Journal of
Applied Sciences

science
alert

ANSI*net*
an open access publisher
http://ansinet.com

Selecting a Variable for Predicting the Diagnosis of PTB Patients
From Comparison of Chest X-ray Images

[1]Omar Mohd. Rijal, [2]Norliza Mohd. Noor and [1]Shee Lee Teng
[1]Institute of Mathematical Science, University Malaya, Lembah Pantai, 59000 Kuala Lumpur, Malaysia
[2]Department of Electrical Engineering, Science and Technology College, UTM City Campus,
Universiti Teknologi Malaysia, Jalan Semarak, 54100 Kuala Lumpur, Malaysia

Abstract: A statistical method of comparing two digital chest radiographs for Pulmonary Tuberculosis (PTB) patients has been proposed. After applying appropriate image registration procedures, a selected subset of each image is converted to an image histogram (or box plot). Comparing two chest X-ray images is equivalent to the direct comparison of the two corresponding histograms. From each histogram, eleven percentiles (of image intensity) are calculated. The number of percentiles that shift to the left (NLSP) when second image is compared to the first has been shown to be an indicator of patients' progress. In this study, the values of NLSP is to be compared with the actual diagnosis (Y) of several medical practitioners. A logistic regression model is used to study the relationship between NLSP and Y. This study showed that NLSP may be used as an alternative or second opinion for Y. The proposed regression model also show that important explanatory variables such as outcomes of sputum test (Z) and degree of image registration (W) may be omitted when estimating Y-values.

Key words: Comparing chest radiograph, pulmonary tuberculosis, monitoring, logistic regression

INTRODUCTION

With the number of Pulmonary Tuberculosis (PTB) cases on the rise (Cruez, 2005; WHO, 2006), greater awareness of the disease and the need for better monitoring approach is essential. Reports by WHO (2004), the global targets for tuberculosis control are to cure 85% of the sputum smear-positive cases detected and to detect 70% of the estimated new sputum smear-positive cases.

Unnikrishnan and Jagannatha (2002) uses the definition, monitoring is a program of supervision designed to monitor the health and activities of an afflicted person. It includes ensuring compliance with an appropriate prescribed course of medication or medical therapy and with the recommendations and orders of the expert doctors. However, in this study, monitoring is defined as the investigation of a patients response to treatment by comparing a series of his chest radiographs.

The study by Rijal *et al.* (2006) applies the seven control point registration (SCPR) method and a resizing technique before the comparison of images is carried out. The correlation R_F^2 defined from the ULFR model (Rijal *et al.*, 2006; Dolby, 1976; Fuller, 1987) was used as a measure of quality of image registration. In this study the variables W is defined as follows;

$$
\begin{aligned}
&\text{If} && R_F^2 \leq 0.5 \text{ then } W = 1, \\
&\text{otherwise} && 0.5 < R_F^2 \leq 0.7 \text{ then } W = 2, \\
&\text{else} && W = 3.
\end{aligned}
$$

Figure 1 shows three box-plots corresponding to three histograms of the appropriate chest radiograph image of the PTB infected area of the same patient taken at three consultation time points. The left-ward shift of the
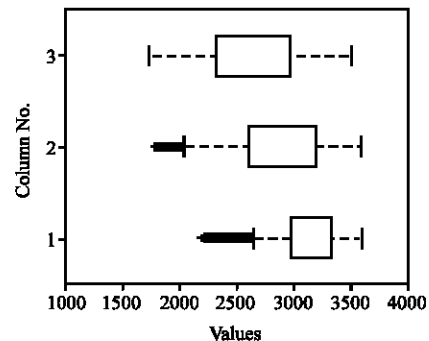

Fig. 1: Lowest box-plot corresponds to the histogram of chest radiograph of the PTB infected area from the first consultation. Similarly top-most box-plot is for the last consultation visit

Corresponding Author: Norliza Mohd. Noor, Department of Electrical Engineering, College of Science and Technology,
UTM City Campus, Universiti Teknologi Malaysia, Jalan Semarak, 54100 Kuala Lumpur, Malaysia
Tel: 603-2615-4589, 6019-327-4854   Fax: 603-2963-4844

histogram (from bottom to top) motivates the definition of NLSP which is the number of percentiles that shifted to the left (for example, second image compared to the first).

