

# 

**Citation:** Tang JR, Mat Isa NA, Ch'ng ES (2016) Evaluating Nuclear Membrane Irregularity for the Classification of Cervical Squamous Epithelial Cells. PLoS ONE 11(10): e0164389. doi:10.1371/journal. pone.0164389

Editor: Silvana Allodi, Federal University of Rio de Janeiro, BRAZIL

Received: May 25, 2016

Accepted: September 23, 2016

Published: October 14, 2016

**Copyright:** © 2016 Tang et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

Funding: This study was partially supported by National Cancer Council Malaysia (MAKNA), Malaysia, under project entitled "Development of an Intelligent Screening System for Cervical Cancer" [http://makna.org.my/]. The fund is received by Nor Ashidi Mat Isa. The study was also partially supported by the Ministry of Higher Education (MOHE), Malaysia under MyPhD Scholarship [https://biasiswa.mohe.gov.my/ MyBrain15/v2/index\_myphd.php]. The fund is **RESEARCH ARTICLE** 

# Evaluating Nuclear Membrane Irregularity for the Classification of Cervical Squamous Epithelial Cells

### Jing Rui Tang<sup>1</sup>, Nor Ashidi Mat Isa<sup>1</sup>\*, Ewe Seng Ch'ng<sup>2</sup>

1 Imaging and Intelligent Systems Research Team, School of Electrical and Electronic Engineering, Universiti Sains Malaysia, Nibong Tebal, Pulau Pinang, Malaysia, 2 Advanced Medical and Dental Institute, Universiti Sains Malaysia, Bertam, Kepala Batas, Pulau Pinang, Malaysia

\* ashidi@usm.my

# Abstract

Pap test involves searching of morphological changes in cervical squamous epithelial cells by pathologists or cytotechnologists to identify potential cancerous cells in the cervix. Nuclear membrane irregularity is one of the morphological changes of malignancy. This paper proposes two novel techniques for the evaluation of nuclear membrane irregularity. The first technique, namely, penalty-driven smoothing analysis, introduces different penalty values for nuclear membrane contour with different degrees of irregularity. The second technique, which can be subdivided into mean- or median-type residual-based analysis, computes the number of points of nuclear membrane contour that deviates from the mean or median of the nuclear membrane contour. Performance of the proposed techniques was compared to three state-of-the-art techniques, namely, radial asymmetric, shape factor, and rim difference. Friedman and post hoc tests using Holm, Shaffer, and Bergmann procedures returned significant differences for all the three classes, i.e., negative for intraepithelial lesion or malignancy (NILM) versus low grade squamous intraepithelial lesion (LSIL), NILM versus high grade squamous intraepithelial lesion (HSIL), and LSIL versus HSIL when the span value equaled 3 was employed with linear penalty function. When span values equaled 5, 7, and 9, NILM versus LSIL and HSIL showed significant differences regardless of the penalty functions. In addition, the results of penalty-driven smoothing analysis were comparable with those of other state-of-the-art techniques. Residual-based analysis returned significant differences for the comparison among the three diagnostic classes. Findings of this study proved the significance of nuclear membrane irregularity as one of the features to differentiate the different diagnostic classes of cervical squamous epithelial cells.

# 1.0 Introduction

Papanicolaou test (Pap test) is a screening test for cervical cancer aiming to identify pre-cancerous and cancerous cells in the cervix. Named after the inventor, George Papanicolaou, Pap test



PLOS ONE

manuscript.
Competing Interests: The authors have declared

that no competing interests exist.

has significantly reduced the mortality rate owing to cervical cancer [1]. The test involves the search of morphological alterations related to malignancy in cervical squamous epithelial cells by pathologists or cytotechnologists. The nuclei of cancerous cells often demonstrate characteristics that are observable under light microscope, such as changes in the nuclear shape, nuclear size, and chromatin distribution [2–6].

The nucleus of non-neoplastic cell is generally round, oval, or bean-shaped. Conversely, nuclear membrane irregularity, an aspect of pleomorphism in terms of variability in shapes, is commonly seen in malignant nucleus [7–16]. Irregularity may appear as nuclear grooving, nuclear molding, or nuclear convolutions [2]. Nuclear shape is controlled by the forces from nuclear envelope, nuclear matrix, and cytoskeleton [17]. Judgment of nuclear membrane irregularity is highly subjective and depends on how individual pathologists interpret and define the shape of the nucleus. Based on the observation of numerous cases, pathologists are trained to differentiate between "normal" and "abnormal" contours of the nuclear membrane. Acknowledging that recognition and discrimination between the "normal" and "abnormal" contours of the nuclear membrane may help in rapid examination of Pap smear, this study aims to propose simple yet effective methods of defining nuclear membrane irregularity and further justify the significance of nuclear membrane irregularity in differentiating the different diagnostic classes of cervical squamous epithelial cells.

This paper presents two techniques to evaluate nuclear membrane irregularity. The first technique assumes that the nuclear membrane contour of a normal cell is smoother than those of the abnormal cell. Therefore, the technique compares the nuclear membrane contour with a "smoothed" contour, which is derived from the original nuclear membrane contour. The second technique is designed based on the assumption that the nucleus of a normal cell is round in shape. The amount of deviation of the nuclear membrane contour from the mean or median value of its nuclear membrane contour will determine to which categories the cell belong (i.e., negative for intraepithelial lesion or malignancy (NILM), low grade squamous intraepithelial lesion (LSIL), or high grade squamous intraepithelial lesion (HSIL)). A total of 600 test images (200 test images from NILM, LSIL, and HSIL, respectively, as attached in S1-S3 Files) were used to evaluate the performance of the proposed techniques. Statistical analysis was performed to seek any significant differences among the three diagnostic classes. Studies have reported various methods of nuclear shape evaluation [18-22]. Performance of the proposed techniques was compared with three techniques of nuclear shape evaluation in the literature, namely, radial asymmetric (RA) [18], shape factor (SF) [18], and rim difference (RD) [23]. Briefly, a technique employing RA initially computes the centroid of the nuclear area to be used as the center to inscribe the largest circle in a nuclear area. RA is the proportion of the nuclear pixels lying outside the inscribed circle [18]. SF, which is mathematically expressed as Eq (1), is another nuclear shape metric in [18]. SF of a perfect circle is one. CYBEST4, which is one of the automated cytologic screening systems, evaluates nuclear shape based on the assumption that the length of the nuclear membrane contour is shorter for regular nucleus compared with irregular nucleus with the same area [23]. Nuclear membrane irregularity is defined in term of RD.

$$SF = \frac{Perimeter^2}{4 \times \pi \times Area} \tag{1}$$

where *Perimeter* and *Area* are the perimeter and area of the nucleus, respectively, and  $\pi$  is a mathematical constant with a value of approximately 3.14159 and represents the ratio of a circle circumference to its diameter.

