

Ranked Features Selection with MSBRG Algorithm and Rules Classifiers for Cervical Cancer

<https://doi.org/10.3991/ijoe.v15i12.10803>

Mohammad Subhi Al-Batah
Jadara University, Irbid, Jordan
albatlah@jadara.edu.jo

Abstract—In this paper, an automatic three-phase cervical cancer diagnosis system is employed which includes feature extraction, feature selection followed by classification. Firstly, the modified seed-based region growing (MSBRG) algorithm is implemented for automatic segmentation and feature extraction. Processes to obtain the threshold values and the initial seed location are carried out automatically using moving k-mean (MKM) algorithm and invariant moment techniques. Secondly, eight attribute evaluators are applied for selecting and ranking the features, which are Correlation-based Feature Selection, Classifier Attribute Evaluator, Correlation Attribute Evaluator, Gain Ratio, Info Gain, OneR, ReliefF, and Symmetrical Uncertainty. Finally, the classification is compared based on five classifiers: Decision Table, JRip, OneR, PART, and ZeroR. The performance of the classifiers is evaluated using 3 test options: the training percentage splits (50% to 98%), the full training data and the cross validation (2-fold to 10-fold). The experimental results prove the capability of the MSBRG algorithm as an automatic feature extraction method. Furthermore, this paper proves the ability of the ranked feature selection methods to select important features of a cervical cell, and favors the Decision Table as the best classifier for cervical cancer classification.

Keywords—Cervical cancer, features extraction, features selection, test options, classification, modified seed based region growing, moving k-means, medical imaging.

1 Introduction

Cervical cancer is known to be the cause of many deaths each year. Cancer of the cervix is the most common gynecological malignancy and the second most common cancer among females after breast cancer [1]. In most cases, cervical cancer takes many years to develop from normal to advanced stages. Therefore, the mortality related to cervical cancer can be reduced through early detection and treatment [2].

Screening tests such as Pap smear used for the detection of the precancerous stage are able to avoid the occurrence of cervical cancer [3]. In general, the test is performed by examining cells using a microscope. However, Pap smear test has many disadvantages such as blurriness, noises, less effective slides preparation and human

error which may lead to false diagnosis. As a solution to the problem, several studies developed a diagnosis systems based on computer vision [4].

Development of a cervical cells vision system involves five major components; data acquisition, data pre-processing, data extraction, features selection, and data classification. Data acquisition converts information from Pap smear slides to digital numeric values (digital cervical cell images) that can be manipulated by a computer system [5].

Data pre-processing involves contrast and quality enhancement of Pap smear images as well as segmentation of regions of interest (ROI) [6]. Many image processing techniques have been used to detect the edges or segment the digital images for certain ROI [7]. In cervical cancer, the techniques are used to segment the image into the nucleus, cytoplasm and background regions. The algorithms include Cubic Spline, Frei Chen, Kirsch, Laplacian, Prewitt, Roberts, Robinson Sobel, Otsu thresholding, and region growing algorithms [8].

Region growing algorithms such as the seed based region growing (SBRG) algorithm offers several advantages over other segmentation techniques [9]. However, two parameters, namely the seed point and the threshold value of the SBRG algorithm need to be determined manually. Also, segmentation by SBRG algorithm needs to be repeated according to the number of regions in an image. Therefore, a modified SBRG algorithm to automatically segment all ROI in an image is proposed [10].

Data extraction concentrates on finding significant features of images, signals or other medical modalities, which are commonly used by physicians during clinical diagnosis. Researchers proposed many features that are suitable for cervical cell classification such as size, saturation, intensity, perimeter, grey level, and colors as: green, blue, and red color [11].

Feature selection selects a small subset of features from the original feature space [12]. Three main techniques are used in feature selection which include filter-based, wrapper based and hybrid-based methods [13]. These methods are categorized based on their criteria of using a learning algorithm [14]. Examples of nonparametric filters are Relief-F, Information gain, Correlation coefficient (Pearson), and Gain ratio [15].

Data classification aims to classify cervical cells into a three-class problem: normal, low-grade squamous intraepithelial lesion (LSIL) and high-grade intraepithelial lesion (HSIL). Recently, many classification algorithms using artificial neural network, fuzzy logic, genetic algorithm and statistical methods have been implemented for cervical cancer diagnosis [16].

In this paper, images are captured from the slides by the Auto Capture System [17]. A total of 500 single cell images are selected from all of these captured for this study. Firstly, the automatic MSBRG algorithm is used to extract features from the images. Features extraction process starts with threshold value finding to discriminate between nucleus, cytoplasm and background. This will follow by centroid location finding as a starting point for segmentation and feature extraction. After the segmentation process is completed, 6 features are extracted from both the nucleus and the cytoplasm. The features are size, grey level, and perimeter of both nucleus and cytoplasm. Secondly, eight attribute evaluators are considered for selecting and ranking these features. Finally, five classifiers are conducted for classifying cervical cell types.

