LOGISTIC REGRESSION SCORING INDEX FOR DETECTION OF INTERSTITIAL LUNG DISEASE

Norliza, M. Noor[1], Ahmad F. Roseli[2], Zuriani Dahlan[3], Rosminah M. Kassim[4], Ashari Yunus[5] and Omar M. Rijal[2]
[1]Razak Faculty of Technology and Informatics, Universiti Teknologi Malaysia, Kuala Lumpur Campus, Malaysia
[2]Institute of Mathematical Sciences, University of Malaya, Kuala Lumpur, Malaysia
[3]UTM Space, Universiti Teknologi Malaysia, Kuala Lumpur Campus, Malaysia
[4]Department of Diagnostic Imaging, Kuala Lumpur Hospital, Malaysia
[5]Institute of Respiratory Medicine, Kuala Lumpur Hospital, Malaysia
E-Mail: norliza@utm.my

ABSTRACT

The decision for a lung transplant requires the knowledge of the severity of interstitial lung disease (ILD). A manual scoring sheet was developed for discriminating ILD cases from non-ILD cases based on eleven indicators selected by the radiologist. The manual scoring sheet requires the radiologists to score the severity of general ILD from visual inspection on the changes of lung tissue using the high resolution computed tomography images. This paper investigates the use of logistic regression in developing a scoring index, $\pi(x)$, to investigate the presence of ILD. The result shows that the scoring index, $\pi(x)$, is robust and has discriminatory potentials. The threshold (c, d) having values (0.7, 0.24) allowed correct classifications of 80% of ILD cases and 61% of non-ILD cases in a total sample of 134 cases. This study suggests that $\pi(x)$ may be used as a simple and practical scoring index in initial investigations for ILD detection.

Keywords: logistic regression, scoring index, interstitial lung disease, detection, confidence probability.

INTRODUCTION

Lungs are among the important organs in the respiratory system. In Malaysia, respiratory disease is the second highest (10.36%) of the principal causes of hospitalization and the second highest for death (19.48%) [1]. A group of lung diseases known as the interstitial lung disease (ILD) has serious consequences on the patients whereby lung transplant may be the only solution to the patients. However, the decision making for lung transplant is non-trivial due to scarcity of lung donors in Malaysia. Important decisions must be made once the ILD has been detected.

The ILD refers to the group of diseases related to the disorder of the lung structure at the level of interstitium, which may affect not only the tissues, but also the air sacs, blood vessels and distal airways [2]. The worst consequence of ILD is death due to respiratory failure [2]. Known causes of ILD include inhalation of some agents known to be harmful to the lung such as asbestos and tobacco smoke, effect of drugs, exposure to radiation and genetic abnormalities. However the causes for certain categories of ILD remain unknown, for example idiopathic pulmonary fibrosis and sarcoidosis [3].

Non-ILD cases include but are not limited to pulmonary tuberculosis, bronchiectasis, bullous disease, lung infection, and collapse of lung. Most of the non-ILD cases can be treated and the patients may recover from the disease.

A suspected ILD patient will generally have his lungs investigated with high resolution computed tomography (HRCT) Thorax images by a radiologist. ILD has various important features or indicators visible on the HRCT Thorax image, where the radiologists visually evaluate each indicator when differentiating ILD cases from non-ILD cases. Due to visual similarity on some of ILD cases and non-ILD cases, some of the indicators may be difficult to differentiate.

Scoring indices has been widely used for disease assessment, prediction and detection [4]-[12]. In some of these studies logistic regression is used to evaluate these scores [4, 7, 8, 9, 10]. In this study, a novel application of the logistic regression model proposes a scoring index $\pi(x)$, which not only indicates the presence or absence of ILD but also gives a degree of certainty by the observer. The selection of indicators when discriminating ILD cases from non-ILD cases uses a simple statistical graphical method [13, 14, 15] and standard procedures when applying logistic regression.

