Artificial Neural Networks for Protein Structure Prediction

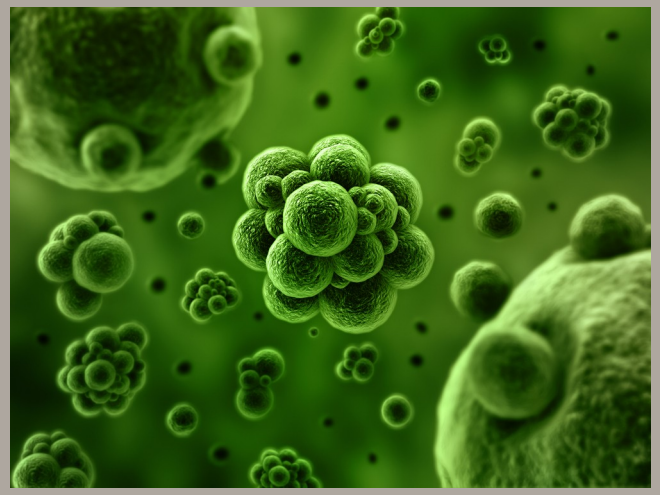
W. Willie Wells

Diseases incapacitate and sometimes kill their victims.

It would be pleasant to be able to engineer therapeutic drugs and enzymes for all known pathogens.

Accurate protein secondary structure prediction saves time and money in the development process of drugs to counteract pathogens.

Though accurate X-ray Diffraction is difficult, slow, and expensive. Solution: utilization of machine learning techniques.

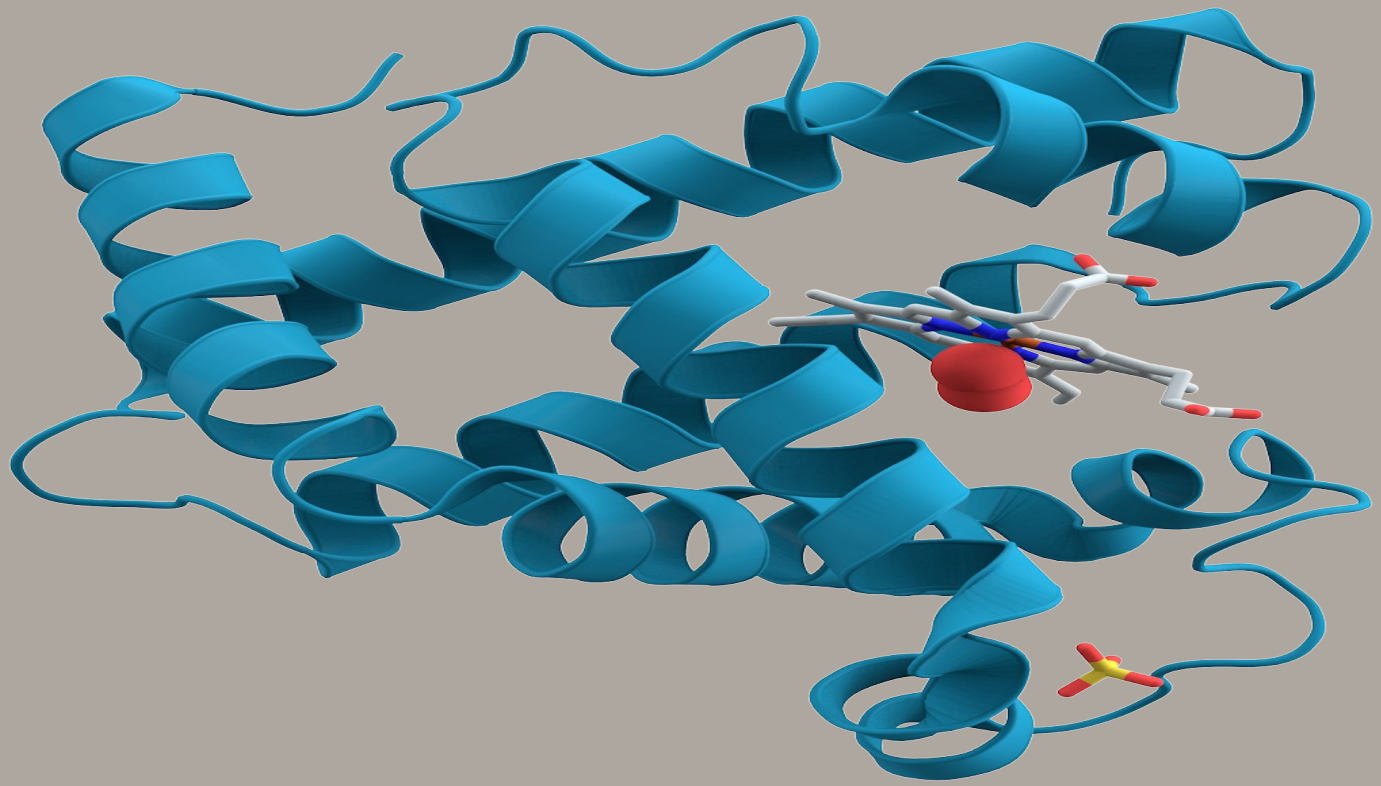


- Basics of protein secondary structure
- Basics of Artificial Neural Networks (ANN)
- How to apply ANN to protein secondary structure prediction

There exists multiple algorithms in use to predict protein secondary structure.

