Karbala International Journal of Modern Science

Volume 7 | Issue 3 Article 5

Two-Dimensional Quantitative Profiling of Cell Morphology with Serous Effusion by Unsupervised Machine Learning Analysis

Safaa Al-Qaysi Ph.D.
Department of Physics, East Carolina University, salkaysi@gmail.com

Ding Dai MD Ph.D.
Department of Pathology, Virginia Commonwealth University Health System

Heng Hong MD Ph.D.
Department of Pathology and Comparative Medicine, Wake Forest School of Medicine, Wake Forest University

Yuhua Wen Ph.D.
Institute for Advanced Optics, Hunan Institute of Science and Technology

X.H. Hu Ph.D.
Department of Physics, East Carolina University

Follow this and additional works at: https://kijoms.uokerbala.edu.iq/home

 Part of the Biology Commons, Chemistry Commons, Computer Sciences Commons, and the Physics Commons

Recommended Citation

Al-Qaysi, Safaa Ph.D.; Dai, Ding MD Ph.D.; Hong, Heng MD Ph.D.; Wen, Yuhua Ph.D.; and Hu, X.H. Ph.D. (2021) "Two-Dimensional Quantitative Profiling of Cell Morphology with Serous Effusion by Unsupervised Machine Learning Analysis," *Karbala International Journal of Modern Science*: Vol. 7 : Iss. 3 , Article 5.
Available at: https://doi.org/10.33640/2405-609X.3120

This Research Paper is brought to you for free and open access by Karbala International Journal of Modern Science. It has been accepted for inclusion in Karbala International Journal of Modern Science by an authorized editor of Karbala International Journal of Modern Science.

University of Kerbala

Two-Dimensional Quantitative Profiling of Cell Morphology with Serous Effusion by Unsupervised Machine Learning Analysis

Abstract

Cytological evaluation of serous effusion specimens is an important part of cancer diagnosis. In this study we performed two-dimensional (2D) morphometric features and clustering analysis for development of useful techniques for identification and differentiation of malignant and begin cells in serous effusion specimens extracted from ten patients with clinical symptoms of pleural and peritoneal effusion. Our findings show that the two-dimensional (2D) morphometric features and clustering analysis are useful techniques for identification and differentiation of malignant and begin cells in serous effusion specimens, which can lead to development of new methods for rapid cells profiling in clinical application.

Keywords

cell morphology, serous effusions, clustering analysis, Gaussian mixture model

Creative Commons License

This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License.

This research paper is available in Karbala International Journal of Modern Science: https://kijoms.uokerbala.edu.iq/home/vol7/iss3/5

1. Introduction

In the human body, serosa is a thin layer of mesothelial cell lining at the surface of pleural and peritoneal cavities [1]. A small amount of lubricating fluid presents inside these cavities. This fluid allows the internal organs to slide over each other [2,3]. Accumulation of excess fluid due to disease conditions such as infection, inflammation, or unbalanced pressure may lead to a benign serous effusion [4]. Benign effusion usually presents mesothelial cells, inflammatory cells, macrophages, and blood elements [5]. The presence of the malignant cells in effusion fluid defines malignant serous effusion that is usually caused by tumor metastasis. The most common types of malignancies that develop malignant serous effusions are lung cancer, breast cancer, lymphoma, gastrointestinal carcinoma, gynecological carcinoma, genitourinary carcinomas, and so on [6−8]. Cytologic evaluation is a sensitive and specific method to identify malignant or benign effusion. It assists in diagnosing the type and source of the malignancy with the help of accessory tests [9,10].

Morphometric features of serous effusions cells can provide useful indicators for detecting malignant cells in conventional cytology as the gold standard cancer diagnosis [11]. Accordingly, the development of morphology based, and label-free methods are very promising for their abilities to achieve simple and less expensive malignancies detection and make significant therapeutic and prognostic implications. For instance, the polarization diffraction imaging flow cytometry (p-DIFC) method showed the capability for the acquisition of high-quality diffraction patterns images for different types of malignant and benign cells and correlate their three-dimensional (3D) morphology by recording spatial distribution of coherent light scatter [12−17]. Other studies have been reported to visualize 3D structures of the cells and to study 3D morphology on major intracellular organelles such as the nucleus, mitochondria, cytoplasm, and cell membrane [18−20]. While investigations of two-dimensional (2D) morphology as a diagnosis of differences between malignant and benign cells are insufficient [5,11,21,22]. Recently, different machine learning algorithms have been used in different areas of cytology such as in gastric, breast, thyroid, urothelial, and effusion cytology to identify benign and malignant cells [23−30]. However, cell area, nuclear area, and nuclear to cell area ratio (A_c, A_n, and Ar_{nc}) have been regarded as the essential markers in this study. We focus our effort on profiling and differentiating malignant and benign cells in serous effusion samples by comparing the results of machine learning analysis of morphometric features with the visual cytopathologic examination.

2. Material and methods

2.1. Cytology

We performed a quantitative study of the 2D morphology of cells extracted from fresh serous effusion specimens. Either pleural fluid or peritoneal fluid was obtained from ten patients. The specimen is arbitrarily designated as P1 to P10. These specimens consisted of six malignant serous fluids and four benign serous fluids (Table 1). The specimens were

Table 1
The total number of both malignant and benign cells extracted from the cytology images of 10 patients.