2 Features Extraction

2.1 Moving k-means algorithm

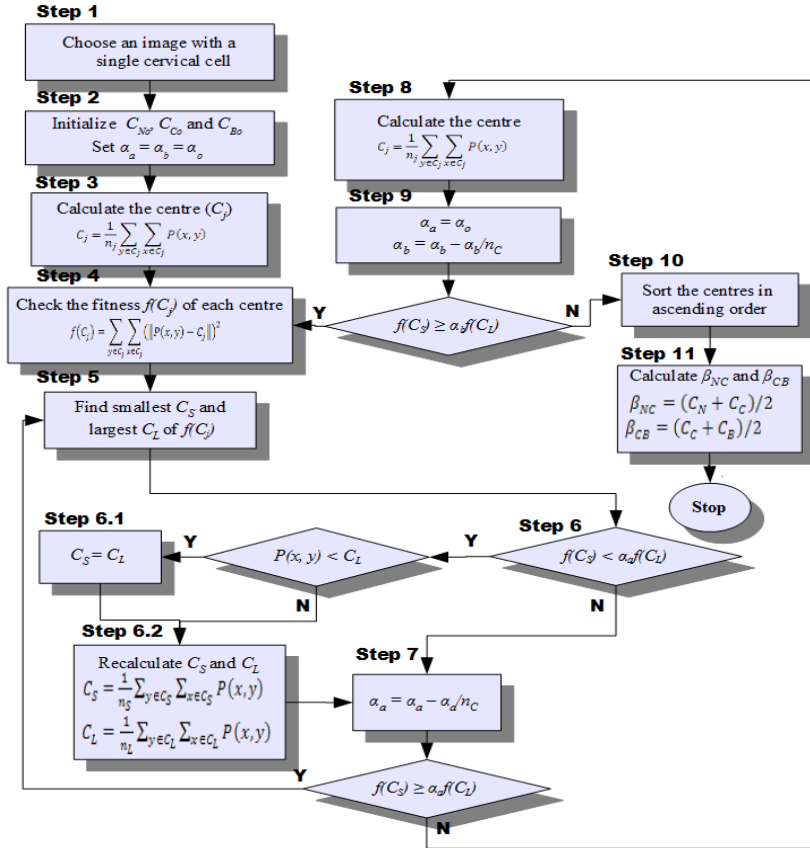


Fig. 1. Flowchart of MKM algorithm

The MKM algorithm [18] is used to find threshold values on cervical cell images. The MKM is described and flowcharted in Fig 1. To discuss the MKM algorithm, consider $P(x,y)$, pixels of one digital image with resolution of $M \times N$, where $x = 0, 1, 2, 3, \dots, M - 1$ and $y = 0, 1, 2, 3, \dots, N - 1$. The image is clustered into n_c clusters (here, $n_c = 3$ to represent the nucleus, cytoplasm and background areas) and has C_j centres (the centres are C_N, C_C and C_B for nucleus, cytoplasm and background of the cervical cell, respectively). Based on the Euclidean distance concept [19], the threshold value, β_{NC} and β_{CB} is calculated where, β_{NC} is the threshold value to differentiate the nucleus-cytoplasm area and, β_{CB} is the threshold value to differentiate the cytoplasm-background area.

2.2 Seed-based region growing (SBRG) algorithm

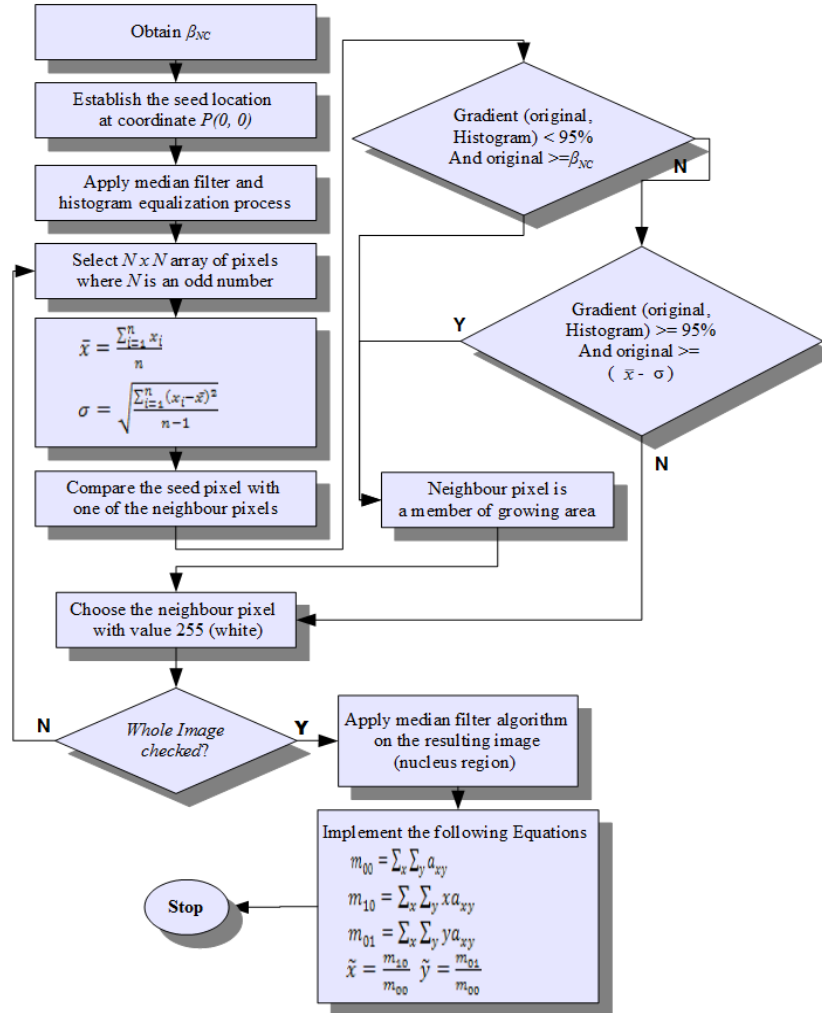


Fig. 2. Flowchart of SBRG algorithm

The centroid or seed location is used as a starting point for segmentation and feature extraction [20]. The centroid of a cervical cell image is located inside the nucleus of a cell. Thus, a process to eliminate the cytoplasm and the background region is needed before the process of obtaining the centroid is applied. The conventional SBRG algorithm based on the invariant moment technique is used as sub-clustering process of the nucleus region. The flowchart of sub-clustering process using SBRG algorithm is shown in Fig 2.

