Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 20 (2013) 379 – 384

Procedia Computer Science

Complex Adaptive Systems, Publication 3
Cihan H. Dagli, Editor in Chief
Conference Organized by Missouri University of Science and Technology
2013- Baltimore, MD.

# Clinical Decision Support System (CDSS) for the Classification of Atypical Cells in Pleural Effusions

David Bassen[a], Saurabh Nayak[a], Xia Chong Li,[a] Mitchell Sam,[a]
Jagmohan Sidhu, MD[b], Martha F. Nelson, RN MS MPA,[b] Walker H. Land, Jr.[a]
[a] Dept. of Bioengineering, Binghamton University Vestal, NY 13902 USA

[b] UHS Wilson Medical Center, Johnson City, NY 13790 USA

**Abstract**

The objective of this research is to develop a prototype Clinical Decision Support System (CDSS) to aid pathologists in correctly discriminating between reactive mesothelial cells and malignant epithelial cells. Currently, there is great difficulty in visually discriminating between cells that are malignant and cells that are otherwise reactive to antigens present in the effusion. Features have been identified, which can correctly discriminate between benign epithelial cells and malignant epithelial cells with a validation **AZ accuracy of ~ 0.934, training AZ of ~ 0.937**. Using these features, the system trained on visually known cases was shown to find discriminating information in the feature subset of the atypical cases by examining probabilities generated from subjecting the system to atypical cells. While these results are preliminary, they do demonstrate that an intelligent CDSS, which has the potential to discriminate between reactive mesothelial cells and malignant epithelial cells, designed using newly developed and /or revised statistical learning theory (SLT) algorithms, has the potential to be used as a second opinion diagnostic aid by physicians, as they deem appropriate.



Keywords: Statistical Learning theory, Pleural effusion, atypical cell classification, clinical decision support systems (CDSS)

## 1. INTRODUCTION

The primary objective of this paper is to describe the operational results for a recently developed CDSS for the classification of atypical cells in pleural effusion. It also has a second objective to provide some background and causes of pleural effusion for the reader unfamiliar with this medical condition.

The two layer membrane consisting of mesothelial cells that lines the serous cavity of the lungs is called the Pleura. In some cases, this thin layer of cells undergoes a 'reactive' hypertrophy (enlargement) which causes the cells to swell up to a cubical (cuboidal) appearance. These cells are referred to as Reactive mesothelial cells (synonyms: Activated mesothelial cells, irritated mesothelial cells, hypertrophic or proliferative mesothelial cells).

Reactive mesothelial cells can be found when there is an infection or an inflammatory response present in a body cavity. This condition can be due to the presence of a bacterial, viral or fungal infection. It can also be the result of trauma or the presence of metastatic tumor. More often than not, these cells detach from the lining of the cavity and


doi:10.1016/j.procs.2013.09.290

introduce themselves into the pleural fluid. These reactive mesothelial cells resemble malignant epithelial cells in that they are almost identical in terms of morphological features. When a sample of pleural effusion is observed under a microscope, it is almost impossible to differentiate between these two types of cells visually.

Under normal conditions, there are certain amounts of fluid present in the pleural cavity. This fluid serves as a lubricant between the two pleural layers, provides nutrients, removes wastes and reduces surface tension. Pleural Effusion occurs when there is an imbalance in the haemostatic mechanism in the system. This imbalance in the mechanism can be caused due to infections, haemorrhage, malignancies and other disorders.

There are two types of Pleural Effusions- Transudative effusion is one that is a result of a systemic disorder that disrupts the balance of fluid production / fluid re-absorption. This effusion is usually pale yellow and clear. Exudative effusion is one that is a result of a problem with the membranes themselves [1]. It is produced by conditions that directly involve the membranes of the particular cavity like infections, inflammations and malignancies. This effusion is of abnormal color and usually bloody, turbid or cloudy.

When pleural fluid is extracted from the body and analysed under a microscope, the atypical cells are observed in order to determine the origin of the effusion. The malignant cells can be differentiated easily from the normal benign cells. However, the Reactive Mesothelial Cells mentioned earlier have multiple characteristics in common with malignant cells. This is the problem that many pathologists face.

While immunohistochemical tests are the preferred method to overcome this problem [2], in the case where there is insufficient cellular material, differentiating between the Reactive Mesothelial cells and Malignant Epithelial cells becomes much more difficult. Cytological examination is another approach which has been used to diagnose pleural effusions, but which has shown to have a wide range of accuracies, and should not be used without support [3]. When available, these methods add extra cost and time to the diagnostic process. When unavailable, diagnosing malignancy in pleural effusion becomes a serious clinical problem. Precedence exists supporting the use of computerized imagery for diagnosis in serous effusions [4].