The next section presents the methodology of the proposed techniques. Section 3 outlines the simulation results followed by the discussions in Section 4. Finally, Section 5 concludes our work.

# 2.0 Methodology

Two techniques, namely, penalty-driven smoothing analysis and residual-based analysis, were proposed to evaluate nuclear membrane irregularity. The study involved three main stages, namely, (1) data acquisition, (2) processing of cervical squamous epithelial cell images, and (3) evaluation of nuclear membrane irregularity. All processing steps of cervical squamous epithelial cell images were performed using MATLAB version R2015a. Details of each stage are presented in the following sub-sections.

# 2.1. Acquisition of Cervical Squamous Epithelial Cell Images

The study was approved by the Human Research Ethics Committee of Universiti Sains Malaysia with the reference code USMKK/PPP/JEPeM[217.4(2.6)]. Human Research Ethics Committee of Universiti Sains Malaysia is listed under the Office for Human Research Protections (OHRP), United States Department of Health and Human Services. The Federal-wide Assurance (FWA) identification number is FWA00007718 and the Institutional Review Board (IRB) number is IRB00004494. ThinPrep slides were borrowed from Penang General Hospital and Tuanku Fauziah Hospital in Malaysia. A total of 102 slides were borrowed (namely, 37 slides from NILM, 42 slides from LSIL, and 23 slides from HSIL). The slides had been previously read and screened by at least a cytotechnologist and a pathologist and formally reported as NILM, LSIL, or HSIL. Cells from NILM, LSIL, and HSIL classes were then individually picked by a cytotechnologist and reconfirmed by a pathologist. The slides were reviewed without knowledge on the patients' background and history; therefore, no consent was obtained from the patients. Cells were selected according to the set of criteria in the Bethesda system [24]. A total of 600 images, consisting of two hundred images for each diagnostic class, were captured from the 102 ThinPrep slides. Images were captured using an Olympus BX43F clinical microscope mounted with a video camera. Every cell image was zoomed with 100× objective with oil immersion.

# 2.2 Processing of Cervical Squamous Epithelial Cell Images

Processing of cervical squamous epithelial cell images included image enhancement and nucleus segmentation. The cervical squamous epithelial cell image that was captured from ThinPrep slide was initially cropped for the nucleus region and then converted from color to gray level image to reduce computational burden. Histogram equalization was then performed to enhance the contrast of the image.

After the image was pre-processed, gradient of the image was computed using the Sobel operator. Mean and standard deviation of the gradient image were computed. The summation and the difference between these mean and standard deviation values were computed as well. If the intensities of the entire gradient image fell in range of the computed difference and the summation, the region consisting of pixels with the intensity equaled to the mean value was taken as nucleus region. Otherwise, the nucleus region was segmented by selecting pixels with intensities that fell in the range of the computed difference and the summation. Morphological closing was employed to fill the small holes in the nucleus region. If more than a single closed region were detected, the region with the largest area was considered as the nucleus. Processing of cervical squamous epithelial cell images is summarized in the flowchart in Fig 1.



Fig 1. Flowchart for the processing of cervical squamous epithelial cell images.

### 2.3 Evaluation of Nuclear Membrane Irregularity

For the evaluation of nuclear membrane irregularity, two techniques were proposed. The first proposed technique, namely, penalty-driven smoothing analysis, smoothed the nuclear membrane contour and employed penalties to the absolute difference between original and smoothed nuclear membrane contours. Averaging filter, a low-pass filter with filter coefficients equaled to the reciprocal of the span, was employed to smooth the nuclear membrane contour. Parameter testing was performed, whereby the span values of 3, 5, 7, and 9 were tested. Absolute difference between the original and smoothed nuclear membrane contour was computed. The ratio between these absolute difference values and the mean value of the absolute difference was obtained. The partitions of nuclear membrane contour that was less than 0.1, fell in

PLOS ONE

Functions	Penalty Value				
	<b>с</b> 1	<b>C</b> 2	<i>C</i> <sub>3</sub>	<i>C</i> <sub>4</sub>	
Linear	1.0	2.0	3.0	4.0	
Quadratic	1.0	4.0	9.0	16.0	
Cubic	1.0	8.0	27.0	64.0	

#### Table 1. Penalty values for different functions.

doi:10.1371/journal.pone.0164389.t001

the range of 0.1 to 0.2, fell in the range of 0.2 to 0.3, and greater than 0.3 were multiplied with different penalty values (namely,  $c_1$ ,  $c_2$ ,  $c_3$ , and  $c_4$  respectively). The penalty values, which were generated from linear, quadratic, or cubic function, were introduced to assign different weights so that a more irregular nuclear membrane contour would receive larger penalty. Penalty values employed for the three functions are as listed in Table 1. The procedure of the proposed penalty-driven smoothing analysis is illustrated in Fig 2.

