5-2016

A Clinical Decision Support System for Malignant Pleural Effusion Analysis

Troy J, Vargason
Binghamton University--SUNY

Joshua Cohn

David Rios

Olivia Schultz

Joseph Cleary

See next page for additional authors

Authors
Troy J, Vargason; Joshua Cohn; David Rios; Olivia Schultz; Joseph Cleary; Dennis Lau; Walker Land; J. David Schaffer; Yinglei Li; Chun-An Chou; Saba A. Syouri; Jagmohan S. Sidhu; Martha F. Nelson; and Xingye Qiao

Proceedings of the 2016 Industrial and Systems Engineering Research Conference
H. Yang, Z. Kong, and MD Sarder, eds.

A Clinical Decision Support System for Malignant Pleural Effusion Analysis

Troy J. Vargason, Joshua Cohn, David Rios, Olivia Schultz, Joseph Cleary, Dennis Lau, Walker H. Land Jr., J. David Schaffer
Department of Biomedical Engineering, State University of New York at Binghamton
Binghamton, NY 13902 USA

Yinglei Li, Chun-An Chou, Saba A Syouri
Department of Systems Science and Industrial Engineering, State University of New York at Binghamton
Binghamton, NY 13902 USA

Jagmohan Sidhu, Martha F. Nelson
UHS Wilson Medical Center
Johnson City, NY 13790 USA

Xingye Qiao
Department of Mathematical Sciences, State University of New York at Binghamton
Binghamton, NY 13902 USA

Abstract

Pleural effusion occurs when fluid accumulates in the pleural cavity surrounding the lung. This condition is commonly caused by infection, but can also be associated with the presence of a metastatic tumor. Samples of pleural fluid are used to analyze the morphologies of mesothelial cells and can typically be used to make a diagnosis between benignity and malignancy. Atypical pleural effusion samples are not easily identified as benign or malignant due to a lack of differentiable visual features, and such a problem has a significant influence in clinicians' decision making. In this paper, the goal is to develop a clinical decision support system (CDSS) using computer imaging and machine learning techniques for diagnosing atypical pleural effusion. The proposed approach involves four steps for analyzing slides of pleural effusion samples: image processing, feature measurement, feature selection, and classification. Processing and measurement of images produced a preliminary data set of 500 samples; each is described by 398 features. A genetic algorithm was applied for feature selection and identified a subset of 39 important features. The experimental results showed that the selected features can distinguish atypical nuclei as benign or malignant with a five-fold cross validation accuracy of 91%.

Keywords: decision support system, pleural effusion, machine learning

1. Introduction

The pleura is a membrane that lines the cavity of the human lung and is comprised of two layers of mesothelial cells. In a healthy individual, the body naturally produces fluid to lubricate the pleura to ensure effective lung expansion and contraction during respiration. However, when an infection, bacterial or otherwise, is present, excess pleural fluid may be produced and can accumulate in the cavity; this condition is known as pleural effusion. Irritation of the pleura during an infection causes its mesothelial cells to enlarge and take on a cube-like shape; mesothelial cells in this state are referred to as "reactive." In more potentially lethal scenarios, pleural effusion can occur in the presence of a metastatic tumor, in which case the swollen mesothelial cells are given a "malignant" classification [1]. During the occurrence of pleural effusion, mesothelial cells often separate from the membrane and float freely in the pleural fluid. If a patient is diagnosed with pleural effusion, clinicians will typically insert a needle behind the rib cage to remove a fluid sample

from the cavity. After a stain is applied to the sample, a clinician will visually inspect it under a high-resolution microscope to determine whether the occurrence of the pleural effusion is due to cancer or a benign condition. The nuclei of benign mesothelial cells have morphological characteristics that are distinct from those of malignant mesothelial cells, so benign and malignant cells are easily differentiated visually by an experienced professional. Difficulties arise if the fluid sample contains reactive mesothelial cells. Reactive cell nuclei have morphological features that are very similar to those of malignant cells, and thus are not easy to visually differentiate from malignant cells. Clinicians are currently unable to make consistently accurate diagnoses of benignity or malignancy based solely on visual inspection of these "atypical" samples. Examples of benign, malignant, and atypical pleural effusion samples are shown in Figure 1. Generally, benign nuclei have a distinct roundness and are more disperse, while malignant nuclei are clumped together and have distorted shapes. Atypical nuclei closely resemble malignant nuclei, but can have additional benign features. These shared features contribute to greater difficulty in visual classification. It should be noted that the smaller purple dots in the images are lymphocytes and are not taken into consideration during classification.