METHODOLOGY

Data collection

The HRCT thorax images of 27 ILD cases and 54 non-ILD cases were taken from the Department of Diagnostic Imaging at the Kuala Lumpur Hospital. The equipment used to obtain the HRCT thorax image is the Siemens Somaton Plus 4 [16]. An accepted practice is to study thirty slices of HRCT thorax images taken at 1 cm interval [17]. In this study, the scoring system was developed from five predetermined slices of HRCT thorax based on information about the anatomy of the lung, namely; aortic arch (level 1), tracheal carina (level 2), pulmonary hilar (level 3), pulmonary venous confluence (level 4), and 1-2cm above the dome of right hemi-diaphragm (level 5), that represent the upper, upper-middle, middle, lower-middle and lower zone of the lung

ARPN Journal of Engineering and Applied Sciences

area. The five levels selected are considered as sufficient in representing the whole lung. After the selection of slices, the radiologists viewed the HRCT Thorax images and recorded the severity of eleven indicators, namely reticular pattern (RP), plural plaques (PP), pleural thickening (PT), fissural thickening (FT), peribronchovascular interstitial thickening (PBIT), bronchial dilation (BD), intra parenchyma nodule (IP nodule), septal and non-septal lines (SNSL), traction bronchiectasis (TRB), ground glass opacity (GGO) and parenchyma consolidation (PC) in the scoring sheets shown in Table-1 (Appendix I). The degree of severity ranges from grade 0 (indicate absent) to grade 3 (indicate very severe). At the end of the score sheet, the radiologist makes a summary of the evaluation and state the presence or absence of ILD and the distribution of disease.

Bivariate plots and logistic regression models investigate the severity scores when seeking indicators of ILD. A discrimination procedure using a scoring index was then developed for the detection of ILD.

Bivariate plots

Bivariate plots involving pairs of indicators may yield two clusters or groups of individuals who may be ILD cases or non-ILD cases. All pair-wise combinations of indicators were investigated. At each level, the degree of severity for an indicator is defined as the sum of the degree of severity for both left and right lung. Each patient, or observation, is therefore represented by a two-dimensional point consisting of the severity of two indicators. Three dimensional or higher dimensional graph were not considered due to the discrete nature of the data.

Logistic regression

The application of 2D plots is compared to a more systemic procedure or method for finding the most probable indicator. A possible solution is the use of logistic regression model where standard software, in particular S-Plus [18] is easily available for use. Define the variable Y such that Y is equal to one indicates ILD present and Y is equal to zero indicates non-ILD present. Further, let x_1, x_2, …, x_{11} represent the scores from the score sheet for the eleven indicators. Denote the expected value of Y for a given x value as $E(Y|x)$. For binary response variable, let $\pi(x) = E(Y|x)$, and the logistic regression model takes the form,

$$\pi(x) = \frac{e^{g(x)}}{1+e^{g(x)}} \qquad (1)$$

A transformation of $\pi(x)$ that is important is the logit transformation, namely,

$$g(x) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots\ldots + \beta_n x_n \qquad (2)$$

The estimates of α, β_1,….., β_n were obtained from S-Plus, giving an estimate of $g(x)$, and consequently an estimate of $\pi(x)$. The expression for $\pi(x)$ in Equation (1) provides the conditional probability that Y is equal to 1 given x. This will be denoted as $P(Y=1\,|\,x)$. After having fit a particular single/multiple (single variable/multivariable) logistic regression model, the process of model assessment is carried out using the following test statistics;

$$G = -2\,ln\left(\frac{\text{likelihood without the variable}}{\text{likelihood with the variable}}\right).$$

Under the hypothesis that a given parameter, say β_1, is equal to zero, the statistic G follows a chi square distribution with one degree of freedom. A detail explanation of the use of G is given in Hosmer and Lemeshow [19].

RESULT

Inferences from bivariate plot

Many possible bivariate plots can be considered. For this investigation a plot of SNSL versus RP, at level 3, (Figure-1) detected ten ILD cases and 34 non-ILD cases. The remaining undetected cases were then investigated with a plot of PBIT versus TRB at level 3 (Figure-2) where eight non-ILD cases were detected. The undetected 17 ILD cases and 12 non-ILD cases were finally investigated with GGO and PBIT, at level 5, (Figure-3) where seven ILD cases and eight non-ILD cases were detected. Using this method, a total of 17 ILD cases (63%) and 50 non-ILD cases (92%) were detected.

The graphical method suggested that only five from the eleven indicators have a discriminatory capability. However, this graphical method is very likely to yield a different result with the use of other combination of indicators. Therefore, there is a need to investigate if indeed the other six indicators can be eliminated. This is more so since this graphical method is tedious, time consuming and the plots are not easily comparable. As such a more systematic approach, in the form of applying a logistic regression model will be considered next.