This brings us to the primary purpose of this paper, which is the description and operation of a preliminary CDSS based on available clinical data collected in this study to aid the physician in the classification of these atypical effusion cells. Our CDSS has been developed, implemented and preliminarily tested. *However, since the developed technology to achieve this objective is proprietary, details or content of the system operation cannot be described and/or discussed.*

## 2. SYSTEM OVERVIEW AND DESCRIPTION

Pleural effusion can house a wide variety of cells which might be benign or malignant in nature. In some cases, pathologists are able to visually differentiate between the two by viewing slides of specifically stained fluid under a light microscope. But, in some cases, certain benign mesothelial cells undergo a reaction to antigens which render them visually similar to malignant epithelial cells. In this case, it is almost impossible to differentiate them visually based on morphological cell features. Moreover, the existing immunochemistry methods are not always helpful and sometimes cannot be used because of lack of adequate cellular material. The generic CDSS depicted in Figure 1 below consists of the pleural effusion data set, the image processing software, software to obtain the features, intelligent processing software, some modeling software and validation processing, and, finally, the output provided by this new CDSS.

An image of the cells is presented to the imaging processing software after slide preparation of pleural effusion sample. Pleural effusion collects at the bottom of the lung and depicted as a white space on chest X-ray (Figure 2).

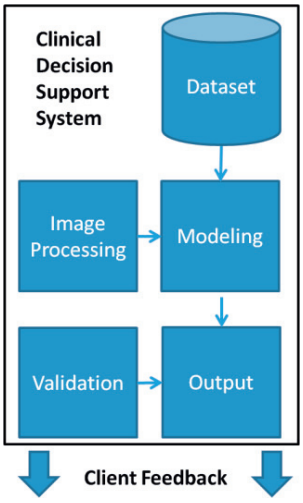



Figure 1. Generic Description of CDSS

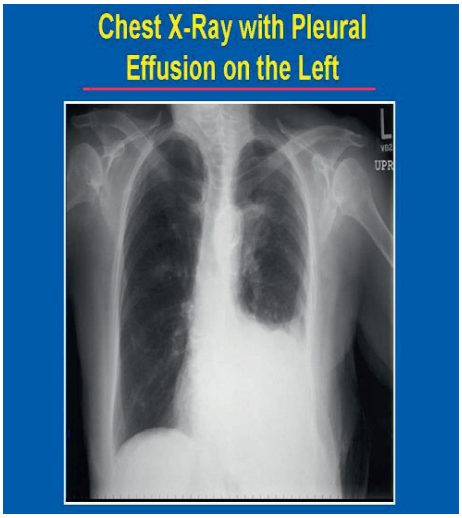



Figure 2. Chest X-ray of pleural effusion

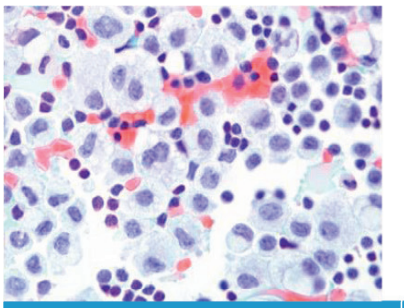
An example of a graphical interface (GUI) image that is processed by the system is depicted in Figure 3, which demonstrates the process of a clinician interacting with the CDSS without the use of typed command lines—a very useful clinical attribute. Here the operator is processing three cells denoted by [9, 2, 3]. Current GUI operational features are: (1) ability to view imported image; (2) ability to "swap" between viewing original and segmented images; (3) test entry field that enables the clinician to type cell number for which cells to be included / excluded from analysis; (4) "buttons," that when activated, will process the analysis for both including / excluding cells.

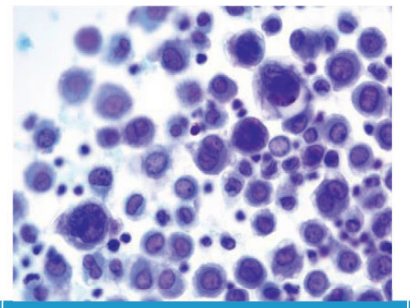
A number of other primary and secondary functions (which are currently depicted by "placeholder buttons") will be completed at a future time (i.e. classify button, which will provide an AZ value along with the status of the cell analysis, intelligent adaptive system train and validate function buttons, and so on).