For a cervical squamous epithelial cell, the centroid of the nucleus region, as represented by coordinate  $(c_x, c_y)$ , was computed as follows:

$$c_x = \frac{\sum_{i=1}^{N} x_i}{N}$$
 2(a)

Algorithm: Penalty-driven smoothing analysis
1: Compute centroid of nucleus, $(c_x, c_y)$ , by using Eq. (2).
2: Compute distance of nuclear membrane contour, $d_i$ , from the centroid of nucleus
3: by using Eq. (3).
4: for span equals to: 3, 5, 7, 9
5: Compute smoothed distance, $d_s$ , according to the rules in Fig. 3.
6: end for
7: Compute absolute difference between $d_i$ and $d_s$ , $diff_i$ , by using Eq. (4).
8: Compute mean, $\mu_{PD}$ of <i>diff</i> and ratio between <i>diff</i> and $\mu_{PD}$ , <i>rat</i> , by using Eqs. (5)
9: and (6).
10: for every i
11: <b>if</b> $rat_i < 0.1$
12: $p_1 = p_1 + 1;$
13: end if
14: <b>if</b> $0.1 < rat_i < 0.2$
15: $p_2 = p_2 + 1;$
16: end if
17: <b>if</b> $0.2 < rat_i < 0.3$
18: $p_3 = p_3 + 1;$
19: end if
20: end for
21: $p_4 = N - p_1 - p_2 - p_3$ ;
22: Update $p_j$ by using Eq. (7).
23: Compute nuclear membrane irregularity value, PD using Eq. (8).
Fig 2. Procedure of the proposed penalty-driven smoothing analysis.

$$c_y = \frac{\sum_{i=1}^{N} y_i}{N}$$
 2(b)

where  $x_i$  and  $y_i$  are the coordinates of *i*-th point of the nuclear membrane contour and the nuclear membrane contour is build-up of *N* points.

The distance of the nuclear membrane contour,  $d_i$ , from the centroid of nucleus was computed using

$$d_i = \sqrt{(c_x - x_i)^2 + (c_y - y_i)^2}$$
 for  $i = 1, 2, \dots, N$  (3)

The smooth distance with different span values was computed according to Fig 3.

if span=3  $d_{\epsilon}(i) = d(i)$ , for i = 1, N $d_{s}(i) = \frac{d(i-1) + d(i) + d(i+1)}{3}$ , for i = 2, 3, ..., N-1End if if span=5  $d_i(i) = d(i)$ , for i = 1, N $d_{s}(2) = \frac{d(1) + d(2) + d(3)}{3}, \text{ for } i = 2, N - 1$  $d_{s}(i) = \frac{d(i-2) + d(i-1) + d(i) + d(i+1) + d(i+2)}{5}, \text{ for } i = 3, 4, \dots, N - 2$ End if if span=7  $d_s(i) = d(i)$ , for i = 1, N $\begin{aligned} & d_{z}(i) = d(i), \text{ for } i = 2, N-1 \\ & d_{z}(2) = \frac{d(i) + d(2) + d(3)}{3}, \text{ for } i = 2, N-1 \\ & d_{z}(i) = \frac{d(i-2) + d(i-1) + d(i) + d(i+1) + d(i+2)}{5}, \text{ for } i = 3, N-2 \\ & d_{z}(i) = \frac{d(i-3) + d(i-2) + d(i-1) + d(i) + d(i+1) + d(i+2) + d(i+3)}{7}, \text{ for } i = 4, 5, ..., N-3 \end{aligned}$ End if if span=9  $d_i(i) = d(i)$ , for i = 1, N $d_s(2) = \frac{d(1) + d(2) + d(3)}{3}$ , for i = 2, N-1 $d_{z}(i) = \frac{d(i-2) + d(i-1) + d(i) + d(i+1) + d(i+2)}{5}, \text{ for } i = 3, N-2$   $d_{z}(i) = \frac{d(i-3) + d(i-2) + d(i-1) + d(i) + d(i+1) + d(i+2) + d(i+3)}{7}, \text{ for } i = 4, N-3$   $d_{z}(i) = \frac{d(i-4) + d(i-3) + d(i-2) + d(i-1) + d(i) + d(i+1) + d(i+2) + d(i+3) + d(i+4)}{9} \text{ for } i = 5, 6, ..., N-4$ End if

#### Fig 3. Computation of smoothed distance by using different span values.

\* Note that  $d(1), d(2), \ldots d(N)$  represent the 1st, 2nd,  $\ldots N$ -th points of the distance of nuclear membrane contour from the centroid;  $d_s$  represents the smoothed distance; and N is the total number of points of the nuclear membrane contour.

After the smoothed distance,  $d_s$ , was obtained using procedure as listed in Fig 3, the absolute difference between the distance of nuclear membrane contour from the centroid and the smoothed distance,  $diff_i$ , was computed using

$$diff_i = |d - d_s| \tag{4}$$

The mean value of the absolute difference,  $\mu_{PD}$ , was computed using

$$\mu_{PD} = \frac{\sum_{i=1}^{N} diff_i}{N}$$
(5)

The ratio between the absolute difference and the mean value of the absolute difference, *rat<sub>i</sub>*, was computed using

$$rat_i = \frac{diff_i}{\mu_{PD}} \tag{6}$$

For normalization, dividing the absolute difference with the mean value of the absolute difference normalized the data for comparisons. Therefore, the calculated ratio can be compared directly for the three diagnostic classes. Partition of nuclear membrane difference,  $p_j$ , that was less than 0.1, fell in the range of 0.1 to 0.2, fell in the range of 0.2 to 0.3, and exceeded 0.3 was computed using

$$p_i = p_j / N \text{ for } j = 1, 2, 3, 4$$
 (7)

where  $p_1$ ,  $p_2$ ,  $p_3$ , and  $p_4$  represent the partition of nuclear membrane difference that is less than 0.1, falls in the range of 0.1 to 0.2, falls in the range of 0.2 to 0.3, and exceeds 0.3, respectively.

Nuclear membrane irregularity, as computed using the penalty driven smoothing analysis, was represented by *PD* and was computed using

$$PD = c_1 p_1 + c_2 p_2 + c_3 p_3 + c_4 p_4 \tag{8}$$

The second technique, namely, residual-based analysis, evaluated nuclear membrane irregularity based on the residuals of nuclear membrane contour. Distance of the nuclear membrane contour from the centroid of the nucleus region was computed. Two types of residual-based analysis were based on the mean and based on the median. For the mean-type residual-based analysis, residuals of the nuclear membrane contour with the mean of the nuclear membrane contour were computed. In contrast, median-type residual-based analysis computed the residuals of the nuclear membrane contour with the median of the nuclear membrane contour. Then, for both techniques, the mean and standard deviation of the residuals were computed to evaluate the nuclear membrane irregularity. Motivated by the idea of the residuals of a perfect circle will be zero, we anticipate that the mean and standard deviation of the residuals for LSIL and HSIL classes will be greater when compared with the residuals for NILM class. Even though the nuclear membrane contour of NILM class might not be a perfect circle, the variation in shape of LSIL and HSIL classes is suspected to be greater. The procedure for the proposed mean- or median-type residual-based analysis is illustrated in Fig 4.