VOL. 14, NO. 5, MARCH 2019 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

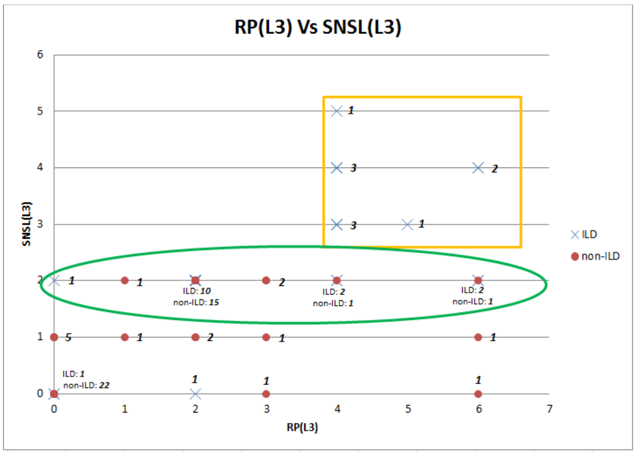

Figure-1. Investigating RP and SNSL as indicators of ILD.

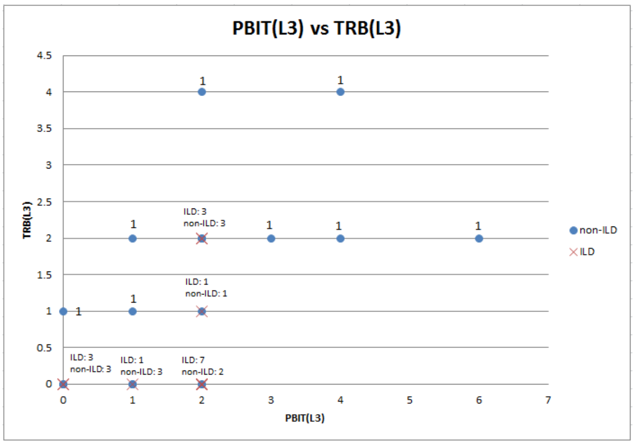

Figure-2. Graph PBIT(L3) Vs TRB(L3).

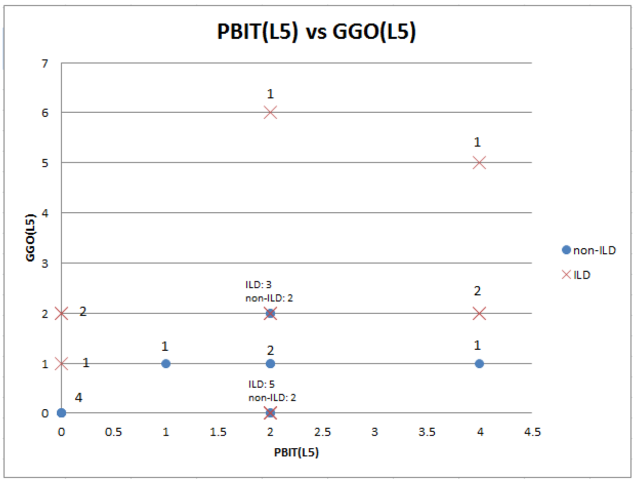

Figure-3. Investigating PBIT and GGO as indicators of ILD.

Inferences from logistic regression model

In anticipation of a large number of parameters to be estimated, the total sample size were increased to 35 ILD cases and 99 non-ILD cases. A control data set consisting of 25 ILD cases and 30 non-ILD cases were used for parameter estimation, in particular obtaining an estimate of $\pi(x)$. A test set consisting of 10 ILD cases and 69 non-ILD cases were used to investigate two threshold values in deciding ILD and non-ILD cases. In particular;

If $\pi(x) > c$ then ILD is detected,

and if $\pi(x) < d$, a non-ILD is detected,

Otherwise, if $d < \pi(x) < c$, no decision is made.

The logistic regression model using RP and PT from level 3, as the indicator gave the best result. The parameter estimates are $\hat{\alpha} = -1.99$, $\hat{\beta}_1 = 0.3874$, $\hat{\beta}_2 = 1.075$. Using the threshold values c = 0.7 and d = 0.24, the discrimination results are shown in Table-2.