2.3 Modified seed-based region growing (MSBRG) algorithm

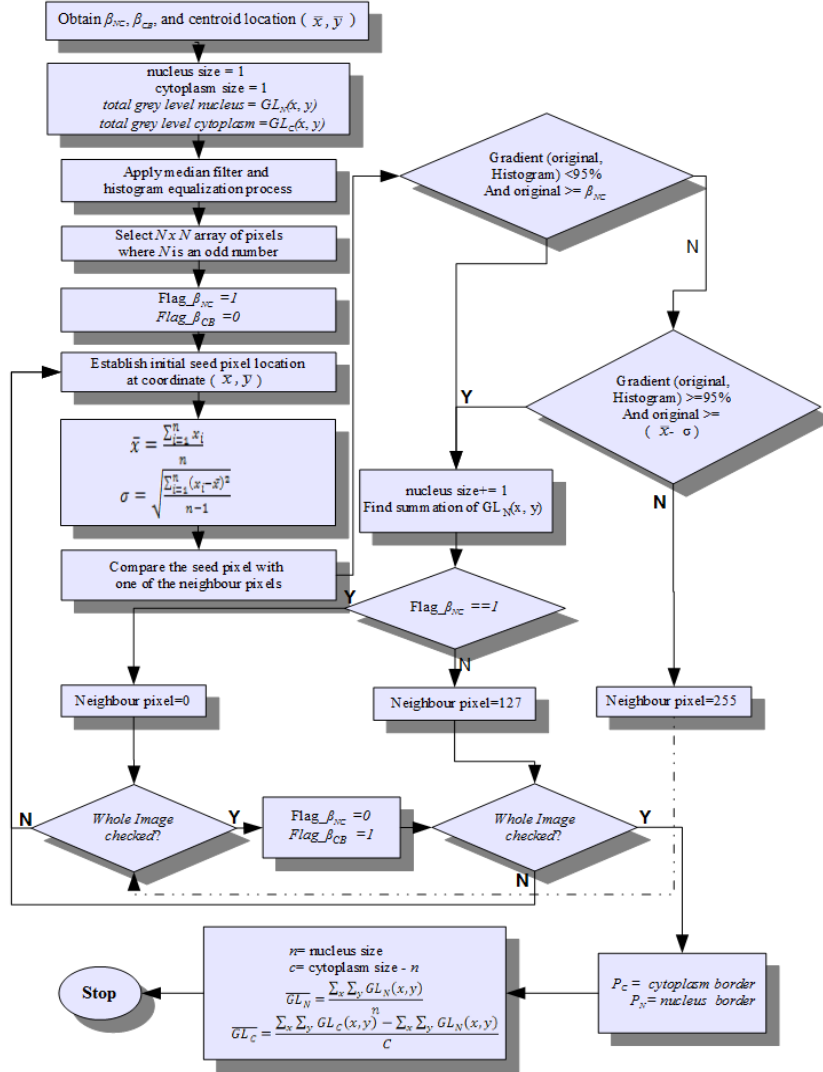


Fig. 3. Flowchart of MSBRG algorithm

The MSBRG algorithm is adopted to perform segmentation and automatically extract features from cervical cell images [21]. The threshold values, β_{NC} and β_{CB} are obtained from MKM clustering algorithm whereas the initial seed location is acquired from the invariant moment technique. The flowchart of MSBRG algorithm is presented in Fig 3.

As shown in Fig 3, the extraction process begins at the centroid where it is automatically accepted as a nucleus region member. The neighboring pixels will then be

examined and accepted as nucleus region members if the growing condition is fulfilled. It will then increase the number of pixels grown by one. At the same time, the grey level value of the new member will be extracted and summed with the total of grey level values of accepted members. After completing the growing process, the total number of nucleus region members will represent the size of the nucleus, n . Pixels that formed the border of the nucleus will be calculated as the perimeter of the nucleus, P_N . Then, the average grey level for the nucleus GL_N region is calculated. Also, the region growing process for extracting cytoplasm features is started by using the same methods in order to calculate the size of the cytoplasm c , perimeter P_C , and average grey level value GL_C .

2.4 Discussion of automatic MSBRG algorithm

The steps to segment the cell image for feature extraction using MSBRG algorithm process are shown in Fig. 4.

