# A Graphical Method of Detecting Pneumonia Using Chest Radiograph

NORLIZA MOHD. NOOR[∞], OMAR MOHD. RIJAL*, S. A. R. ABU-BAKAR **,
MOHD. IQBAL*, GAN CHEW PENG*
[∞]Dept. of Elect. Eng., College of Science and Technology, Universiti Teknologi Malaysia, MALAYSIA
* Institute of Mathematical Science, University of Malaya, MALAYSIA
**Faculty of Elect. Eng., Universiti Teknologi Malaysia, MALAYSIA
norliza@ic.utm.my, omarrija@um.edu.my, syed@fke.utm.my,
iqbal1510@gmail.com, ganchewpeng@yahoo.com

***Abstract:-*** An important ingredient of health care is the correct initial diagnosis of chest ailments using images of chest radiograph. This paper develops a simple graphical method to aid the initial screening and discrimination of pneumonia patients (PNEU) from pulmonary tuberculosis (PTB) patients, lung cancer (LC) patients and normal healthy individuals (NL). Approximate confidence regions using principal component methods on selected texture measures detect and discriminate PNEU from PTB, LC and NL. A brief simulation study indicates that the probability ellipsoid is robust to mild deviation from normality with the presence of two outliers. The main result of this study is that the PNEU-ellipsoid when applied to test data is capable of detecting pneumonia in the sense that 100% of NL, 85% of LC and 65% of PTB will be rejected when using contrast texture measures. Using the combination of twelve features gives similar results. Membership of PNEU-ellipsoid obtained from the chest radiograph image can be used as a useful first stage detection of pneumonia.

***Keywords:-*** Image manipulation and recognition, medicine, texture measures, statistical methods

## 1 Introduction

Despite rapid advances in medical imaging technology, the conventional chest radiograph is still an important ingredient in the diagnosis of lung ailments [1, 2, 3]. In Malaysia, government hospitals perform the diagnosis using radiograph films simply out of economic considerations.

The chest X-ray of a patient is frequently used as the initial indicator of the existence of lung diseases. It is well-known that diseases like pulmonary tuberculosis and lung cancer may also be detected from the chest X-ray by the experienced medical practitioner who is normally a consultant. The efficiency of detection clearly depends on the ability of the medical practitioner to differentiate the three major lung diseases; pneumonia (PNEU), pulmonary tuberculosis (PTB) and lung cancer (LC), and hence, image based detection methods have been developed [4, 5, 6].

Respiratory infections are the most common of all infections. The most serious respiratory infection, pneumonia, carries the highest mortality rates amongst all infectious diseases and is the sixth leading cause of death in people over 65 years of age [7]. In 2004, 58,564 people died of pneumonia in the United States of America, [8].

About one million new cases of lung cancer have been detected annually [9] and two million deaths worldwide are due to tuberculosis every year [10]. Tuberculosis hinders socioeconomic development; 75% of people with tuberculosis are in the economically productive age group of 15-54 years. Ninety five percent of all cases and 99% of deadly cases occur in developing countries [11]. In Malaysia, the tuberculosis problem is making a comeback partly due to the failure in diagnosing PTB patients seeking treatment for continuous cough [12].

From [6, 13], a graphical method based on the use of the Andrews Curve was proposed for initial screening. For a given chest X-ray, a region of interest (ROI) (Fig. 1) is selected from which a set of line profiles are chosen. Each line profile may be interpreted as a signal (Fig. 2) which in turn is subjected to the Daubechies 4 transformation. The average of these signals in the form of a vector of Daubechies coefficients represents the ROI. This average vector is then represented as an Andrews' Curve. Given two patients, hence two average vectors, we will have two Andrews' Curves. The vertical separation between two Andrews' Curves is equivalent to the Euclidean distance between the two average vectors. Fig. 3 shows that for a given $t$ value (along horizontal axis) three distinct clusters is clearly seen and henceforth the probability of classification for each disease type may be estimated.

This graphical method [13] was shown to have 85% chance of correct detection when discriminating between PTB, LC and NL (see Fig. 3). However, the inclusion of a sample of pneumonia patients resulted in significant overlapping of the Andrews' curve representing the pneumonia patients with PTB samples, (see Fig. 4).

To remedy this problem of overlapping of Andrews' curve, a new approach using probability ellipsoids of texture measures is proposed in the next section. The combined use of texture measures and principal components method has been used in other application but is novel in this study, [14].

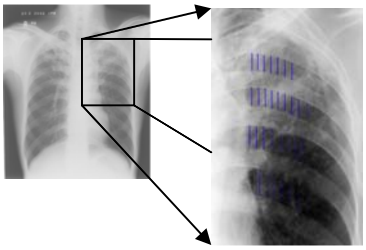


Fig.1: Line profiles taken on the region of interest of a PTB patient (Source: Malaysian Institute of Respiratory Medicine).

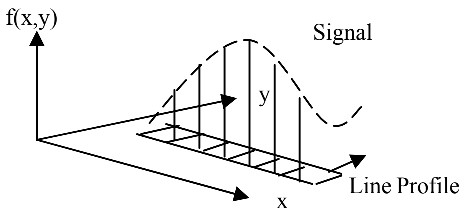



Fig. 2: A Line profile: A two-dimensional light intensity function f(x, y), where x and y denotes spatial coordinates.

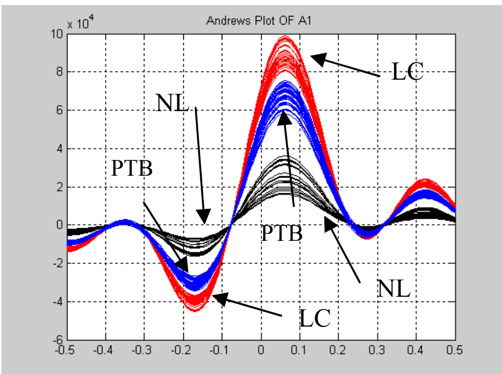



Fig. 3: Andrews' curve 20 normal lung (NL), 40 pulmonary tuberculosis (PTB) patients and 30 lung cancer (LC) patients plot over the range of –0.5 to 0.5.