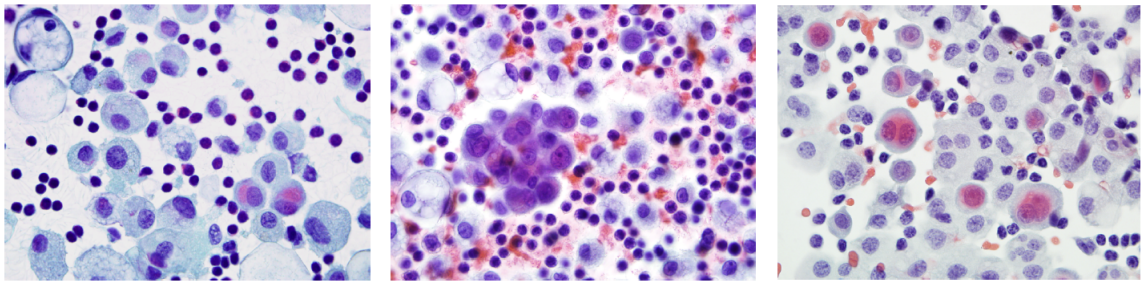

Figure 1: Sample images of stained pleural fluid from approximately 200 images. From left to right: (a) Benign, (b) Malignant, (c)Atypical.

To distinguish atypical samples of pleural effusion, clinicians may use a variety of tests such as analysis of immunohistochemical markers [2] or cytological examination [3]. However, these techniques have a wide range of accuracies and are not always available for on-site use. The use of extra diagnostic tests also requires the patient to spend more money and is a time-consuming process for doctors and clinicians. In addition, an invasive biopsy may be necessary to obtain an acceptable sample of pleural effusion from the patient, in which case the patient's physical health could be further compromised. There are many shortcomings associated with current methods for classifying atypical pleural effusion samples. The development of a faster and more accurate testing procedure would save money and time for patients, doctors, and clinicians. This paper details the process and experimental results of a clinical decision support system (CDSS) that applies machine learning and computer imaging techniques to classify atypical cell samples in pleural effusion quickly and effectively.

2. Methods and System Overview

At the core of the CDSS are four steps for analyzing slides of pleural effusion. The four primary constituents of the CDSS are: (1) image processing; (2) feature measurement; (3) feature selection; and (4) classification. Each component involves the use of different techniques that contribute to the system's operation.

Image processing involves the analysis of pleural effusion sample images and requires the identification of individual nuclei. For use in a clinical setting, it also must be easily learned and simple to use. A graphical user interface (GUI) software package was developed in MATLAB. A view of the current GUI can be seen in Figure 2. The GUI allows the user to interact with the image to identify nuclei. Several GUI functions are available for the user to process an image. The GUI allows for the importing of a pleural effusion sample image, which is then displayed on the axes in the interface. The user can employ an automatic filtering function, which attempts to identify all the nuclei present in the image as accurately as possible, or they can drag a selection box to manually identify specific nuclei of interest. Nuclei identified by the software are overlaid with a highlighting mask. Another function allows for selected nuclei to be unrecognized by the software. One of the most important abilities of the GUI is that which enables the user to divide nuclei that may be clustered or slightly overlapping; this is achieved by drawing a segmented line between two nuclei. When image processing is complete, the sample can be exported to a Portable Network Graphics (PNG) file that contains all highlighted nuclei overlaid on a black background. The black background allows for easier nucleus

identification during the feature measurement step and reduction of noise in each measurement.