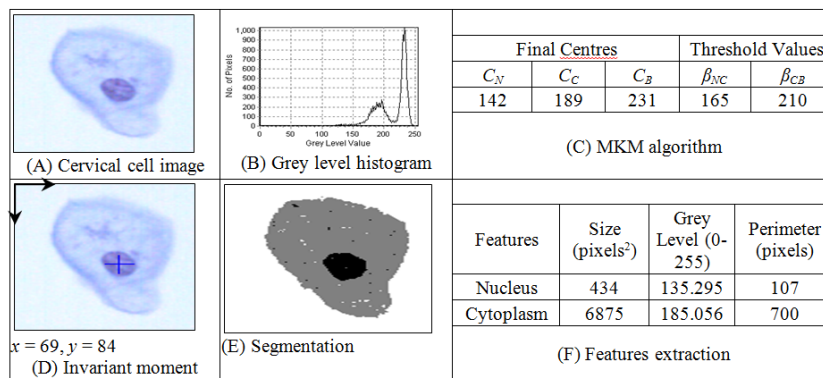


Fig. 4. Example for cervical cancer cell segmentation and features extraction

- From the all cervical cell images, one image is selected arbitrarily for discussion. The image is shown in Fig 4 (A) along with the grey level histogram in Fig 4 (B).
- The MKM algorithm is used to get the threshold value between the nucleus and the cytoplasm, β_{NC} and also the threshold value between the cytoplasm and the background of the cell, β_{CB} . The final centre for each cluster: nucleus, C_N , cytoplasm, C_C and background, C_B needs to be determined first. Fig 4 (C) shows the final centres and threshold values for the image.
- Invariant moment method is involved to get the centroid location or the initial seed for region growing as shown in Fig 4 (D). The centroid is located inside the nucleus region which has a minimal distance from any pixel inside the nucleus region. Thus, the process to obtain the centroid location only involves the nucleus region. Sub-clustering is needed to eliminate the cytoplasm and the background region of the cervical cell image. The β_{NC} value from the previous step has been used to eliminate the two regions.

- The MSBRG algorithm is applied to simultaneously segment and extract features. The algorithm checked all the pixels inside the image, starting at the centroid pixel of the cell. The process continues towards the image's border through all the 8 neighbouring pixels nearby. Then, images are segmented into three main regions; nucleus, cytoplasm and background of the cell. For observation purpose, the nucleus, cytoplasm and background regions are converted to black (0), grey (127) and white (255), respectively. Fig 4 (E) shows the segmentation result of the image.
- The MSBRG algorithm is used to extract features of cervical cells. The algorithm reads the value of the pixels while segmentation process is being executed. The size and grey level value of the nucleus and cytoplasm are updated each time a new pixel that belongs to nucleus and cytoplasm clusters is examined. Pixels that form the nucleus-cytoplasm or cytoplasm-background borders are marked and updated during the extraction process. Features extracted from the image are shown in Fig 4 (F). Size and perimeter results are in micrometer (μm) whereas average grey level result is in the range of 0 to 255.

3 Selected Attributes

In this paper, eight attribute evaluators are proposed for solving feature selection problems in cervical cancer data [22]. The evaluation mode in all the evaluators is based on full training data. The CFS subset Evaluator is applied based on the Best First and Greedy Stepwise search method, while other evaluators are applied based on the Ranker search method. The original six features are given as nucleus size (n), cytoplasm size (c), nucleus average grey level value (GL_N), cytoplasm average grey level value (GL_C), nucleus perimeter (P_N), and cytoplasm perimeter (P_C).

Table 1 shows the results of the selected and ranked attributes using the eight evaluators. From the analysis, we can notice the following results:

- The CFS Subset Evaluator selected 4 useful features; c , GL_C , P_N , and P_C . The n , and GL_N are not selected using CFS algorithm.
- The Classifier Attribute Evaluator ranked the importance of the features as; P_C , P_N , c , GL_N , GL_C , and n .
- OneR Attribute Evaluator arranged the attributes as; P_C , c , P_N , GL_C , n , and GL_N .
- ReliefF Attribute Evaluator organised the attributes as; c , P_C , GL_C , GL_N , P_N , and n .
- Two evaluators; Correlation Attribute Evaluator, and Gain Ratio Attribute Evaluator, ordered the features as; c , GL_C , P_C , P_N , n , and GL_N .
- Two evaluators; Info Gain Attribute Evaluator, and Symmetrical Uncertainty Attribute Evaluator ranked the attributes as; c , P_C , GL_C , P_N , n , and GL_N .

In addition, the selected and ranked attributes using the eight evaluators are analyzed and compared. From the analysis, it can be noticed that the two features; n , and GL_N are not selected using CFS subset Evaluator, and also ranked as the last two features using five other attribute evaluators: OneR Attribute Evaluator, Correlation Attribute Evaluator, Gain Ratio Attribute Evaluator, Info Gain Attribute Evaluator,

and Symmetrical Uncertainty Attribute Evaluator. Also, it can be seen that four features are considered important which are: c , GL_C , P_N , and P_C . Thus, features selection techniques can be used in the cervical cancer and only four important features are considered for the classification stage.

Table 1. Selected and ranked attributes

| Attribute Evaluator | Ranked attributes |
|--------------------------------|---|
| CfsSubsetEval | 4 Selected attributes: c , GL_C , P_N , P_C |
| ClassifierAttributeEval | P_C , P_N , c , GL_N , GL_C , n |
| CorrelationAttributeEval | 0.675 c , 0.635 GL_C , 0.516 P_C , 0.455 P_N , 0.432 n , 0.253 GL_N |
| GainRatioAttributeEval | 0.4118 c , 0.4067 GL_C , 0.3962 P_C , 0.3379 P_N , 0.2884 n , 0.0646 GL_N |
| InfoGainAttributeEval | 0.678 c , 0.5449 P_C , 0.3456 GL_C , 0.2675 P_N , 0.2093 n , 0.0493 GL_N |
| OneRAttributeEval | 85 P_C , 84.8 c , 80 P_N , 79.2 GL_C , 78.2 n , 73 GL_N |
| ReliefFAttributeEval | 0.2528 c , 0.13 P_C , 0.0709 GL_C , 0.0397 GL_N , 0.0287 P_N , 0.0222 n |
| SymmetricalUncertAttributeEval | 0.5043 c , 0.4507 P_C , 0.3653 GL_C , 0.2917 P_N , 0.2367 n , 0.0546 GL_N |

4 Classification Results and Analysis

After feature selection, classifiers are applied on the reduced datasets to see the effect of this phase. In this paper, classification accuracy of cervical cancer is compared with five different classifiers: Decision Table, JRip, OneR, PART, and ZeroR [23-25].

