. *Artificial Intelligence in Medicine*, 50(1), 13-21. https://doi.org/10.1016/j.artmed.2010.04.006
13. Depeursinge, A., Iavindrasana, J., Cohen, G., Platon, A., Poletti, P. A., & Müller, H. (2008, June). **Lung tissue classification in HRCT data integrating the clinical context**, *in 2008 21st IEEE International Symposium on Computer-Based Medical Systems* (pp. 542-547).
14. Haralick, R. M., Shanmugam, K., & Dinstein, I. H. (1973). **Textural features for image classification**, *IEEE Transactions on Systems, Man, And Cybernetics*, 3(6), 610-621. https://doi.org/10.1109/TSMC.1973.4309314
15. Byra, M., Dobruch-Sobczak, K., Piotrzkowska-Wróblewska, H., & Nowicki, A. (2017). **Added value of morphological features to breast lesion diagnosis in ultrasound**, a*rXiv preprint arXiv*, 1706.01855.