

## **Disease Diagnosis Using Machine Learning Techniques: A Review and Classification**

Mehrbakhsh Nilashi<sup>a,\*</sup>, Neda Ahmadi<sup>b</sup>, Sarminah Samad<sup>c</sup>, Leila Shahmoradi<sup>d,e</sup>, Hossein Ahmadi<sup>f</sup>, Othman Ibrahim<sup>g</sup>,  
Shahla Asadi<sup>h,\*</sup>, Rusli Abdullah<sup>h</sup>, Rabab Ali Abumalloh<sup>i</sup>, Elaheh Yadegari<sup>j</sup>

<sup>a</sup>School of Computing, Faculty of Engineering, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor, Malaysia

<sup>b</sup>Shahid Chamran University of Ahvaz, Faculty of Engineering, Department of Computer Engineering, Ahvaz, Iran

<sup>c</sup>Department of Business Administration, Collage of Business and Administration, Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia

<sup>d</sup>Halal Research Center of IRI, FDA, Tehran, Iran

<sup>e</sup>Health Information Management Department, School of Allied Medical Sciences, Tehran University of Medical Sciences, Tehran, Iran

<sup>f</sup>Department of Information Technology, University of Human Development, Sulaymaniyah, Iraq

<sup>g</sup>Department of Software Engineering & Information System, Faculty of Computer Science & Information Technology, University Putra Malaysia, Selangor 43400, Malaysia

<sup>h</sup>Azman Hashim International Business School, Universiti Teknologi Malaysia (UTM), Skudai, Johor 81310, Malaysia

<sup>i</sup>Computer Department, Community College, Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia

<sup>j</sup>Center for Software Technology and Management (SOFTAM), Faculty of Information Science and Technology (FTSM), Universiti Kebangsaan Malaysia (UKM), 43600 Bangi, Selangor, Malaysia

\* Corresponding author email address: [nilashidotnet@hotmail.com](mailto:nilashidotnet@hotmail.com)

### **Abstract**

In this research, we reviewed and classified academic conference and journal papers; which used data mining techniques in disease classification and diagnosis based on public medical datasets published between 2007 and 2019. The results of this review demonstrated that the application of data mining techniques in disease classification has experienced a dramatic rise in recent years. The finding of this paper also revealed that there was minimal focus on developing methods using incremental version of data mining techniques. We hope that this research will provide useful information about various data mining techniques, their application in disease diagnosis, and help researchers in developing medical decision support systems with insights into the state-of-the-art of development methods.

Keywords: Data Mining, Public Medical Datasets, Diseases Diagnosis, Literature Survey, UCI, PRISMA

### **1. Introduction**

Data mining has been applied effectively in developing decision support systems to discover patterns in big datasets (Han et al., 2011). The potential of the data mining as an efficient disease classification tool was identified by World Health Organization (WHO) (Nilashi et al., 2016b). The use of data mining techniques in knowledge discovery for disease classification has been recognized as one of the interesting and important topics addressed by researchers (Ahmadi et al., 2018; Cauchi et al., 2015; Gopalakrishnan et al., 2010; Nilashi et al., 2016a).

There are many of supervised machine learning techniques for prediction and classification tasks (Kotsiantis et al., 2007; Nilashi et al., 2014a; Nilashi et al., 2014b; Nilashi et al., 2015a; Nilashi et al., 2015b). Using these techniques, several studies have developed methods for disease diagnosis and have been evaluated using real-world datasets (Çalışır and Doğantekin, 2011; Nilashi et al., 2019a; Nilashi et al., 2017a; Nilashi et al., 2017c; Polat, 2012). Hence, the classification of these studies based on applied machine learning techniques for investigating different types of disease diagnosis and reporting their outputs are important tasks; which this research aims to perform. Additionally, further researches should be