The images of cervical cancer have been collected at Hospitals of Jordan and Malaysia. The dataset consists of 500 single cells, where 376 of them are normal, 79 are LSIL, and the other 45 are HSIL. Classifiers were trained with similar sets of data and comparison of the results was done. Four important features for cervical cancer are selected which are size of cytoplasm c , grey level of cytoplasm GL_C , perimeter of nucleus P_N , and perimeter of cytoplasm P_C . The accuracy of all classifiers is measured based on 3 test options; training percentage splits (50% to 98%), full training data and cross validation (2-fold to 10-fold) [26-29].

Initially, the analysis was conducted by varying percentages of dataset distribution coupled with training and test datasets in order to make sure that the best positive sets of data provided the best results [30-32]. Generally, the results show that the accuracy of the classifiers is so low when using training percentage splits (1% to 49%). Thus, the results of training percentage splits (50% to 98%) are only included in this paper.

In addition, the results show that the accuracy of the classifiers using features selection performed better results when compared with those applied directly on the original features. Thus, the results in this paper are those based on the selected features only.

The classification accuracies using the 5 classifiers based on the 4 selected features in the three different test options are tabulated in Table 2. The average accuracy of the

5 classifiers conducted in this paper can be ordered as follows: Decision Table (88.6%), OneR (88.3%), Jrip (87.4%), PART (87.0%), and ZeroR (77.3%). We can see clearly that the Decision Table is the winner or the best classifier of selection compared to other classifiers. Also, the comparison results show that the lowest average accuracy among the classifiers conducted is reached by ZeroR.

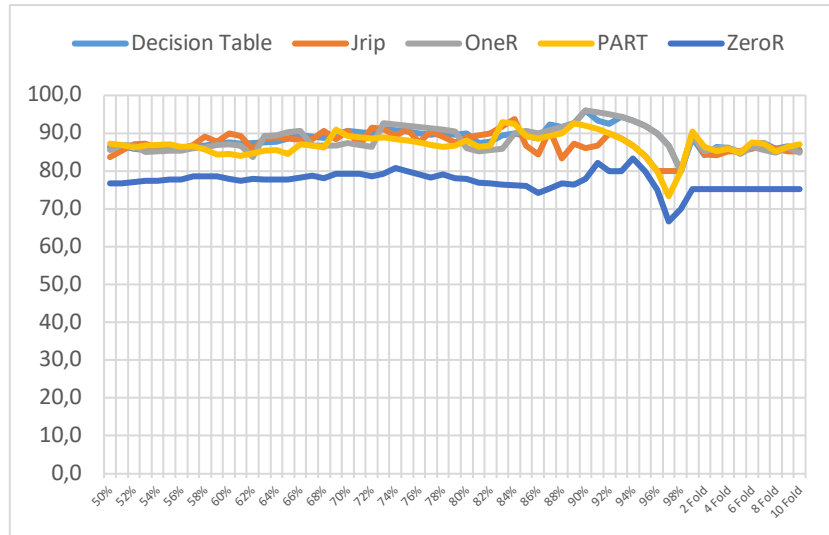


Fig. 5. Classification accuracy for the five classifiers using the three test options.

Fig. 5 illustrates a comparison for the classification accuracy of the 5 classifiers using all test options. As can be seen in Fig. 5, the 5 classifiers have different classification results using different training splits. It was noticed that the maximum accuracy achieved using Decision Table was 96.0% with 90% training splits, while the maximum accuracy of the Jrip was 93.8% using 84% training splits, the OneR accuracy was 95.7% using 90% training splits, the PART accuracy was 92.9% using 83% training splits, and ZeroR accuracy was 83.3% using 94% training splits. Again, the comparison results of maximum accuracy prove that the Decision Table can classify cervical cancer types with high accuracy.

Furthermore, the results show that the Decision Table exceeded the accuracy compared to other classifiers. Decision Table outperformed other classifiers with a difference of average accuracy percentage equal to 0.3%, 1.2%, 1.6%, and 11.3% for OneR, Jrip, PART, and ZeroR, respectively. Also, Decision Table outperformed other classifiers in terms of the maximum accuracy by 0.3% as for OneR, 2.2% for Jrip, 3.1% for PART, and 11.7% for ZeroR. Outcomes consistently demonstrate the effectiveness of Decision Table for classifying cervical cancer task.