Patient ID	Effusion Type	Status	Diagnosis (Origin)	No. of Malignant Cells	No. of Benign Cells	Total cells/Patient
P-1	Pleural	Malignant	Adenocarcinoma (Lung)	65 (45.8 %)	77 (54.2 %)	142
P-2	Pleural	Malignant	Adenocarcinoma (Lung)	27 (28.4 %)	68 (71.6 %)	95
P-3	Pleural	Malignant	Ovarian carcinoma (Ovary)	24 (47.1 %)	27 (52.9 %)	51
P-4	Peritoneal	Malignant	Ovarian carcinoma (Ovary)	12 (50.0 %)	12 (50.0 %)	24
P-5	Pleural	Malignant	Adenocarcinoma (Lung)	48 (46.6 %)	55 (53.4 %)	103
P-6	Pleural	Benign	Benign	0 (0.0 %)	129 (100 %)	129
P-7	Pleural	Benign	Benign	0 (0.0 %)	170 (100 %)	170
P-8	Pleural	Malignant	NSC lung cancer (Lung)	66 (45.5 %)	79 (53.4 %)	145
P-9	Pleural	Benign	Benign	0 (0.0 %)	58 (100 %)	58
P-10	Peritoneal	Benign	Benign	0 (0.0 %)	149 (100 %)	149
Total cells/Group				**242 (22.7 %)**	**824 (77.3 %)**	**1066**

https://doi.org/10.33640/2405-609X.3120
2405-609X/© 2021 University of Kerbala. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

collected between July 2017 and August 2018 in the Department of Pathology at the Brody School of Medicine, East Carolina University (ECU). The study was approved by the institutional review board (IRB) at the school of medicine in ECU. All specimens were prepared by the standardized method of cytospin slides and stained by the conventional Diff-Quik (DQ) and Papanicolaou (Pap) liquid-based technique. The diagnosis was made by cytopathologists in the department of pathology, ECU. Patient's clinical history and cytopathologic features of the cases were reviewed. The cytology slides were imaged under a bright field microscope (Olympus, BX43 Upright Microscope, Tokyo, Japan) of 40x objective magnification and saved as "jpg." image files format with dimensions of 2448 × 1920 pixels and bit depth 24. These images were evaluated by cytopathologists for the generally known diagnostic indicator features associated with malignant cells such as single and cluster cells population, cell size, nuclear to cell ratio, vacuolated cytoplasm, multinucleation, increased nuclear size, nuclear shape, and presence of prominent nucleoli [3,4]. Fig. 1 presents examples of acquired microscopic images that include different types of cells. A total of 1066 cells were randomly marked and identified into two categories, 242 (22.7 %) malignant cells and 824 (77.3 %) benign cells. These cells were further subjected to quantitative measures for this study by using image processing software. Table 1 shows details of the clinical diagnosis of each serous effusion specimen and the number of cells extracted from each one.

2.2. Two-dimensional features

We used Fiji (ImageJ distribution focuses on biological image analysis) to calculate marked cells' 2D morphometric features [31,32]. These features are characterized by A_c parameter and related A_n parameter. The spatial dimensions of Fiji were calibrated using a stage micrometer to provide accurate distances measure. The cells of interest were segmented into two regions of interest (ROIs), cell and nucleus, either manually or automatically. We used the rectangular selection tool from the Fiji user interface menu to select the cell that needs to be analyzed.

We either manually creates a user-defined ROI around the single cell or nucleus using the freehand selection tool or use histogram analysis of the pixel intensity and threshold method to detect the boundaries automatically. Once the segmentation is completed, the parameters of A_c and A_n are determined in a unit of μm^2. The values of mean and Standard Deviations

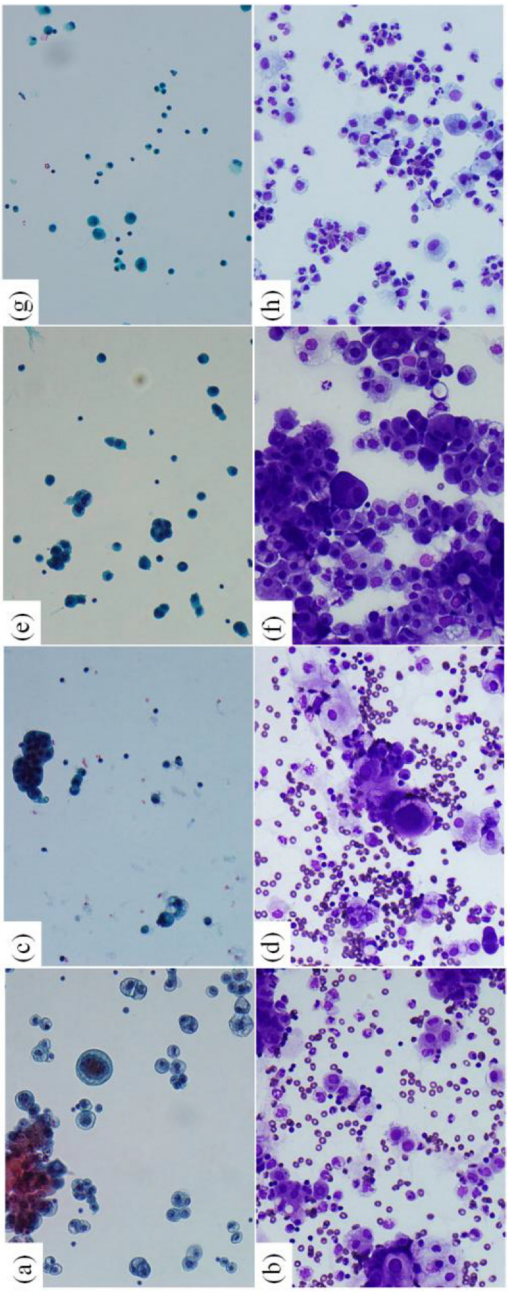

Fig. 1. Representative images (40x magnification) of four serous effusion specimens from four different patients. These images show variety of kinds of cells. The top panel is Pap stain, and the bottom panel is Diff-Quik stain. P1: a and b; P2: c and d; P5: e and f; P6: g and h. P1, P2, and P5 are malignant specimens, P6 is a benign specimen.

