Proceedings of the 2005 IEEE
Engineering in Medicine and Biology 27th Annual Conference
Shanghai, China, September 1-4, 2005

A STATISTICAL COMPARISON OF DIGITAL X-RAY IMAGES FOR MTB PATIENTS

Omar Mohd. Rijal*, Norliza Mohd. Noor†, Hamidah Shaban§, Shee Lee Teng*
omarrija@um.edu.my, norliza@citycampus.utm.my
* Institute of Mathematical Science, University Malaya
†Diploma Program Studies, Universiti Teknologi Malaysia
§Institute of Respiratory Disease, Kuala Lumpur Hospital

ABSTRACT

A common practice in medical diagnosis and patient management is the comparison of two chest radiographs images. The difference between two digital images at two time points is a measure of the effect of treatment on the patient. Two measures of similarity, the ordinary regression coefficients, R_S^2 and coefficients of determination for the Unreplicated linear functional relationship model (ULFR), R_F^2, are used to compare images for the particular case of Mycobacterium Tuberculosis (MTB). Our results suggest that a series of R^2 values indicates gradual declining trends with values falling within a band. New patients with a series of R^2 values falling within this band may be consider as making a good or acceptable recovery.

INTRODUCTION

The common radiograph film is still an important tool in the diagnostic process for lung ailment despite rapid advances in medical imaging technology, see Middlemiss [1] and Moores [2]. In Malaysia, government hospitals perform the major part of detection using radiographs films simply out of economic considerations. Problems associated with the visual interpretation (and comparison) of standard chest radiograph films are well known. These remarks motivate a need to create objective methods in particular for comparing two or more digital radiograph images.

Two similarity measures firstly, coefficient of determination (COD) for the Unreplicated linear functional relationship (ULFR) model, labeled R_F^2 [3], [4]; and secondly the ordinary regression coefficient [5] were proposed in [6]. The fundamental idea of using correlations is that as patient's health improves the values of R_S^2 and R_F^2 between selected images should decrease. A series of R^2-values over time is used as an approximate indicator of effect of treatment. For each time point, a set of $k - R^2$ values corresponding to k patients are standardized as follows:

$$R^{2*} = \frac{R^2 - \overline{R^2}}{S_a} \quad \text{(at each time point for } k \text{ patients)}$$

where $\overline{R^2}$ is the mean of R^2 and S_a is the standard deviation of R^2. We defined Band(l,m) as the range of R^2-values with max R^2 equal l, and minimum R^2-values equal m.

In this study, let {A(i,j); i = 1, …, M, j= 1, …, N} represent the digital X-ray image of a patient on his first visit to the hospital. Let {B(i,j); i = 1, …, M, j= 1, …, N} represent the same patients image at a later prescribed time point.

BRIEF REVIEW OF ULFR AND R_F^2

Re-label the observations (or experimental values) of {A(i,j); i = 1, …, M, j= 1, …, N} as $y_1, y_2, \ldots, y_{MN}$, the observations of {B(i,j); i = 1, …, M, j= 1, …, N} as $x_1, x_2, \ldots, x_{MN}$, and the true A(i,j) and B(i,j) values will be denoted by $Y_1, Y_2, \ldots, Y_{MN}$ and $X_1, X_2, \ldots, X_{MN}$, respectively. We look at two regression models to study the relationship between y_i and x_i.

We first look at the ordinary simple linear regression (SL) model [5] of the dependent variable, y_i and explanatory variable, x_i:

$$y_i = \alpha_s + \beta_s x_i + \varepsilon_i, \quad i = 1, 2, \ldots, MN \qquad (1)$$

where the maximum likelihood estimators (MLE) and COD are given as follows:

$$\hat{\alpha}_s = \bar{y} - \hat{\beta}_s \bar{x}, \quad \hat{\beta}_s = \frac{S_{xy}}{S_{xx}}$$

and $R_s^2 = \dfrac{\hat{\beta}_s S_{xy}}{S_{yy}}$ (2)

where the R_s^2 is the proportion of variation explained by explanatory variable x and

$$\bar{y} = \frac{\sum y_i}{MN}, \bar{x} = \frac{\sum x_i}{MN},$$

$$S_{yy} = \sum (y_i - \bar{y})^2,$$

$$S_{xx} = \sum (x_i - \bar{x})^2$$

and $S_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y})$.