Table-2. Discrimination properties of $\pi(x)$ using RP and PT from level 3 with $c = 0.7$ and $d = 0.24$.

ILD (n=10)			Non-ILD (n=69)		
correct classification $\hat{\pi} > c$	**Misclassification $\hat{\pi} < d$**	**No decision $d < \hat{\pi} < c$**	**correct classification $\hat{\pi} < d$**	**Misclassification $\hat{\pi} > c$**	**No decision $d < \hat{\pi} < c$**
8 (80%)	1 (10 %)	1 (10 %)	42 (60.87%)	6 (8.7%)	21 (30.43%)

DISCUSSIONS

Indicators from the graphical method differ from those obtained from the logistic regression method. The former method arbitrarily chooses indicators but the logistic regression method performed the same task systematically in particular, RP and PT, were the optimal choices. This probably explains why the bivariate plots and the logistic regression model detected 63% and 80% of the ILD-cases, respectively. However, the logistic regression model performed worse in detecting non-ILD cases, which is most likely due to the choice of the threshold - *d*. Since the main goal of this study is the

VOL. 14, NO. 5, MARCH 2019 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

detection of ILD-cases, distributional properties of $\pi(x)$ is now considered.

Distribution of $\pi(x)$ for ILD and non-ILD cases

The potential of using $\pi(x)$ as a scoring index in this study is further considered with the distribution of $\pi(x)$ and the robustness of $\pi(x)$. Figure-4 shows the distribution of $\pi(x)$ values for ILD and non-ILD cases using box-plots. The distributions of ILD and non-ILD shows that the median are far apart with some overlapping at the tail of non-ILD.

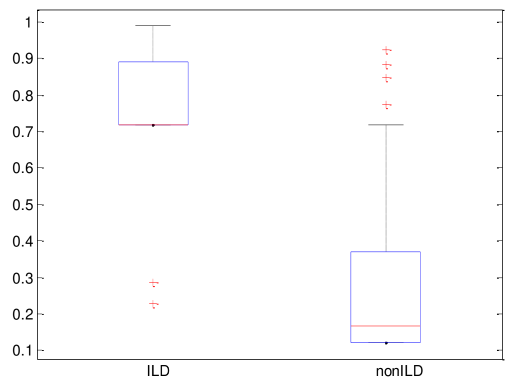

Figure-4. Box plot of $\pi(x)$ values for ILD and non-ILD cases.

Non-parametric test for quality of median

Both samples of $\pi(x)$ values for ILD and non-ILD cases do not have the univariate normal distribution. As an additional indicator of group separation, the two sample Wilcoxon rank sum test investigate equality of medians.

Let $x_i = \pi_i(x)$ values for ILD, $i = 1,2,......,10$, and $y_j = \pi_j(x)$ values for non-ILD $j = 1,2,......,69$. Let, $N = 79$. These N values test, H_0: median of x observation equal to median of y observation, versus, H_1: medians are not equal.

Arrange all 79 $\pi(x)$ values in ascending order and ranked the smallest observation with number one and the largest observation with number 79.

Wilcoxon rank sum statistic, W = 655, is the sum of the rank of the small sample.

If both samples come from the same continuous distribution, then W has:

Mean, $\mu_w = \dfrac{n_1(N+1)}{2} = \dfrac{10(79+1)}{2} = 400$

Standard deviation, $\sigma_w = \sqrt{\dfrac{n_1 n_2(N+1)}{12}} = \sqrt{\dfrac{(10)(69)(79+1)}{12}} = 67.823$

where n_1 is sample size for small group, i.e. group of $\pi(x)$ value for ILD.

We reject H_0 if the value test statistic, W far from its mean, μ_w.

$$Z = \frac{W - \mu_w}{\sigma_w} = \frac{655 - 400}{67.823} = 3.76$$

Since Z greater than 1.96, the hypothesis for equality of median may be rejected, suggesting again significant separation of ILD from non-ILD. This is further supported by the p-value being less than 0.001.

The box-plots and Wilcoxon rank sum test strongly suggest significant separation for the ILD and non-ILD cases and in particular, the thresholds c = 0.7 and d = 0.24 give reasonably good discrimination results.