(STD) of all cells A_c and A_n along with Ar_{nc} are presented in Table 2. It can be noticed that malignant cells tend to have large A_c and A_n compared to benign cells, which indicates that cellular and nuclear size do correlate to cell type and morphology. Statically, we used one-way analysis of variance (ANOVA) method for comparison between the two categories of cells employing Minitab19 statistical software.

2.3. Clustering analysis

The 2D parameters calculated from the Fiji software are analyzed for cells clustering using two different machine learning techniques, Gaussian Mixture Model (GMM) and Hierarchical Clustering (HC). These two techniques used previously in the study of the 3D morphology of effusion cells and showed good results since they are assuming the size of cells and nucleus follow the Normal distribution, which gives rapid converges and accurate classification [17,33,34]. In general, the GMM sorts a given input data into an assigned number of clusters based on Gaussian probability density functions using methods of maximum−likelihood parameter estimation, and expectation-maximization iteration [33,35]. A MATLAB integrated function (fitgmdst) was used for this job [36]. The GMM clustering outcomes are overly sensitive to the initial input values of mean and covariance matrix elements. These values are randomly obtained for each assigned cluster and used for calculating the probability density functions [37]. As a result, the outcomes of GMM clustering were not consistent every time we run the clustering algorithm. To avoid such problem, HC algorithm was introduced before GMM to produce an initial set of parameters for obtaining stable clustering outcomes based on the minimum distance between the nearest pair of clusters [38].

2.4. Statistical analysis

One-way ANOVA is applied to compare morphometric parameters among the clusters. $P < 0.01$ is considered a significant difference.

3. Results

We have performed quantitative measurement of cell and nuclear area on 1066 serous effusion cells selected from the cytology slides acquired from serous effusion specimens of 10 patients. These cells were evaluated by routine cytopathological examination, excluding multinucleated cells, red blood cells, and degenerate cells. The overall distributions of measured cells are 242 (22.7 %) malignant cells and 824 (77.3 %) benign cells, Table 1. The means and standard deviations (STD) of all parameters are summarized in Table 2 according to patients and cell status.

Among the three parameters analyzed, A_c was maximum in the malignant cases with a mean and STD value of (229.5 ± 240.7), followed by benign cases with mean and STD value of (49.62 ± 43.03). The A_n was found to be lowest in benign cases with mean and STD value of (23.26 ± 16.33), compared to those of malignant (97.25 ± 136.0). Correspondingly, the highest Ar_{nc} ratio was observed with mean and STD value of (0.51 ± 0.149) in benign cases and (0.42 ± 0.15) in malignant cases. The A_c, A_n, and Ar_{nc} were found to be highly significant parameters in distinguishing benign versus malignant cases with

Table 2
2D morphological parameters for the two categories of cells.

Patient ID	Malignant Cells			Benign Cells		
	A_c (μm^2) Mean ± STD	A_n (μm^2) Mean ± STD	Ar_{nc} Mean ± STD	A_c (μm^2) Mean ± STD	A_n (μm^2) Mean ± STD	Ar_{nc} Mean ± STD
P-1	373.0 ± 341.0	157.4 ± 212.3	0.39 ± 0.16	44.21 ± 25.37	14.33 ± 13.38	0.31 ± 0.19
P-2	298.3 ± 293.8	116.3 ± 123.9	0.41 ± 0.15	54.27 ± 70.00	23.13 ± 20.71	0.48 ± 0.12
P-3	179.3 ± 61.75	89.76 ± 36.97	0.51 ± 0.15	36.08 ± 6.80	18.50 ± 3.81	0.52 ± 0.07
P-4	146.2 ± 60.61	84.73 ± 44.07	0.57 ± 0.14	29.10 ± 4.45	16.26 ± 2.80	0.57 ± 0.10
P-5	185.0 ± 173.8	80.92 ± 115.5	0.42 ± 0.12	35.36 ± 22.85	14.97 ± 6.29	0.46 ± 0.10
P-6	0	0	0	79.71 ± 47.31	34.39 ± 17.81	0.47 ± 0.13
P-7	0	0	0	69.43 ± 56.70	33.00 ± 22.15	0.53 ± 0.12
P-8	125.8 ± 69.20	47.11 ± 32.97	0.40 ± 0.13	30.16 ± 4.55	17.96 ± 3.06	0.46 ± 0.10
P-9	0	0	0	32.22 ± 6.79	18.24 ± 5.03	0.57 ± 0.11
P-10	0	0	0	31.62 ± 9.32	17.96 ± 4.87	0.59 ± 0.14
Total cells/Group	229.5 ± 240.7	97.25 ± 136.0	0.42 ± 0.15	49.62 ± 43.03	23.26 ± 16.33	1.51 ± 0.149

Table 3
The total number of both malignant and benign cells extracted from the cytology images of 10 patients and classified into three clusters.

