Cervical single cell of squamous intraepithelial lesion classification using shape features and extreme learning machine

by R Riries

PAPER · OPEN ACCESS

Cervical single cell of *squamous intraepithelial lesion* classification using shape features and extreme learning machine

To cite this article: R Riries et al 2021 *J. Phys.: Conf. Ser.* **1816** 012081

View the article online for updates and enhancements.

Cervical single cell of *squamous intraepithelial lesion* classification using shape features and extreme learning machine

R Riries[1], Winarno[1], C N Asiah[1], A Y Putri[1], A B Suksmono[2], I S Sitanggang[3], N A Setiawan[4], A S Rahaju[5] and E H Kusumastuti[5]

[1]Departement of Physics, Faculty of Science and Technology, Universitas Airlangga, Surabaya, Indonesia
[2]School of Electrical Engineering and Informatics, Institut Teknologi Bandung, Bandung, Indonesia
[3]Department of Computer Science, IPB University, Bogor, Indonesia
[4]Electrical and Information Engineering, Universitas Gadjah Mada, Jogjakarta
[5]Department of Anatomical Pathology, Faculty of Medicine, Universitas Airlangga, Dr. Soetomo General Academic Hospital, Universitas Airlangga Hospital

E-mail: riries-r@fst.unair.ac.id

Abstract. Cervical cancer is an abnormal growth of cells found on the cervix. In general, cervical cancer is identified early by doing a pap smear test. However, this examination is still manually performed by doctors and the results are still subjective. Therefore, this study aims to determine the classification of Squamous Intraepithelial Lesion automatically from cervical single cells. The classification of those Squamous Intraepithelial Lesion includes normal cervical cells, Low-Grade Squamous Intraepithelial Lesion (LSIL), and High-Grade Squamous Intraepithelial Lesion (HSIL). We used Extreme Learning Machine (ELM) as a classifier and tried to compare the ELM's performances with Backpropagation Neural Network method. We used 225 data and 3 classes include normal, LSIL, and HSIL. The classification was carried out by manual cropping and segmentation as the image pre-processing and the feature extraction was based on shape features consisting of Circularity, Semi Major and Minor Axis Length, Equivalent Diameter, Average Radius, and Compactness. This study concluded that Squamous Intraepithelial Lesion classification by using ELM had better performances than Backpropagation Neural Network. The highest accuracy result of 96.67% was obtained in Backpropagation training, while the highest accuracy in ELM's training was 100% when both methods were tried by using 225 data.

1. Introduction

Cancer is a disease that arises from the abnormal growth of body tissue cells that turn into cancer cells. From several types of cancer, cervical cancer ranks fourth in incidence after breast cancer in developing countries [1]. Cervical cancer is generally caused by the Human Papilloma Virus (HPV) where this virus causes changes in cell DNA. This causes cell growth to occur continuously so that early detection is necessary. In addition, the cause of delay in the diagnosis of cervical cancer is a symptom that is not visible and only obvious when it is in its final stage. To detect cervical cancer,

generally, an early examination or pap smear can be done. Pap smear is very important in reducing the incidence of cervical cancer [2]. This examination requires medical personnel to get an accurate diagnosis. However, medical personnel still analyse the result visually, so the results are subjective. Therefore it is necessary to have an automatic analysis in carrying out the process of diagnosing cervical cancer as a second opinion for doctors, so that it is expected that an accurate diagnosis of cervical cancer cells can be established.