However, as pointed out by [4]. The assumption that the explanatory variable can be measured exactly may not be realistic in many situations. The estimates of explanatory variable may contain measurement error arising from the techniques or instruments used or trying to quantify a variable that has no physical dimension. In these cases, the explanatory variable is subject to error.

Suppose that now the X and Y are two linearly related unobservable variables (see [3] and [7])

$$Y_i = \alpha + \beta X_i \quad (3)$$

and the two corresponding random variables x and y are observed with error δ and ε respectively

$$\left.\begin{aligned} x_i &= X_i + \delta_i \\ y_i &= Y_i + \varepsilon_i \end{aligned}\right\} \quad i = 1, 2, \ldots, MN \quad (4)$$

where δ_i and ε_i are mutually independent and normally distributed random variables. Equation (3) and (4) are known as the ULFR model when there is only one relationship between the two variables X and Y. It can be shown that the maximum likelihood estimators when the ratio of the error variances is equal to one, $\dfrac{\sigma_\varepsilon^2}{\sigma_\delta^2} = \lambda = 1$, are given as follows:

$$\hat{\alpha}_F = \bar{y} - \hat{\beta}_F \bar{x} \quad (5)$$

$$\hat{\beta}_F = \frac{(S_{yy} - S_{xx}) + \{(S_{yy} - S_{xx})^2 + 4S_{xy}^2\}^{1/2}}{2S_{xy}} \quad (6)$$

$$\hat{\sigma}_\delta^2 = \frac{1}{MN-2}\left[\sum (x_i - \hat{X}_i)^2 + \frac{1}{\lambda}\sum (y_i - \hat{\alpha} - \hat{\beta}\hat{X}_i)^2\right] \quad (7)$$

and $$\hat{X}_i = \frac{x_i + \hat{\beta}(y_i - \hat{\alpha})}{\lambda + \hat{\beta}^2} \quad (8)$$

The equation in (3) and (4) can be written as

$$\begin{aligned} y_i &= \alpha_F + \beta_F x_i + (\varepsilon_i - \beta_F \delta_i) \\ &= \alpha_F + \beta_F x_i + V_i \quad \text{for } i = 1, \ldots, MN \quad (9) \end{aligned}$$

where the error of the model, V_i is normally distributed. The residual sum of squares and the regression sum of squares are given as follows:

$$S_E = \frac{S_{yy} - 2\hat{\beta}S_{xy} + \hat{\beta}^2 S_{xx}}{1 + \hat{\beta}^2}$$

and

$$\begin{aligned} S_R &= S_{yy} - S_E \\ &= S_{yy} - \frac{S_{yy} - 2\hat{\beta}S_{xy} + \hat{\beta}^2 S_{xx}}{1 + \hat{\beta}^2} \end{aligned}$$

Therefore, the COD for ULFR can be defined as

$$R_F^2 = \frac{S_R}{S_{yy}} = \frac{\hat{\beta}_F S_{xy}}{S_{yy}} \quad (10)$$

and it can be shown that $0 \le R_s^2 \le R_F^2 \le 1$.

IMAGE REGISTRATION

Areas of the lung infected by MTB shows up in the X-ray film as white 'snow flakes' or 'cloud' either centred into a specific area or disperse around the lung depending on the level of infection. Since the disease is airborne, the infected area normally starts at the upper part of the lung and gradually moves toward the lower lung as the infections grow.