Robustness of $\pi(x)$

A leave-one-out cross validation method was applied to investigate the robustness of $\pi(x)$ with respect to variation of indicator values from different cases. Since the estimates of $\pi(x)$ were derived from the control data set consisting of 55 cases, the leave-one-out method was carried out as follows;

Firstly, a case (control data) is withdrawn and the parameter α, β_1 and β_2 are estimated with the remaining 54 cases. Using the current model, 79 estimates of $\pi(x)$ was obtained from each of the test cases. The mean and box plot of these 79 estimates were obtained. Secondly, reinsert the first case and withdraw another case from the control data set and repeat the whole process again. Continue reinserting and withdrawing until 55 means and 55 box plots were obtained. The robustness of $\pi(x)$ is reflected by the sampling properties of the 55 estimates of $\pi(x)$ with consistent means of $\hat{\pi}(x)$ (see Figure-5), and similar box-plots of $\hat{\pi}(x)$ values (see Figure-6).

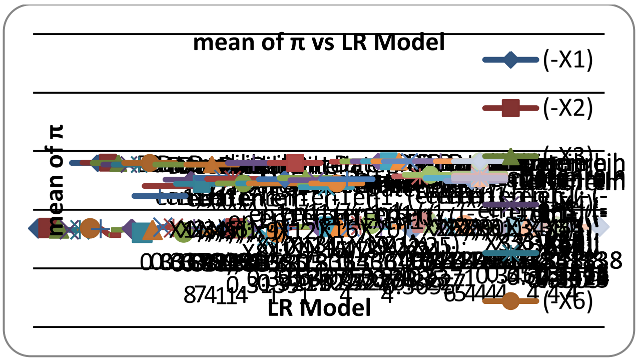

Figure-5. Average of $\hat{\pi}(x)$. The left most point is the average of 79 $\hat{\pi}(x)$ values when the first case in the control data set was removed.

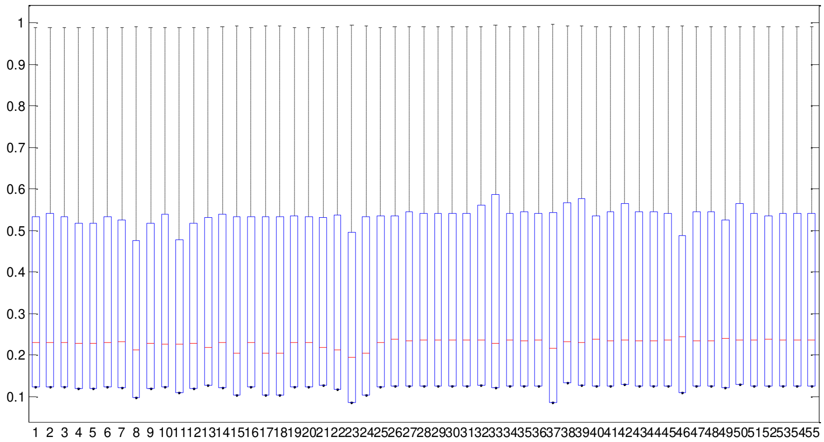

Figure-6. Variations of box plots of $\hat{\pi}(x)$. The left most box plot is the distribution of 79 $\hat{\pi}(x)$ values when the first case in the control data set was removed.

CONCLUSIONS

The statistic $\pi(x)$ from the logistic regression model may be used as a scoring index in initial investigations of ILD. This scoring index is robust and its distribution for ILD and non-ILD cases is significantly different, strongly suggesting that it can be used with confidence. This study shows that the logistic regression model using RP and PT gave the best discrimination result. The threshold c = 0.7 implies that the radiologist classified 80% of ILD cases with a confidence probability of at least 0.70. Similarly, the threshold d = 0.24 shows a confidence probability of at least 0.76 for detecting 61% of non-ILD cases in a total sample of 134 cases. However, no generalization may be made at this stage as to the use of $\pi(x)$ as the scoring index since more samples of ILD need to be studied and the experiments should be repeated with other radiologists.

ACKNOWLEDGMENT

We acknowledge the contribution from Dato' Dr. Abdul Razak bin Abdul Muttalif, The Insitute of Respiratory Medicine, Kuala Lumpur, and Dr. Hajjah Zaleha binti Abd. Manaf, Head of Radiology Department (Diagnostic Imaging), Kuala Lumpur Hospital. This work was supported by Ministry of Education Malaysia, Universiti Teknologi Malaysia and University of Malaya.

