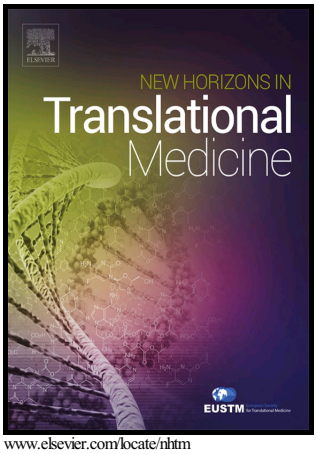
Improving disease diagnosis by a new hybrid model

Bikash Kanti Sarkar



www.elsevier.com/locate/nhtm

PII: S2307-5023(17)30023-1
DOI: http://dx.doi.org/10.1016/j.nhtm.2017.07.001
Reference: NHTM49

To appear in: *New Horizons in Translational Medicine*

Received date: 20 June 2017
Revised date: 5 July 2017
Accepted date: 6 July 2017

Cite this article as: Bikash Kanti Sarkar, Improving disease diagnosis by a nev hybrid model, *New Horizons in Translational Medicine* http://dx.doi.org/10.1016/j.nhtm.2017.07.001

Improving disease diagnosis by a new hybrid model

Dr. Bikash Kanti Sarkar*

Department of Computer Science and Engineering Birla Institute of Technology, Mesra, Ranchi, India

bk_sarkarbit@hotmail.com

*Corresponding author. Dr. Bikash Kanti Sarkar

ABSTRACT

Knowledge extraction is an important part of e-Health system. However, datasets in health domain are highly *imbalanced*, *voluminous*, *conflicting* and *complex* in nature, and these can lead to erroneous diagnosis of diseases. So, designing accurate and robust clinical diagnosis models for such datasets is a challenging task in data mining. In literature, numerous standard intelligent models have been proposed for this purpose but they usually suffer from several drawbacks like lack of *understandability*, incapability of operating *rare cases*, inefficiency in making *quick* and *correct* decision, etc. In fact, specific health application using standard intelligent methods may not satisfy multiple criteria. However, recent research indicates that hybrid intelligent methods (integrating several standard ones, can achieve better performance for health applications. Addressing the limitations of the existing approaches, the present research introduces a new hybrid predictive model (integrating C4.5 and PRISM learners) for diagnosing effectively the diseases (instead of any specific disease) in comprehensible way by the practitioners with better prediction results in comparison to the traditional approaches. The empirical results (in terms of *accuracy*, *sensitivity* and *false positive rate*) obtained over fourteen benchmark datasets demonstrate that the model outperforms the base learners in almost all cases. The performance of the model also claims that it can be good alternative to the specialized learners (each designed for specific disease) published in the literature. After all, the presented intelligent system is effective in undertaking medical data classification task.

Keywords: classifier; C4.5; PRISM; hybrid-model; disease-prediction; accurate