Research on the identification of cervical cancer cells has been carried out in previous studies using SVM, k-NN and ANN to identify normal and abnormal cells. The feature extraction method used is the Gray Level Cooccurence Matrix (GLCM). The accuracy results of this study are 86% with SVM, 70% with KNN, and 65% with ANN. Subsequent research used morphological feature extraction and image classification using the k-nearest neighbor (kNN) method, this study generated the accuracy of 82.9% with 5 Fold Cross validation [3]. Another research employed the random forest method and resulted in 81.71% of accuracy. To identify cells and feature extraction, the study used the Gray Level Cooccurence Matrix, local binary pattern and tamura [4]. Subsequent research used the multilayer perceptron method to identify normal and abnormal cells. This study uses the extraction method of morphological features such as size, shape and texture with an accuracy of 85.05% [5]. Another research employed the fuzzy min-max neural network classification method to classify normal cells, low-grade squamous intra-epithelial lesions (LSIL) and high-grade squamous intra-epithelial lesions (HSIL) while the feature extraction method used is adaptive fuzzy. moving k-means (AFMKM). The level of accuracy obtained in this study is 75% [6].

We used the Extreme Learning Machine (ELM) method and compared the result with the *Backpropagation*, that is the commonly used image classification method. The ELM has the advantage of being able to work optimally even on complex functions with linear and non-linear data [7]. ELM has a good learning ability in generating image classification with accurate results. In this study, the ELM method was used for image classification of Squamous Intraepithelial Lesions automatically from cervical single cells.

2. Method

2.1 Datasets
Cervical cancer Pap smear image data were obtained from Dr. Soetomo General Hospital. The data taken is an image of Squamous Intraepithelial Lesion consisting of Normal class, LSIL (Low-Grade SIL), HSIL (High-Grade SIL) with jpg format. A total of 225 cervical cancer Pap smear image data were used. Training process used 180 data and 45 data for the testing process.
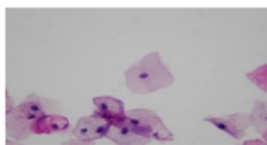

Figure 1. Normal Pap Smear Image
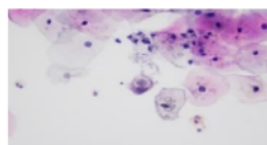
Figure 2. LSIL (Low Grade Squamous Intraephitelial Lesion) Pap Smear Image
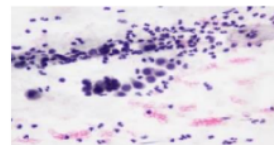
Figure 3. HSIL (High-Grade Squamous Intraephitelial Lesion) Pap Smear Image

2.2 Feature Extraction
The shape features to be extracted are: Circularity, Semi Major and Minor Axis Length, Equivalent Diameter, Average Radius, Compactness. Cervical cancer cells have a similar shape. However, abnormal cervical cancer cells have a shape that is larger and more irregular than normal cells. At the feature extraction stage, there are 8 parameter calculations in the nucleus and cytoplasm shown in Table 1:

Table 1. Feature Shape [8]

Feature	**Shape**
Semi minor-Axis length	b = shortest distance between nuclei centroid and boundary
Semi major-Axis length	a = longest distance between nuclei centroid and boundary
Average Radius	Avr = average distance between nuclei centroid and boundary
Eccentricity	$c = \sqrt{a^2 - b^2}$; $e = c\,/\,a$
Equivalent Diameter	$ED = \dfrac{4 \times Area}{\pi}$
Perimeter	P = sum of pixels in nuclei boundary
Circularity	$C = \dfrac{4 \times Area}{\pi}$
Compactness	$Cp = P^2\,/\,Area$

2.3 ELM Classification

Extreme Learning Machine (ELM) is a supervised algorithm in the Artificial Neural Network group which is a type of single hidden layer feedforward network (SFLN). The ELM algorithm was introduced by by Huang et al. in 2006. ELM can perform the process of generalizing data in a relatively short time compared to the Support Vector Machine algorithm and the Backpropagation algorithm [9].