Patient ID	C-1		C-2		C-3		total
	Malignant cell	Benign cell	Malignant cell	Benign cell	Malignant cell	Benign cell	
P-1	48 (34 %)	0 (0 %)	17 (12 %)	34 (24 %)	0 (0 %)	43 (30 %)	142
P-2	16 (17 %)	3 (3 %)	8 (8 %)	15 (16 %)	3 (3 %)	50 (53 %)	95
P-3	8 (16 %)	0 (0 %)	16 (31 %)	4 (8 %)	0 (0 %)	23 (45 %)	51
P-4	4 (17 %)	0 (0 %)	8 (33 %)	0 (0 %)	0 (0 %)	12 (50 %)	24
P-5	11 (11 %)	0 (0 %)	37 (36 %)	6 (6 %)	0 (0 %)	49 (48 %)	103
P-6	0 (0 %)	10 (8 %)	0 (0 %)	76 (59 %)	0 (0 %)	43 (33 %)	129
P-7	0 (0 %)	14 (8 %)	0 (0 %)	71 (42 %)	0 (0 %)	85 (50 %)	170
P-8	5 (3 %)	0 (0 %)	57 (39 %)	1 (1 %)	4 (3 %)	78 (54 %)	145
P-9	0 (0 %)	0 (0 %)	0 (0 %)	3 (5 %)	0 (0 %)	55 (95 %)	58
P-10	0 (0 %)	1 (1 %)	0 (0 %)	10 (7 %)	0 (0 %)	138 (93 %)	149
total	**120**		**363**		**583**		**1066**

$p < 0.0001$ based on one-way ANOVA method. To gain more insight on identifying the measured cells, it is useful to use clustering analysis algorithms.

All the measured data imported to in-house developed MATLAB code for clustering calculations. The imported data is in the form of a 1066×3 matrix. The 1066 rows correspond to the malignant and benign cells sequence, and the three columns correspond to the 2D parameters of A_c, A_n, and Ar_{nc}. The calculated parameters were analyzed for cell classification by HC + GMM clustering techniques. These techniques are used to classify the measured data into three clusters, malignant cell cluster (C-1), benign cell cluster (C-3), and mixture cell cluster (C-2). Table 3 lists the results of clustering and corresponding numbers of C-1, C-2, and C-3. From this table, one can observe that clustering processes performed well for identifying the cells based on their 2D morphology. Among the three clusters, about 76.7 % of the cells clustered in C-1 are malignant cells, and 98.8 % of the cells clustered in C-3 are benign cells. C-2 has a mixed portion of 39.4 % malignant cells and 60.6 % benign cells. The confusion matrix of the clustering is shown in Table 4.

Fig. 2(a), (b), and (c) present charts for the mean and 95 % Confidence Interval (CI) values of A_c, A_n, and Ar_{nc} of the three clusters of serous effusion cells. The results show monotonous changes of area for cell and nucleus, Fig. 2(a), (b). While we note in Fig. 2 (c) that the small cell cluster C-3 exhibits the opposite trend of area ratio change between cell and nucleus.

We further analyzed the clustering results using two parameters A_c and Ar_{nc}, in 2D parameter space for the measured data as shown in Fig. 3. The distribution of malignant cells and benign cells in the two parameters space for each patient shown in Fig. 3(a). One can find from this figure that most of benign cells form all patients are found on the left side of the graph specifically in the region of small cell area and while the malignant cells are spread in the bigger region of relatively large cell area with a certain number of benign cells and malignant cells are overlapped in the middle region. In contrast, both types of cells span a wide range of nuclear to cell area ratio. Fig. 3(b) presents the distribution of both types of cells in terms of clustering results, and clearly demonstrates the satisfactory results of the HC + GMM clustering method. In comparison with measurements

Table 4
The confusion matrix of cells clustering.

Clusters	C-1 Large cell cluster (A_c: 380.26 ± 306.05 μm^2 A_n: 165.08 ± 185.72 μm^2 95 % CI: 361.1, 399.4 μm^2)	C-2 Medium cell cluster (A_c: 107.11 ± 44.99 μm^2 A_n: 42.87 ± 19.23 μm^2 95 % CI: 96.51, 117.71 μm^2)	C-3 Small cell cluster (A_c: 31.21 ± 6.03 μm^2 A_n: 16.99 ± 5.28 μm^2 95 % CI: 23.147, 39.272 μm^2)	Total
Malignant cells	92 (76.7 %)	143 (39.4 %)	7 (1.2 %)	242
Benign cells	28 (23.3%)	220 (60.6 %)	576 (98.8 %)	824
Total	120	363	583	1066