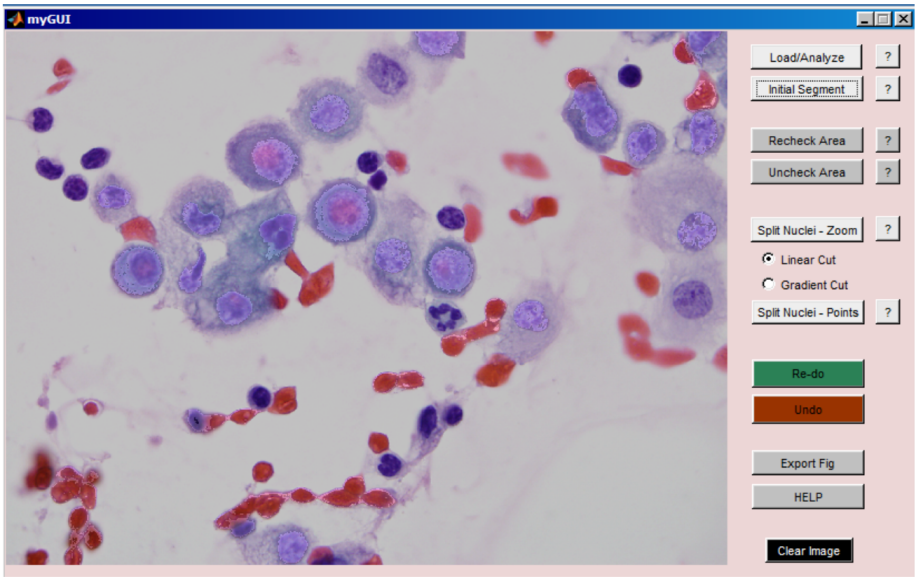

Figure 2: The graphical user interface in MATLAB.

Measurement of features for each nucleus was achieved by using CellProfiler open-source software [4]. Within the program, an algorithm was designed to import a processed image, recognize and assign a number to each individual nucleus, and record measurements. The measurements, which include data pertaining to nucleus texture, coloration, shape, and size, are exported to a comma-separated values (CSV) file with each row corresponding to a numbered nucleus and each column being a separate measurement for a nucleus. An image showing the numbering of each nucleus is also exported. The complete data set was compiled by manually combining all exported CSV files. Data was normalized between 0.1 and 0.9 using tanh() normalization. Ground truths were manually assigned to each atypical nucleus based on assessment of corresponding patient outcomes due to cancer. That is, nuclei in an atypical sample image were classified as benign if the corresponding patient has not died from or been diagnosed with cancer. Similarly, atypical nuclei were classified as malignant if the corresponding patient was diagnosed with cancer or died from cancer.

Feature selection was performed in two ways. First, coarse statistical methods were used to obtain a general understanding of the distribution of the data. Statistical t-tests were used to test for differences between benign and malignant samples within individual features. With this method, only features with significant differences were included in the trimmed data set. Variance pruning was then applied to the remaining features, which involved only the features with the highest coefficients of variation being included in the further reduced data set. For finer feature selection, a genetic algorithm (GA) as described by [5]. It was used with a linear SVM as the validation classifier. It was expected that the GA would produce a subset of features that allowed for more accurate classification than would simple statistical tests.

Several classifiers were used for training and validation of the reduced data sets. First, a random forest classifier was used. A support vector machine (SVM) with different kernels was also employed. These are all described in [6] and [7]. The data was evaluated using five-fold cross validation with a randomized split of the training and validation data. Ten trials were performed for each reduced data set with each classifier.

The four steps of the CDSS method are currently discrete processes and were performed in separate computer applications for purposes of obtaining a preliminary data set and identifying an optimal subset of features. However, future iterations of the CDSS will have feature measurement and classification integrated with the GUI to form one cohesive software package for classifying atypical samples of pleural effusion.