REFERENCES

[1] Health Facts. 2012. Ministry of Health Malaysia, July 2012, p. 7.

[2] Ryu JH, Daniels CE, Hartman TE, Yi ES. 2007. Diagnosis of Interstitial Lung Diseases. Mayo Clinic Proceedings. 82(2): 976.

[3] American Thoracic Society, European Respiratory Society. American Thoracic Society/European Respiratory Society International Multidisciplinary Consensus Classification of the Idiopathic Interstitial Pneumonias. This joint statement of the American Thoracic Society (ATS), and the European Respiratory Society (ERS) was adopted by the ATS board of directors, June 2001 and by the ERS Executive Committee, June 2001. Am J Respir Crit Care Med 2002; 165: 277-304.

[4] Gibbons RJ, Lee KL, Pryor D, Harel FE Jr, Coleman RE, Cobb FR, Rosati RA, Jones RH. 1983. The use of radionuclide angiography in the diagnosis of coronary artery disease - a logistic regression analysis. Circulation. 68(4): 740-746.

[5] Zimran A, Gross E, West C, Sorge J, Kubitz M, Beutler E. 1989. Prediction of severity of Gaucher's disease by identification of mutation at DNA level. The Lancet. 334(8659): 349-352.

[6] Benacerraf BR, Neuberg D, Bromley B, Frigoletto FD Jr. 1992. Sonogrphic scoring index for prenatal detection of chromosomal abnormalities. J Ultrasound in Medicine. 11(9): 449-458.

[7] Le Gall JR, Klar J, Lemeshow S, Saulnier F, Alberti C, Artigas A, Teres D. 1996. The logistic organ dysfunction system: A new way to assess organ dysfunction in the intensive care unit. J Am Med Assoc. 276(10): 802-810. doi:10.1001/jama.1996.03540100046027

[8] Budoff MJ, Georgiou D, Brody A, Agatston AS, Kennedy J, Wolfkiel C, Stanford W, Shields P, Lewis RJ, Janowitz WR, Rich S, Brundage BH. 1996. Ultrafast computed tomography as a diagnosis

modality in the detection of coronary artery disease: A multicenter study. Circulation. 93: 898-904.

[9] Tierney MC, Szalai JP, Snow WG, Fisher RH, Nores A, Nadon G, Dunn E, St. George-Hyslop PH. 1996. Prediction of probable Alzheimer's disease in memory-impaired atients: A prospective longitudinal study. Neurology. 46: 661-665.

[10] Pollito 2n JM, Childs B, Mellits ED, Tokayer AZ, Harris ML, Bayless TM. 1996. Crohn's disease: Influence of age at diagnosis on site and clinical type of disease. Gastreonterology. 111(3): 580-586.

[11] Oga T, Nishimura K, Tsukino M, sato S, Hajiro T. 2003. Analysis of the factors related to mortality in chronic obstructive pulmonary disease: Role of exercise capacity and health status. Am J Respir Crit Care Med, 167:544-549. doi:10.1164/rccm.200206-583OC.

[12] Fowkes FGR, Murray GD, Butcher I, Heald CL, Lee RJ, Chambless LE, Folsom AR, Hirsch AT, Dramaix M, deBacker G., Wautrecht J-C, Kornitzer M, Newman AB, Cushman M, Sutton-Tyrrell K, Lee AJ, Price JF, d'Agostino RB, Murabito JM, Norman PE, Jamrozik K, Curb JD, Masaki KH, Rodríguez BL, Dekker JM, Bouter LM, Heine RJ, Nijpels G, Stehouwer CDA, Ferrucci L, McDermott MM, Stoffers HE, Hooi JD, Knottnerus JA, Ogren M, Hedblad B, Witteman JC, Breteler MMB, Hunink MGM, Hofman A, Criqui MH, Langer RD, Fronek A, Hiatt WR, Hamman R, Resnick HE, Guralnik J, McDermott MM. 2008. Ankle brachial index combined with Framingham risk score to predict cardiovascular events and mortality: A meta-analysis. J. Am Med Assoc. 300(2): 197-208. doi:10.1001/jama.300.2.197

[13] Johnson RA, and Wichern DW. 2007. Applied Multivariate Statistical Analysis, 6th ed. New Jersey: Pearson International Edition.

[14] Andrews, DF. 1972. Plots of high dimensional data. Biometrics. 28: 125-136.