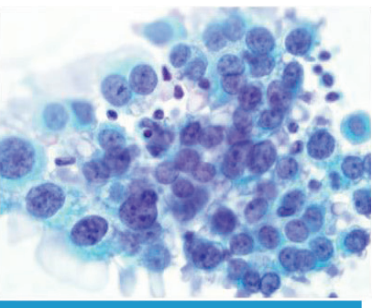

Figure 3. Graphical User Interface (GUI) of the CDSS showing segmented nuclei

## 3. DATA PROCESSING AND PRELIMINARY RESULTS

Figure 4 depicts examples of cell types contained in the image for this preliminary data set, which were collected and provided by Dr. Jagmohan Sidhu. This data set contained benign, malignant and atypical cases (which can be either benign or malignant). These atypical cells cannot be differentiated by visual means or by using *immunochemistry* because of the lack of adequate cellular material. At the time of this writing, our data set is comprised of 928 usable differentiated nuclei as depicted in Figure 4.







| Benign Cases | Malignant Cases | Atypical Cases |
|---|---|---|
| 17 Patients | 19 Patients | 14 Patients |
| 18 Images | 56 Images | 26 Images |
| 269 Usable Nuclei | 456 Usable Nuclei | 203 Usable Nuclei |

Figure 4: The system was preliminarily tested using a data set size of 928 segmented and usable nuclei.

The intelligent processing systems for both the normal and atypical benign / malignant data set were trained and validated by an Evolutionary Programming /Evolutionary Strategies (EP /ES) hybrid process using five–fold cross validation. The results are depicted in Table 1 below.

**Table 1. Performance for Five Trials of 5-fold Cross Validation**

| Evaluation type | AZ values for 5 trials of 5-fold cross validation | | | | | AVG | SD |
|---|---|---|---|---|---|---|---|
| | Trial 1 | Trial 2 | Trial 3 | Trial 4 | Trial 5 | | |
| Training fold | 0.933 | 0.936 | 0.940 | 0.939 | 0.938 | 0.937 | 0.003 |
| Validation fold | 0.930 | 0.928 | 0.938 | 0.939 | 0.933 | 0.934 | 0.005 |

Table 1 demonstrates that the EP /ES training of the intelligent processor provided training and validation accuracies for a clinically differentiable benign / malignant subset of the data set (a subset of data for which the pathologist can define benign and malignant epithelial cells) of 0.93 for both training and validation using a five-fold cross validation process using computer generated features.

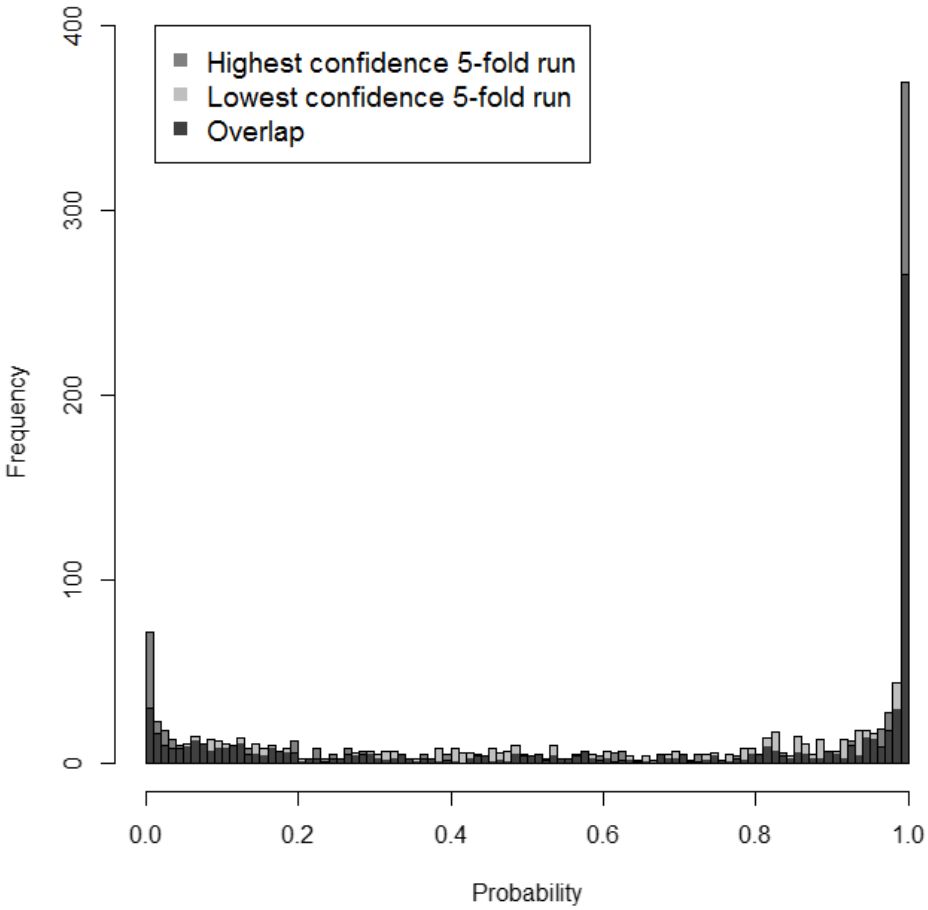



Figure 5. Probabilities generated from the most and least confident 5-fold trials

Figure 5 contains two histograms which are super imposed. To generate the intermediate grey histogram, probabilities were aggregated from all folds of the 5-fold run which classified the most cells in the high confidence range (83%). To generate the light grey histogram, probabilities were aggregated from all folds of the 5-fold run which classified the least cells in the high confidence range (70%). The two histograms are transparent such that darkest regions are overlapping regions of both histograms. Since there are 202 atypical cells, there are 1010 probabilities associated with each histogram. In total, 5050 probabilities were calculated for the five 5-fold runs.

