Pattern Recognition for Human Diseases Classification in Spectral Analysis

ABSTRACT

Pattern recognition is a multidisciplinary area that received more scientific attraction during this period of rapid technological innovation. Today, many real issues and scenarios require pattern recognition to aid in the faster resolution of complicated problems, particularly those that cannot be solved using traditional human heuristics. One common problem in pattern recognition is dealing with multidimensional data, which is prominent in studies involving spectral data such as ultraviolet visible (UV/Vis), infrared (IR), and Raman spectroscopy data. UV/Vis, IR, and Raman spectroscopy are well-known spectroscopic methods that are used to determine the atomic or molecular structure of a sample in various fields. Typically, pattern recognition consists of two components: exploratory data analysis and classification method. Exploratory data analysis is an approach that involves detecting anomalies in data, extracting essential variables, and revealing the data's underlying structure. On the other hand, classification methods are techniques or algorithms used to group samples into a predetermined category. This article discusses the fundamental assumptions, benefits, and limitations of some well-known pattern recognition algorithms including Principal Component Analysis (PCA), Kernel PCA, Successive Projection Algorithm (SPA), Genetic Algorithm (GA), Partial Least Square Regression (PLS-R), Linear Discriminant Analysis (LDA), K-Nearest Neighbors (KNN), Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), Partial Least Square-Discriminant Analysis (PLS-DA) and Artificial Neural Network (ANN). The use of UV/Vis, IR, and Raman spectroscopy for disease classification is also highlighted. To conclude, many pattern recognition algorithms have the potential to overcome each of their distinct limits, and there is also the option of combining all of these algorithms to create an ensemble of methods.