3. Data Processing

Using the four-step method, 61 images of atypical pleural effusion collected from United Health Services were processed and analyzed. This produced a data set of 500 nuclei described by 394 features. Using the aforementioned class assessment technique, 265 benign nuclei and 235 malignant nuclei were identified. Although the master data set contained 394 features, many of these measurements were related to the location or orientation of each nucleus, features which are not relevant to the classification of pleural effusion. Thus, an initial trimming of the data was performed, which reduced the number of relevant features to 132. Coarse feature selection was then applied to the 132 remaining features. Statistical t-tests were first used to test for differences between benign and malignant classes on each feature. Using a significance level of $p < 0.05$ with Bonferroni correction meant that only features with $p < 0.0004$ were included in the reduced data set. This resulted in 61 significant features remaining. A coefficient of variation was then calculated for each feature. Figure 3 shows these coefficients sorted by decreasing value. It was determined that subsets of the 10, 30, and 50 most variable features would be used in the final, coarsely-selected data sets. The respective cutoff points for these subset sizes were variation coefficients of 0.55, 0.24, and 0.10.

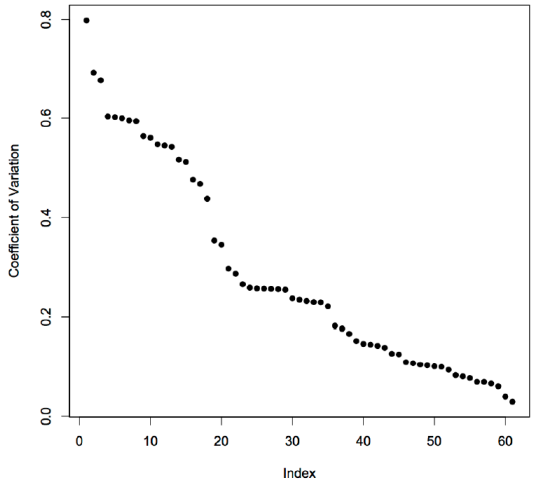

Figure 3: Coefficients of variation, sorted in descending order, for the 61 features remaining after t-test pruning.

A GA-SVM was used to perform fine feature selection with the 132 feature data set. The program was run for 900,000 trials, where 90% of the data were used for training and 100% were used for validation in each trial. After approximately 200,000 trials, the GA converged on 39 features that consistently produced validation Az (area under the receiver operating characteristic curve, or AUC-ROC) values above 0.90 when evaluated in the linear SVM. The names of these features are listed in Table 1, with more detailed descriptions of each feature given by [4].

4. Experimental Results

The four reduced data sets (three coarse, one fine) were evaluated with four classifiers (random forest with 100 trees, linear SVM, SVM with 3rd degree polynomial kernel, and SVM with radial basis function), each using ten trials of randomized five-fold cross validation. Az was used as the primary measure of performance. Az ranges from 0.5 to 1, where a value of 0.5 indicates random guessing by the system and a value of 1 denotes perfect accuracy. Therefore, achieving an Az value of 1 is the goal of the CDSS. The measurement also takes diagnostic sensitivity and specificity into account, so it provides a comprehensive indication of how well the system is able to discriminate between benign and malignant samples of pleural effusion. Average Az results of the four classifier evaluations are shown in Table 2. The data sets containing the fewest features tended to produce lower accuracies in the classifiers. The 39 feature set outperformed the 50 feature set when evaluated in the SVMs with 3rd degree polynomial and radial basis function kernels. Most Az values produced by using the 39 feature set were close to 0.9, and the best average result obtained with this reduced data set was 0.908976 (≈ 0.91) when using a linear SVM.

