Chapter 5
Differential Diagnosis of Erythematous Squamous Diseases With Feature Selection and Classification Algorithms

Aydın Çetin
Gazi University, Turkey

Tuba Gökhan
Gazi University, Turkey

ABSTRACT

In this chapter, the differential diagnosis of erythematous diseases was determined using data mining and machine learning algorithms. In this chapter, data mining and its application to differential diagnosis of erythematous squamous diseases were discussed. A dermatology dataset from UCI Machine Learning Repository was used for the study. The dataset consists of 366 data items with 34 attributes. Initially, feature selection was made, and then classification was performed by using various algorithms. The number of attributes has been reduced from 34 to 19 as a result of the integration of the correlation-based filter methods and various heuristic search methods. The evaluation results show that Naive Bayes has 100% success rate in classification of psoriasis, seborrheic dermatitis, lichen planus, rose disease, chronic dermatitis, and pityriasis rubra pilaris diseases with 19 attributes selected with feature extraction algorithms.

INTRODUCTION

In primary health care institutions where dermatologic diseases are frequent, there are frequent cases of inaccurate diagnosis or no diagnosis. Incorrect diagnosis causes wrong treatment. Inaccuracies in diagnosing diseases also lead to problems such as difficulties in disease management and inadequate information of patients, unnecessary referrals and laboratory examinations. In the field of medicine, it is important to examine the theoretical knowledge and disease management in order to be able to obtain

DOI: 10.4018/978-1-5225-4769-3.ch005

a successful diagnosis in the diagnosis and treatment of common dermatological diseases (Yilmaz, 2016). The differential diagnosis of erythematous squamous disease is a real problem in dermatology. Erythematous squamous disease, psoriasis, seborrheic dermatitis, lichen planus, Pityriasis Rosea, chronic dermatitis and pityriasis rubra pilaris have similar clinical features. Biopsy is usually carried out for diagnosis of the disease. However, the histopathological features of these diseases are similar and even while these diseases show the characteristics of another disease at earlier stages, they can show their own characteristics when disease progresses (Güvenir, Demiröz, & İlter, 1998).

Today, as the amount of information on the world increases day by day with the data being stored in the digital environment, the number of databases is also increasing at a similar or even higher rate. This increase makes it difficult to retrieve meaningful information. Therefor there is a good reason to believe that smart data analysis will be more pervasive in medicine. Data mining is a multidisciplinary field that serves as a bridge between many technical fields such as database technology, statistics, artificial intelligence, machine learning, pattern recognition and data visualization. Data mining is applied in many fields such as astronomy, biology, finance, marketing, insurance, medicine (Ozekes, 2003).

In this chapter, Data mining and its application to differential diagnosis of erythematous squamous diseases were discussed. A dermatology dataset from UCI Machine Learning Repository which was shared by Prof. Dr. Nilsel İlter from Gazi University and Altay Güvenir from Bilkent University, Turkey, was used for the study (Güvenir et al, 1998). The dataset consists of 366 data items with 34 attributes. Initially, feature selection was made and then, classification was performed by using various algorithms. The number of attributes has been reduced from 34 to 19 as a result of the integration of the correlation based filter methods and various heuristic search methods.

DATA MINING

Data Mining is the process of obtaining previously unknown, valid and applicable information from large data sources. With data mining relationships between the variables can be deduced and future predictions can be made (Kaya & Özel, 2014). The process of obtaining raw data to knowledge using data mining is shown in Figure 1.

As shown in Figure 1, in the data mining process, the process begins with obtaining the data of the probing that is desired to be examined first. Noisy and inconsistent data is removed from the data set by data cleansing. After selecting and extracting the attributes on the data warehouse, the appropriate models for the data and probing are created, and the suitability and adequacy of these models are evaluated. Intelligent methods are applied to capture data patterns while models are being created. In the pattern evaluation step, interesting patterns representing information obtained according to predetermined criteria are defined. In the information presentation step, the obtained information is presented to the user (Kaya & Özel, 2014).

FEATURE SELECTION

The complexity and size of databases in the real world has prompted researchers to reduce data sizes in the data mining. Size reduction can basically be done in two ways. First, to create a random sample space from existing data, and second, to reduce the amount of available attributes. There are two main