The ELM method has a different mathematical model from the feedforward neural network. The mathematical model of ELM is simpler and more effective. For N different number of input pairs and output targets (x_i, t_i), with $xi = [x_{i1}, x_{i2}, ..., x_{in}]\,T \in \boldsymbol{R}n$ and $ti = [t_{i1},t_{i2},...,t_{in}]T \in \boldsymbol{R}m$, SLFNs standard with the number of hidden nodes as much as $\tilde{N}$ and the activation function $g\,(x)$ can be modelled mathematically as follows [10]:

$$\sum_{i=1}^{\tilde{N}} \beta_i g_i(x_j) = \sum_{i=1}^{\tilde{N}} \beta_i g_i(w_i . x_j + b_i) = o_j, j = 1,2, ..., N \qquad (2.5)$$

with

$wi = [w_{i1}, w_{i2},....]T$ is a weight vector connecting the ith hidden node and the input nodes.
$\beta i = [\beta_{i1}, \beta_{i2},....]T$ is a weight vector connecting the ith hidden node and the output nodes.
bi is *threshold* from *hidden node* to-i.
$w_i \,.\, x_j$ is *inner product* from w_i and x_j

Standard SLFNs with $\tilde{N}$ hidden nodes and the activation function (x) are assumed to be able to estimate N of this sample with an error rate of 0 which means $\sum_{j=1}^{N} \|oj - tj\| = 0$, so there is β_i, w_i, dan b_i such that:

$$\sum_{i=1}^{\tilde{N}} \beta_i g_i(w_i . x_j + b_i) = t_j, \; j = 1,2, ..., N \qquad (2.6)$$

The equation above can be written simply as:

$$H\beta = T \qquad (2.7)$$

with

$$H = \begin{bmatrix} g(w_1.x_1 + b_1) & ... & g(w_{\tilde{N}}.x_1 + b_{\tilde{N}}) \\ \vdots & \vdots & \vdots \\ g(w_1.x_N + b_1) & ... & g(w_{\tilde{N}}.x_1 + b_{\tilde{N}}) \end{bmatrix} \qquad (2.8)$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_N^T \end{bmatrix} \qquad (2.9)$$

$$T = \begin{bmatrix} t_1^T \\ \vdots \\ t_{\tilde{N}}^T \end{bmatrix} \qquad (2.10)$$

H in the equation (2.8) is the hidden layer output matrix of the neural network. $g\,(wi.\,xj + bi)$ shows the output of hidden neurons related to the input xj. β is the matrix of the output weight and T the matrix of the target. In ELM, the input weight and hidden bias are determined randomly, so that the output weight associated with the hidden layer can be determined from the equation:

$$\beta = H^\dagger T \qquad (2.11)$$

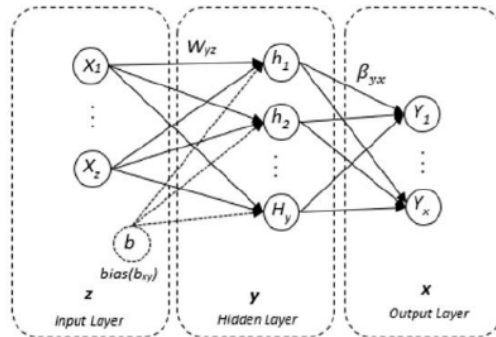The structure of the ELM that can be observed is as shown in Figure 4:

Figure 4. Extreme Learning Machine Structure [11]

Data normalization is carried out so that the range between data is not too far away so that when the data is processed, accurate results will be obtained.

3. Result and Discussion

3.1 Segmentation
Prior to the extraction of shape features, manual segmentation was carried out through cropping. In one field of view there are several cells, so cropping is done to make some data where there are several cells that are indicated of the same type. The dimensions of the cropping result are 431 x 432 pixels shown in Figure 5.
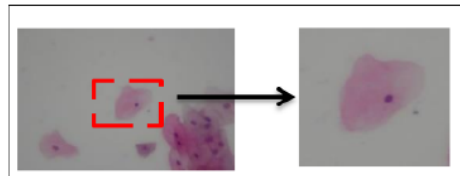

Figure 5. Before and after cropping

4

The resulting image from manual segmentation in the form of RGB is converted into a ycbrc image. This is done to differentiate between the nucleus and cytoplasmic images so that they can be selected and circled in the nucleus and cytoplasm by using the regionprops toolbox. After that the program can calculate 8 parameters from the shape of the cytoplasmic cells and the shape of the nucleus cells. Then the results of the average value for calculation of parameter calculations can be seen in Figure 6 and Figure 7.