Table 2. Classification results using four selected and ranked attributes

| Training Splits | Decision Table | Jrip | OneR | PART | ZeroR |
|-----------------|----------------|------|------|------|-------|
| 50% | 86.4 | 83.6 | 85.6 | 87.2 | 76.8 |
| 51% | 86.5 | 85.3 | 86.9 | 86.9 | 76.7 |
| 52% | 85.8 | 87.1 | 86.7 | 86.3 | 77.1 |
| 53% | 85.5 | 87.2 | 85.1 | 86.8 | 77.4 |
| 54% | 85.7 | 86.1 | 85.2 | 87.0 | 77.4 |
| 55% | 85.8 | 86.7 | 85.3 | 87.1 | 77.8 |
| 56% | 85.9 | 85.9 | 85.5 | 86.4 | 77.7 |
| 57% | 86.5 | 87.0 | 86.0 | 86.5 | 78.6 |
| 58% | 86.7 | 89.0 | 86.2 | 85.7 | 78.6 |
| 59% | 87.3 | 87.8 | 86.8 | 84.4 | 78.5 |
| 60% | 87.5 | 90.0 | 87.0 | 84.5 | 78.0 |
| 61% | 87.2 | 89.2 | 86.7 | 84.1 | 77.4 |
| 62% | 87.4 | 85.8 | 83.7 | 84.7 | 77.9 |
| 63% | 87.6 | 88.6 | 89.2 | 85.4 | 77.8 |
| 64% | 87.8 | 88.9 | 89.4 | 85.6 | 77.8 |
| 65% | 88.6 | 88.6 | 90.3 | 84.6 | 77.7 |
| 66% | 89.4 | 88.2 | 90.6 | 87.1 | 78.2 |
| 67% | 89.1 | 88.5 | 86.7 | 86.7 | 78.8 |
| 68% | 88.8 | 90.6 | 86.7 | 86.3 | 78.1 |
| 69% | 89.7 | 88.4 | 86.7 | 91.0 | 79.4 |
| 70% | 90.7 | 90.7 | 87.3 | 89.3 | 79.3 |
| 71% | 90.3 | 86.9 | 86.9 | 89.0 | 79.3 |
| 72% | 90.0 | 91.4 | 86.4 | 88.6 | 78.6 |
| 73% | 91.1 | 91.1 | 92.6 | 88.9 | 79.3 |
| 74% | 90.8 | 89.2 | 92.3 | 88.5 | 80.8 |
| 75% | 90.4 | 91.2 | 92.0 | 88.0 | 80.0 |
| 76% | 90.0 | 87.5 | 91.7 | 87.5 | 79.2 |
| 77% | 89.6 | 90.4 | 91.3 | 87.0 | 78.3 |
| 78% | 90.0 | 89.1 | 90.9 | 86.4 | 79.1 |
| 79% | 89.5 | 87.6 | 90.5 | 86.7 | 78.1 |
| 80% | 90.0 | 89.0 | 86.0 | 88.0 | 78.0 |
| 81% | 87.4 | 89.5 | 85.3 | 86.3 | 76.8 |
| 82% | 87.8 | 90.0 | 85.6 | 86.7 | 76.7 |
| 83% | 89.4 | 91.8 | 85.9 | 92.9 | 76.5 |
| 84% | 90.0 | 93.8 | 90.0 | 92.5 | 76.3 |
| 85% | 89.3 | 86.7 | 90.7 | 89.3 | 76.0 |
| 86% | 88.6 | 84.3 | 90.0 | 88.6 | 74.3 |
| 87% | 92.3 | 90.8 | 90.8 | 89.2 | 75.4 |
| 88% | 91.7 | 83.3 | 91.7 | 90.0 | 76.7 |
| 89% | 92.7 | 87.3 | 92.7 | 92.7 | 76.4 |
| 90% | 96.0 | 86.0 | 95.7 | 92.0 | 78.0 |
| 91% | 93.3 | 86.7 | 95.6 | 91.1 | 82.2 |
| 92% | 92.5 | 90.0 | 95.0 | 90.0 | 80.0 |
| 93% | 94.3 | 88.6 | 94.3 | 88.6 | 80.0 |
| 94% | 93.3 | 86.7 | 93.3 | 86.7 | 83.3 |
| 95% | 92.0 | 84.0 | 92.0 | 84.0 | 80.0 |
| 96% | 90.0 | 80.0 | 90.0 | 80.0 | 75.0 |

| | | | | | |
|-----------------|------|------|------|------|------|
| 97% | 86.7 | 80.0 | 86.7 | 73.3 | 66.7 |
| 98% | 80.0 | 80.0 | 80.0 | 80.0 | 70.0 |
| 100% full Train | 88.8 | 90.0 | 89.2 | 90.4 | 75.2 |
| 2 Fold | 84.2 | 84.4 | 85.2 | 86.4 | 75.2 |
| 3 Fold | 86.4 | 84.2 | 86.0 | 85.2 | 75.2 |
| 4 Fold | 86.0 | 85.2 | 86.2 | 85.8 | 75.2 |
| 5 Fold | 85.0 | 85.2 | 84.6 | 84.8 | 75.2 |
| 6 Fold | 87.0 | 85.8 | 86.4 | 87.6 | 75.2 |
| 7 Fold | 87.4 | 86.8 | 85.6 | 87.2 | 75.2 |
| 8 Fold | 85.8 | 85.8 | 84.8 | 85.2 | 75.2 |
| 9 Fold | 86.6 | 85.2 | 86.2 | 86.4 | 75.2 |
| 10 Fold | 85.6 | 85.2 | 84.8 | 87.0 | 75.2 |
| AVERAGE | 88.6 | 87.4 | 88.3 | 87.0 | 77.3 |
| MAX | 96.0 | 93.8 | 95.7 | 92.9 | 83.3 |

5 Conclusion

In this paper, an intelligent system is developed to categorize cervical cells into three groups: normal, LSIL and HSIL. The system consists of three stages: feature extraction, feature selection, and classification. In the feature extraction stage, automatic MSBRG algorithm is implemented. In feature selection stage, eight attribute evaluators are considered for selecting and ranking the features. The selected features; size of cytoplasm, grey level of cytoplasm, perimeter of nucleus, and perimeter of cytoplasm were found to be the most appropriate features for recognition of cervical cancer. Also, the results indicate that feature selection not only improved the efficiency of the classification process but also its accuracy was enhanced. In the classification stage, five classifiers are conducted for classifying the cervical cancer. The analysis of classifiers was performed using different percentages of distribution of the dataset for training and testing datasets. The results show that the maximum accuracy classification is obtained via Decision Table with an accuracy of 96.0% compared to 93.8% using Jrip, 95.7% using OneR, 92.9% using PART, and 83.3% using ZeroR. The current paper can be extended using more samples for a complete diagnosis of cervical cancer. Also, we intend to investigate the work further with additional features extraction, and other features selection methods (like Genetic algorithm).