#### Algorithm: Mean- or median-type residual-based analysis

- 1: Compute centroid of nucleus,  $(c_x, c_y)$ , by using Eq. (2).
- 2: Compute distance of nuclear membrane contour, d<sub>i</sub>, from the centroid of
- 3: nucleus using Eq. (3).
- 4: /\* Mean-type \*/
- Compute mean value of d<sub>i</sub>, d<sub>mean</sub>, by using Eq. (9).
- Compute residual between d<sub>i</sub> and d<sub>mean</sub> resi<sub>mean</sub>, by using Eq. (10).
- Compute mean, µ<sub>mean</sub> of resi<sub>mean</sub>, by using Eq. (11).
- Compute standard deviation, σ<sub>mean</sub> of resi<sub>mean\_i</sub>, by using Eq. (12).
- 9: /\* Median-type\*/
- Compute median value of d<sub>i</sub>, d<sub>median</sub>, by using Eq. (13).
- 11: Compute residual between di and dimedian resimedian, by using Eq. (14).
- 12: Compute mean, μ<sub>median</sub> of resi<sub>median</sub>, by using Eq. (15).
- Compute standard deviation, σ<sub>median</sub> of resi<sub>median</sub>, by using Eq. (16).

Fig 4. Procedure of the proposed mean- or median-type residual-based analysis.

doi:10.1371/journal.pone.0164389.g004

The mean value for the distance of nuclear membrane contour,  $d_{mean}$ , was computed as follows:

$$d_{mean} = \frac{\sum_{i=1}^{N} d_i}{N} \tag{9}$$

Residuals between the mean value for the distance of nuclear membrane contour and the distance of nuclear membrane contour, *resi<sub>mean\_i</sub>*, were computed using

$$resi_{mean\_i} = |d_i - d_{mean}| \text{ for } i = 1, 2, \cdots, N$$
(10)

The mean value of the residuals of the mean,  $\mu_{mean}$ , was computed using

$$\mu_{mean} = \frac{\sum_{i=1}^{N} resi_{mean\_i}}{N}$$
(11)

Standard deviation of the residuals of the mean,  $\sigma_{mean}$ , was computed using

$$\sigma_{mean} = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} \left( resi_{mean\_i} - \mu_{mean} \right)^2}$$
(12)

For median-type residual-based analysis, the median value of the distance of nuclear membrane contour,  $d_{median}$ , was computed using

$$d_{median} = \{(N+1) \div 2\} \text{th value}$$
(13)

where the values for distance of nuclear membrane contour are sorted in ascending order.

Residuals between the median value for the distance of nuclear membrane contour and the distance of nuclear membrane contour, *resi*<sub>median</sub>, were computed using

$$\operatorname{resi}_{\operatorname{median}_{i}} = |d_{i} - d_{\operatorname{median}}| \text{ for } i = 1, 2, \cdots, N$$
(14)

The mean value of the residuals of the median,  $\mu_{median}$ , was computed using

$$\mu_{median} = \frac{\sum_{i=1}^{N} resi_{median\_i}}{N}$$
(15)

Standard deviation of the residuals of the median,  $\sigma_{median}$ , was computed using

$$\sigma_{median} = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} \left( resi_{median_i} - \mu_{median} \right)^2}$$
(16)

Graphical illustration for penalty-driven smoothing analysis and residual-based analysis are presented in Fig 5. The original contours of the cells are drawn in black, and the computed contours of the cells are drawn in red. For the penalty-driven smoothing analysis as shown in Fig 5A, the smoothed contour followed closely with the contour of the cell. By contrast, for the residual-based analysis, as shown in Fig 5B, the computed shape is a circle with radius equals the mean or median value of the distance of the nuclear membrane contour.

## 3.0 Results

Examples of nuclear membrane contours of three cervical squamous epithelial cells from three classes are illustrated in Fig 6. Fig 6 shows that the nucleus of cervical squamous epithelial cell from NILM class presents a more regular contour as compared with nuclei from LSIL and HSIL classes. Nuclear membrane contour smoothed using different span values for cells in Fig 6 are shown in Fig 7. As shown in Fig 7, if the nuclear membrane contour is originally smooth (such as NILM cell), smoothing exerts minimal effect on the contour. As a result, the difference between the original and smoothed contours will be small. For LSIL and HSIL cells, the irregularity is greater compared with the NILM cell. The difference between the original and smoothed contours will be large.



Fig 5. Graphical illustration of (a) penalty-driven smoothing analysis and (b) mean- or median-type residual-based analysis on the same cell.





As described in the procedure of the proposed penalty-driven smoothing analysis in Fig 2, smoothing is performed on the distance of nuclear membrane contour from the centroid of nucleus. An example of distance smoothing for a LSIL cell is illustrated in Fig 8. Smoothed



Fig 7. Smoothed nuclear membrane contour of (a) NILM, (b) LSIL, and (c) HSIL by using span = (i) 3, (ii) 5, (iii) 7, and (iv) 9.



Fig 8. (a) Input nucleus. (b) Original distance of nuclear membrane contour from the centroid of nucleus. Smoothed distance using span = (c) 3, (d) 5, (e) 7, and (f) 9.

PLOS ONE







doi:10.1371/journal.pone.0164389.g009

distance at different span values (that is, span equals 3, 5, 7, and 9) for LSIL cell in Fig 6B–6I are shown in Fig 8C to 8F. The smoothed nuclear membrane contour (illustrated in Fig 7) can be reconstructed based on the smoothed distance.