5. Discussion and Future Directions

As expected, the data sets with fewer features exhibited the worst performance across the four classifiers. The 39 feature subset produced accuracies that were similar to those of the 50 feature data set, but it was anticipated to perform

Table 1: Names of the 39 features obtained from using a GA-SVM for feature selection

1. AreaShape_Area	77. Intensity_UpperQuartileIntensity_OrigBlue
2. AreaShape_Compactness	80. Intensity_UpperQuartileIntensity_OrigRed
5. AreaShape_FormFactor	81. Neighbors_FirstClosestDistance_Expanded
10. AreaShape_MedianRadius	82. Neighbors_NumberOfNeighbors_Expanded
20. AreaShape_Zernike_3_3	95. RadialDistribution_FracAtD_OrigGreen_3of4
22. AreaShape_Zernike_4_2	100. RadialDistribution_FracAtD_OrigRed_4of4
26. AreaShape_Zernike_5_5	101. RadialDistribution_MeanFrac_OrigBlue_1of4
27. AreaShape_Zernike_6_0	106. RadialDistribution_MeanFrac_OrigGray_2of4
28. AreaShape_Zernike_6_2	107. RadialDistribution_MeanFrac_OrigGray_3of4
33. AreaShape_Zernike_7_5	110. RadialDistribution_MeanFrac_OrigGreen_2of4
35. AreaShape_Zernike_8_0	114. RadialDistribution_MeanFrac_OrigRed_2of4
38. AreaShape_Zernike_8_6	117. RadialDistribution_RadialCV_OrigBlue_1of4
40. AreaShape_Zernike_9_1	119. RadialDistribution_RadialCV_OrigBlue_3of4
42. AreaShape_Zernike_9_5	122. RadialDistribution_RadialCV_OrigGray_2of4
43. AreaShape_Zernike_9_7	125. RadialDistribution_RadialCV_OrigGreen_1of4
51. Intensity_LowerQuartileIntensity_OrigGreen	127. RadialDistribution_RadialCV_OrigGreen_3of4
62. Intensity_MaxIntensity_OrigGray	128. RadialDistribution_RadialCV_OrigGreen_4of4
63. Intensity_MaxIntensity_OrigGreen	130. RadialDistribution_RadialCV_OrigRed_2of4
64. Intensity_MaxIntensity_OrigRed	130. RadialDistribution_RadialCV_OrigRed_3of4
67. Intensity_MeanIntensity_OrigGreen	

Table 2: Average Az results from ten trials of five–fold cross validation for each classifier.

Classifier	10 Features (Coarse)	30 Features (Coarse)	50 Features (Coarse)	39 Features (Fine)
Random Forest (100 Trees)	0.86	0.89	0.90	0.90
SVM –Linear Kernel	0.81	0.88	0.91	0.91
SVM –3rd Degree Polynomial	0.80	0.82	0.83	0.85
SVM –Radial Basis Function	0.81	0.86	0.88	0.89

better than it actually did. The use of fewer features to achieve similar results, however, shows that an equal, if not greater, amount of information content can be found with less features. Additional features may simply contribute extra noise. To improve classifier accuracies in the future, more feature selection and pattern recognition techniques will be explored. The methods used in the development of the CDSS are subject to several sources of error. For one, the GUI does not always perfectly identify nuclei without including excess cytoplasm or other undesired microscopic structures. This means that the measured features, such as coloration and shape qualities, may not accurately reflect the traits of the nucleus alone. A remedy to this would be to refine the GUI's filtering methods and improve its sensitivity. Even if the user accurately identifies a nucleus, CellProfiler is not always capable of knowing exactly where a nucleus begins and where it ends. This also contributes extra noise, and may be avoided by modifying the designed CellProfiler algorithm. Additionally, the technique of assigning benign and malignant classes to samples based on later patient cancer outcomes is inherently error-prone. Any incorrectly classified nuclei will make it difficult for classifiers to achieve higher accuracies. Further clinician input would be useful for determining which nuclei have been incorrectly classified in the current pleural effusion data set.

During coarse feature selection, the inclusion of the entire data set when identifying preferred features may have introduced bias into the training and validation process. Although the intention was to gain a preliminary idea of the variability and class separation associated with each feature of the data set, it might be more fruitful to perform these feature selection techniques with a form of five-fold cross validation. That is, when calculating coefficients of variation and performing t-tests, only 80% of the samples might be used. This 80% would then be used to train a classifier, and the remaining 20% of the data would be the validation set.