After registration and resizing (Hill *et al.*, 2001; Gonzalez *et al.*, 2004; Zitova and Flusser, 2003), a subset of two images, say

$$\{I\ (i,\ j);\ i = 1,...,M,\ j = 1,...,N\}$$

and

$$\{K\ (i,\ j);\ i = 1,...,M,\ j = 1,...,N\}$$

were obtained such that $I(\cdot,\cdot)$ and $K(\cdot,\cdot)$ represent identical regions at two different time point, say t and t+1.

Henceforth eleven percentile of the image intensity values for each of $I(\cdot,\cdot)$ and $K(\cdot,\cdot)$ were calculated and compared. For example if the median of intensity values for $K(\cdot,\cdot)$ is less than the corresponding median for $I(\cdot,\cdot)$, this suggests that the image histogram has shifted to the left. This leftward shift suggests a decrease in intensity of snowflakes and hence in turn indicate a positive response to treatment. The total number of percentile shifting to the left is defined as NLSP. Estimate for NLSP was obtained from a computer aided diagnostic for monitoring PTB (CAD-M-PTB) system (Rijal *et al.*, 2006).

Selection of hospital: The data or images were collected randomly from three types of hospitals recommended by Mahyuddin (2006). They are the District Hospital Tanjung Karang (HTK), Kuala Selangor, Hospital Tengku Ampuan Rahimah (HTAR), Klang and The Institute of Respiratory Medicine (IPR), Hospital Kuala Lumpur. The HTK and HTAR were selected because

*   They have a dedicated chest clinic. The hospitals were located in areas of high occurrence of PTB cases and therefore a high number of PTB cases can be obtained easily.
*   The medical officers are experienced in the PTB cases.
*   The hospitals have their own particular PTB treatment days. We may obtain and get latest information about the progress of the PTB patients accordingly.

A total of 5 volunteer medical officers were involved in this study. The hospital visits were limited as follows due to time consideration:

    Week 1: HTK
    Week 2: HTAR
    Week 3 and 4: IPR

MATERIALS AND METHODS

The experiment: In this experiment, medical records (patient wallet) at each hospital are randomly selected. Then for each selected patient wallet/file, the medical officer uses the standard procedure of direct visual interpretation of x-ray and analyses the information (e.g., sputum test) from patient's file, (Shaban, 2005). The same patient will then be subjected to the diagnostic procedure of the CAM-PTB system where the statistic NLSP was calculated.

Only the confirmed PTB cases (37 from IPR, 29 from HTK and 16 from HTAR) are used for data analysis. The data is analyzed separately because IPR is a referral centre to which other hospital like HTAR, must report to. HTK is a smaller district hospital where respiratory specialist from HTAR will do their routine weekly visit.

Each doctor was required to make one of four responses in a survey form which are as follow:

*   Patient's chest x-ray Worse, y = 1
*   Patient's chest x-ray No change, y = 2
*   Patient's chest x-ray Improvement, y = 3
*   Patient's chest x-ray Stable, y = 4

Increasing y-values indicated good progress of treatment. We relabeled y = 1, y = 2 into category 0 and y = 3, y = 4 into category 1. The combination of dependent variable y into two categories helps reduce the complication of interpreting the four responses.

As stated in the introduction section, the NLSP values (X) ranged from 0 to 11.

To model the relationship between y and x, a simple logistic regression function is then proposed. According to Menard (2002) and Hosmer and Lemeshow (1989), it has become, in many fields, especially in the health sciences, the standard method of analysis when studying the relationship between a (binary) discrete response/dependent variable and one or more explanatory variables.

Since the medical doctor incorporates information from the patients file before confirming his diagnosis, this study also consider the relationship of sputum test results, Z with that of Y. Finally the effect of image registration, W, is also investigated.