Spread of data for the first proposed technique, the penalty-driven smoothing analysis with four different span values, and three penalty functions are presented in Figs 9–12. For the second technique, the spread of the residuals of the nuclear membrane contour for NILM, LSIL, and HSIL classes from the mean and median values of residuals are presented as boxplots in Figs 13 and 14, respectively. Boxplots of other comparison techniques, namely, RA, SF, and RD, are presented in Fig 15A to 15C, respectively.

For the penalty-driven smoothing analysis, when the span value equaled 3, the median values of the data from NILM class were the minimum among the three classes for all penalty functions. When the span values increased from 5 to 9, both penalty functions and span values demonstrated minimal effect on the spread of data for LSIL and HSIL classes (Figs 10 to 12). Boxplots show similar pattern when employing different penalty functions regardless of span values. NILM class exhibited the largest median value when the span values exceeded 3. Compared with LSIL and HSIL classes, the NILM class showed the smallest range for the data spread. Boxplots in Figs 13 and 14 show that both the mean and standard deviation values of the mean- and median-type residual-based analyses have the smallest range for the NILM class. In addition, residuals of nuclear membrane contour for both techniques were larger for LSIL and HSIL classes, revealing that the shape of the nuclear membranes deviated more in









Fig 11. Penalty-driven smoothing analysis with span = 7 and (a) linear, (b) quadratic, and (c) cubic penalty functions.

these cells. Minimal overlapping of data occurred for the standard deviation of residuals for both techniques, indicating that NILM could be separated from LSIL and HSIL classes based on the standard deviation of residuals. The boxplots in Figs 13 and 14 appear to be similar as the mean and median of the residuals were not significantly different. Other comparison techniques, such as RA, SF, and RD in Fig 15, demonstrated similar trends to those in the proposed techniques, in which the range of the data was the smallest for the NILM class. Results of the proposed techniques were comparable with those of comparison techniques. Overall, NILM can be separated from LSIL and HSIL classes based on the two proposed techniques or other state-of-art techniques. However, overlapping data can be seen for the LSIL and HSIL classes, resulting in difficulty of separating the two classes based on nuclear membrane irregularity.

To investigate whether these data different significantly, Friedman test was employed for multiple comparisons by using KEEL Software tool [25] with the significance level,  $\alpha$  equals 0.05 [26]. Friedman test is a non-parametric procedure that investigates the significance of differences between multiple ranks through ranking of the algorithms, where 1st rank is given to the best performing algorithm. Here, a lower value returned for both proposed techniques is considered a superior value. If Friedman test returned a *p*-value less than 0.05, the measurements are significantly different. Friedman test returned statistically significant differences for all comparisons, except for the penalty-driven smoothing analysis with span value equaled 3 using cubic penalty function (*p*-value = 0.5461) as listed in Table 2. Rejection of the null hypothesis leads to post hoc analysis, which aims to obtain pairwise comparisons that yield





doi:10.1371/journal.pone.0164389.g012

doi:10.1371/journal.pone.0164389.g011



**Fig 13.** Mean-type residual-based analysis. Boxplot of (a) mean and (b) standard deviation of residuals. doi:10.1371/journal.pone.0164389.g013

differences. Holm, Shaffer, and Bergmann procedures were further applied, and the results are summarized in Table 2. These tests were selected based on previous recommendations [26,27], whereby Nemenyi's test is not recommended owing to its conservative nature, but the use of Shaffer and Holm procedures are strongly recommended. In contrast, Bergmann procedure is the best performing approach despite being computationally expensive. Details of Holm, Shaffer, and Bergmann procedures are presented in <u>S1–S14</u> Tables. Holm and Shaffer procedures reject all hypotheses if the corresponding *p*-values are smaller than the adjusted  $\alpha$ 's.

From Table 2, for comparisons of the penalty-driven smoothing analysis using span values equal to 5, 7, and 9, the NILM and LSIL classes and the NILM and HSIL classes presented significant differences. However, LSIL and HSIL classes were not significantly different. A span value of 3 with linear penalty function returned significant difference for all comparisons. By using span value of 3, only one comparison (that is, NILM versus HSIL) returned significant difference when quadratic penalty function was employed, but no significant difference could be found when cubic penalty function was employed. The second proposed technique, i.e., mean- or median-type residual-based analysis, returned significant differences for the comparisons among the three datasets. The three state-of-the-art techniques gave similar results as the majority results from the penalty-driven smoothing analysis, where NILM versus LSIL and









doi:10.1371/journal.pone.0164389.g015

NILM versus HSIL class were statistically different, but not for the LSIL versus HSIL class. Based on simulation results, it can be concluded that the diagnostic criteria on nuclear membrane for NILM, LSIL, and HSIL classes could be measured in a quantitative way to reduce vagueness and further increase the reproducibility of judgment.

### 4.0 Discussion

The nuclei of cervical squamous epithelial cells are nearly spherical, which change to oblate spheroidal due to cells flattening during smear preparation. Under microscope, nuclei appear to have near circular profile (that is, round or oval) [18]. Alteration in nuclear shape is either due to changes in the nuclear lamina or by forces from the cytoplasm [28,29]. Abnormality in nuclear shape is correlated to malignancy [11]. In practice, classification of cervical squamous epithelial cells is performed based on the combinations of several diagnostic criteria [30]. This study focused on one of the features, namely, nuclear membrane irregularity, in differentiating the different diagnostic classes of cervical squamous epithelial cells. The diagnostic criteria of the nuclear membrane in one of the reporting standards, the Bethesda system, are listed in Table 3.

As shown in Table 3, the diagnostic criteria for nuclear membrane are qualitative in nature. Judgment of the descriptive terms, such as smooth and regular, depends on the individual pathologists and is highly subjective depending on his or her skills and experience. Moreover, degree of irregularity is defined through terms such as "quite" and "very". As a result, discrepancies between individual pathologists are unavoidable, and the diagnostic results may lack of reproducibility. Hence, this study suggests techniques to measure the irregularity in a quantitative way.