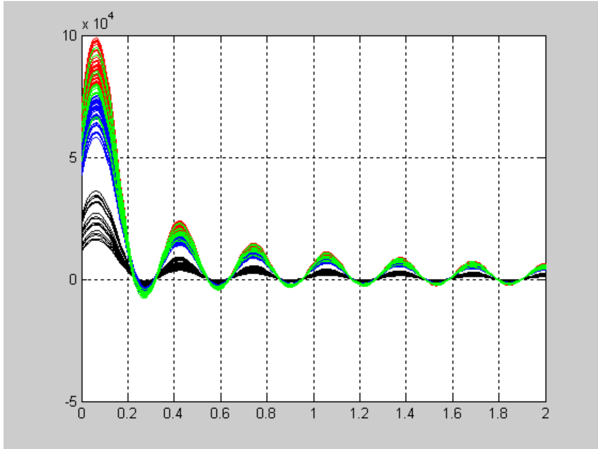


Fig.4 The above plot illustrate Andrews Curve for LC (red curve), pneumonia (green curve), PTB (blue curve) and normal lung (black curve).

## 2 Methodology

Cases that arrived at the IPR may be considered a random sample since an individual case may come from any of the Malaysian hospitals or clinics. The chest X-ray films from IPR were then digitized into DICOM format using the X-ray Film Scanner Kodak LS 75.

A sample of one hundred images were concurrently read and interpreted for the presence of either PNEU only, PTB only and LC only (no other secondary medical condition) by two independent pulmonologists who are trained to apply the WHO guideline, and the affected region (ROI) was identified (Fig. 5). This process was repeated for the normal sample of another thirty individuals and each member of the sample was identified as a healthy individual.

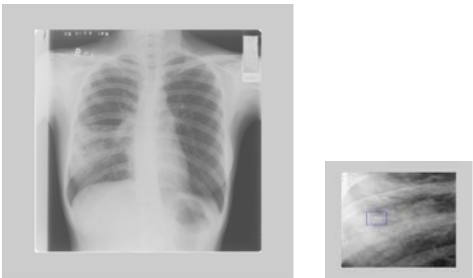
Each of the ROI were subjected to the two dimensional Daub4 wavelet transform, (Fig. 6). The wavelet transform convert the image into four subsets, labeled LL, LH, HL and HH representing the trend, horizontal, vertical and diagonal detail coefficients.

Twelve texture measures in each of LL, LH, HL and HH yields 48 descriptors or features that will be used to detect pneumonia, [5, 15, 16, 17]. The twelve texture measures used were;

(i) Mean Energy
(ii) Entropy
(iii) Contrast
(iv) Homogeneity
(v) Standard deviation of value
(vi) Standard deviation of mean of energy

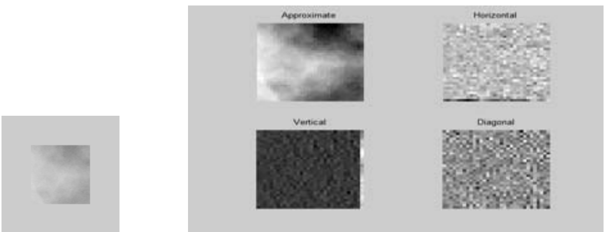
(vii) Maximum wavelet coefficient value
(viii) Minimum wavelet coefficient value
(ix) Maximum value of energy
(x) Maximum row sum energy
(xi) Maximum column sum energy
(xii) Average number of zero-crossings

Appendix I give the definitions of the above texture measures.



(a) (b)

Fig. 5 (a) Chest X-ray of a pneumonia patient,
(b) A subset image of the infected area.
(Source: The Institute of Respiratory Medicine, Malaysia)




(a) (b)

Fig. 6 (a) Region of interest, (b) the transformed image where four image subset was formed

For a given texture measure (example entropy), one value of entropy was calculated for each of LL, LH, HL and HH. Therefore each ROI for a given individual may be represented by a four dimensional feature vector. A sample of PNEU patients ($n_1$ = 30), PTB patients ($n_2$ = 40), LC patients ($n_3$ = 30), and a sample of normals ($n_4$ = 30) yields 130 four-dimensional vectors. Principal component analysis was carried out on each of the samples separately, whereby a biplot of the first two principal components was derived (see, Fig.7(a)-(e)). The proportion of explained variance for the first two principal components was at least 80%. Each bivariate sample of principal component was tested and shown to be normally distributed. Without loss of generality, let $\underline{x} \sim N_2(\underline{\mu}, \Sigma)$ represent the bivariate normal distribution for one of the group. Let,

$$y = (\underline{x} - \underline{\mu})^T \Sigma^{-1} (\underline{x} - \underline{\mu}),$$

then $y$ has a univariate chi-square distribution, [18]. Further, the equation $y = c$ is the equation of an ellipsoid (ellipse in two dimension) where the value c may be selected from the chi-square table. In particular when $P(y < c) = 0.95$, then $y = c$ may be called the 95% confidence ellipse.

Two of the texture measures, minimum value and the combination of twelve texture measures showed ellipsoids non-overlapping for the pneumonia samples (Fig. 7(d) and 7(e)). However, Fig. 7(b) and 7(c) suggests minimal overlap between PNEU and LC, and PNEU and PTB, respectively. This promising observation will only be useful if the probability ellipsoids are robust. In the next section, robustness of the ellipsoid will be investigated for effects of non-normality, small samples and outliers.