Brief review of logistic regression: The logistic regression model has been used in statistical analysis for many years. It does not assume linearity of relationship between the independent variables and the dependent, does not require normally distributed variables, does

not assume homoscedasticity and in general has less stringent requirements (Menard, 2002). A simple logistic regression function, Model (A), is given as follows (Hosmer and Lemeshow, 1989; Neter *et al.*, 1989):

$$E\{Y_i\} = \pi_i = \frac{\exp(\beta_0 + \beta_1 X_i)}{1 + \exp(\beta_0 + \beta_1 X_i)} \tag{1}$$

where, $E\{Y_i\}$ is viewed as a conditional mean of the doctor's response, given the value of NLSP (X_i). Both parameters β_0 and β_1 will be estimated by using maximum likelihood estimation.

The logarithm of the likelihood functions is given as:

$$\log_e L(\beta_0, \beta_1) = \sum_{i=1}^{n} Y_i (\beta_0 + \beta_1 X_i) - \sum_{i=1}^{n} \log_e \left[1 + \exp(\beta_0 + \beta_1 X_i) \right]$$

To maximize the likelihood function, we take partial derivatives with respect to β_0 and β_1, set these equal to zero, replace β_0 and β_1 with b_0 and b_1, respectively, yields the likelihood equations:

$$\sum_{i=1}^{n} Y_i - \sum_{i=1}^{n} \frac{\exp(b_0 + b_1 X_i)}{1 + \exp(b_0 + b_1 X_i)} = 0 \tag{2}$$

$$\sum_{i=1}^{n} Y_i X_i - \sum_{i=1}^{n} \frac{X_i \exp(b_0 + b_1 X_i)}{1 + \exp(b_0 + b_1 X_i)} = 0 \tag{3}$$

Since the Eq. 2 and 3 are nonlinear in the parameter estimates b_0 and b_1, a numerical solution specifically designed for logistic regression in MATLAB software program is used to estimate the parameters. Once the maximum likelihood estimates b_0 and b_1 are found, the fitted response function can be obtained as follows:

$$\hat{\pi}_i = \frac{\exp(b_0 + b_1 X_i)}{1 + \exp(b_0 + b_1 X_i)} \tag{4}$$

To obtain the multiple logistic regression model, we simply replace $\beta_0 + \beta_1 X$ in Eq. 1 by $\beta_0 + \beta_1 X_i + \beta_2 Z_i$, where for example Z_i is the value of sputum test results. The multiple logistic response function, Model (B), is then given as:

$$E\{Y\} = \frac{\exp(\beta'V)}{1 + \exp(\beta'V)} \tag{5}$$

where $\quad \beta_{3\times1} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_3 \end{bmatrix}$ and $V_{3\times1} = \begin{bmatrix} 1 \\ X \\ Z \end{bmatrix}$

The procedure for estimating the regression parameters in multiple regression is the same as explained in simple logistic regression.

Goodness-of-fit test for model (A): As the case for all regression models, the aptness of the logistic regression model needs to be checked before it is accepted for use. Since our data is in the ungroup binary category, two common methods: Hosmer-Lemeshow statistic and analogues of the R^2-statistic are suggested by Hosmer and Lemeshow (1989) and Collet (1952) to summarize model adequacy. However, Imon (2007) commended that our sample size is small in three hospitals, respectively, thus the Hosmer-Lemeshow statistic may not be appropriate for testing the accuracy of the fitted model. Only the analogues of the R^2-statistic is then applied to evaluate the fit of the model.

Several analogues to the linear regression R^2 have been proposed for logistic regression. For general use, Nagelkerke (1991) recommended the statistic R_N^2 and Menard (2002) suggested the statistic R_L^2.

For statistic R_N^2, it is defined as:

$$R_N^2 = \frac{1 - \left(\dfrac{L_O}{L_M} \right)^{\frac{2}{n}}}{1 - (L_O)^{\frac{2}{n}}} \tag{6}$$

where, L_O is the likelihood function for the model that contains only the intercept, L_M is the likelihood function that contains all the predictors (NLSP or sputum results) and n is the total number of binary observations (tested PTB patients) that make up the data base.