The progress of an MTB patient towards a given treatment may be defined as the decline or reduction in the brightness of 'snow flakes' when two X-rays are compared at different time points. Since all patients considered had highly developed stages of MTB, showing bright white snow flakes, subsequent images (after treatment) will exhibit a decline in the intensity of the white snow flakes.

Image registration proves to be a problem for chest radiograph. It is therefore important that the region of interest of the X-ray image (selected manually under expert advice) which shows the infected area be aligned when comparing two images. Instead of using pixel coordinates to define the location of the infected area, position of the rib bone is used. This is done by reference to the highest point of a rib. Figure 1 shows point P is the highest vertical point for the 4th rib; and R the corresponding point for the 6th rib. Point X is mid-way between the extreme points of rib-cage on horizontal line through R. The two horizontal lines through P and R and the vertical line through X define three sides of the area of interest. Finally by fixing the area of the rectangle, the point w is automatically determined. Although done

manually, this is not a particularly restrictive procedure when carried out in the MATLAB environment.

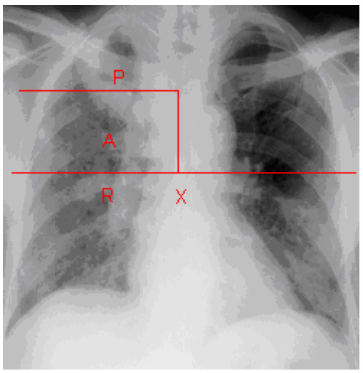

Figure 1: Image registration for chest X-ray.

A DESCRIPTION OF THE EXPERIMENTS

For each patient, his series of visits to the hospital and consequently the chest x-ray images obtained are labeled A,B, C, D etc. The minimum treatment period for MTB is 6 months and the progress is monitored every 2 months via clinical test (usually the sputum test) and chest x-ray images. In this study the chest x-ray images are compared pair-wise, that is first and the second chest x-ray (AB), first and third chest x-ray (AC) resulting in a series of R^2 values corresponding to AB, AC, AD, …. and so forth.

The medical interpretations or visual interpretation of the image was done by a medical expert from Institute of Respiratory Disease [7]. This stage of the experiment was then followed by associating the medical interpretations with the numerical R^2–values obtained from the scanned X-ray films. The X-ray films were scanned into 16 bit DICOM file using Kodak LS 75 X-ray film scanner.

SOME RESULTS

A total of 24 male patients and 25 female patients were used in this study. Table 1 summarizes the main results from Figure 2, 3, 4 and 5.

The Band (*l,m*) for male are narrower than for the female, suggesting male patients are more similar with respect to recovery rates. Secondly, performance of R_F^2 and R_S^2 are similar, possibly because only confirmed patients were studied.

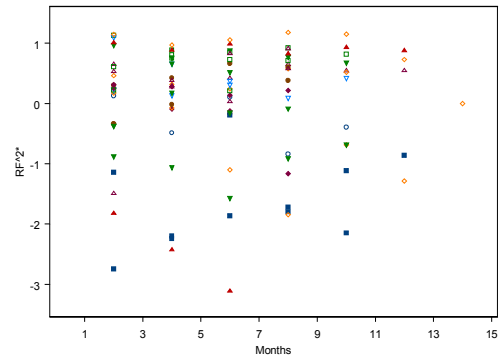

Figure 2: Graph of standardized R_F^{2*} for male MTB.

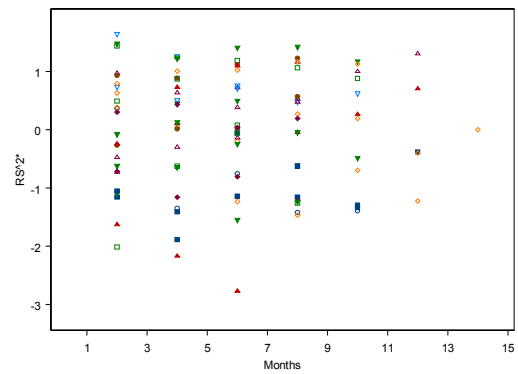

Figure 3: Graph of standardized R_S^{2*} for male MTB.