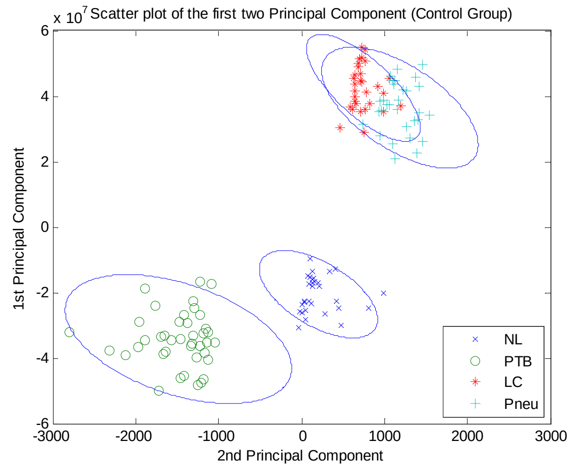



Fig. 7(a) Mean of Energy

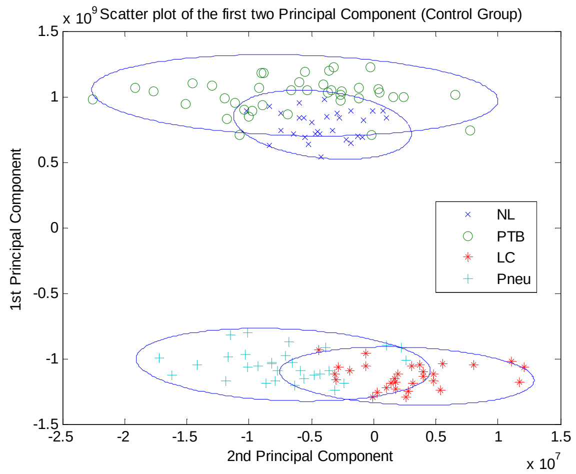



Fig. 7(b) Contrast

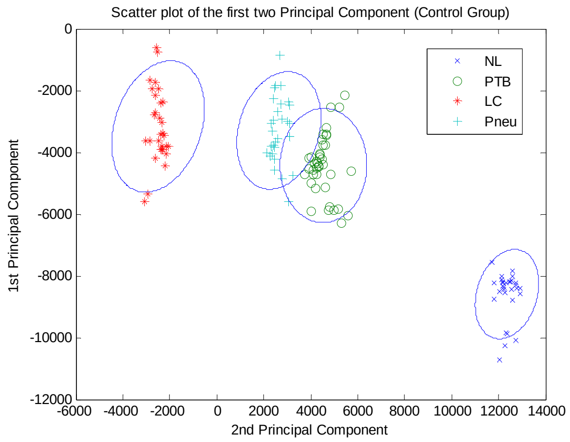



Fig. 7(c) Entropy

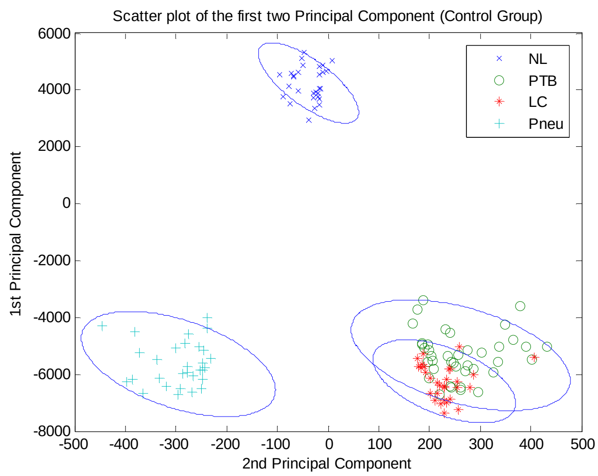



Fig. 7(d) Min Value

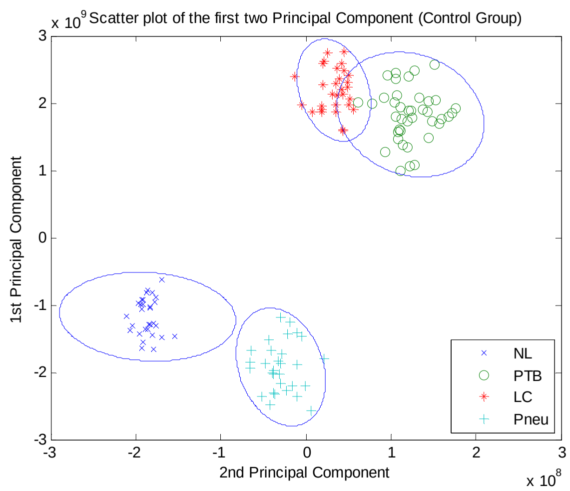



Fig. 7(e) Combination of 12 texture measures

Fig. 7(a)-(e) Plots of texture ellipsoids, PNEU (cyan-+), PTB(green -o), LC (red-*), and NL(blue-x).

## 3 Investigation of Robustness of the Confidence Ellipse

In practice, given a random sample $\underline{x}_1, \underline{x}_2, \cdots, \underline{x}_n$, the parameters $\underline{\mu}$ and $\Sigma$ are estimated by $\bar{\underline{x}}$ and $S/(n-1)$ where

$$\bar{\underline{x}} = \frac{1}{n}(\underline{x}_1 + \cdots + \underline{x}_n) \quad (1)$$

and

$$S = \sum_{r=1}^{n} (\underline{x}_r - \bar{\underline{x}})(\underline{x}_r - \bar{\underline{x}})^T \quad (2)$$

Let

$$y^* = (\underline{x} - \bar{\underline{x}})^T \left( \frac{1}{n-1} S \right)^{-1} (\underline{x} - \bar{\underline{x}}) \quad (3)$$

Then the ellipsoid $y^* = c$ defines a confidence region.

The following simulation study investigates whether $y^*$ has a good approximation to the chi-squared distribution taking into consideration effects of sample size, outlier and deviations from normality of $\underline{x}$. A sample of $y^*$ will be generated (simulated) and its distributions tested using the Kolmogorov-Smirnov test for a given set of condition (example small sample size), [19, 20]. An explanation of the Kolmogorov-Smirnov test is given in Appendix II. When the Kolmogorov-Smirnov test states that $y^*$ is distributed as the univariate Chi-squared distribution, we consider the required approximation to be a good one. Henceforth, $y^* \le c$ may be used as the required confidence region.

## 4 Simulation Study

### CASE 1 Investigating the Effect of Sample Size for Normal Data

Given two populations $\prod_j$, $\underline{x} \sim N_2(\underline{\mu}_j, \Omega_j)$, $j = 1,2$ where $\prod_1$ denote the pneumonia population and $\prod_2$ denote the population of normal healthy individuals. Applying a linear transformation $\underline{x} \to A\underline{x} + \underline{b} = \underline{w}$ converts both distribution into canonical form given as follows:

$$\prod_1 : \underline{w} \sim N_2(\underline{\mu}, D) \text{ and } \prod_2 : \underline{w} \sim N_2(\underline{0}, I)$$

where $D = \text{diag}(d_1, d_2)$,
$I$ is the identity matrix,

$\underline{0} = (0,\cdots,0)^T$ and $\underline{\mu}$ is a bivariate vector.

A proof of this result is given in [21].

Since the focus is on the robust properties of the distribution of a diseased group, for example, pneumonia, therefore only the probability ellipse for the pneumonia will be generated.

This result implies that each component of $\underline{w}_j$ ($j$ = 1, .., n) can be generated separately since $D$ implies independence of the vector components.

The choice of parameter values are as follows;

(i) $n$=15,20,21,22,23,24,25,26,27,28,29,30,35,40

(ii) $\underline{\mu} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

(iii) $D = \left( \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix} \right)$

The choice of (i) is to verify the validity of using n=30. Using zero means in (ii) is simply stating that the center of the ellipse is treated as the origin. The choice for (iii) investigates some effect of different variances.