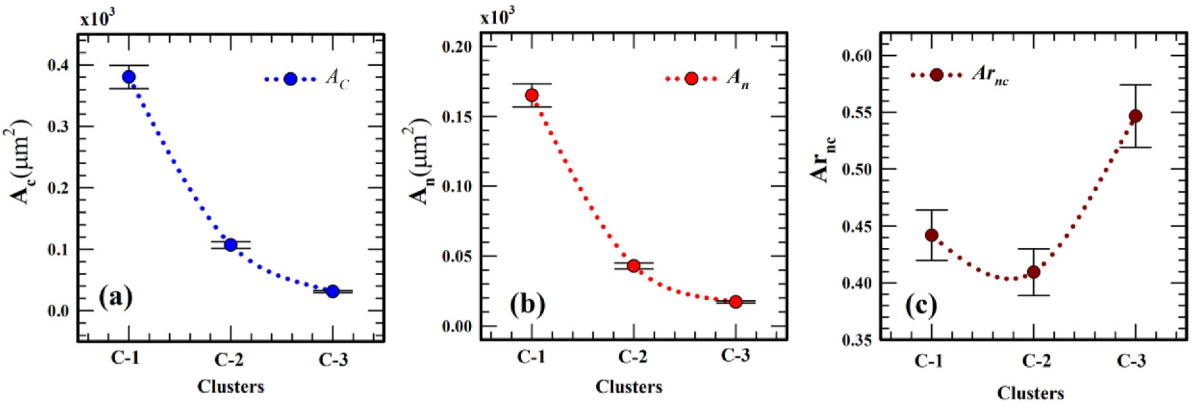

Fig. 2. This figure shows the mean values of cell area, nucleus area, and the area ratio of nucleus to the cell with the 95 % confidence interval (CI) bars of three clusters resulted from GMM clustering process for the two-dimensional morphometric data.

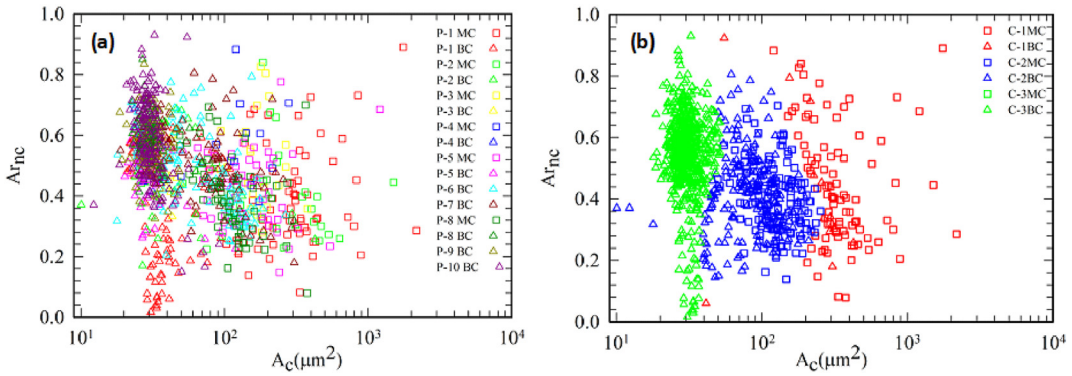

Fig. 3. Scatter plots of A_c and Ar_{nc} of selected cells from cytology images of three patients: (a) malignant cells (MC) and benign cells (BC) of 10 patients marked in colors; (b) results of all analyzed cells clustered by HC+GMM.

summarized in Table 3, these results provide educated guess on the nature of cells, which are either malignant cells or benign cells based the parameters of cell and nuclear area value.

4. Discussion

The detection of malignant cells in serous effusion specimens is thus critical for both therapeutic and prognostic purposes [39]. However, in cytologic preparation of isolated cells, malignancy is sometimes difficult to detect on simple morphologic grounds, and malignant cells are frequently difficult to differentiate from particular cells in effusions such as macrophages and reactive mesothelial cells [4]. This reflects not just inherent problems in cytological diagnosis, such as sample issues and the presence of particular cells, but also systematic bias owing to cytologists' properly cautious approach, who would rather underdiagnose than overdiagnose certain conditions. Despite morphological skill in identification, the employment of certain additional procedures is frequently necessary, regardless of the fact that it is typically time consuming and costly [40]. Quantitative profiling of cell morphology has the potential to be a useful approach in these circumstances, both in terms of diagnostic accuracy and as the foundation for an automated screening program.

Recently, employing image analysis software has been seen as a means of conducting quantitative cytology surveys. This technique is capable of highlighting indistinguishable cellular variations. ImageJ is a popular open-source image analysis program capable of reading most common image file formats used in the biomedical field, supporting image manipulations, and preforming basic operations and calculations [41]. Nikousefat et al. conducted a study to test the usefulness of the ImageJ program in the classification of lymphoma and hyperplasia with fine needle aspiration from mandibular lymph nodes of dogs with

lymphadenopathy. The data obtained for nuclear morphometric features (diameter, circularity, area, and perimeter) revealed a significant difference between the control group and the hyperplasia and lymphoma groups. The cytology diagnosis technique used in this study was rapid, simple, and cheap and the assessment by image analysis software can increase diagnostic accuracy [32]. In 2015, Raghavan et al. published two independent studies that discussed the use of ImageJ and three of its plug-ins to analyze the nuclear parameters in different kinds of carcinoma. The parameters analyzed in these studies were the nuclear area, perimeter, and circularity. The results obtained were statistically analyzed and suggested that morphometric analysis of nuclear parameters using ImageJ is helpful in the grading of tumors and in assessing their prognosis [42,43].