Pathologists and cytotechnologists examine and observe the cervical squamous epithelial cells in three-dimensional view by adjusting the focus of the microscope. Abnormality in nuclear membrane for LSIL and HSIL cells, which are characterized as nuclear grooving, nuclear molding, or nuclear convolutions, is aggregated in the two-dimensional image. As such, irregularity in shape yields abrupt change in intensities in the two-dimensional image; a nuclear grooving, which is highlighted and shown in the nuclear membrane contour is illustrated in Fig 16. Capturing images of cervical squamous epithelial cells under microscope can be imagined as capturing an inflatable ball (that is, the NILM cell) and a dented ball (that is, LSIL or HSIL cell). Nuclear membrane irregularity, which appears similar to a dented ball, demonstrates changes in intensity gradient in the two-dimensional image.

The current existing techniques rely on the assumption that the nuclear shape is originally round or symmetrical in shape. By applying the same concept, the proposed mean- or median-

No	Technique		<i>p</i> -value	Classes with Significant Difference	
1	Penalty-driven smoothing analysis				
	3 Linear		5.546E-7	NILM versus HSIL, NILM versus LSIL, LSIL versus HSIL	
	3	Quadratic	4.954E-2	NILM versus HSIL	
	3	Cubic	5.461E-1	Nil*	
	5	Linear	6.587E-11	NILM versus HSIL, NILM versus LSIL	
	5	Quadratic	5.809E-11	NILM versus HSIL, NILM versus LSIL	
	5	Cubic	7.650E-11	NILM versus HSIL, NILM versus LSIL	
	7	Linear	4.404E-11	NILM versus HSIL, NILM versus LSIL	
	7	Quadratic	7.623E-11	NILM versus HSIL, NILM versus LSIL	
	7	Cubic	9.359E-11	NILM versus HSIL, NILM versus LSIL	
	9	Linear	6.525E-11	NILM versus HSIL, NILM versus LSIL	
	9	Quadratic	6.632E-11	NILM versus HSIL, NILM versus LSIL	
	9	Cubic	9.692E-11	NILM versus HSIL, NILM versus LSIL	
2	Mean-type residual- based analysis				
	Mean		5.278E-11	NILM versus HSIL, NILM versus LSIL, LSIL versus HSIL	
	Standard Deviation		1.057E-10	NILM versus HSIL, NILM versus LSIL, LSIL versus HSIL	
3	Median-type residual-based analysis				
	Mean		6.264E-11	NILM versus HSIL, NILM versus LSIL, LSIL versus HSIL	
	Stand	dard Deviation	1.165E-10	NILM versus HSIL, NILM versus LSIL, LSIL versus HSIL	
4	RA		8.823E-11	NILM versus HSIL, NILM versus LSIL	
5	SF 1		1.028E-10	NILM versus HSIL, NILM versus LSIL	
6	RD		1.399E-10	NILM versus HSIL, NILM versus LSIL	

 Table 2. P-values of Friedman test and the summary of classes with significant difference. P-value

 greater than 0.05 is italicized.

\* Nil: None of the classes present significant difference.

doi:10.1371/journal.pone.0164389.t002

type residual-based analysis is designed based on the deviation of nuclear membrane contour from the circle with the radius of mean or median value of the distance of nuclear membrane contour. Simulation results of the proposed mean- or median-type residual-based analysis returned statistical difference among the three classes (i.e., NILM, LSIL, and HSIL). However, there is a concern that the relationship between "round" or "symmetry" and the "regular" as defined by the Bethesda system (Table 3) has yet to be verified. Therefore, the uncertainty in

Table 3. Description for the nuclear membrane in the Bethesda system [30].

Class	Page	Description		
NILM	38, 42	Nuclei have smooth contours.		
	40, 50	Nuclear contours are regular.		
LSIL	138	Contour of nuclear membranes is variable ranging from smooth to very irregular with notches.		
HSIL	147	Contour of the nuclear membrane is quite irregular and frequently demonstrates prominent indentations or grooves.		



Fig 16. Circles highlight the irregularity in nuclear membrane contour.

justifying the nuclear shape based on the assumption of "round" or "symmetrical" motivated us to propose the penalty-driven smoothing analysis.

Without making the assumption that the initial shape of the nucleus is round or symmetrical, the proposed penalty-driven smoothing analysis evaluates nuclear membrane irregularity by comparing the original nuclear membrane contour with the smoothed nuclear membrane contour, which is derived from the original nuclear membrane contour. If the nuclear membrane contour is originally smooth, the smoothed profile will be extremely close to the original nuclear membrane contour, resulting in small difference for irregularity measurement. This is a novel approach in evaluating nuclear membrane irregularity as the irregularity is defined based on the smoothness rather than the roundness or symmetrical property.

Different span values were tested in study to investigate the effect of different degrees of smoothing on the evaluation of nuclear membrane irregularity. A suitable span value should yield a smoothed nuclear membrane contour that is loyal to the original nuclear membrane contour and still capable of capturing the point-to-point variance on nuclear membrane contour. A larger span value takes more points into account to smooth the nuclear membrane contour. However, when the span value is extremely large, the smoothed contour would deviate markedly from the original nuclear membrane contour. Although the smoothing effect with different span values are subtle to the naked eye (as shown in Fig 7), results from statistical analysis revealed that span value of 3 was the optimum choice. More specifically, when a span value of 3 was employed with linear penalty function, the proposed penalty driven smoothing analysis showed significant difference among the three diagnostic classes. When span values equaled 5, 7, and 9, NILM class could be separated from LSIL and HSIL classes, but the latter classes were hardly be separated from one another solely based on nuclear membrane irregularity.

For the comparison techniques, RA studies the shape of nucleus via asymmetry, whereas SF studies the degree of deviation from an ideal circle. RA is capable to capture the differences among near-spherical, oval, and irregular shapes of nuclei, whereas SF can hardly distinguish nuclei of oval and irregular shapes [18]. RD could be seen as a variant to the proposed meanor median-type residual-based analysis, which assumes that the length of the nuclear membrane contour is shorter for a regular nucleus. RD defines nuclear membrane irregularity by comparing the length of the nuclear membrane contour with the length of round (regular) nuclei with the same area. These three comparison techniques evaluate nuclear membrane irregularity from different perspectives and thus serve a suitable comparison to our proposed techniques. Simulation results showed that the proposed techniques in this study yielded comparable outputs with the comparison techniques.

