A Multi-level Classification Framework for Multi-site Medical Data: Application to the ADHD-200 Collection

Sarah Itani, Fabian Lecron, Philippe Fortemps

Expert Systems with Applications, August 2017
DOI: https://doi.org/10.1016/j.eswa.2017.08.044.

Abstract

Recently, the culture of sharing medical data has emerged impressively, reducing significantly the barrier to the development of medical research accordingly. As open-access large datasets result from this significant initiative, data mining techniques can be considered for the development of interpretable expert systems to help in diagnosis. However, the collaborative effort of information gathering yields heterogeneous databases because of technical and geographical factors. Indeed, on the one hand, the harmonization of protocols for data collection is still missing. On the other hand, cultural and social factors impact locally both the epidemiology and etiology of a given disease. Ignoring these factors could weaken the credibility of studies based on multi-site data. Thereby, our work tackles the development of computer-aided diagnosis systems relying on heterogeneous data. For such a purpose, we propose a multi-level approach (inspired by multi-level statistical modeling) based on decision trees (in the sense of machine learning). This framework is applied on the public ADHD-200 collection for the study of Attention Deficit Hyperactivity Disorder (ADHD).