**CASE 2 Investigating Effect of Non-normality**

A random bivariate sample $\underline{x}_1, \underline{x}_2 \cdots, \underline{x}_n$, is generated from a sample with known measure of skewness ($\beta$) or kurtosis ($\gamma$). This is achieved by generating random variables from a mixture distribution. Suppose, $f(x) = \lambda N(a,b) + (1-\lambda) N(-a,bc)$ where N(.,.) denote the univariate normal distribution. It can be shown that when $\lambda = 0.5$, $\mathrm{E}(x) = 0$, $\mathrm{E}(x^2) = a^2 + \frac{b}{2}(1+c)$,

$$\mathrm{E}(x^3) = \frac{3ab}{2}(1-c)$$

and $\mathrm{E}(x^4) = a^4 + 3a^2 b(1+c) + \frac{3b^2}{2}(1+c^2)$.

Henceforth we calculate

$$\beta = \frac{\mathrm{E}(x^3)}{[\mathrm{E}(x^2)]^{3/2}} \text{ and } \gamma = \frac{\mathrm{E}(x^4)}{[\mathrm{E}(x^2)]^2}.$$

The choice of parameter values are shown in Table 1(a)-(b). A graphic illustration of changing $\gamma$ and changing $\beta$ is given in Fig. 8 and Fig. 9.

Table 1(a) Changing $\gamma$ when $\beta$=0.0

| | | | |
|---|---|---|---|
| $a$=0.0 | $b$=1.0 | $c$=1.0 | $\gamma$=3.0 |
| $a$=0.0 | $b$=0.667 | $b$=2.0 | $\gamma$=3.33 |
| $a$=0.0 | $b$=1.450 | $c$=0.379 | $\gamma$=3.61 |
| $a$=0.633 | $b$=0.600 | $c$=1.0 | $\gamma$=2.66 |

Table 1(b) Changing $\beta$ when $\gamma$=3.0

| | | | |
|---|---|---|---|
| $a$=0.0 | $b$=1.0 | $c$=1.0 | $\beta$=0.0 |
| $a$=0.512 | $b$=0.952 | $c$=0.550 | $\beta$=0.33 |
| $a$=0.589 | $b$=0.936 | $c$=0.395 | $\beta$=0.5 |
| $a$=0.509 | $b$=0.529 | $c$=1.800 | $\beta$=-0.32 |

The values $a$=0.0, $b$=1.0 and $c$=1.0 implies the normal case.

**CASE 3 Investigating Effect of Non-normality and Outliers**

The experiments in case 2 was repeated but outliers where added as follows;

(i) $x_0 = \begin{pmatrix} 4\sqrt{\mathrm{E}(x^2)} \\ 0 \end{pmatrix}$

(ii) $x_0 = \begin{pmatrix} 5\sqrt{\mathrm{E}(x^2)} \\ 0 \end{pmatrix}$

(iii) use (i) and (ii).

# 5 Estimated Misclassification Probability

As a measure of overlap for the confidence ellipses, we estimate the misclassification probability of the form P(A|B) where A and B are two different disease types. Without loss of generality estimation of P(PNEU|PTB) requires counting the number of PTB patients that fall into PNEU confidence ellipse. This estimation process is repeated for all diseases in a pair-wise manner.

Since the focus of the study is on detecting PNEU, three error probabilities P(PNEU|NL), P(PNEU|PTB) and P(PNEU|LC) were estimated as follows;

(i) Given a random sample of PNEU represented as vectors of principal components $\underline{x}_1, \underline{x}_2, \cdots, \underline{x}_n$, define the PNEU-ellipsoid as $D(\underline{x}) = c$ (see Fig. 7) where

$$D(\underline{x}) = (\underline{x} - \bar{\underline{x}})^T \left( \frac{1}{n-1} S \right)^{-1} (\underline{x} - \bar{\underline{x}}).$$

The definition of $\bar{\underline{x}}$ and $S$ is given in equation (1) and equation (2). The set of vectors $\underline{x}_1, \underline{x}_2, \cdots, \underline{x}_n$ are the control data.

(ii) Let a random sample of NL be represented as vectors of principal components $\underline{w}_1, \underline{w}_2, \cdots, \underline{w}_q$.

(iii) Let a random sample of PTB be represented as vectors of principal components, $\underline{y}_1, \underline{y}_2, \cdots, \underline{y}_m$.

(iv) Let a random sample of LC be represented as vectors of principal components $\underline{z}_1, \underline{z}_2, \cdots, \underline{z}_k$.

Note that $\underline{w}_1, \underline{w}_2, \cdots, \underline{w}_q$, $\underline{y}_1, \underline{y}_2, \cdots, \underline{y}_m$, and $\underline{z}_1, \underline{z}_2, \cdots, \underline{z}_k$ are the test data.

Henceforth, without loss of generality P(PNEU|NL) may be estimated by counting the proportion of times $D(\underline{w}_j) < c$. Therefore estimates of P(PNEU|PTB) and P(PNEU|LC) are similarly derived.

Table 2 shows the estimates of the three error probabilities for all texture measures considered.

# 6 Results and Discussion

The Andrews' curve has several disadvantages, the main being the inability to detect and differentiate pneumonia from PTB and LC, [10]. Also the feature vectors do not represent the whole ROI, and decisions have to be made before selecting the line profiles. Finally a suitable *t*-value for the Andrews' curve must be selected before performing discrimination, making the use of Andrews' curve difficult. As such a new method involving texture measures is proposed.

The probability ellipsoid of individual texture measures showed two cases where the pneumonia ellipsoid does not overlap with the other disease type. Table 2 shows the misclassification probability when the PNEU-ellipsoid were considered.

The simulation exercise investigates the robustness of the ellipsoid. The effect of sample sizes, non-normality and presence of outliers on the Chi-square approximation was investigated. Table 3 showed that when data is normal without any outliers, a sample size as small as fifteen guarantees that the approximation is good. Increasing skewness (without outliers) appears to have no effect on the Chi-square approximation (Table 4). Increasing kurtosis, generally give the same results as for skewness except for the case *n*=26, gamma = 3.61.

For most cases considered either one or two outliers do not seriously affect the approximation (Table 5 and Table 6).

The significant effect of kurtosis was only observed for the case *n*=43, gamma=3.61, where the approximation was rejected when both previous outliers were present.

The simulation results above strongly suggest that the patients when represented by its principal components will fall inside the probability ellipsoid, and Fig. 7 tends to support this remarks.

# 7 Conclusion

The texture ellipsoids clearly appear to be better than the Andrews' curve in its ability to detect pneumonia. The PNEU-ellipsoid is capable of detecting pneumonia in the sense that 100% of NL, 85% of LC and 65% of PTB will be rejected when using contrast texture measures. Using the combination of twelve features gives similar results followed closely by minimum value texture measure. Membership of PNEU-ellipsoid obtained from the chest radiograph image can be used as a useful first stage detection of pneumonia.