Likewise, in the present study, we used Fiji (ImageJ distribution) to extract both cell and nuclear morphological parameters such as A_c, A_n, and Ar_{nc}. These three parameters were measured for identifying malignant and benign cells in effusion specimens. Furthermore, these parameters were found to be the most reliable criteria for classifying the C-1, C-2, and C-3 clusters. Two independent studies [5,44] emphasized the significance of A_c and A_n as morphometric variables. Arora et al. compared morphometric analysis of cells in 100 effusion samples and revealed that the observed mean values for benign cells were (185.7 ± 43.69) and (58.39 ± 12.22) versus (274.65 ± 61.01) and (120.12 ± 16.28) for malignant cells [5]. In the present study, the mean values of A_c and A_n in 10 effusion samples were (49.62 ± 43.03) and (23.26 ± 16.33) for benign cells, and (229.5 ± 240.7) and (97.25 ± 136.0) for malignant cells respectively. The reasons behind the differences in the measured values in both studies belong to the total number of samples, the variety of diagnosed cases, and the magnification power of the microscope and imaging system. Athanassiadou et al. discovered cytomorphometry of effusion smears to be an improved approach for distinguishing benign or malignant cells from suspicious/atypical reactive cells. They measured two nuclear morphometric variables (nuclear major axis length and nuclear area) using a computerized image analysis system [44]. However, there are significant differences in nuclei and cytoplasmic areas between benign and malignant cells, and we discovered that quantified Ar_{nc} was another useful measure in distinguishing these two groups ($P < 0.001$).

Separately from quantifying cell and nuclear features, machine learning analysis has been found to be an incredibly useful technique for segmentation, identification, and differentiation of malignant and benign cells not only in serous effusion but also in thyroid, breast, and urine cytology specimens. Some of the earlier studies suggested that a combination of cell and nuclear morphometric features and supervised machine learning algorithms of artificial neural network (ANN) or deep learning would be valuable tools in the analysis of cytologic data, while other studies proposed unsupervised machine learning algorithms of adaptive thresholding and Gaussian mixture clustering segmentation method as a cytological diagnostic tool. The findings of these studies revealed a high level of accuracy [23–30,45]. In this study, we adopted the integrated algorithm of HC + GMM clustering techniques, a form of unsupervised machine learning in which a computer can be used to cluster effusion cells into three clusters (C-1, C-2, and C-3) based on their similarity in the measured parameter space without prior knowledge of cell types. The three clusters were defined as the malignant cell cluster (C-1), the benign cell cluster (C-3), and the mixed cell cluster (C-2). The results showed that the clustering process functioned effectively to identify the cells. 76.7 % of the cells in cluster C-1 are malignant, whereas 98.8 % of those in cluster C-3 are benign. C-2 includes a mixed percentage of 39.4 % malignant cells and 60.6 % benign cells. These results deepen our understanding of cytological diagnosis in terms of 2D morphology, which allows future development of morphology tools for cytology.

5. Conclusion

In this study, we intend to quantify 2D morphological differences of cells extracted from 10 specimens of malignant and benign serous effusion and sort them automatically into three different clusters using unsupervised machine learning tools of Hierarchical Clustering and Gaussian mixture model (GMM). The results associate the nature of cells to the morphological parameters of the cell and nuclear area value. The performance of clustering is evaluated using a confusion matrix (Table 4). The results show that 62.6 % of the cells well characterized by the clustering algorithms: 76.7 % of the cells in the large cell cluster C-1 are malignant cells, and 98.8 % of the cells in the small cell cluster C-3 are benign cells. The medium cell cluster C-2 contains a mixture of malignant and benign cells, 39.4 % and 60.6 % respectively. However, quantitative morphology analysis allows clustering of serous effusion cells objectively with a machine learning tool requiring no training. While our current

result can identify malignant cells with high specificity, the sensitivity is still not high enough for practical clinical use. This study is an encouraging beginning for future development of new methods of research to rapidly profile perfusion cells by morphometric features for the identification of malignant cells in clinical samples.

Acknowledgements

Authors are thankful to East Carolina University for supported and facilitated this study.

References

[1] S. Mutsaers, Mesothelial cells: their structure, function and role in serosal repair, Respirology 7 (2002) 171–191, https://doi.org/10.1046/j.1440-1843.2002.00404.x.

[2] R. Light, Pleural Disease, sixth ed., Lippincott Williams & Wilkins, Philadelphia, 2013.

[3] H. Ehya, Effusion cytology, Clin. Lab. Med. 11 (1991) 443–467. PMID: 1873966.

[4] E. Cibas, B. Ducatman, Cytology: Diagnostic Principle and Clinical Correlates, third ed., Saunders, Philadelphia, 2009.

[5] B. Arora, S. Setia, B. Rekhi, Role of computerized morphometric analysis in diagnosis of effusion specimens, Diagn. Cytopathol. 34 (2006) 670–675, https://doi.org/10.1002/dc.20541.

[6] J. Kastelik, Management of malignant pleural effusion, Lung 191 (2013) 165–175, https://doi.org/10.1007/s00408-012-9445-1.

[7] J. Thomas, A. Musani, Malignant pleural effusions: a review, Clin. Chest Med. 34 (2013) 459–471, https://doi.org/10.1016/j.ccm.2013.05.004.

[8] D. Das, Serous effusions in malignant lymphomas: a review, Diagn. Cytopathol. 34 (2006) 335–347, https://doi.org/10.1002/dc.20432.

[9] W. Johnston, The malignant pleural effusion: a review of cytopathologic diagnoses of 584 specimens from 472 consecutive patients, Cancer 56 (1985) 905–909, https://doi.org/10.1002/1097-0142(19850815)56:4<905::AID-CNCR2820560435>3.0.CO;2-U.