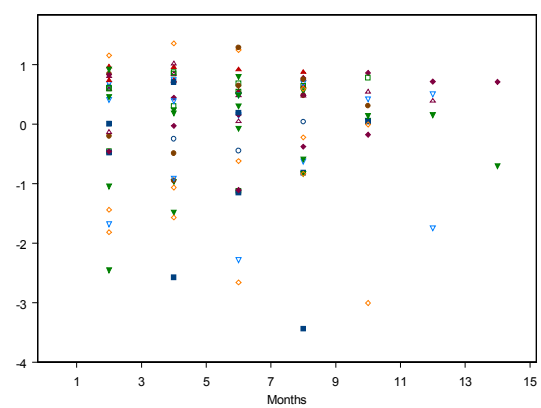

Figure 4: Graph of standardized R_F^{2*} for female MTB patients.

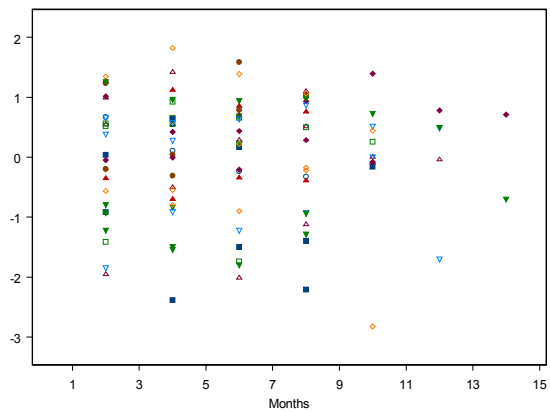

Figure 5: Graph of standardized R_S^{2*} for female MTB patients.

Table 1: Values for *l* and *m* for Band(*l,m*).

	R_F^{2*}	R_S^{2*}
Male	(1.2, -2.8)	(1.5, -2.9)
Female	(1.4, -3.5)	(1.8, -2.9)

DISCUSSION

All cases considered were MTB patient who were studied from their first visit to the clinic until being confirmed by the hospital medical expert [7] as being completely recovered. The expert's confirmation is based on variables such as weight (being stabilized), lungs being clear, chest X-ray shows improvement, not sweating and good appetite.

As such the Band (*l,m*) was shown to be an indicator for MTB recovery. However work still has to be done for understanding inter-patients recovery rates. There were cases of reversal of trends, for example patients developing secondary infections or existence of scarring of lung tissues. The band also does not explicitly account for the patients who make complete recovery in a shorter time period. These issues form the area of subsequence work for this project whereby the image histogram of say A and B are compared. Here the mean, variance, skewness, kurtosis and percentile from the image histogram will be used as possible measures of recovery rate. These statistics measure the leftward shifts of the image histogram indicating a decline in the intensity of the white snow flakes.

REFERENCES

[1] Middlemiss, H, "Radiology of the future in developing countries", British Journal of Radiology, 55, pp. 698-699, 1982.

[2] Moores, B.M., "Digital X-ray Imaging", IEE Proceedings, Vol. 134, part A, Number 2, Special Issues On Medical Imaging, 1987.

[3] Dolby, G. R., "The Ultrastructural Relation; A Synthesis of the Functional and Structural Relations", BIOMETRIKA, 63, pp. 30-50, 1976.

[4] Fuller, W. A., Measurement Error Models, New York, John Wiley, 1987.

[5] Weisberg, S., Applied Linear Regression, New York, John Wiley, 1980.

[6] Chang, Y. F., Rijal O. M. and Noor N. M., "Image Quality and Noise Evaluation", The Proceedings of the 7th International Symposium On Signal Processing and Its Applications (ISPAA 2003), Paris, France, 2003.

[7] Shaban, H, Medical Consultant for Respiratory Disease, Private Conversation, Institute of Respiratory Disease, Kuala Lumpur Hospital.