A hybrid intelligent system for medical data classification

Manjeevan Seera\textsuperscript{a}, Chee Peng Lim\textsuperscript{b,}\textsuperscript{*}
\textsuperscript{a}Intel Technology, Penang, Malaysia
\textsuperscript{b}Centre for Intelligent Systems Research, Deakin University, Australia

\textbf{A B S T R A C T}

In this paper, a hybrid intelligent system that consists of the Fuzzy Min–Max neural network, the Classification and Regression Tree, and the Random Forest model is proposed, and its efficacy as a decision support tool for medical data classification is examined. The hybrid intelligent system aims to exploit the advantages of the constituent models and, at the same time, alleviate their limitations. It is able to learn incrementally from data samples (owing to Fuzzy Min–Max neural network), explain its predicted outputs (owing to the Classification and Regression Tree), and achieve high classification performances (owing to Random Forest). To evaluate the effectiveness of the hybrid intelligent system, three benchmark medical data sets, viz., Breast Cancer Wisconsin, Pima Indians Diabetes, and Liver Disorders from the UCI Repository of Machine Learning, are used for evaluation. A number of useful performance metrics in medical applications which include accuracy, sensitivity, specificity, as well as the area under the Receiver Operating Characteristic curve are computed. The results are analyzed and compared with those from other methods published in the literature. The experimental outcomes positively demonstrate that the hybrid intelligent system is effective in undertaking medical data classification tasks. More importantly, the hybrid intelligent system not only is able to produce good results but also to elucidate its knowledge base with a decision tree. As a result, domain users (i.e., medical practitioners) are able to comprehend the prediction given by the hybrid intelligent system; hence accepting its role as a useful medical decision support tool.

\textsuperscript{*}Corresponding author. Tel.: +61 3 5227 3307.
\textsuperscript{E-mail address:} chee.lim@deakin.edu.au (C.P. Lim).

1. Introduction

Research in computerized intelligent systems for medical applications is an important and exciting domain. In general, a physician typically accumulates his/her knowledge based on patients' symptoms and the confirmed diagnoses. In other words, prognostic relevance of symptoms towards certain diseases and diagnostic accuracy of a patient are highly dependent on a physician's experience (Meesad & Yen, 2003). As medical knowledge and treatment therapy progress rapidly, e.g. the occurrence of new diseases and the availability of new drugs, it is challenging for a physician to keep up-to-date with all recent knowledge and development in clinical practice (Meesad & Yen, 2003). On the other hand, with the advent of computing technologies, it is now relatively easy to acquire and store a lot of information digitally, e.g. in dedicated databases of electronic patient records (Favlopoulos & Delopoulos, 1999). As such, the deployment of computerized medical decision support systems becomes a viable approach to assisting physicians to swiftly and accurately diagnose patients (Chabat, Hansell, & Yang, 2000). Nevertheless, numerous issues have to be overcome before a useful medical decision support system can be developed and deployed, which include decision making in the presence of uncertainty and imprecision (Tsipouras, Voglis, & Fotiadis, 2007). While medical experts' knowledge and experience is important, ranging from assessing a patient's condition to making a diagnosis, advances in machine learning (Kwiatkowska, Atkins, Ayas, & Ryan, 2007) techniques have opened up the way for medical practitioners to exploit computerized intelligent systems for decision support in their workplace, e.g. surgical imagery and X-ray photography (Isola, Carvalho, & Tripathy, 2012). When treating a patient, a physician first needs to narrow down the suspected disease to the root cause (out of a list of probable causes with similar symptoms) using his/her knowledge and experience, and then confirms the diagnosis by performing a number of tests (Isola \textit{et al.}, 2012). Concomitantly, computerized intelligent systems can be useful in assisting the physician to arrive at an informed decision quickly, e.g. by learning from similar past cases in a large database of electronic patient records and inferring the diagnosis for the current patient with proper justifications. The advantages of using such intelligent systems include increasing diagnosis accuracy and, at the same time, reducing time and costs associated with patient treatment (Çomak, Polat, Günes, & Arslan 2007).