In summary, irregularity can be computed through analysis of the variations in nuclear membrane contour. The proposed techniques in this study are simple yet effective for the evaluation of nuclear membrane irregularity. The issue on the assumption on initial nucleus shape is specifically addressed in this study. The proposed techniques can be employed independently to differentiate the different diagnostic classes of cervical squamous epithelial cells.

# 5.0 Conclusion

Nuclear membrane irregularity is one of the morphological changes related to malignancy. This study proposed two techniques, namely, penalty-driven smoothing analysis and residualbased analysis, to evaluate nuclear membrane irregularity. The former employs different penalty values to a more irregular nuclear membrane contour. Residual-based analysis, which consists of two types of analyses (that is, mean- and median-type), evaluates nuclear membrane irregularity through analyzing the residuals of the nuclear membrane contour from the mean or median value of the nuclear membrane contour. Statistical analyses using Friedman tests returned significant difference with *p*-value less than 0.05 for all comparisons except when the span value was 3 with cubic penalty function. Further tests using Holm, Shaffer, and Bergmann procedures for penalty-driven smoothing analysis returned significant differences for all three classes when the span value of 3 was employed with linear penalty function. Comparisons of NILM versus LSIL and HSIL were significant, but that of LSIL versus HSIL was not significantly different when span values equaled 5, 7, and 9 for all penalty functions. The optimum span value was 3 with the linear penalty function. The residual-based analysis produced significant differences for the three diagnostic classes. The proposed techniques addressed the issue on the assumption on nucleus shape. Findings from this study proved the significance of nuclear membrane irregularity in differentiating the different diagnostic classes of cervical squamous epithelial cells.

## **Supporting Information**

S1 File. Dataset of NILM test images. (RAR)
S2 File. Dataset of LSIL test images. (RAR)

**S3 File. Dataset of HSIL test images.** (RAR)

S1 Table. Family of hypotheses ordered by *p*-value and adjusting of  $\alpha$  by Holm and Shaffer procedures, considering an initial  $\alpha = 0.05$  for penalty-driven smoothing analysis with span = 3.

(DOC)

S2 Table. Adjusted *p*-value (APVs) by Holm, Shaffer's static, and Bergmann-Hommel's dynamic (Berg) for penalty-driven smoothing analysis with span = 3. (DOC)

S3 Table. Family of hypotheses ordered by *p*-value and adjusting of  $\alpha$  by Holm and Shaffer procedures, considering an initial  $\alpha = 0.05$  for penalty-driven smoothing analysis with

**span = 5.** (DOC)

S4 Table. Adjusted *p*-value (APVs) by Holm, Shaffer's static, and Bergmann-Hommel's dynamic (Berg) for penalty-driven smoothing analysis with span = 5. (DOC)

S5 Table. Family of hypotheses ordered by *p*-value and adjusting of  $\alpha$  by Holm and Shaffer procedures, considering an initial  $\alpha = 0.05$  for penalty-driven smoothing analysis with span = 7.

(DOC)

S6 Table. Adjusted *p*-value (APVs) by Holm, Shaffer's static, and Bergmann-Hommel's dynamic (Berg) for penalty-driven smoothing analysis with span = 7. (DOC)

S7 Table. Family of hypotheses ordered by *p*-value and adjusting of  $\alpha$  by Holm and Shaffer procedures, considering an initial  $\alpha = 0.05$  for penalty-driven smoothing analysis with span = 9.

(DOC)

S8 Table. Adjusted *p*-value (APVs) by Holm, Shaffer's static, and Bergmann-Hommel's dynamic (Berg) for penalty-driven smoothing analysis with span = 9. (DOC)

S9 Table. Family of hypotheses ordered by *p*-value and adjusting of  $\alpha$  by Holm and Shaffer procedures, considering an initial  $\alpha = 0.05$  for mean-type residual-based analysis. (DOC)

**S10** Table. Adjusted *p*-value (APVs) by Holm, Shaffer's static, and Bergmann-Hommel's dynamic (Berg) for mean-type residual-based analysis. (DOC)

S11 Table. Family of hypotheses ordered by *p*-value and adjusting of  $\alpha$  by Holm and Shaffer procedures, considering an initial  $\alpha = 0.05$  for median-type residual-based analysis. (DOC)

S12 Table. Adjusted *p*-value (APVs) by Holm, Shaffer's static, and Bergmann-Hommel's dynamic (Berg) for median-type residual-based analysis. (DOC)

S13 Table. Family of hypotheses ordered by *p*-value and adjusting of  $\alpha$  by Holm and Shaffer procedures, considering an initial  $\alpha = 0.05$  for RA, SD and RD techniques. (DOC)

S14 Table. Adjusted *p*-value (APVs) by Holm, Shaffer's static, and Bergmann-Hommel's dynamic (Berg) for RA, SD and RD techniques. (DOC)

## Acknowledgments

This study was partially supported by National Cancer Council Malaysia (MAKNA), Malaysia, under project entitled "Development of an Intelligent Screening System for Cervical Cancer" and the Ministry of Higher Education (MOHE), Malaysia, under MyPhD Scholarship. Credits are given to Dr. Yasmin Sulaiman, Dr. Soo Peng Yen, Dr. Farisyah Syakirah Mohd Zain, and

Dr. Hairulhasliza Harun for providing the ThinPrep slides. Sincere thanks to Mr. Mohd Sobriey Zakaria for his assistance on capturing the sample images.

### **Author Contributions**

Conceptualization: JRT NAMI ESC.

Data curation: JRT NAMI ESC.

Formal analysis: JRT NAMI ESC.

Funding acquisition: JRT NAMI.

Investigation: JRT NAMI ESC.

Methodology: JRT NAMI ESC.

Project administration: JRT NAMI ESC.

Resources: JRT NAMI ESC.

Software: JRT NAMI.

Supervision: NAMI ESC.

Validation: JRT ESC.

Visualization: JRT NAMI ESC.

Writing - original draft: JRT NAMI ESC.

Writing - review & editing: JRT NAMI ESC.

