Expert Systems with Applications 39 (2012) 1862–1868

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Diagnosis of chest diseases using artificial immune system

Orhan Er [a], Nejat Yumusak [b], Feyzullah Temurtas [a,*]

[a] Bozok University, Department of Electrical and Electronics Engineering, 66200 Yozgat, Turkey
[b] Sakarya University, Department of Computer Engineering, 54187 Adapazari, Turkey

ARTICLE INFO

Keywords:
Chest disease diagnosis
Artificial immune system

ABSTRACT

Chest diseases are one of the greatest health problems for people living in the developing world. Millions of people are diagnosed every year with a chest disease in the world. Chronic obstructive pulmonary, pneumonia, asthma, tuberculosis, lung cancer diseases are most important chest diseases and these are very common illnesses in Turkey. In this paper, a study on chest diseases diagnosis was realized by using artificial immune system. We obtained the classification accuracy with artificial immune system 93.84%. The result of the study was compared with the results of the previous similar studies reported focusing on chest diseases diagnosis. The chest diseases dataset were prepared from a chest diseases hospital's database using patient's epicrisis reports.

1. Introduction

There are a number of other chest diseases which should challenge the ingenuity and determination of progressive-minded physicians. Diseases of the various organs and structures of the chest represent a large enough group and a complexity of problems to completely engage the talent and ability of interested physicians. Competence and efficiency in handling these conditions are more readily acquired when studied as interrelated items. The very nature of the close anatomic and functional connections between the organs potentially involved implies that only through a correlated understanding of each problem is it possible to diagnose and treat diseases of the chest.

Chest diseases have some phenomenon such as air pollution various infections, and smoking habit have recently increased its risk factors drastically. Thus, millions of people are diagnosed every year with a chest disease in the world (Baemani, Monadjemi, & Moallem, 2008). Tuberculosis (TB), chronic obstructive pulmonary disease (COPD), pneumonia, asthma, lung cancer diseases are the most important chest diseases which are very common illnesses in the world (MedHelp, http://www.medhelp.org/Medical-Dictionary/Terms/2/8964.htm (last accessed: 18 March 2009)).

Tuberculosis is an infectious disease, caused in most cases by microorganisms called "*Mycobacterium tuberculosis*". The microorganisms usually enter the body by inhalation through the lungs. They spread from the initial location in the lungs to other parts of the body via the blood stream, the lymphatic system, via the airways or by direct extension to other organs tuberculosis develops in the human body in two stages. The first stage occurs when an individual who is exposed to micro-organisms from an infectious case of tuberculosis becomes infected (tuberculous infection), and the second is when the infected individual develops the disease (tuberculosis) (Enarson, Rieder, Arnadottir, & Trébucq, 2000). TB is a major cause of illness and death worldwide and globally, 9.2 million new cases and 1.7 million deaths from tuberculosis occurred in 2006 (Royal College of Physicians of London, 2006; World Health Organization, 2008).

COPD is a disease state characterized by airflow limitation that is not fully reversible. The airflow limitation is usually both progressive and associated with an abnormal inflammatory response of the lungs to noxious particles or gases (Celli & MacNee, 2004). Clinically, patients with COPD experience shortness of breath (dyspnea) and cough, productive of an excess of mucus. There may also be wheeze (Jeffery, 1998). According to the World Health Organization (WHO) data is found 600 million patients who have COPD and every year 2.3 million persons die because of COPD in the world (Sönmez & Uzaslan, 2006).

Pneumonia is an inflammation or infection of the lungs most commonly caused by a bacteria or virus. Pneumonia can also be caused by inhaling vomit or other foreign substances. In all cases, the lungs' air sacs fill with pus, mucous, and other liquids and cannot function properly. This means oxygen cannot reach the blood and the cells of the body effectively. According to the World Health Organization (WHO) data, every year approximate 2.4 million persons die because of pneumonia (Global Action Plan for the Prevention, 2007).

Asthma is a chronic disease characterized by recurrent attacks of breathlessness and wheezing. During an asthma attack, the lining of the bronchial tubes swell, causing the airways to narrow and reducing the flow of air into and out of the lungs. Recurrent asthma symptoms frequently cause sleeplessness, daytime fatigue,

* Corresponding author. Tel.: +90 532 207 56 33.
E-mail address: temurtas@gmail.com (F. Temurtas).

0957-4174/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2011.08.064

reduced activity levels and school and work absenteeism. Asthma has a relatively low fatality rate compared to other chronic diseases. WHO estimates that 300 million people currently suffer from asthma. Asthma is the most common chronic disease among children (http://www.who.int/en/ (last accessed: 18 March 2009)).

Lung cancer is a disease of uncontrolled cell growth in tissues of the lung. This growth may lead to metastasis, which is the invasion of adjacent tissue and infiltration beyond the lungs. The vast majority of primary lung cancers are carcinomas of the lung, derived from epithelial cells. Lung cancer, the most common cause of cancer-related death in men and the second most common in women, is responsible for 1.3 million deaths worldwide annually (http://www.who.int/en/ (last accessed: 18 March 2009)).

It is possible to improve the post-treatment lung function by early diagnosis, better treatment and follow up. The appropriate implementation of these methods, which are so important in the early diagnosis of chest, not only improves the opportunity for treatment of chest disease but also has an effect on decreasing deaths from this disease.

