Chapter 24
Towards Integrating Data Mining With Knowledge-Based System for Diagnosis of Human Eye Diseases:
The Case of an African Hospital

Nilamadhab Mishra
https://orcid.org/0000-0002-1330-4869
School of Computing Science and Engineering, VIT Bhopal University, India

Johny Melese Samuel
Debre Berhan University, Ethiopia

ABSTRACT

The eye is the most important sensory organ of vision function. But some eye diseases can lead to vision loss, so it is important to identify and treat eye disease as early as possible. Eye care professionals can help protect their patients from vision loss or blindness by recognizing common eye diseases and recommending for an eye exam. Eye diseases with early detection, treatment, and appropriate follow-up care, vision loss, and blindness from eye disease can be prevented or delayed. In this study, rule-based eye disease identification and advising the knowledge-based system are projected. The projected system is targeting using hidden knowledge extracted by employing the extraction algorithm of data mining. To identify the best prediction model for the diagnosis of eye disease, four experiments for four classification algorithms were performed. Finally, the researchers decided to use the rules of the J48 pruned classification algorithm for further use in the development of a knowledge base of KBS because it exhibited better performance with a 98.5% evaluation result.

DOI: 10.4018/978-1-7998-2742-9.ch024

INTRODUCTION

Eye problems have been recognized worldwide as one of the major public health problems, particularly in developing countries where 90% of the blind live and international actions to prevent avoidable blindness has been gaining momentum over the last decade. According to the world health organization (WHO), about 37 million people are blind and 124 million people have low vision worldwide [**HeMavatHi et al, 2014; World Health Organization, 2006; Aemero et al, 2015; Abdulkerim, 2013]**. A large proportion of low vision (91.2%) and blindness (87.4%) are due to avoidable (either preventable or treatable) causes. Females and rural residents carry greater risk for eye problems. The burden of eye disease is believed to pose huge economic and social impacts on individuals, society, and the nation at large [**Fayyad et al, 1996, August; Prentzas et al, 2007; Oprea, 2006**]. A computer-based system (expert system), over-dependence on human experts, can be minimized. Knowledgebase (KBS) benefits the individual by providing a high-quality decision within a given time frame and facilitating job security and personal development [**Shiferaw et al, 2015; Berhane et al, 2007; Akerkar et al, 2010**]. Also, artificial expertise (AE) has some features that make it more beneficial over human expertise such as permanent, easy to transfer, easy to document, consistent, and affordable [**Schreiber et al, 1993; Datta et al, 2011**].

Classification is the process of classifying a data instance into one of several predefined categorical classes based on the training set containing known observations. A regression task begins with data instances in which the target values are known. The relationships between predictors and the target are summarized in a regression model that can be applied to different data instances in which the target values are unknown [**Covington et al, 1996; Fayisa et al, 2018; 15. Brose et al, 2009]**. Classification is the derivation of a function or model which determines the class of an object based on its attributes. A set of objects is given as the training set in which every object is represented by a vector of attributes along with its class. A classification function or model is constructed by analysing the relationship between the attributes and the classes of the objects in the training set. Such a classification function or model can be used to classify future objects and develop a better understanding of the classes of the objects in the database [**Han et al, 2011; Morgan, 2006**]. As mentioned in [**Jackson, 2002; Phyu, 2009, March; Serapião et al, 2013]** and [**Quinlan, 2014; Tayel et al, 2013**], classification is also called supervised learning. It is called supervised learning because it works on labeled attributes in which there is a specially chosen attribute and the aim is to use the data given to predict the values of that attribute for instances that have not yet been seen. The chosen attributes in classification are categorical such as 'high', 'low' or medium', [**Esseynew, 2011; Sasikumar et al, 2007; Mishra et al, 2019; Achour et al, 1999**]. Classification is a two-step process [**Prasad et al, 2012; DeKock, 2005**] consisting of model construction and model usage. In the first step, a classifier is built describing a predetermined or labeled set of data classes or concepts. This is the learning step (or training phase), where a classification algorithm builds the classifier by analysing or learning from a training set made up of database instances and their associated class labels. This step is called model construction. Generally, classification is a process of construction model that defines data class and used to predict the class of objects whose class label is unknown. It finds out the relationship between predictor value and the target value. The model is based on the analysis of a set of training data. The data; historical, for classification is typically divided into two datasets: one for building the model; the other for testing the model. [Dokas, 2005, September; Schmoldt et al,2012]Thus the various classification approaches can be employed on medical data for obtaining specific information and disease diagnosis. Decision tree, Byes classifier, neural network, support vector machine, and rule-based learning are some of the classification data mining techniques. The general objective of this