**Machine learning techniques in identifying genetic markers for diseases**

**ABSTRACT**

The approach of new technological developments in the genetic disease repository has facilitated genetic disease treatment. In the post-genomic time gene detection, which causes genetically excessive diseases, is one of the greatest deterrent tasks. Complex diseases are frequently very heterogeneous and make biological markers difficult to identify. Markers commonly depend on the Machine Learning Algorithms to define, but their success completely depends on the quality and dimensions of the data present. In the machine learning area, computers are promised to support people and analyze large and complex data systems primarily for the production of practically enhanced algorithms. A quantitative measure of gene functional similarities was obtained by employing different semantic similarity measures. To infer the hidden functional similarities between ASD genes, various types of machine learning classifiers were built on quantitative semantic similarity matrices of ASD and non-ASD genes. The classifiers trained and tested on ASD and non-ASD gene functional similarities outperformed previously reported ASD classifiers. For example, a Random Forest (RF) classifier achieved an AUC of 0. 80 for predicting new ASD genes, which was higher than the reported classifier (0.73). Additionally, this classifier was able to predict 73 novel ASD candidate genes that were enriched for core ASD phenotypes, such as autism and obsessive-compulsive behavior. In addition, predicted genes were also enriched for ASD co-occurring conditions, including Attention Deficit Hyperactivity Disorder (ADHD). We also developed a KNIME workflow with the proposed methodology which allows users to configure and execute it without requiring machine learning and programming skills. Machine learning is an effective and reliable technique to decipher ASD mechanism by identifying novel disease genes, but this study further demonstrated that their performance can be improved by incorporating a quantitative measure of gene functional similarities.Scource from -1)HTTP://pmc.ncbi.nlm.nih.gov/articles/PMC6287949/<HTTP://pmc.ncbi.nlm.nih.gov/articles/PMC6287949/>

2)<https://ieeexplore.ieee.org/document/9640723> and the ASD gene prediction data is in the source GitHub Id - <https://github.com/Shalini-Sur/gene__prediction_data.>