Artificial immune system (AIS) is a new artificial intelligence (AI) technique which is beginning to mature through the collaborative effort of many interdisciplinary researchers (Andrews & Timmis, 2005). By modeling some metaphors existing in natural immune system or by inspiring from these metaphors, successful applications have being conducted in AI literature. Classification is among these and there have been some promising studies in this branch of AIS. Considering medical diagnosis as an application domain for AISs, there are several studies like (Castro, Coelho, Caetano, & Von Zuben, 2005; Er, Sertkaya, Temurtas, & Tanrikulu, 2009; Hamaker & Boggess, 2004; Polat, Sahan, & Gunes, 2006; Polat, Sahan, Kodaz, & Gunes, 2005; Sahan, Polat, Kodaz, & Gunes, 2005) in AIS literature.

Artificial immune systems (AIS) can be defined as abstract or metaphorical computational systems developed using ideas, theories, and components, extracted from the immune system. Most AIS aim at solving complex computational or engineering problems, such as pattern recognition, classification, elimination, and optimization. The AIS has been formed on the basis of the working principles of the natural immune system found in the human body (Engin & Döyen, 2004). The biological immune system (BIS) is a subject of great research interest because of its powerful information processing capabilities; in particular, understanding the distributed nature of its memory, self-tolerance and decentralized control mechanisms from an informational perspective, and building computational models believed to better solve many science and engineering problems (Dasgupta, 2006). Chest disease diagnosis via proper interpretation of the chest diseases data is an important classification problem. And, AIS can provide an alternative, efficient way for solving chest disease diagnosis problems.

In this paper, a comparative study on chest diseases diagnosis was realized by using artificial immune systems. The chest diseases dataset were prepared by using patient's epicrisis reports from a chest diseases hospital's database. The study aims also to provide machine learning based decision support system for contributing to the doctors in their diagnosis decisions.

2. Method

2.1. Data source

In order to perform the research reported in this manuscript, the patient's epicrisis taken from Diyarbakir Chest Diseases Hospital from southeast of Turkey was used. The dataset were prepared using these epicrisis reports. The study included 357 patients suffering from variety of respiratory diseases and 38 healthy subjects. The patients were classified as having tuberculosis, COPD, pneumonia, asthma, lung cancer diseases or not sick. Tuberculosis disease were diagnosed in 50 patients, COPD disease were diagnosed in 71 patients, pneumonia disease were diagnosed in 60 patients, asthma disease were diagnosed in 44 patients and lung cancer disease were diagnosed in 32 patients. In dataset we have 100 people who have healthy features.

All samples have 38 features. These features are (Laboratory examination): complaint of cough, body temperature, ache on chest, weakness, dyspnoea on exertion, rattle in chest, pressure on chest, sputum, sound on respiratory tract, habit of cigarette, leucocyte (WBC), erythrocyte (RBC), trombosit (PLT), hematocrit (HCT), hemoglobin (HGB), albumin2, alkalen phosphatase 2 L, alanin aminotransferase (ALT), amylase, aspartat aminotransferase (AST), bilirubin (total + direct), CK/creatine kinase total, CK-MB, iron (SERUM), gamma-glutamil transferase (GGT), glukoz, HDL cholesterol, calcium (CA), blood urea nitrogen (BUN), chlorine (CL), cholesterol, creatinin, lactic dehydrogenase (LDH), potassium (K), sodium (NA), total protein, triglesid, uric acid. Diagnostic tests of each patient were recorded by an attending physician.

2.2. Previous studies

There have been several studies reported focusing on chest disease diagnosis problem using artificial neural network and artificial immune system structures as for other clinical diagnosis problems. These studies have applied different structures to the various chest diseases diagnosis problem using their various dataset (Aliferis, Hardin, & Massion, 2002; Ashizawa et al., 2005; Coppini, Miniati, Paterni, Monti, & Ferdeghini, 2007; El-Solh, Hsiao, Goodnough, Serghani, & Grant, 1999; Er & Temurtas, 2008; Er, Temurtas, & Tanrikulu, 2010; Er, Yumusak, & Temurtas, 2010; Er et al., 2009; Hanif, Lan, Daud, & Ahmad, 2009; Paul, Ben, Thomas, & Robert, 2004; Santos, Pereira, & Seixas, 2004).

El-Solh et al., used generalized regression neural network (GRNN) using clinical and radiographic information to predict active pulmonary tuberculosis at the time of presentation at a health-care facility that is superior to physicians' opinion (El-Solh et al., 1999). The input patterns were formed by 21 distinct parameters which were divided into three groups: demographic variables, constitutional symptoms, and radiographic findings. The output of the GRNN provided an estimate of the likelihood of active pulmonary tuberculosis. The authors utilized a 10-fold cross-validation procedure to train the neural networks. The authors reported approximately 92.3% diagnosis accuracy (El-Solh et al., 1999). Er et al., used multilayer, and generalized regression neural networks for diagnosis of tuberculosis (Er, Temurtas, et al., 2010). They used 38 features for the diagnosis and reported approximately 93.3% diagnosis accuracy for GRNN and 95% diagnosis accuracy for MLNN with LM algorithm and two hidden layer.

Aliferis et al., used KNN, Decision Tree Induction, Support Vector Machines and Feed-Forward Neural Networks for classify non-small lung cancers. The primary goal of their study was to develop machine learning models that classify non-small lung cancers according to histopathology types and to compare several machine learning methods in this learning task. The best multi-gene model found had a leave-one-out accuracy of 89.2% with Feed-Forward Neural Networks (Aliferis et al., 2002).

Ashizawa et al., used the MLNN with one hidden layer and they used BP training algorithm for diagnosis of COPD disease (Ashizawa et al., 2005). They used 26 features for the diagnosis. The authors reported approximately 90% diagnosis accuracy. Coppini et al., used the MLNNs with one and two hidden layers and they used BP with momentum as the training algorithm for diagnosis of COPD disease (Coppini et al., 2007). The authors utilized a 10-fold cross-validation procedure to train the neural networks. The