6 Acknowledgement

This research was supported by Jadara University, Jordan. I express my thanks to Doctor Malak Fuad Subhi Al-Battah, a Neurology resident in King Abdullah University Hospital (KAUH), who checked the manuscript. I also thank Prof. Nor ashidi Mat Isa working at Universiti Sains Malaysia for selecting the appropriate datasets and comments that greatly improved the paper.

7 References

- [1] Lehmann U. Lobular breast cancer - the most common special subtype or a most special common subtype? *Breast Cancer Research*. Springer Nature; 2015 Jul 28;17(1). <https://doi.org/10.1186/s13058-015-0606-z>
- [2] Montz FJ. Moderate cervical dysplasia had half the cancer risk of severe dysplasia and most cases regressed to normal. *Evidence-based Obstetrics & Gynecology*. Elsevier BV; 2000 Mar;2(1):23. <https://doi.org/10.1054/ebog.2000.0115>
- [3] Domingo EJ, Noviani R, Noor MRM, Ngelangel CA, Limpaphayom KK, Van Thuan T, et al. Epidemiology and Prevention of Cervical Cancer in Indonesia, Malaysia, the Philippines, Thailand and Vietnam. *Vaccine*. Elsevier BV; 2008 Aug;26:M71–M79. <https://doi.org/10.1016/j.vaccine.2008.05.039>
- [4] Lallas A, Argenziano G. Artificial intelligence and melanoma diagnosis: ignoring human nature may lead to false predictions. *Dermatology Practical & Conceptual*. Derm101.com; 2018 Oct 31;249–51. <https://doi.org/10.5826/dpc.0804a01>
- [5] William W, Ware A, Basaza-Ejiri AH, Obungoloch J. A pap-smear analysis tool (PAT) for detection of cervical cancer from pap-smear images. *BioMedical Engineering OnLine*. Springer Nature; 2019 Feb 12;18(1). <https://doi.org/10.1186/s12938-019-0634-5>
- [6] Isa NAM, Mustafa N, Zamli KZ, Mashor MY. Improvement of contrast enhancement technique for cervical cell of Pap smear images by reducing the effect of unwanted background information. *2006 International Conference on Computing & Informatics*. IEEE; 2006 Jun; <https://doi.org/10.1109/icoci.2006.5276546>
- [7] Jusman Y, Mat Isa NA, Adnan R, Othman NH. Intelligent classification of cervical pre-cancerous cells based on the FTIR spectra. *Ain Shams Engineering Journal*. Elsevier BV; 2012 Mar;3(1):61–70. <https://doi.org/10.1016/j.asej.2011.12.002>
- [8] Mat-Isa NA, Mashor MY, Othman NH. An automated cervical pre-cancerous diagnostic system. *Artificial Intelligence in Medicine*. Elsevier BV; 2008 Jan;42(1):1–11. <https://doi.org/10.1016/j.artmed.2007.09.002>
- [9] Chen H-S, Chang T-T. Color Conversion Technology of Four-Primary Color Images Developed on Wide Color Gamut Red, Green, Blue Monitor. *Journal of Imaging Science and Technology*. Society for Imaging Science & Technology; 2009;53(6):060502. <https://doi.org/10.2352/j.imagingsci.technol.2009.53.6.060502>
- [10] Bolón-Canedo V, Sánchez-Marroño N, Alonso-Betanzos A. A review of feature selection methods on synthetic data. *Knowledge and Information Systems*. Springer Science and Business Media LLC; 2012 Mar 30;34(3):483–519. <https://doi.org/10.1007/s10115-012-0487-8>
- [11] Hira ZM, Gillies DF. A Review of Feature Selection and Feature Extraction Methods Applied on Microarray Data. *Advances in Bioinformatics*. Hindawi Limited; 2015; 2015:1–13. <https://doi.org/10.1155/2015/198363>
- [12] Singh RK, Sivabalakrishnan M. Feature Selection of Gene Expression Data for Cancer Classification: A Review. *Procedia Computer Science*. Elsevier BV; 2015; 50:52–7. <https://doi.org/10.1016/j.procs.2015.04.060>
- [13] Hameed SS, Hassan R, Muhammad FF. Selection and classification of gene expression in autism disorder: Use of a combination of statistical filters and a GBPSO-SVM algorithm. Liu B, editor. *PLOS ONE*. Public Library of Science (PLoS); 2017 Nov 2;12(11): e0187371. <https://doi.org/10.1371/journal.pone.0187371>
- [14] Quteishat A., Al-batah M., Al-mofleh A., Alnabelsi S.H. Cervical Cancer Diagnostic System Using Adaptive Fuzzy Moving K-means Algorithm and Fuzzy MIN-MAX Neural Network. *Journal of Theoretical and Applied Information Technology*; 2013; 57(1):48-53.