For statistic R_L^2, it is defined as:

$$R_L^2 = \frac{(\ln L_O - \ln L_F)}{(\ln L_O - \ln L_F) + (\ln L_F - \ln L_S)}$$

$$= \frac{(\ln L_O - \ln L_F)}{(\ln L_O - \ln L_S)} \tag{7}$$

where, L_O is the log-likelihood function for the model that contains only the intercept, L_F is the log-likelihood function that contains all the predictors (NLSP or sputum results) and L_S is the log-likelihood function from the saturated model. In Eq. 8, the value of the log-likelihood L_S could be easily obtained from the definition of deviance as below:

$$D = -2\ln \left(\frac{L_F}{L_S} \right) = -2\{\ln L_F - \ln L_S\} \tag{8}$$

$$\ln L_S = \ln L_F + \frac{1}{2}D \qquad (9)$$

Both R_N^2 and R_L^2 is then multiplied by 100 to show the percentage of accuracy the model predict the dependent variable.

Comparing linear logistic models: In examining the effect of including terms Z in Model (B), it is important to recognize the change in deviance (Neter *et al.*, 1989; Collet, 1952);

Let

$$D_1 = -2\left\{\ln \hat{L}_{c1} - \ln \hat{L}_S\right\}$$

$$D_2 = -2\left\{\ln \hat{L}_{c2} - \ln \hat{L}_S\right\}$$

and

$$D_e = D_1 - D_2 = -2\left\{\ln \hat{L}_{c1} - \ln \hat{L}_{c2}\right\} \qquad (10)$$

where, $\hat{L}_{c1}$ is the maximum likelihood under Model (A), $\hat{L}_{c2}$ is the maximum likelihood under Model (B), given in Eq. 5 and $\hat{L}_S$ is the maximum likelihood under the saturated model. It is known that $D_e \sim \chi^2 (d_{D1} - d_{D2})$ where d_{D1} and d_{D2}

is the degree of freedom, respectively for model (A) and model (B). If $D_e < \chi^2 (d_{D1} - d_{D2})$ for a given level of significance ($\alpha = 0.05$), the Model (A) and Model (B) are considered similar, in other words choose the simpler model.

If the reduction in deviance is highly significant, the term Z should be added in model (A). Otherwise, we may ignore the variable.

The leave-one-out method (LOU): Since the sample size is important in deciding the choice of model, the robustness of the model is investigated by using the LOU method. Suppose there are n observations (x_1, y_1), (x_2, y_2),..., (x_n, y_n). The LOU method is as follows;

* Omit point (x_1, y_1), estimate β_0, β_1 and $E\{Y\}$.
* Omit point (x_2, y_2), replace with (x_1, y_1) and again estimate β_0, β_1 and $E\{Y\}$.
* Repeat the process until (x_n, y_n) is omitted.

Some of the results are shown in Table 1 for HTK. For large values of X, $E\{Y\}$ is close to 1 (as expected). However for X values less than seven, $E\{Y\}$ takes values around 0.5. This suggests that Model (A) should be recommend for use for values of X greater or equal to seven.

RESULTS AND DISCUSSION

Imon (2007) noted that in practice the coefficient of determination, R_N^2 and R_L^2 with values approximately 0.5 is considered reasonably high. Thus, the goodness-of-fit test showed the simple logistic model fits the data in HTK and HTAR. But, the model is not fitted well in IPR data. The reason might be the biased data since IPR is a referral centre and the chances that a patient referred to this hospital is very likely a positive PTB case.

In Table 2, R_L^2 is compared to R_N^2. The former is suitable as a measure of goodness-of-fit because Eq. 7 does not critically depend on n, whilst the latter is suitable as a measure of goodness-of-fit because Menard (2002) noted that it is a common measure used in the current versions of SPSS and SAS statistical package.

