

## A Soft Computing Method for Mesothelioma Disease Classification

Mehrbakhsh Nilashi <sup>a,c,\*</sup>, Hossein Ahmadi <sup>b</sup>, Leila Shahmoradi <sup>b,\*</sup>, Maryam Salahshour <sup>a</sup>, Othman Ibrahim <sup>a</sup>

<sup>a</sup>Faculty of Computing, Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia

<sup>b</sup>Health Information Management Department, 5th Floor, School of Allied Medical Sciences, Tehran University of Medical Sciences, No #17, Farredanesh Alley, Ghods St, Enghelab Ave, Tehran, Iran

<sup>c</sup>Department of Computer, Lahijan Branch, Islamic Azad University, Lahijan, Iran

\* Corresponding authors email addresses: [nilashidotnet@hotmail.com](mailto:nilashidotnet@hotmail.com); [lshahmoradi@tums.ac.ir](mailto:lshahmoradi@tums.ac.ir)

### Abstract

Malignant Mesothelioma (MM) is a rare but highly aggressive tumour. The aim of this study is to improve the classification accuracy of MM disease by developing an intelligence system using machine learning techniques. Our method is developed through clustering, noise removal and classification approaches. Accordingly, we use Expectation Maximization (EM), Principal Component Analysis (PCA) and Support Vector Machine (SVM) for clustering, noise removal and classification tasks, respectively. We also develop the proposed method for incremental situation by applying the incremental PCA and incremental SVM for incremental learning of data. Experimental results on a malignant pleural mesothelioma disease dataset show that proposed method remarkably improves the accuracy of prediction and reduce computation time in relation to the non-incremental approaches. The hybrid intelligent system can assist medical practitioners in the healthcare practice as a decision support system.

Keywords: Malignant mesothelioma, Clustering, Incremental PCA, Incremental SVM, Machine Learning.

### 1. Introduction

Malignant Mesothelioma (MM) is an aggressive cancer of the serous membranes with a poor prognosis (Gemba et al., 2013; Spugnini et al., 2006). MM is a fatal tumor originating from the mesothelial tissue. There are two major localizations of MM: the pleura and peritoneum. The pleural form of malignant mesothelioma is the most common type accounting for more than 70% of all mesothelioma cases (Suzuki, 1981). Exposure to asbestos is a risk factor independent of tumor localization. The development of a mesothelioma is a lengthy process, the tumor appearing 25-60 years after asbestos exposure. There is no definitive standard of care for this disease and it has been shown that individual modalities such as chemotherapy, radiotherapy and surgery have been failed to prolong survival (Gemba et al., 2013).

This research aims to develop an intelligent system for Mesothelioma disease diagnosis using machine learning methods. Moreover, since in medical datasets constantly new information is available, hence, it is desirable to incrementally update the once trained models to reduce computation time in classifying the data. The proposed method in the study at hand supports incremental updates and re-learning of data and is more efficient in memory requirement. Overall, in comparison with research efforts found in the literature, in this research:

- Expectation Maximization (EM) is used for data clustering the data (Nilashi et al., 2015b; Nilashi et al., 2017; Nilashi et al., 2016a; Nilashi et al., 2016b; Nilashi et al., 2016c; Nilashi et al., 2016d).
- Support Vector Machine (SVM) is used for data classification (Nilashi et al., 2016d).
- Principal Component Analysis (PCA) is used for dimensionality reduction and dealing with the multi-collinearity problem in the experimental data (Nilashi et al., 2015a, Nilashi et al., 2015c).
- Incremental techniques, Incremental SVM (ISVM) and Incremental PCA (IPCA), are used for incremental learning.

Our study at hand is organized as follows: In Section 2, the research methodology is explained. In Section 3, the evaluation of method is presented. Finally, we conclude our work in Section 4.

### 2. Methodology of research

In the present study, PCA, EM, and SVM methods are used. In this study, EM clustering is used as an unsupervised classification method to cluster the data of experimental dataset into similar groups. We propose to rely on SVM to learn the classification models. One reason for this choice is that SVM and incremental SVM has