[10] J. Wong, D. Pitlik, F. Abdul-Karim, Cytology of pleural, peritoneal and pericardial fluids in children: a 40-year summary, Acta Cytol. 41 (1997) 467–473, https://doi.org/10.1159/000332540.

[11] N. Scott, J. Sutton, C. Gray, Morphometric diagnosis of serous effusions: refinement of differences between benign and malignant cases by use of outlying values and larger sample size, J. Clin. Pathol. 42 (1989) 607–612, https://doi.org/10.1136/jcp.42.6.607.

[12] K. Jacobs, J. Lu, X. Hu, Development of a diffraction imaging flow cytometer, Opt. Lett. 34 (2009) 2985–2987, https://doi.org/10.1364/OL.34.002985.

[13] K. Dong, Y. Feng, K. Jacobs, J. Lu, S. Brock, L. Yang, F. Bertrand, M. Farwell, X. Hu, Label-free classification of cultured cells through diffraction imaging, Biomed. Opt. Express 2 (2011) 1717–1726, https://doi.org/10.1364/BOE.2.001717.

[14] Y. Feng, N. Zhang, K. Jacobs, W. Jiang, L. Yang, Z. Li, J. Zhang, J. Lu, X. Hu, Polarization imaging and classification of Jurkat T and Ramos B cells using a flow cytometer, Cytom. Part A 85 (2014) 817–826, https://doi.org/10.1002/cyto.a.22504.

[15] W. Jiang, J. Lu, L. Yang, Y. Sa, Y. Feng, J. Ding, X. Hu, Comparison study of distinguishing cancerous and normal prostate epithelial cells by confocal and polarization diffraction imaging, J. Biomed. Opt. 21 (2015) 1–10, https://doi.org/10.1117/1.JBO.21.7.071102, 071102.

[16] H. Wang, Y. Feng, Y. Sa, J. Lu, J. Ding, J. Zhang, X. Hu, Pattern recognition and classification of two cancer cell lines by diffraction imaging at multiple pixel distances, Pattern Recogn. 61 (2017) 234–244, https://doi.org/10.1016/j.patcog.2016.07.035.

[17] S. Al-Qaysi, H. Hong, Y. Wen, J. Lu, Y. Feng, X. Hu, Profiling pleural effusion cells by a diffraction imaging method, Proc. SPIE 10497, Imaging, Manip. Anal. Biomol. Cells, Tissues XVI (2018) 1–7, https://doi.org/10.1117/12.2288386, 104971J.

[18] W. Choi, C. Fang-Yen, K. Badizadegan, S. Oh, N. Lue, R. Dasari, M. Feld, Tomographic phase microscopy, Nat. Methods 4 (2007) 717–719, https://doi.org/10.1038/nmeth1078.

[19] T. Kim, R. Zhou, M. Mir, S. Babacan, S. Carney, L. Goddard, G. Popescu, White-light diffraction tomography of unlabelled live cells, Nat. Photonics 8 (2014) 256–263, https://doi.org/10.1038/nphoton.2013.350.

[20] C. Rasmi, S. Padmanabhan, K. Shirlekar, K. Rajan, R. Manjithaya, V. Singh, P. Mondal, Integrated light-sheet imaging and flow-based enquiry (iLIFE) system for 3D in-vivo imaging of multicellular organism, Appl. Phys. Lett. 111 (1–5) (2017) 243702, https://doi.org/10.1063/1.5009782.

[21] I. Saha, P. Dey, H. Vhora, R. Nijhawan, Role of DNA flow cytometry and image cytometry on effusion fluid, Diagn. Cytopathol. 22 (2000) 81–85, https://doi.org/10.1002/(SICI)1097-0339(200002)22:2<81::AID-DC4>3.0.CO;2-C.

[22] M. Ambroise, P. Jothilingam, A. Ramdas, Utility of nuclear morphometry in effusion cytology, Asian Pacific J. Cancer Prev. 15 (2014) 6919–6922, https://doi.org/10.7314/APJCP.2014.15.16.6919.

[23] M. Kowal, P. Filipczuk, A. Obuchowicz, J. Korbicz, Computer-aided diagnosis of breast cancer using Gaussian mixture cytological image segmentation, J. Med. Info. Technol. 17 (2011) 1563–1572. ISSN 1642-6037.

[24] A. Ippolito, M. De Laurentiis, G. La Rosa, A. Eleuteri, R. Tagliaferri, S. De Placido, R. Vigneri, A. Belfiore, Neural network analysis for evaluating cancer risk in thyroid nodules with an indeterminate diagnosis at aspiration cytology: identification of a low-risk subgroup, Thyroid 14 (2004) 1065–1071, https://doi.org/10.1089/thy.2004.14.1065.

[25] P. Dey, R. Logasundaram, K. Joshi, Artificial neural network in diagnosis of lobular carcinoma of breast in fine-needle aspiration cytology, Diagn. Cytopathol. 41 (2013) 102–106, https://doi.org/10.1002/dc.21773.

[26] C. Muralidaran, P. Dey, R. Nijhawan, N. Kakkar, Artificial neural network in diagnosis of urothelial cell carcinoma in urine cytology, Diagn. Cytopathol. 43 (2015) 443–449, https://doi.org/10.1002/dc.23244.