The simulation study shows that the choice of a sample size of thirty (*n*=30) ensures that the confidence ellipse is reliable in the sense that two non-overlapping confidence ellipse strongly suggest two well separated groups. These remarks also hold in the presence of mild skewness and kurtosis. However, the presence of outliers can make the Chi-square approximation unacceptable.

Even though the ellipsoids allow the detection of PNEU, they are not capable of detecting PTB and LC. In practice the following detection method is recommended; firstly investigate for PNEU using the ellipsoids, and if the patient is not identified as PNEU then use the Andrews' curve to investigate for PTB and LC.

**Acknowledgements**
We would like to acknowledge the contribution from Datin Dr. Hjh Aziah Ahmad Mahyuddin, Director of

The Institute of Respiratory Medicine, Kuala Lumpur. This research was funded under an E-Science Fund from the Malaysian Ministry of Science, Technology, and Innovation, and Universiti Teknologi Malaysia.

*References*

[1] B. Ginnekien, B. M ter H. Romeny, and M. A. Viergever, Computer Aided Diagnosis in Chest Radiography: A Survey. *IEEE Trans. Medical Imaging*, *vol. 20*, no.12, 2001, pp. 1228-1241.

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

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

[4] K. Le, Lung X-ray Image Analysis For Automated Detection of Early Cancer and Tuberculosis, *WSEAS Transactions on Information Science and Applications*, 3(12), pp. 2347-2354, 2006.

[5] L. L.G. Oliveira *et al*, Computer-aided diagnosis in chest radiography for detection of childhood pneumonia, *Int. J. Med. Inform*. (2007), doi:10.1016/j.ijmedinf.2007.10.010, 2007.

[6] O. M. Rijal, N. M. Noor, A. Hussin, E. L. Ong. Discriminant Functions and Multiresolution Analysis (MRA) For Disease Detection, *WSEAS Transactions on Mathematics*, 5(5), pp. 605-608, 2006.

[7] Centers for Disease Control and Prevention, National Center for Health Statistics: Preliminary Data for 2005. September 2007. USA.

[8] *Improving Life, One Breath at a Time*,, Lung Disease Data: 2008, American Lung Association, New York. pp 68, 2008.

[9] WHO, International Agency for Research on Cancer. Edited by B.W. Steward and P. Kleihues. (2003).*World Cancer Report*. Geneva: WHO.

[10] WHO, The global plan to stop TB 2006-2015: Action for life towards a world free of tuberculosis. Geneva:WHO.

[11] L. Knight, Tuberculosis and sustainable development. Geneva, World Health Organization, (document WHO/CDS/STB/2000 4: Stop TB Initiatives Series).

[12] TB a Problem Once Again. *New Straits Time article*, 29 March 2008.

[13] Rijal OM, Noor NM, Hussin A, Yunus A, Ong EL, Using Statistical Features to Verify the Gold Standard for Pulmonary Tuberculosis Detection, *Inter. Journal of Comp. Assisted Radiology and Surgery*, 3(1):S426-428, 2008.

[14] K. A. Wang *et al*, Implementation Of An Image Retrieval System Using Wavelet Decomposition And Gradient Variation, *WSEAS Transactions on Computers*, 7(6), pp. 724-734, 2008.

[15] M. Unser, Texture classification and segmentation using wavelet frames, *IEEE Trans. Image Proc.*, *vol. 4*, no.11, 1995, pp. 1549-1560.

[16] A. Sengur, I. Turkoglu, M. C. Ince, Wavelet packet neural networks for texture classification, *Expert Systems with Application* , 32(2007) 527-533.

[17] M. F. A. Fauzi and P. H. Lewis, Automatic texture segmentation for content-based image retrieval application, *Pattern Analysis Application*, Springer, 9:307-323, 2006.

[18] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*. 6th Ed. (New Jersey: Prentice Hall, pp. 149-209, 2007).

[19] J. Murdoch and J. A. Barnes, *Statistical Tables*, 4th Ed., (New York, pp.28, 1998).

[20] R.V. Hogg and E. A. Tanis, *Probability and Statistical Inference*, (New York: MacMillan, 1977, pp. 281-285).

[21] C. P. Gan, *Detection of Pneumonia Using Principal Component Analysis*, unpublished M. Sc. Project Thesis, University of Malaya, 2009.

**Appendix I: Definition of Texture Measures Considered**

Suppose the matrix $(C_{jk})$ represent any of LL, LH, HL and HH, then twelve texture measures are defined as follows;

(i) Mean Energy, $E = \frac{1}{N}\sum_j \sum_k |C_{jk}|^2$

Mean of energy is the average of energy or power spectrum of the wavelet coefficients.

(ii) $\text{Entropy} = -\frac{1}{N^2}\sum_j \sum_k |C_{jk}|^2 \log |C_{jk}|^2$

where $\mu = \frac{1}{N^2}\sum_j \sum_k C_{jk}$ is the mean.

(iii) $\text{Contrast} = \sum_j \sum_k (j-k)^2 C_{jk}$

where a high value for contrast means coarse texture and a low value for contrast means less coarseness of the texture.

(iv) Homogeneity, $H = \sum_j \sum_k \frac{C_{jk}}{1+|j-k|}$

(v) Standard deviation of value

$$STDV = \sqrt{\frac{1}{N^2}\sum_j \sum_k (C_{jk} - \mu)}$$

(vi) Standard deviation of mean of energy

$$STDE = \sqrt{\frac{1}{N^2}\sum_j \sum_k (C_{jk} - \mu)^2}$$

(vii) Maximum wavelet coefficient value

$$\max = \max(C_{jk})$$

(viii) Minimum wavelet coefficient value

$$\min = \min(C_{jk})$$

(ix) Maximum value of energy

$$E_{\max} = \max\left(\sum_j \sum_k |C_{jk}|^2\right)$$

(x) Maximum row sum energy,

$$Maxrse = \sum_k^p |C_{jk}|^2$$

(xi) Maximum column sum energy

$$Maxcse = \sum_j^p |C_{jk}|^2$$

(xii) Average number of zero-crossings

**Appendix II: Kolmogorov-Smirnov Test**

The Kolmogorov-Smirnov Test statistic is defined by

$$D_n = \sup_x [|F_n(x) - F_o(x)|]$$

where

$$F_n(x) = \begin{cases} 0, & x < y_1 \\ \frac{k}{n}, & y_k \le x < y_{k+1} \quad, k = 1,2,\cdots,n-1. \\ 1, & y_n \le x \end{cases}$$

and $y_1 < y_2 < \cdots < y_n$ are the order statistics of a random sample $x_1, x_2, \cdots, x_n$. Clearly the empirical distribution $F_n(x)$ is to be compared or tested against the assumed distribution $F_o(x)$. The Appendix Table VII in [20] list critical values of $D_n$.

