A Novel Application of Machine Learning Algorithms on Toxin Proteins to Predict Molecular Function

Kulkarni, Ameya (School: Coppell High School)
Muvva, Praneeth (School: Coppell High School)

Animal venoms consist of toxin peptides and proteins, and drugs derived from venoms are unique in their ability to interfere with key biological processes in the body, such as blood clotting, nerve transmission, and muscle contraction. Previous research has yielded the ability to record the amino acid sequence through high throughput techniques, resulting in an abundance of unreviewed toxin proteins without specific functions attributed to them. Extending the application of machine learning to accurately predict the molecular function of venom peptides from amino acid sequence and structural data would remove a bottleneck in venomics research, allowing researchers to identify the potential pharmaceutical applications of venom peptides. Using data from a variety of toxin databases, we were able to build multiple machine learning algorithms to classify toxin molecular function based on amino acid sequence. After extensive testing of multiple different machine learning models, we determined that an artificial neural network was the best for toxin classification. Utilizing amino acid sequence and protein structural data, a neural network was trained on multi-label classification of 86 possible molecular function classes. The prediction framework was able to predict toxin function with an accuracy of 95.584%. Our results demonstrate the potential of machine learning algorithms to predict the molecular function of venom peptides and facilitate the identification of new types of toxins. Further research utilizing our algorithm can enhance our understanding of venom peptides and their mechanisms of action, leading to the development of more targeted and effective medicinal drugs.

Awards Won:

University of Texas at Dallas: Scholarship of $5,000 per year, renewable for up to four years

University of Texas at Dallas: Scholarship of $5,000 per year, renewable for up to four years