#### References

- 1. Sherris J, Herdman C, Elias C (2001) Cervical cancer in the developing world. Western Journal of Medicine 175: 231–233. PMID: 11577044
- Dey P (2010) Cancer nucleus: Morphology and beyond. Diagnostic Cytopathology 38: 382–390. doi: 10.1002/dc.21234 PMID: 19894267
- Talukdar NA, Deb D, Roy S (2014) Automated blood cancer detection using image processing based on fuzzy system. International Journal of Advanced Research in Computer Science and Software Engineering 4: 467–472.
- Sero JE, Sailem HZ, Ardy RC, Almuttaqi H, Zhang T, Bakal C (2015) Cell shape and the microenvironment regulate nuclear translocation of NF-κB in breast epithelial and tumor cells. Molecular Systems Biology 11: n/a-n/a.
- 5. Chen S, Zhao M, Wu G, Yao C, Zhang J (2012) Recent advances in morphological cell Image analysis. Computational and Mathematical Methods in Medicine 2012: 10.
- Tang JR, Mat Isa NA, Ch'ng ES (2015) A fuzzy-c-means-clustering approach: Quantifying chromatin pattern of non-neoplastic cervical squamous cells. PLoS ONE 10: e0142830. doi: 10.1371/journal. pone.0142830 PMID: 26560331
- 7. Jogai S, Al-Jassar A, Temmim L, Dey P, Adesina AO, G AH (2005) Fine needle aspiration cytology of the thyroid. Acta Cytologica 49: 483–488.
- Dey P, Luthra UK, Sheikh ZA (1999) Fine needle aspiration cytology of Langerhans cell histiocytosis of the thyroid. Acta Cytologica 43: 429–431. PMID: 10349374
- Rajesh L, Dey P, Joshi K (2003) Fine needle aspiration cytology of lobular breast carcinoma. Acta Cytologica 47: 177–182. PMID: 12685185
- Gisselsson D, Björk J, Höglund M, Mertens F, Dal Cin P, Åkerman M, et al. (2001) Abnormal nuclear shape in solid tumors reflects mitotic instability. The American Journal of Pathology 158: 199–206. doi: 10.1016/S0002-9440(10)63958-2 PMID: 11141493
- 11. Zink D, Fischer AH, Nickerson JA (2004) Nuclear structure in cancer cells. Nature Reviews Cancer 4: 677–687. doi: 10.1038/nrc1430 PMID: 15343274

- Li Y, Naveed H, Liang J, Xu LX. Effects of mechanical properties on tumor invasion: Insights from a cellular model; 2014 26–30 Aug. 2014. pp. 6818–6821.
- Thiran JP, Macq B (1996) Morphological feature extraction for the classification of digital images of cancerous tissues. IEEE Transactions on Biomedical Engineering 43: 1011–1020. doi: 10.1109/10. 536902 PMID: 9214818
- Chaddad A, Tanougast C, Golato A, Dandache A (2013) Carcinoma cell identification via optical microscopy and shape feature analysis. Journal of Biomedical Science and Engineering 6: 1029– 1033.
- 15. Wee A, Sampatanukul P (2015) Cytohistology of hepatocellular carcinoma. Diagnostic Histopathology 21: 470–480.
- Ginzburg S, Uzzo R, Al-Saleem T, Dulaimi E, Walton J, Corcoran A, et al. (2014) Coexisting hybrid malignancy in a solitary sporadic solid benign renal mass: Implications for treating patients following renal biopsy. The Journal of Urology 191: 296–300. doi: 10.1016/j.juro.2013.07.059 PMID: 23899990
- Imbalzano KM, Cohet N, Wu Q, Underwood JM, Imbalzano AN, Nickerson JA (2013) Nuclear shape changes are induced by knockdown of the SWI/SNF ATPase BRG1 and are independent of cytoskeletal connections. PLoS ONE 8: e55628. doi: 10.1371/journal.pone.0055628 PMID: 23405182
- Nikonenko AG, Bozhok YM (2015) Simple computational technique to quantify nuclear shape asymmetry Cytometry Part A 37: 309–314.
- Panno JP (1988) Symmetry analysis of cell nuclei. Cytometry 9: 195–200. doi: 10.1002/cyto. 990090302 PMID: 3378457
- Thiran JP, Macq B, Mairesse J. Morphological classification of cancerous cells; 1994 13–16 Nov 1994. pp. 706–710 vol. 703.
- Al-Mamun MA, Ravenhill L, Hossain MA, Farid DM, Bass R. Fractal and image analysis of cytoskeletal changes in tumour cells due to the effects of maspin; 2014 1–4 June 2014. pp. 209–212.
- Xing F, Yang L. Unsupervised shape prior modeling for cell segmentation in neuroendocrine tumor; 2015 16–19 April 2015. pp. 1443–1446.
- Tanaka N, Ikeda H, Ueno T, Mukawa A, Watanabe S, Okamoto K, et al. (1987) Automated cytologic screening system (CYBEST model 4): an integrated image cytometry system. Applied Optics 26: 3301–3307. doi: 10.1364/AO.26.003301 PMID: 20490057
- 24. Solomon D, Nayar R (2004) The Bethesda system for reporting cervical cytology. New York: Springer.
- Alcalá-Fdez J, Sánchez L, García S, del Jesus MJ, Ventura S, Garrell JM, et al. (2009) KEEL: a software tool to assess evolutionary algorithms for data mining problems. Soft Computing 13: 307–318.
- 26. Derrac J, García S, Molina D, Herrera F (2011) A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms. Swarm and Evolutionary Computation 1: 3–18.
- García S, Herrera F (2009) An extension on "statistical comparisons of classifiers over multiple data sets" for all pairwise comparisons. Journal of Machine Learning Research 9: 2677–2694.
- Webster M, Witkin KL, Cohen-Fix O (2009) Sizing up the nucleus: nuclear shape, size and nuclearenvelope assembly. Journal of Cell Science 122: 1477–1486. doi: <u>10.1242/jcs.037333</u> PMID: 19420234
- Stewart CL, Roux KJ, Burke B (2007) Blurring the boundary: The nuclear envelope extends its reach. Science 318: 1408–1412. doi: 10.1126/science.1142034 PMID: 18048680
- 30. Nayar R, Wilbur DC (2015) The Bethesda system for reporting cervical cytology: Springer. 342 p.