## Appendix III: Example of Increasing Skewness and Kurtosis

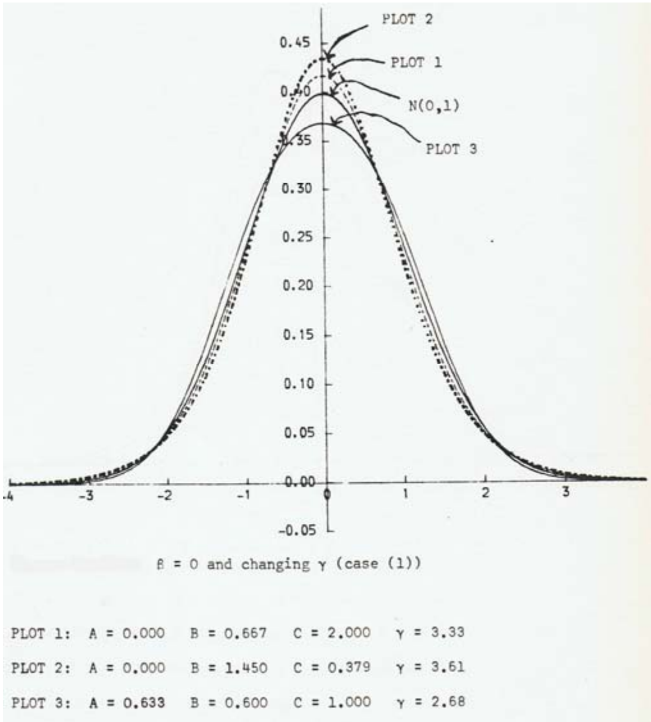



Fig.8 An illustration of increasing kurtosis.

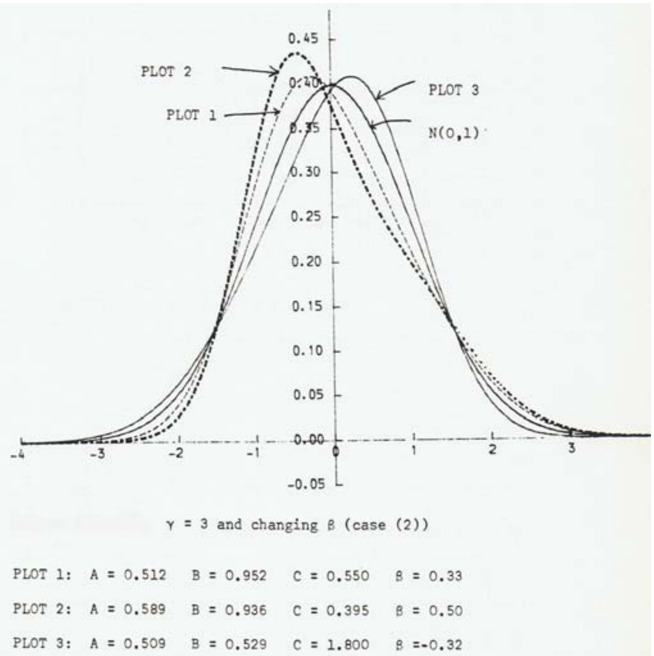



Fig. 9 An illustration of increasing skewness.

Table 2 Probability of misclassifying NL, PTB or LC as PNEU using test data

| Texture Measure | P(PNEU\|NL) | P(PNEU\|PTB) | P(PNEU\|LC) |
|---|---|---|---|
| Mean of Energy | 0 | 0.55 | 0.3 |
| Entropy | 0 | 0.65 | 0.3 |
| Stdev Value | 0 | 0.65 | 0.2 |
| Stdev Energy | 0 | 0.65 | 0.2 |
| Max. Value | 0 | 0.4 | 0.25 |
| Min. Value | 0.05 | 0.45 | 0.35 |
| Max Energy | 0 | 0.45 | 0.25 |
| Maxrse | 0 | 0.55 | 0.25 |
| Maxcse | 0 | 0.65 | 0.25 |
| Zero-Crossing | 0 | 0.65 | 0.55 |
| Contrast | 0 | 0.35 | 0.15 |
| Homogeneity | 0 | 0.6 | 0.25 |
| 12Features | 0 | 0.45 | 0.1 |

Table 3 Kolmogorov-Smirnov test for Chi-squared approximation when data is normal.

| | | $\underline{x} \sim N_2(\underline{\mu}, \Sigma), \underline{\mu} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ | | $\underline{x} \sim N_2(\underline{\mu}, \Sigma), \underline{\mu} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma = \begin{pmatrix} 4 & 0 \\ 0 & 1 \end{pmatrix}$ | |
|---|---|---|---|---|---|
| | | sample | | sample | |
| *n* | critical value | KS Statistic | decision | KS Statistic | decision |
| 15 | 0.338 | 0.1407 | Accept | 0.1694 | Accept |
| 20 | 0.294 | 0.1151 | Accept | 0.137 | Accept |
| 21 | 0.27<x<0.294 | 0.1048 | Accept | 0.1112 | Accept |
| 22 | 0.27<x<0.294 | 0.1107 | Accept | 0.1962 | Accept |
| 23 | 0.27<x<0.294 | 0.141 | Accept | 0.1059 | Accept |
| 24 | 0.27<x<0.294 | 0.0851 | Accept | 0.1116 | Accept |
| 25 | 0.27 | 0.0901 | Accept | 0.127 | Accept |
| 26 | 0.24<x<0.27 | 0.1372 | Accept | 0.0943 | Accept |
| 27 | 0.24<x<0.27 | 0.1221 | Accept | 0.1416 | Accept |
| 28 | 0.24<x<0.27 | 0.1748 | Accept | 0.1078 | Accept |
| 29 | 0.24<x<0.27 | 0.0717 | Accept | 0.1763 | Accept |
| 30 | 0.24 | 0.1675 | Accept | 0.0899 | Accept |
| 35 | 0.23 | 0.1111 | Accept | 0.1539 | Accept |
| 40 | 0.215 | 0.0853 | Accept | 0.0845 | Accept |

