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

**ABSTRACT**

Advancements in genetic research technologies have significantly improved the diagnosis and treatment of genetic diseases. In the post-genomic era, identifying genes responsible for genetically complex disorders remains a major challenge due to the heterogeneity of such conditions and the difficulty in identifying reliable biomarkers. Machine learning (ML) algorithms have become instrumental in addressing this challenge, as their success heavily relies on the quality and dimensionality of available data. ML techniques allow the analysis of large, intricate datasets to design optimized algorithms that facilitate gene detection.

In this study, various semantic similarity measures were employed to quantify gene functional similarities. By leveraging these measures, machine learning classifiers were developed to identify hidden functional links between Autism Spectrum Disorder (ASD) and non-ASD genes using quantitative semantic similarity matrices. The trained classifiers showed significant improvements over previously reported ASD predictors. For instance, a Random Forest (RF) classifier achieved an AUC of 0.80 in predicting novel ASD genes, surpassing earlier classifiers (AUC 0.73). The RF classifier also identified 73 potential ASD candidate genes associated with core ASD traits such as autism and obsessive-compulsive behavior, as well as comorbid conditions like Attention Deficit Hyperactivity Disorder (ADHD).

Additionally, a KNIME workflow implementing the proposed methodology was developed, enabling users without programming expertise to configure and execute the analysis. This research demonstrates that incorporating quantitative measures of gene functional similarities enhances the performance of ML-based gene prediction models. ML continues to be a powerful tool for uncovering the genetic mechanisms underlying ASD and other complex disorders.

**Sources:**

1. [PMC Article](http://pmc.ncbi.nlm.nih.gov/articles/PMC6287949" \t "_new)
2. [IEEE Explore Document](https://ieeexplore.ieee.org/document/9640723" \t "_new)
3. ASD gene prediction data available at [GitHub](https://github.com/Shalini-Sur/gene__prediction_data" \t "_new).