3.2 Future Extraction

Feature extraction is the stage of taking important characters (features) in cells after the image segmentation process. In this study, using shape features consisting of 8 parameters of the shape features. Among others, namely: Circularity, Semi Major and Minor Axis Length, Equivalent Diameter, Average Radius, Compactness.
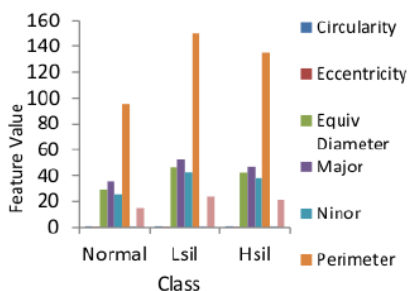

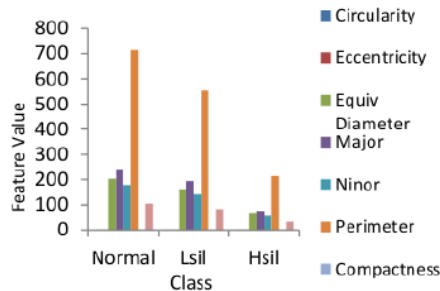**Figure 6.** Shape feature extraction average in nuclei

Figure 7. Shape feature extraction average in cytoplasm

The results of segmentation and feature extraction on the image show the difference between normal epithelial cells and abnormal epithelial cells in size and shape. Figure 7 depicted in the graph shows that the average value of normal epithelial cells (nuclei) and abnormal epithelial cells has a similar shape. Nor in the regularity of its form. The graph of each parameter increases, this proves that in normal nucleus cells the size and shape is smaller than the size and shape of abnormal nucleus cells. So, it can be concluded from the extraction results that the higher the abnormal class the nucleus, the bigger the shape and size of the nucleus. Abnormal epithelial cells have a more irregular shape than normal epithelial cells. On the other hand, in cytoplasmic cells, the more abnormal cervical cells are, the smaller their shape and size, or even until they are not visible as shown in Figure 8, the graph of each parameter shows a decrease, this proves that the size and shape of the cytoplasm is getting smaller than that. normal cells to hsil class.

3.3 Training and Testing Process

In this study, k-fold cross validation was used as a method of dividing data into k sub-data, namely 5-fold cross validation It can be seen that the results shown in Table 2 of ELM have an accuracy of 100% on fold 1 with a training time of 6 seconds. Backpropagation has accuracy of 96.67% with 12 seconds of Training Time.

As for the results of the ELM testing process, it is known that the accuracy of ELM is done by doing 5-Fold Cross Validation in order to get a pretty good validation accuracy and the validation results are obtained at 95%. Meanwhile, based on Table 2, the backpropagation training stage is the result of program testing for normal images, LSIL, and HSIL. The highest accuracy at the testing stage is in the 4th fold, which is 93.89%. It can be seen that ELM and backpropagation neural networks are

able to distinguish pap smear images in normal, LSIL, and HSIL classes by using shape feature extraction.

Table 2. ELM and Backpropagation Training Process Result

Fold-n	ELM			Backpropagation		
	Accuracy of Training (%)	Training Time(s)	Accuracy of Testing (%)	Accuracy of Training (%)	Training Time(s)	Accuracy of Testing (%)
1	100	6	95	92.22	9	92.22
2	90	0	84	93.89	9	93.33
3	88	0	88	95.56	9	93.33
4	87	1	77	95.56	9	93.89
5	91.6	0	80	96.67	12	86.67

4. Conclusion

The classification of Squamous Intraepithelial Lesions using ELM has a better performance than Backpropagation Neural Network. The result of ELM training time is 6 seconds and Backpropagation for 12 seconds. It can be seen that the ELM computation is faster so that the running time of the program during training and testing does not last long. It is also known that the highest accuracy results of 96.67% were obtained in Backpropagation training, while the highest accuracy in ELM training was 100% when both methods were tried using 225 data.