Table 1: Analysis of Leave-One-Out method (LOU) for HTK data

Patient excluded	Patients' data		β_0	β_1	$E\{Y\}$
	X	Y			
1	1	0	0.26735	0.55645	0.56644
2	0	0	0.92650	0.46477	0.71636
3	1	0	0.26735	0.55645	0.56644
4	1	0	0.26735	0.55645	0.56644
5	1	0	0.26735	0.55645	0.56644
6	1	0	0.26735	0.55645	0.56644
7	0	0	0.92650	0.46477	0.71636
8	0	1	0.91629	3.29760	0.98543
9	1	5	0.49307	0.49046	0.95004
10	1	7	0.50157	0.50503	0.98265
11	1	10	0.50261	0.51702	0.99657
12	1	11	0.50245	0.51885	0.99799
13	1	11	0.50245	0.51885	0.99799
14	1	11	0.50245	0.51885	0.99799
15	1	11	0.50245	0.51885	0.99799
16	1	11	0.50245	0.51885	0.99799
17	1	11	0.50245	0.51885	0.99799
18	1	11	0.50245	0.51885	0.99799
19	1	11	0.50245	0.51885	0.99799
20	1	11	0.50245	0.51885	0.99799
21	1	11	0.50245	0.51885	0.99799
22	1	11	0.50245	0.51885	0.99799
23	1	11	0.50245	0.51885	0.99799
24	1	11	0.50245	0.51885	0.99799
25	1	11	0.50245	0.51885	0.99799
26	1	11	0.50245	0.51885	0.99799
27	1	11	0.50245	0.51885	0.99799
28	1	11	0.50245	0.51885	0.99799
29	1	11	0.50245	0.51885	0.99799

Table 2: Results of R_N^2 and R_L^2 for three hospitals

Hospital	Goodness-of-Fit test (%)	
	R_N^2	R_L^2
IPR	25.50	23.24
HTK	48.54	40.43
HTAR	53.15	43.84

Table 3: Analysis of deviance table and test statistic for IPR, HTK and HTAR

Term fitted in model	IPR			HTK			HTAR		
	D	Test statistic (χ^2)	χ^2	D	Test statistic (χ^2)	χ^2	D	Test statistic (χ^2)	χ^2
β_0+X	7.0580	-	-	11.4908	-	-	6.7712	-	-
β_0+X+Z	6.9766	0.0814	3.841	10.4960	0.9948	3.841	6.7712	0.0000	3.841
β_0+X+W	6.7032	0.3548	3.841	10.9018	0.5890	3.841	6.7726	0.0014	3.841
$\beta_0+X+Z+W$	6.6754	0.3826	5.991	7.7366	3.7542	5.991	6.7726	0.0014	5.991

D: Deviance

Table 4: Prediction interval for E(Y) from several selected x_0 values for HTK

x_0	Lower limit	Upper limit	Length of prediction interval
0	0.4785	0.7483	0.2697
1	0.6087	0.8329	0.2241
2	0.7250	0.8932	0.1682
3	0.8169	0.9335	0.1166
4	0.8830	0.9593	0.0763
5	0.9273	0.9754	0.0481
6	0.9556	0.9852	0.0296
7	0.9732	0.9912	0.0180
8	0.9839	0.9948	0.0108
9	0.9904	0.9969	0.0065
10	0.9943	0.9982	0.0039
11	0.9966	0.9989	0.0023

Nevertheless, the general conclusion is the same, namely the simple logistic model fits the data 'better' for HTK and HTAR.

Table 3 shows that the change of deviance in the three hospitals, respectively is not significant when the variable Z is added to Model (A). This result suggests that the visual interpretation of the chest X-ray film may be carried out independently of the outcome of the sputum test. In view of the possible errors when conducting sputum test, for example, a patient not being able to give a good sputum sample, omission of variable Z would mean simplifying the diagnosis process. The results of Table 3 also suggest that it is possible to omit the variable W. The omission of W implies that two X-ray films need not be perfectly aligned before any comparison can be carried out.