As it currently stands, the CDSS is a collection of independent computer applications that were manually connected for data collection and analysis. Before use in a clinical setting, the entire process must be automated and easily accessible. This will be achieved by integrating CellProfiler source code and the classifiers with the MATLAB GUI code. Functions for measuring features of selected individual nuclei will also be explored and may be implemented in future versions of the CDSS. Finally, the use of computed tomography (CT) images from corresponding patients may produce input when diagnosing malignant pleural effusion. In past studies, information was extracted from CT scans of pleural effusion for helping with the diagnosis of ovarian cancer [8]. If CT imaging is determined to be useful for identifying other forms of cancer, functionality may be incorporated into the CDSS for analyzing CT scans and measuring features such as volume or density of pleural fluid.

6. Conclusion

A CDSS for the classification of atypical samples of pleural effusion is in development and is based on a four-step method involving image processing, feature measurement, feature selection, and classification. Experimental results have been promising and show that there are features that can identify atypical cells as being benign or malignant. Currently, a 39 feature data subset identified by the use of a GA-SVM provides the most encouraging indication of separability. Although current techniques are subject to computational and user error, the CDSS and its individual tools will continue to be refined to provide better performance. Future testing with different feature selection techniques and classifier algorithms is expected to allow for more accurate classification of atypical pleural effusion.

Acknowledgements

This work was funded by the Howard Hughes Medical Institute through the Binghamton University HHMI Interdisciplinary Research Program. The authors thank UHS Wilson Medical Center for providing pleural effusion sample images for this research. Thanks is also extended to Dr. Guy German and Dr. Craig Laramee of Department of Biomedical Engineering, State University of New York at Binghamton, for their assistance with the graphical user interface and statistical analyses, respectively.

References

[1] Bassen, D., Nayak, S., Li, X.C., Sam, M., Sidhu, J., Nelson, M.F. and Land, W.H., 2013. Clinical Decision Support System (CDSS) for the Classification of Atypical Cells in Pleural Effusions. Procedia Computer Science, 20, pp.379-384.

[2] Hasteh, F., Lin, G.Y., Weidner, N. and Michael, C.W., 2010. The use of immunohistochemistry to distinguish reactive mesothelial cells from malignant mesothelioma in cytologic effusions. Cancer cytopathology, 118(2), pp.90-96.

[3] Abouzgheib, W., Bartter, T., Dagher, H., Pratter, M. and Klump, W., 2009. A prospective study of the volume of pleural fluid required for accurate diagnosis of malignant pleural effusion. CHEST Journal, 135(4), pp.999-1001.

[4] Carpenter, A.E., Jones, T.R., Lamprecht, M.R., Clarke, C., Kang, I.H., Friman, O., Guertin, D.A., Chang, J.H., Lindquist, R.A., Moffat, J. and Golland, P., 2006. CellProfiler: image analysis software for identifying and quantifying cell phenotypes. Genome biology, 7(10), p.R100.

[5] Schaffer, J.D., Janevski, A. and Simpson, M.R., 2005, November. A genetic algorithm approach for discovering diagnostic patterns in molecular measurement data. In Computational Intelligence in Bioinformatics and Computational Biology, 2005. CIBCB'05. Proceedings of the 2005 IEEE Symposium on (pp. 1-8). IEEE.

[6] Hastie, T., Tibshirani, R., Friedman, J. and Franklin, J., 2005. The elements of statistical learning: data mining, inference and prediction. The Mathematical Intelligencer, 27(2), pp.83-85.

[7] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V. and Vanderplas, J., 2011. Scikit-learn: Machine learning in Python. The Journal of Machine Learning Research, 12, pp.2825-2830.

[8] Kim, K.W., Choi, H.J., Kang, S., Park, S.Y., Jung, D.C., Cho, J.Y., Cho, K.S. and Kim, S.H., 2010. The utility of multi-detector computed tomography in the diagnosis of malignant pleural effusion in the patients with ovarian cancer. European journal of radiology, 75(2), pp.230-235.