[15] Mardia KV, Kent JT and Bibby JM. 1979. Multivariate Analysis, London: Academic Press.

[16] 1997. Somaton Plus 4, Siemens, Forchheim, Germany.

[17] Sluimer IC, van Waes PF, Viergever MA. 2003. Computer-aided diagnosis in high resolution CT of the lungs. Med Phys. 30: 3081-3090.

[18] S-Plus Software, TIBCO Software Inc.

[19] Hosmer D W, Lemeshow S. and Sturdivant R.X. 2013. Applied Logistic Regression, New York: Wiley Publisher.

APPENDIX I

Table-1. Score Sheet, Level 1 For Degree of Severity For Eleven Features Changes/Signs Of ILD.

Patient's ID:		**Reader's Code:**	
Zone : Level 1 (Aortic Arch)		Slice No:	
Changes/Signs of ILD	**Score** (Please tick ☑ the alphanumeric that represent the extent of the disease)	Right Lung	Left Lung
1. Coarseness score for reticular pattern (RP)	Grade 0: Absent Grade 1: Fine intralobular fibrosis predominating Grade 2: Microcystic pattern with airspace less than 3 mm in diameter Grade 3: Larger cysts 3-6 mm in diameter	☐ ☐ ☐ ☐	☐ ☐ ☐ ☐
2. Circumscribed pleural plaques (PP)	Grade 0 : Absent or 0mm Grade 1 : <5mm Grade 2 : 5-10mm Grade 3 : >10 mm	☐ ☐ ☐ ☐	☐ ☐ ☐ ☐
3. Pleural thickening (PT)	Grade 0 : Absent Grade 1: Minimal Grade 2 : Moderate Grade 3 : Extensive	☐ ☐ ☐ ☐	☐ ☐ ☐ ☐
4. Extent of fissural thickening (FT)	N/A : Absence of fissures Grade 0 : Absent or pencil thin line < 1mm Grade 1 : Slight thickening 1 - 2 mm Grade 2 : Significant thickening > 2mm	☐ ☐ ☐ ☐	☐ ☐ ☐ ☐
5. Peribronchovascular interstitial thickening (PBIT)	Grade 0 : Absent Grade 1 : Mild Grade 2 : Moderate Grade 3 : Severe	☐ ☐ ☐ ☐	☐ ☐ ☐ ☐
6. Severity of bronchial dilation (BD)	Grade 0 : Absent Grade 1 : Mild Grade 2 : Slightly bigger than adjacent vessel Grade 3 : Severe dilatation	☐ ☐ ☐ ☐	☐ ☐ ☐ ☐
7. Intra parenchyma nodules (IP nodule)	Grade 0 : Absent Grade 1 : Micro nodule size less than 7mm Grade 2 : Small nodule size 7 - 10mm Grade 3 : Large nodule size more than 10mm	☐ ☐ ☐ ☐	☐ ☐ ☐ ☐
8. Septal and non-septal line (SNSL)	Grade 0 : Absent Grade 1 : Fine networks of lines 0.8-1.5mm thick separated by a distance of 2 to 3mm Grade 2 : Network of 1 to 3 mm thick septa of intermediate sized lattice elements separated by a distance of 6 to 10mm Grade 3 : Network of 1 to 5mm thick lines delineating large spaces of 15 -25mm in diameter arranged in a polyhedral pattern, around a vessel of 1 -3mm diameter.	☐ ☐ ☐ ☐	☐ ☐ ☐ ☐
9. Traction bronchiectasis (TRB)	Grade 0 : Absent Grade 1: Minimal Grade 2 : Moderate Grade 3 : Extensive	☐ ☐ ☐ ☐	☐ ☐ ☐ ☐
10. Ground Glass Opacity(GGO)	Grade 0 : Absent Grade 1: Minimal or Not normal Grade 2 : Moderate Grade 3 : Extensive	☐ ☐ ☐ ☐	☐ ☐ ☐ ☐
11. Parenchyma consolidation (PC)	Grade 0 : Absent Grade 1 : Minimal or questionable parenchymal consolidation Grade 2 : Marked attenuation of air spaces relative to surrounding parenchyma Grade 3 : Definite parenchymal consolidation	☐ ☐ ☐ ☐	☐ ☐ ☐ ☐