[27] K. Win, S. Choomchuay, K. Hamamoto, M. Raveesunthornkiat, Artificial neural network based nuclei segmentation on cytology pleural effusion images, in: 2017 International

Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS), 2018-Janua, 2017, pp. 245−249, https://doi.org/10.1109/ICIIBMS.2017.8279748.

[28] S. Aboobacker, D. Vijayasenan, S. Sumam David, P.K. Suresh, S. Sreeram, A deep learning model for the automatic detection of malignancy in effusion cytology, in: 2020 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC), 2020, pp. 1−5, https://doi.org/10.1109/ICSPCC50002.2020.9259490.

[29] A. Barwad, P. Dey, S. Susheilia, Artificial neural network in diagnosis of metastatic carcinoma in effusion cytology, Cytometry B Clin. Cytometry 82 (2012) 107−111, https://doi.org/10.1002/cyto.b.20632.

[30] A. Pouliakis, C. Margari, N. Margari, C. Chrelias, D. Zygouris, C. Meristoudis, I. Panayiotides, P. Karakitsos, Using classification and regression trees, liquid-based cytology and nuclear morphometry for the discrimination of endometrial lesions, Diagn. Cytopathol. 42 (2014) 582−591, https://doi.org/10.1002/dc.23077.

[31] J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J. Tinevez, D. White, V. Hartenstein, K. Eliceiri, P. Tomancak, A. Cardona, Fiji: an open-source platform for biological-image analysis, Nat. Methods 9 (2012) 676−682, https://doi.org/10.1038/nmeth.2019.

[32] Z. Nikousefat, M. Hashemnia, A. Karamyan, Potential of the ImageJ software in the differentiation of various types of canine lymphadenopathies with fine needle aspiration cytology (FNAC), Comp. Clin. Pathol. 27 (2018) 643−648, https://doi.org/10.1007/s00580-018-2641-5.

[33] W. Zhao, C. Wang, H. Chen, M. Chen, S. Yang, High-speed cell recognition algorithm for ultrafast flow cytometer imaging system, J. Biomed. Opt. 23 (1−8) (2018), 046001, https://doi.org/10.1117/1.JBO.23.4.046001.

[34] S. Al-Qaysi, Profiling Effusion Cells by Quantitative Analysis of Morphology and Diffraction Imaging Patterns, East Carolina University, North Carolina, 2019.

[35] R. Chellappa, A. Veeraraghavan, N. Ramanathan, C. Yam, M. Nixon, A. Elgammal, J. Boyd, J. Little, N. Lynnerup, P. Larsen, D. Reynolds, Gaussian mixture models, in: Encyclopedia of Biometrics, Springer, Boston, 2009, pp. 659−663.

[36] MathWorks Documentation, Fit Gaussian mixture model to data. https://www.mathworks.com/help/stats/fitgmdist.html.

[37] C. Biernacki, G. Celeux, G. Govaert, Choosing starting values for the EM algorithm for getting the highest likelihood in multivariate Gaussian mixture models, Comput. Stat. Data Anal. 41 (2003) 561−575, https://doi.org/10.1016/S0167-9473(02)00163-9.

[38] J. Blömer, K. Bujna, Adaptive seeding for Gaussian mixture models, in: J. Bailey, L. Khan, T. Washio, G. Dobbie, J. Huang, R. Wang (Eds.), Advances in Knowledge Discovery and Data Mining, PAKDD 2016, Lecture Notes in Computer Science, vol. 9652, Springer, Cham, 2016, https://doi.org/10.1007/978-3-319-31750-2_24.

[39] S. Monte, H. Ehya, W. Lang, Positive effusion cytology as the initial presentation of malignancy, Acta Cytol. 31 (1987) 448−452. ISSN: 00015547.

[40] S. Mohanty, Serous effusions: diagnosis of malignancy beyond cytomorphology. An analytic review, Postgrad. Med. 79 (2003) 569−574, https://doi.org/10.1136/pmj.79.936.569.

[41] M. Abràmoff, P. Magalhães, S. Ram, Image processing with ImageJ, Biophot. Int. 11 (2005) 36−42. ISSN: 10818693.

[42] R. Vijayashree, K. Ramesh Rao, A semi - automated morphometric assessment of nuclei in pap smears using ImageJ, J. Evol. Med. Dent. Sci. 4 (2015) 5363−5370, https://doi.org/10.14260/jemds/2015/784.

[43] R. Vijayashree, K. Ramesh Rao, An ImageJ based semi-automated morphometric assessment of nuclei in oncopathology, Int. J. Sci. Stud. 3 (2015) 189−194, https://doi.org/10.17354/ijss/2015/475.

[44] P. Athanassiadou, N. Kavantzas, P. Davaris, M. Gonidi, A. Liossi, L. Nakopoulou, E. Petrakakou, Diagnostic approach of effusion cytology using computerized image analysis, J. Exp. Clin. Cancer Res. 21 (2002) 49−56. ISSN: 03929078.

[45] C. Fragopoulos, A. Pouliakis, C. Meristoudis, E. Mastorakis, N. Margari, N. Chroniaris, N. Koufopoulos, A. Delides, N. Machairas, V. Ntomi, K. Nastos, I. Panayiotides, E. Pikoulis, E. Misiakos, Radial basis function artificial neural network for the investigation of thyroid cytological lesions, J. Thyroid Res. (2020) 1−14, https://doi.org/10.1155/2020/5464787.