References

[1] Kamariyah, Nurlinawati, Yusnilawati 2019 Pendidikan Kesehatan Dan Pemeriksaan Iva Pada Ibu-Ibu Pus Sebagai Upaya Deteksi Dini Ca.Cervik Di Wilayah Kerja Puskesmas Sungai Duren Kabupaten Muaro Jambi **3**(1)

[2] Mastutik G et.al. 2015 Skrining Kanker Serviks dengan Pemeriksaan Pap Smear di Puskesmas Tanah Kali Kedinding Surabaya dan Rumah Sakit Mawadah Mojokerto *Majalah Obstetri & Ginekologi* **23**(2) 54–60

[3] Sharma M, Singh S K, Agrawal P and Madaan V 2016 Classification of clinical dataset of cervical cancer using knn *Indian Journal of Science and Technology* 1-5

[4] Ashok B, Nair S H, Puviarasan N dan Aruna P 2016 Diagnosis of cervical cancer using global and local shape features *International Journal of Innovative Research in Computer and Communication Engineering* 19709-19718l

[5] Hemalatha K and Rani D U 2016 Improvement of multilayer perceptron classfication on cervical pap smear data with feature extraction *International Journal of Innovative Research in Science, Engineering and Technology* 20419-20424

[6] Quteishat A, Al-Batah M and Al-Mofleh A 2013 Cervical cancer diagnostic system using adaptive fuzzy moving k-means algorithm and fuzzy min-max neural network *Journal of Theoretical and Applied Information Technology* **57**(1) 48-53

[7] Hidayah U R, Cholissodin I and Adikara P P 2019 *Klasifikasi Penyakit Kanker Serviks dengan Extreme Learning Machine* **3**(7)

[8] Kurniawan R, Sasmito D E K and Suryani F 2013 *Ekstraksi dan Seleksi Fitur untuk Klasifikasi Sel Epitel dengan Sel radang pada Citra PapSmear*

[9] Huang G B, Zhu Q Y and Siew C K 2006 *Extreme Learning Machine: theory and applications*. *Neurocomputing* **70**(1) 489-501.

[10] Fikriya Z A, Irawan M I and Soetrisno 2017 *Implementasi Extreme Learning Machine untuk Pengenalan Objek Citra Digital* **6**(1)

[11] Irawan D P D, Cholissodin I and Santoso Edy 2018 Klasifikasi Risiko Gagal Ginjal Kronis Menggunakan Extreme Learning Machine *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer Universitas Brawijaya*

Cervical single cell of squamous intraepithelial lesion classification using shape features and extreme learning machine

6 Ajaya Kumar Barik, Sanoop Pavithran M, Jijo Lukose, Rekha Upadhya, Muralidhar V Pai, V.B. Kartha, Santhosh Chidangil. " spectroscopy: optical fiber probes for clinical applications ", Expert Review of Medical Devices, 2022
Publication

1 %

7 Nan Liu, Zhiping Lin, Zhixiong Koh, Guang-Bin Huang, Wee Ser, Marcus Eng Hock Ong. "Patient Outcome Prediction with Heart Rate Variability and Vital Signs", Journal of Signal Processing Systems, 2010
Publication

1 %

8 Guang-Bin Huang. "Protein sequence classification using extreme learning machine", Proceedings 2005 IEEE International Joint Conference on Neural Networks 2005, 2005
Publication

1 %

9 Rohit Kundu, Soham Chattopadhyay. "Deep features selection through genetic algorithm for cervical pre-cancerous cell classification", Multimedia Tools and Applications, 2022
Publication