Protein secondary structure is highly non-linear and involves complex interactions.

Hydrogen bonding patterns determine protein secondary structure.



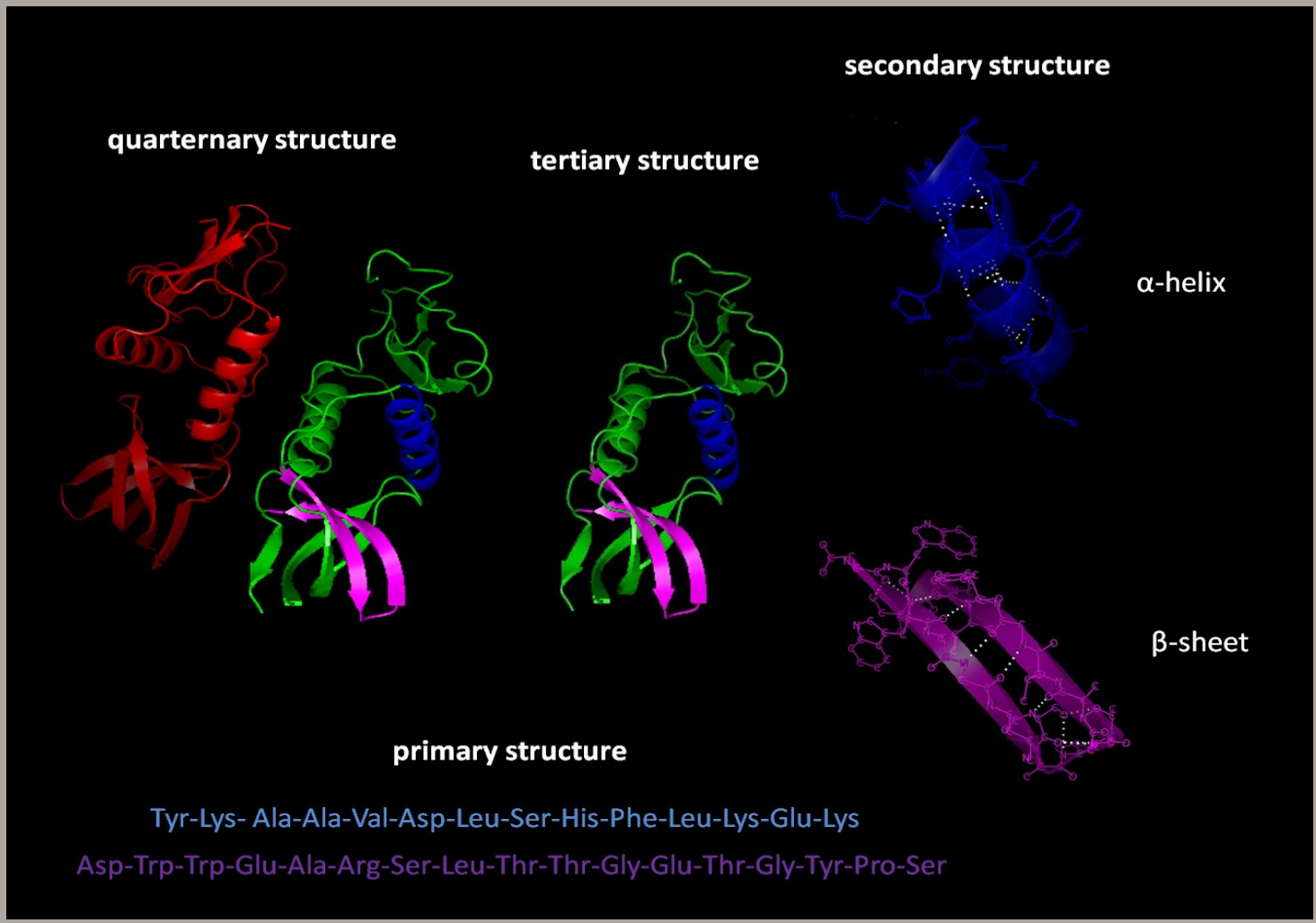
Types of Protein Secondary Structures:

α-helix (H)

β-sheet (E = extended strand)

coil or loop (C or L)

turn (T)