Table 4 Kolmogorov-Smirnov test for Chi-square approximation when data is skewed (non-normal) with one outlier

$$outlier = \begin{pmatrix} d \\ 0 \end{pmatrix},\ d_1 = 4 * \sqrt{a^2 + \frac{b}{2}(1+c)} \qquad d_2 = 5 * \sqrt{a^2 + \frac{b}{2}(1+c)}$$

| $n$ + outlier $d$ | beta | KS Statistic for $n + d_1$ | critical value | Decision | KS Statistic for $n + d_2$ | Decision |
|---|---|---|---|---|---|---|
| 16 | 0 | 0.2570 | 0.328 | Accept | 0.2194 | Accept |
| | 0.33 | 0.1383 | | Accept | 0.1413 | Accept |
| | 0.5 | 0.2186 | | Accept | 0.0931 | Accept |
| | -0.32 | 0.1040 | | Accept | 0.1073 | Accept |
| 21 | 0 | 0.1379 | 0.27<x<0.294 | Accept | 0.1874 | Accept |
| | 0.33 | 0.1239 | | Accept | 0.1857 | Accept |
| | 0.5 | 0.1350 | | Accept | 0.1109 | Accept |
| | -0.32 | 0.1335 | | Accept | 0.1157 | Accept |
| 22 | 0 | 0.1421 | 0.27<x<0.294 | Accept | 0.1830 | Accept |
| | 0.33 | 0.0821 | | Accept | 0.2203 | Accept |
| | 0.5 | 0.1544 | | Accept | 0.1536 | Accept |
| | -0.32 | 0.1262 | | Accept | 0.1372 | Accept |
| 23 | 0 | 0.1143 | 0.27<x<0.294 | Accept | 0.1590 | Accept |
| | 0.33 | 0.0730 | | Accept | 0.2144 | Accept |
| | 0.5 | 0.1071 | | Accept | 0.1381 | Accept |
| | -0.32 | 0.1119 | | Accept | 0.1197 | Accept |
| 24 | 0 | 0.2212 | 0.27<x<0.294 | Accept | 0.0980 | Accept |
| | 0.33 | 0.0937 | | Accept | 0.1734 | Accept |
| | 0.5 | 0.1410 | | Accept | 0.1144 | Accept |
| | -0.32 | 0.0958 | | Accept | 0.1270 | Accept |
| 25 | 0 | 0.1033 | 0.27 | Accept | 0.2810 | Reject |
| | 0.33 | 0.0923 | | Accept | 0.1315 | Accept |
| | 0.5 | 0.1430 | | Accept | 0.1537 | Accept |
| | -0.32 | 0.1149 | | Accept | 0.1188 | Accept |
| 26 | 0 | 0.2689 | 0.24<x<0.27 | Reject | 0.1443 | Accept |
| | 0.33 | 0.1141 | | Accept | 0.2076 | Accept |
| | 0.5 | 0.1543 | | Accept | 0.1457 | Accept |
| | -0.32 | 0.0770 | | Accept | 0.1397 | Accept |
| 27 | 0 | 0.2416 | 0.24<x<0.27 | Reject | 0.2995 | Reject |
| | 0.33 | 0.0863 | | Accept | 0.2273 | Accept |
| | 0.5 | 0.1301 | | Accept | 0.1424 | Accept |
| | -0.32 | 0.1370 | | Accept | 0.0910 | Accept |
| 28 | 0 | 0.1677 | 0.24<x<0.27 | Accept | 0.2165 | Accept |
| | 0.33 | 0.1046 | | Accept | 0.2065 | Accept |
| | 0.5 | 0.1835 | | Accept | 0.1538 | Accept |
| | -0.32 | 0.1471 | | Accept | 0.1024 | Accept |
| 29 | 0 | 0.1689 | 0.24<x<0.27 | Accept | 0.1089 | Accept |
| | 0.33 | 0.0755 | | Accept | 0.1466 | Accept |
| | 0.5 | 0.0506 | | Accept | 0.2047 | Accept |
| | -0.32 | 0.1533 | | Accept | 0.1375 | Accept |
| 30 | 0 | 0.1235 | 0.24 | Accept | 0.1954 | Accept |
| | 0.33 | 0.0876 | | Accept | 0.1534 | Accept |
| | 0.5 | 0.1327 | | Accept | 0.1310 | Accept |
| | -0.32 | 0.1203 | | Accept | 0.1207 | Accept |
| 31 | 0 | 0.1274 | 0.23<x<0.24 | Accept | 0.1917 | Accept |
| | 0.33 | 0.1039 | | Accept | 0.2453 | Reject |
| | 0.5 | 0.0763 | | Accept | 0.1539 | Accept |
| | -0.32 | 0.2066 | | Accept | 0.1247 | Accept |
| 36 | 0 | 0.1788 | 0.2267 | Accept | 0.1029 | Accept |
| | 0.33 | 0.0903 | | Accept | 0.1020 | Accept |
| | 0.5 | 0.0898 | | Accept | 0.1460 | Accept |
| | -0.32 | 0.1069 | | Accept | 0.1004 | Accept |
| 41 | 0 | 0.1011 | 0.2124 | Accept | 0.1785 | Accept |
| | 0.33 | 0.1820 | | Accept | 0.1018 | Accept |
| | 0.5 | 0.1040 | | Accept | 0.1113 | Accept |
| | -0.32 | 0.0913 | | Accept | 0.1426 | Accept |

Table 5 Kolmogorov-Smirnov test for Chi-squared approximation when data is skewed (non-normal) with 2 outliers

$$d_3 = \begin{pmatrix} 4*\sqrt{a^2 + \frac{b}{2}(1+c)} & 5*\sqrt{a^2 + \frac{b}{2}(1+c)} \\ 0 & 0 \end{pmatrix}$$