1 %

10 proteomesci.biomedcentral.com
Internet Source

1 %

11 K.T. Efendiev, P.M. Alekseeva, A.A. Shiryaev, A.S. Skobeltsin, I.L. Solonina, A.S. Fatyanova,

1 %

I.V. Reshetov, V.B. Loschenov. "Preliminary low-dose photodynamic exposure to skin cancer with chlorin e6 photosensitizer", Photodiagnosis and Photodynamic Therapy, 2022
Publication

12   Nurhayati, Indratmo Soekarno, Iwan K. Hadihardaja, M. Cahyono. "A study of hold-out and k-fold cross validation for accuracy of groundwater modeling in tidal lowland reclamation using extreme learning machine", 2014 2nd International Conference on Technology, Informatics, Management, Engineering & Environment, 2014   1 %
Publication

13   www.co.oneida.ny.us   1 %
Internet Source

14   doaj.org   1 %
Internet Source

15   ijmrr.medresearch.in   1 %
Internet Source

16   polymerscience.imedpub.com   1 %
Internet Source

17   International Journal of Clothing Science and Technology, Volume 26, Issue 5 (2014-09-16)   <1 %
Publication

jmss.mui.ac.ir

18   Internet Source   <1%

19   Bing Liu, Shi-Xiong Xia, Fan-Rong Meng, Yong Zhou. "Manifold regularized extreme learning machine", Neural Computing and Applications, 2015   <1%
Publication

20   Hieu Trung Huynh, Yonggwan Won. "Hematocrit estimation from compact single hidden layer feedforward neural networks trained by evolutionary algorithm", 2008 IEEE Congress on Evolutionary Computation (IEEE World Congress on Computational Intelligence), 2008   <1%
Publication

21   Lei Chen. "Universal Approximation and QoS Violation Application of Extreme Learning Machine", Neural Processing Letters, 10/2008   <1%
Publication

22   Wasswa William, Andrew Ware, Annabella Habinka Basaza-Ejiri, Johnes Obungoloch. "A review of Image Analysis and Machine Learning Techniques for Automated Cervical Cancer Screening from pap-smear images", Computer Methods and Programs in Biomedicine, 2018   <1%
Publication

23 pcmp.springeropen.com
Internet Source
<1%

24 www.iaeng.org
Internet Source
<1%

25 www.spiedigitallibrary.org
Internet Source
<1%

26 Ajaya Kumar Barik, Sanoop Pavithran M, Mithun N, Muralidhar V Pai et al. "A micro-Raman spectroscopy study of inflammatory condition of human cervix: Probing of tissues and blood plasma samples", Photodiagnosis and Photodynamic Therapy, 2022
Publication
<1%

27 Chao Wang, Jianhui Wang, Shusheng Gu. "Deep Network Based on Stacked Orthogonal Convex Incremental ELM Autoencoders", Mathematical Problems in Engineering, 2016
Publication
<1%

28 Hieu Trung Huynh, Nguyen H. V., Minh-Tuan T. Hoang, Yonggwan Won. "Outlier Treatment for SLFNs in Classification", 2007 International Conference on Computational Science and its Applications (ICCSA 2007), 2007
Publication
<1%

29 doku.pub
Internet Source
<1%

30	e-journal.unair.ac.id Internet Source	<1 %
31	www.science.gov Internet Source	<1 %
32	"Plenary Oral Free Paper Sessions", Virchows Archiv, 2011 Publication	<1 %
33	Sung K. Chang, M. Yusoff Dawood, Gregg Staerkel, Urs Utzinger, E. Neely Atkinson, Rebecca R. Richards-Kortum, Michele Follen. "Fluorescence spectroscopy for cervical precancer detection: Is there variance across the menstrual cycle?", Journal of Biomedical Optics, 2002 Publication	<1 %

Exclude quotes   On
Exclude bibliography   On

Exclude matches   Off

Cervical single cell of squamous intraepithelial lesion classification using shape features and extreme learning machine

FINAL GRADE

/100

GENERAL COMMENTS

Instructor