Figure 5 shows the confidence values (probabilities) of the CDSS system classifying atypical cases after being trained and validated on the visually known cases quantified in this study, where 0.0 represents a purely benign case and 1.0 a purely malignant case. While no ground truth currently exists to validate this operation in the same fashion as with visually identifiable cells, these results demonstrate that the patterns seen in the visually generated feature subset exist within the available atypical cell population. Note that a randomly trained system yields a uniform distribution of probabilities as expected.

**Table 2. Confidence of PNN Configuration 1 in Classifying Atypical Cells**

| Range | Percentage of atypical cells in bin | | | | | AVG | SD |
|---|---|---|---|---|---|---|---|
| | Trial 1 | Trial 2 | Trial 3 | Trial 4 | Trial 5 | | |
| High confidence | 83% | 81% | 70% | 73% | 77% | 77% | 5% |
| Low confidence | 12% | 14% | 20% | 18% | 15% | 16% | 3% |
| No confidence | 6% | 5% | 10% | 9% | 7% | 7% | 2% |

Table 2. indicates the percentages of cells classified with 0.2 or lesser probability and 0.8 or greater probability, between 0.2 and 0.4 probability and between 0.6 and 0.8 probability, and between 0.6 and 0.7 probability. Each set of intervals defining confidence ranges of high, low, and no confidence, respectively. Note that for all 5050 probabilities, 49.6% fall above 0.95 or below 0.05.

Table 2. provides a complete view of system confidence for all trials to support the point made by figure 5. In both depictions, the confidence of the system is measured as function of how well it generates two distinct peaks on the probability distribution.

As a whole, this preliminary research has provided the following contributions: (1) generation of a quantified dataset derived from the nuclei of cells found in pleural effusion images; (2) an intelligent system capable of classifying visually known cases which was validated using ROC analysis (AZ around 0.90); (3) an intelligent system with the potential to classify atypical cells; (4) an assembled CDSS capable of accepting images from new cases and rendering output to a clinician. In doing so, this project has laid a framework for the quantitative use of pleural effusion images in the development of second opinion clinical decision support systems.

Further progress will need to be made in conjunction with new data collection methods which allow a ground truth mapping to be made from other types of effusion tests or clinical outcomes to the visually indistinguishable atypical cells.

## 4. CONCLUSIONS

The purpose of this initial research effort is to develop a prototype CDSS to aid pathologists in correctly discriminating reactive benign mesothelial from malignant epithelial cells. Images of stained pleural effusion slides were collected by Dr. Jagmohan Sidhu. The project has provided: (1) generation of a quantified dataset derived from the nuclei of cells found in pleural effusion images; (2) an intelligent system capable of classifying visually known cases which was validated using ROC analysis (AZ around 0.90); (3) an intelligent system with the potential to classify atypical cells; (4) an assembled CDSS capable of accepting images from new cases and rendering output to a clinician. In doing so, this project has laid a framework for the quantitative use of pleural effusion images in the development of second opinion clinical decision support systems.

Internal recommendations have been provided to guide future work with the intention of further validating the clinical usefulness of the system, its theoretical underpinnings, as well as to increase the quality and proper use of current and future data. Clearly, more data accompanied by patient outcomes and other experimental information will be critical to validating and developing the CDSS.

## References

1 Broaddus C, Light RW. Pleural effusion. In: Mason RJ, Broaddus CV, Martin TR, et al, eds. Textbook of Respiratory Medicine. 5th ed. Philadelphia, Pa: Saunders Elsevier;2010:chap 73.

2. Hatteh, F., et.al. *" The Use of Immunohistochemistry to distinguish reactive mesothelial cells from malignant mesothelioma in cytologic effusions"* Cancer Cytopathology, pp 90-96, April 26, 2010

3. Abouzgheib, W., T. Bartter, H. Dagher, M. Pratter, and W. Klump. "A Prospective Study of the Volume of Pleural Fluid Required for Accurate Diagnosis of Malignant Pleural Effusion." *Chest* 135.4 (2009): 999-1001. Print.

4. da Silva VD, Prolla JC, Diehl AR, Baldo MF, Muller RL: Comparison of conventional microscopy and digitized imaging for diagnosis in serous effusions. Anal Quant Cytol Histol 1997; 19(3):202-206.