| $n$ + outlier $d_3$ | beta | KS Statistic for $n + d_3$ | critical value | Decision |
|---|---|---|---|---|
| 17 | 0 | 0.1240 | 0.318 | Accept |
| | 0.33 | 0.0813 | | Accept |
| | 0.5 | 0.0993 | | Accept |
| | -0.32 | 0.1930 | | Accept |
| 22 | 0 | 0.1377 | 0.27<x<0.294 | Accept |
| | 0.33 | 0.0730 | | Accept |
| | 0.5 | 0.1256 | | Accept |
| | -0.32 | 0.1268 | | Accept |
| 23 | 0 | 0.2045 | 0.27<x<0.294 | Accept |
| | 0.33 | 0.0953 | | Accept |
| | 0.5 | 0.0733 | | Accept |
| | -0.32 | 0.1730 | | Accept |
| 24 | 0 | 0.1206 | 0.27<x<0.294 | Accept |
| | 0.33 | 0.0872 | | Accept |
| | 0.5 | 0.0877 | | Accept |
| | -0.32 | 0.1039 | | Accept |
| 25 | 0 | 0.1804 | 0.27 | Accept |
| | 0.33 | 0.0917 | | Accept |
| | 0.5 | 0.2237 | | Accept |
| | -0.32 | 0.1246 | | Accept |
| 26 | 0 | 0.1038 | 0.24<x<0.27 | Reject |
| | 0.33 | 0.1349 | | Accept |
| | 0.5 | 0.0825 | | Accept |
| | -0.32 | 0.1215 | | Accept |
| 27 | 0 | 0.1354 | 0.24<x<0.27 | Accept |
| | 0.33 | 0.1714 | | Accept |
| | 0.5 | 0.1236 | | Accept |
| | -0.32 | 0.1273 | | Accept |
| 28 | 0 | 0.1045 | 0.24<x<0.27 | Accept |
| | 0.33 | 0.1462 | | Accept |
| | 0.5 | 0.1191 | | Accept |
| | -0.32 | 0.1156 | | Accept |
| 29 | 0 | 0.1407 | 0.24<x<0.27 | Accept |
| | 0.33 | 0.1229 | | Accept |
| | 0.5 | 0.1242 | | Accept |
| | -0.32 | 0.1418 | | Accept |
| 30 | 0 | 0.1745 | 0.24 | Accept |
| | 0.33 | 0.0591 | | Accept |
| | 0.5 | 0.0992 | | Accept |
| | -0.32 | 0.0740 | | Accept |
| 31 | 0 | 0.2665 | 0.23<x<0.24 | Reject |
| | 0.33 | 0.0965 | | Accept |
| | 0.5 | 0.1367 | | Accept |
| | -0.32 | 0.1131 | | Accept |
| 32 | 0 | 0.1133 | 0.23<x<0.24 | Accept |
| | 0.33 | 0.1088 | | Accept |
| | 0.5 | 0.1326 | | Accept |
| | -0.32 | 0.1584 | | Accept |
| 37 | 0 | 0.2289 | 0.220621133 | Reject |
| | 0.33 | 0.1232 | | Accept |
| | 0.5 | 0.1290 | | Accept |
| | -0.32 | 0.1503 | | Accept |
| 42 | 0 | 0.1316 | 0.207398056 | Accept |
| | 0.33 | 0.1135 | | Accept |
| | 0.5 | 0.1359 | | Accept |
| | -0.32 | 0.1493 | | Accept |

Table 6 Kolmogorov-Smirnov test for Chi-squared approximation when data has kurtosis (non-normal) with 2 outliers

$$d_3 = \begin{pmatrix} 4*\sqrt{a^2 + \frac{b}{2}(1+c)} & 5*\sqrt{a^2 + \frac{b}{2}(1+c)} \\ 0 & 0 \end{pmatrix}$$

| $n$ + outlier $d_3$ | gamma | KS Statistic for $n + d_3$ | critical value | Decision |
|---|---|---|---|---|
| 12 | 3 | 0.1484 | 0.375 | Accept |
| | 3.33 | 0.1440 | | Accept |
| | 3.61 | 0.1088 | | Accept |
| | 2.68 | 0.2455 | | Accept |
| 17 | 3 | 0.2147 | 0.318 | Accept |
| | 3.33 | 0.0975 | | Accept |
| | 3.61 | 0.0906 | | Accept |
| | 2.68 | 0.1195 | | Accept |
| 22 | 3 | 0.1817 | 0.27<x<0.294 | Accept |
| | 3.33 | 0.1179 | | Accept |
| | 3.61 | 0.0924 | | Accept |
| | 2.68 | 0.2367 | | Accept |
| 23 | 3 | 0.1047 | 0.27<x<0.294 | Accept |
| | 3.33 | 0.2008 | | Accept |
| | 3.61 | 0.1194 | | Accept |
| | 2.68 | 0.1396 | | Accept |
| 24 | 3 | 0.2700 | 0.27<x<0.294 | Accept |
| | 3.33 | 0.1391 | | Accept |
| | 3.61 | 0.1536 | | Accept |
| | 2.68 | 0.1914 | | Accept |
| 25 | 3 | 0.1631 | 0.27 | Accept |
| | 3.33 | 0.2483 | | Accept |
| | 3.61 | 0.1050 | | Accept |
| | 2.68 | 0.1099 | | Accept |
| 26 | 3 | 0.1044 | 0.24<x<0.27 | Accept |
| | 3.33 | 0.2047 | | Accept |
| | 3.61 | 0.1047 | | Accept |
| | 2.68 | 0.1108 | | Accept |
| 27 | 3 | 0.1570 | 0.24<x<0.27 | Accept |
| | 3.33 | 0.2494 | | Accept |
| | 3.61 | 0.1508 | | Accept |
| | 2.68 | 0.1417 | | Accept |
| 28 | 3 | 0.1221 | 0.24<x<0.27 | Accept |
| | 3.33 | 0.1685 | | Accept |
| | 3.61 | 0.1200 | | Accept |
| | 2.68 | 0.1114 | | Accept |
| 29 | 3 | 0.2034 | 0.24<x<0.27 | Accept |
| | 3.33 | 0.1858 | | Accept |
| | 3.61 | 0.2417 | | Accept |
| | 2.68 | 0.1887 | | Accept |
| 30 | 3 | 0.1546 | 0.24 | Accept |
| | 3.33 | 0.1390 | | Accept |
| | 3.61 | 0.1709 | | Accept |
| | 2.68 | 0.0966 | | Accept |
| 31 | 3 | 0.1396 | 0.23<x<0.24 | Accept |
| | 3.33 | 0.2040 | | Accept |
| | 3.61 | 0.1483 | | Accept |
| | 2.68 | 0.1051 | | Accept |
| 32 | 3 | 0.2227 | 0.23<x<0.24 | Accept |
| | 3.33 | 0.1948 | | Accept |
| | 3.61 | 0.1417 | | Accept |
| | 2.68 | 0.1765 | | Accept |
| 38 | 3 | 0.2216 | 0.220621133 | Reject |
| | 3.33 | 0.1281 | | Accept |
| | 3.61 | 0.1118 | | Accept |
| | 2.68 | 0.1265 | | Accept |
| 43 | 3 | 0.1451 | 0.207398056 | Accept |
| | 3.33 | 0.0830 | | Accept |
| | 3.61 | 0.2337 | | Reject |
| | 2.68 | 0.1084 | | Accept |