

A Predictive Method for Mesothelioma Disease Classification Using Naïve Bayes Classifier

Mehrbakhsh Nilashi ^{a,b,*}, Morteza Zamani Roudbaraki ^b, Mohammadreza Farahmand ^c

^a Faculty of Computing, Universiti Teknologi Malaysia, Johor, Malaysia

^b Department of Computer Engineering, Lahijan Branch, Islamic Azad University, Lahijan, Iran

^c Department of Computer Science, Abarkouh Branch, Islamic Azad University, Abarkouh, Iran

* Corresponding author email address: nilashidotnet@hotmail.com

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) and Naive Bayes (NB) for clustering and classification tasks, respectively. Experimental results on a malignant pleural mesothelioma disease dataset show that proposed method remarkably improves the accuracy of mesothelioma's disease prediction. The hybrid intelligent system can assist medical practitioners in the healthcare practice as a decision support system.

Keywords: Malignant Mesothelioma disease diagnosis, Clustering, Naïve Bayes, 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. Due to the association with asbestos exposure this tumor is often the result of industrial practices (Lanphear and Buncher, 1992; Marinaccio et al., 2007). According to the recent studies, BRCA1-associated protein-1 (BAP1) germline gene mutations have been identified and associated with a cancer syndrome that includes Malignant Pleural Mesothelioma (MPM), ocular or cutaneous melanoma, and other cancers (Iacono et al., 2015; Testa et al., 2011; Cheung et al., 2013). MM is associated with poor patient survival due to difficulties in diagnosis and treatment options currently available (Sekido, 2008). This disease is largely unresponsive to traditional cancer treatments such as chemotherapy or radiation (Spugnini et al., 2006). It has caused 15,000–20,000 deaths annually worldwide (Zervos et al., 2008; Er et al., 2015). 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).

Machine learning deals with the development of technologies which allow machines to learn (Crown, 2015). The challenge is to create algorithms that can take a group of patterns (on a broader range, the existing knowledge) and automatically make new inferences from the initial information, with or without human intervention.

From the machine learning perspective, classification is the problem of identifying a set of observations into several categories, basing on the training result of a subset of observations whose belonging category is known. The unsupervised learning is defined as cluster analysis. It is also called clustering. Clustering is a process of putting a set of observations into several reasonable groups according to certain measure of similarity within each group (Nilashi et al., 2015). The clustering problem has been addressed in many diseases diagnosis systems. This reflects its broad appeal and usefulness as one of the steps in exploratory health data analysis.

There is a vast sea of different techniques and algorithms used in data mining especially for supervised machine learning techniques, therefore, selecting appropriate techniques is an important task in developing the disease diagnosis systems with a reasonable accuracy. In addition, many methods and algorithms have been developed for diseases diagnosis by machine learning techniques (Nilashi et al., 2018; Nilashi et al., 2017a; Nilashi et al., 2017b; Nilashi et al., 2017c; Nilashi et al., 2017d; Nilashi et al., 2017e; Nilashi et al., 2016a; Nilashi et