

Classifying Autism Spectrum Disorder (ASD) Causal Genes Based On Intrinsic Disordered Regions

Shula Shazman

Mathematics and Computer Science Department, The Open University of Israel
Raanana, 4353701 Israel

Autism spectrum disorder (ASD) is a complex neurodevelopmental disorder with a strong genetic basis, yet only a small fraction of potentially causal genes are known with strong genetic evidence from sequencing studies. The method developed in this study is a complementary machine-learning approach for classifying ASD causal genes based on intrinsically disordered regions (IDRs) found in proteins that are the product of genes associated with ASD. The method successfully distinguished ASD causal genes from non-causal genes with Area Under the Curve (AUC) = 0.82. The method uses artificial neural networks (ANN) and other machine learning classifiers to classify the ASD causal genes. Furthermore, analyzing the results reveals that the most important feature by which to classify ASD causal genes is related to the number of IDRs in the protein. Finally, those results indicate that the contribution of IDRs must be taken into consideration when deciphering the ASD mechanism.