Based on a sample of 29 patients for HTK and a sample of 16 patients for HTAR, the following models were obtained:

$$HTK : E\{Y\} = \frac{\exp(0.5018 + 0.5223X)}{1 + \exp(0.5018 + 0.5223X)}$$

$$HTAR : E\{Y\} = \frac{\exp(0.4208 + 3.4619X)}{1 + \exp(0.4208 + 3.4619X)}$$

The LOU method for HTK (Table 1) shows that the β_1 parameter is not very different from the value 0.5223, except when patient number eight is omitted (this is a patient where his chest radiograph shows lung scarring). Hence Model (A) gives robust estimates of β_1. However for β_0, estimates are reliable for X values larger than seven. Similar remarks may be made for the HTAR model.

The estimate for β_0 and β_1 for HTK and HTAR suggest that this parameter values are geographically dependent. Comparison of these models is not trivial. Instead of performing any hypothesis test on the existing data, prediction interval for E{Y} was derived. For values of $X \geq 7$ narrow prediction intervals were obtained as shown in Table 4.

CONCLUSION

The results from the experiment suggested that the simple logistic regression Model (A) is appropriate to model the relationship between the doctors diagnosis (or response) (Y) and NLSP (X) for hospital HTK and HTAR. Further the LOU method suggests that the logistic regression model may predict Y-values (doctors' response) accurately for X-values greater than seven, but should be treated with caution otherwise.

Since films scanners are affordable by most hospitals, the CAD-M-PTB system used may be easily applied to estimate the statistics NLSP which in turn can give an objective assessment of the patient's progress to treatment. This facility would be most suitable when doctors involved are not in the category of specialist or consultant.

ACKNOWLEDGMENTS

REFERENCES

Collet, D., 1952. Modelling Binary Data. 2nd Edn. Chapman and Hall/CRC.

Cruez, A.F., 2005. TB Returns as Number One Infectious Killer Disease. New Straits Times, NSTP Sdn. Bhd.

Dolby, G.R., 1976. The ultrastructural relation; A synthesis of the functional and structures relations. Biometrika, 63 (1): 30-50.

Fuller, W.A., 1987. Measurement Error Models. Wiley, New York.

Gonzalez, R.C., R.E. Woods and S.L. Eddins, 2004. Digital Image Processing Using MATLAB. Pearson Education.

Hill, D.L., G.P. Batchelor, M. Holden and D.J. Hawkes, 2001. Medical image registration. Phys. Med. Biol., 46: R1-R45.

Hosmer, D.W. and S. Lemeshow, 1989. Applied Logistic Regression. Wiley, New York.

Imon, A.H.M.R., 2007. Private Conversation. Professor of Statistics, University of Rajshahi, Bangladesh.

Mahyuddin, A.A., 2006. Private conversation. Medical consultant for respiratory disease, director. Institute of Respiratory Medicine, Kuala Lumpur Hospital.

Menard, S., 2002. Applied Logistic Regression Analysis. 2nd Edn. Sage.

Nagelkerke, N.J.D., 1991. A note on a general definition of the coefficient of determination. Biometrika, 78 (3): 691-692.

Neter, J., W. Wasserman and M.H. Kutner, 1989. Applied Linear Regression Models. 2nd Edn. Richard D. Irwin.

Rijal, O.M., N.M. Noor and L.T. Shee, 2006. A statistical method for comparing chest radiograph images of MTB patients. WSEAS. Trans. Sig. Process, 2 (7): 1017-1023.

Shaban, H., 2005. Private Conversation. Medical Consultant for Respiratory Disease, Selangor Medical Center.

Unnikrishnan, K.P. and P.S. Jagannatha, 2002. Monitoring of tuberculosis control programme-recording, reporting and supervision. National Tuberculosis Institute, Bulletin.

WHO, 2004. Global tuberculosis control, surveillance, planning, financing. World Health Organization, Geneva.

WHO, 2006. The global plan to stop TB 2006-2015: Action for life towards a world free of tuberculosis. World Health Organization, Geneva.

Zitova, B. and J. Flusser, 2003. Image registration methods: A survey. Image Vision Comput., 21: 977-1000.