- [15] Noor NRM, Isa NAM, Mashor MY, Othman NH, Zamli KZ, Al-Batah MS. Automatic Glass-Slide Capturing System for Cervical Cancer Pre-Screening Program. *American Journal of Applied Sciences*. Science Publications; 2008 May 1;5(5):461–7. <https://doi.org/10.3844/ajassp.2008.461.467>
- [16] Hung W-L, Yang M-S, Chen D-H. Parameter selection for suppressed fuzzy c-means with an application to MRI segmentation. *Pattern Recognition Letters*. Elsevier BV; 2006 Apr;27(5):424–38. <https://doi.org/10.1016/j.patrec.2005.09.005>
- [17] Sing JK, Basu DK, Nasipuri M, Kundu M. Self-adaptive RBF neural network-based segmentation of medical images of the brain. *Proceedings of 2005 International Conference on Intelligent Sensing and Information Processing, 2005*. IEEE; <https://doi.org/10.1109/icisip.2005.1529496>
- [18] Mustafa N. Features Extraction of Cervical Cell Based on ThinPrep Images for Cervical Cancer Diagnostic System. MSc, School of Electrical and Electronic Engineering, Universiti Sains Malaysia, Malaysia, 2007.
- [19] Mustafa N., Mat-Isa N.A., Mashor M.Y. Automated Multicells Segmentation of Thin-Prep® Image Using Modified Seed Based Region Growing Algorithm. *Biomedical Soft Computing and Human Sciences*; 2009; 14(2): 41-47.
- [20] Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. The WEKA data mining software. *ACM SIGKDD Explorations Newsletter*. Association for Computing Machinery (ACM); 2009 Nov 16;11(1):10. <https://doi.org/10.1145/1656274.1656278>
- [21] Kohavi R. The power of decision tables. *Machine Learning: ECML-95*. Springer Berlin Heidelberg; 1995;174–89. <https://doi.org/10.14569/ijacsa.2011.021212>
- [22] S I, S.M. N, N. G. Preprocessor Agent Approach to Knowledge Discovery Using Zero-R Algorithm. *International Journal of Advanced Computer Science and Applications*. The Science and Information Organization; 2011;2(12). <https://doi.org/10.14569/ijacsa.2011.021212>
- [23] Gaines BR, Compton P. Induction of ripple-down rules applied to modeling large databases. *Journal of Intelligent Information Systems*. Springer Science and Business Media LLC; 1995 Nov;5(3):211–28. <https://doi.org/10.1007/bf00962234>
- [24] Al-Batah MS, Zaqaibeh BM, Alomari SA, Alzboon MS. Gene Microarray Cancer Classification using Correlation Based Feature Selection Algorithm and Rules Classifiers. *International Journal of Online and Biomedical Engineering (iJOE)*. International Association of Online Engineering (IAOE); 2019 May 14; 15(08):62. <https://doi.org/10.3991/ijoe.v15i08.10617>
- [25] Al-Batah MS. Integrating the Principal Component Analysis with Partial Decision Tree in Microarray Gene Data. *IJCSNS International Journal of Computer Science and Network Security*; 2019 March; 19(3): 24-29.
- [26] Al-Batah MS. Automatic Diagnosis System for Heart Disorder using ESG Peak Recognition with Ranked Features Selection. *International journal of circuits, systems and signal processing*; 2019; 13:391-398.
- [27] karimBaareh A, F. Sheta A, S. Al-Batah M. Feature based 3D Object Recognition using Artificial Neural Networks. *International Journal of Computer Applications*. Foundation of Computer Science; 2012 Apr 30;44(5):1–7. <https://doi.org/10.5120/6256-8402>
- [28] Sheta AF, Baareh A, Al-Batah M. 3D object recognition using fuzzy mathematical modeling of 2D images. 2012 *International Conference on Multimedia Computing and Systems*. IEEE; 2012 May; <http://dx.doi.org/10.1109/icmcs.2012.632011> <https://doi.org/10.1109/icmcs.2012.6320118>

- [29] Al-Batah MS, Mrayyen S, Alzaqebah M. Arabic Sentiment Classification using MLP Network Hybrid with Naive Bayes Algorithm. *Journal of Computer Science. Science Publications*; 2018 Aug 1; 14(8):1104–14. <https://doi.org/10.3844/jcssp.2018.1104.1114>
- [30] Mrayyen S, Al-Batah MS, Alzaqebah M. Investigation of Naive Bayes Combined with Multilayered Perceptron for Arabic Sentiment Analysis and Opinion Mining. *International Journal of Mathematical Models and Methods in Applied Sciences*; 2018; Volume 12. <https://doi.org/10.3844/jcssp.2018.1104.1114>

8 Author

Mohammad Subhi Al-Batah received his PhD in Computer Science/ Artificial Intelligence from the University of Science Malaysia in 2009. After working as an assistant professor (from 2009) in the Dept. of Computer Science, Jadara Univ. in Jordan, he has been an associate professor at Jadara Univ. since 2014. He worked as a dean of Faculty of Science and Information Technology from 2015-2018. In 2019, he is the director of the Academic Development and Quality Assurance Center. His research interests include Image Processing, Artificial Intelligence, Medical Analysis, Real Time Classification and Software Engineering, E-mail: dralbatah@gmail.com.

Article submitted 2019-05-05. Resubmitted 2019-06-13. Final acceptance 2019-06-26. Final version published as submitted by the authors