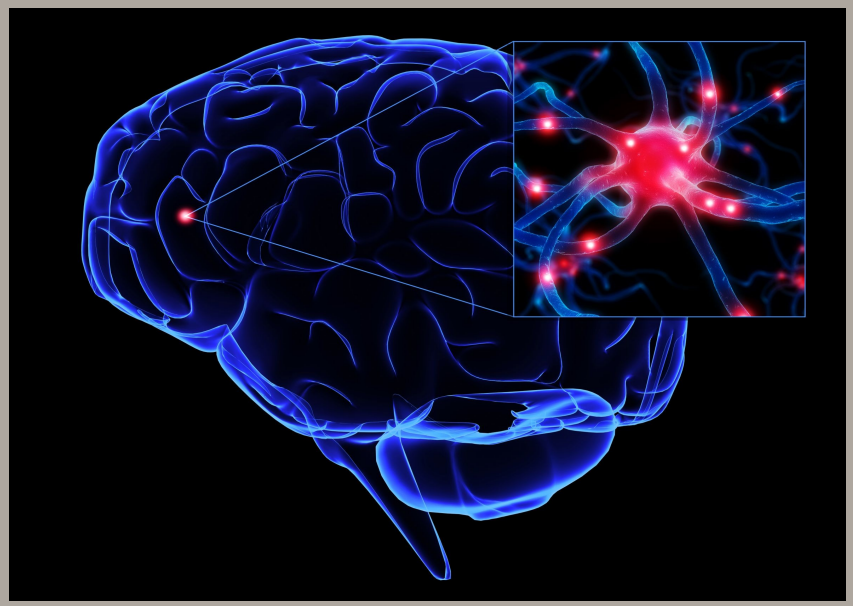


Artificial neural networks are modeled after the network of neurons in the brain.

Each artificial neuron has a weight (w_{ij}) associated with each input (x_j).

All inputs to a single neuron are weighted, summed and passed to an activation function (f) to produce an output (y_i).

$$y_i = f(\Sigma_j w_{ij} x_j), \text{ for all } y_i \text{ in ANN}$$



The learning process during the training phase:

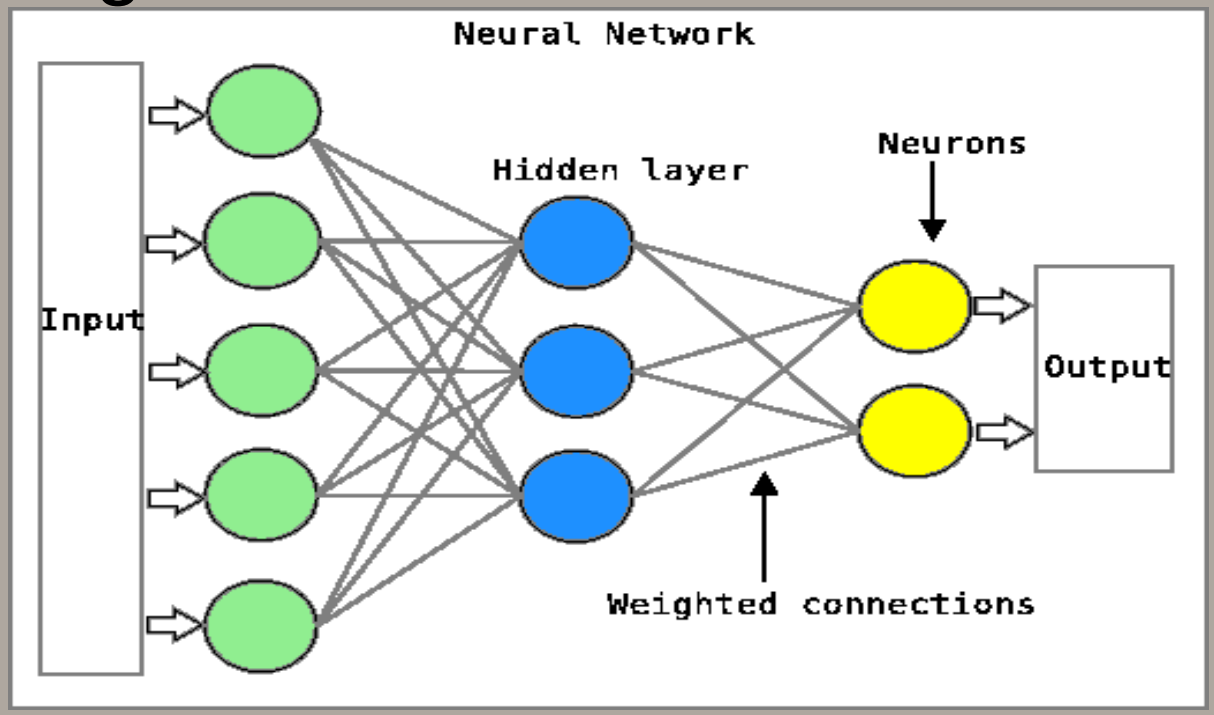
Compare output with desired output

Backward propagate the error (E)

Update weights via $w_{ij} + \Delta w_{ij}$

$\Delta w_{ij} = -\alpha * \delta E/\delta w_{ij}$

Where α is the learning rate.



Perceptron

Feed Forward (Multi-layer Perceptron)

Radial Basis Function

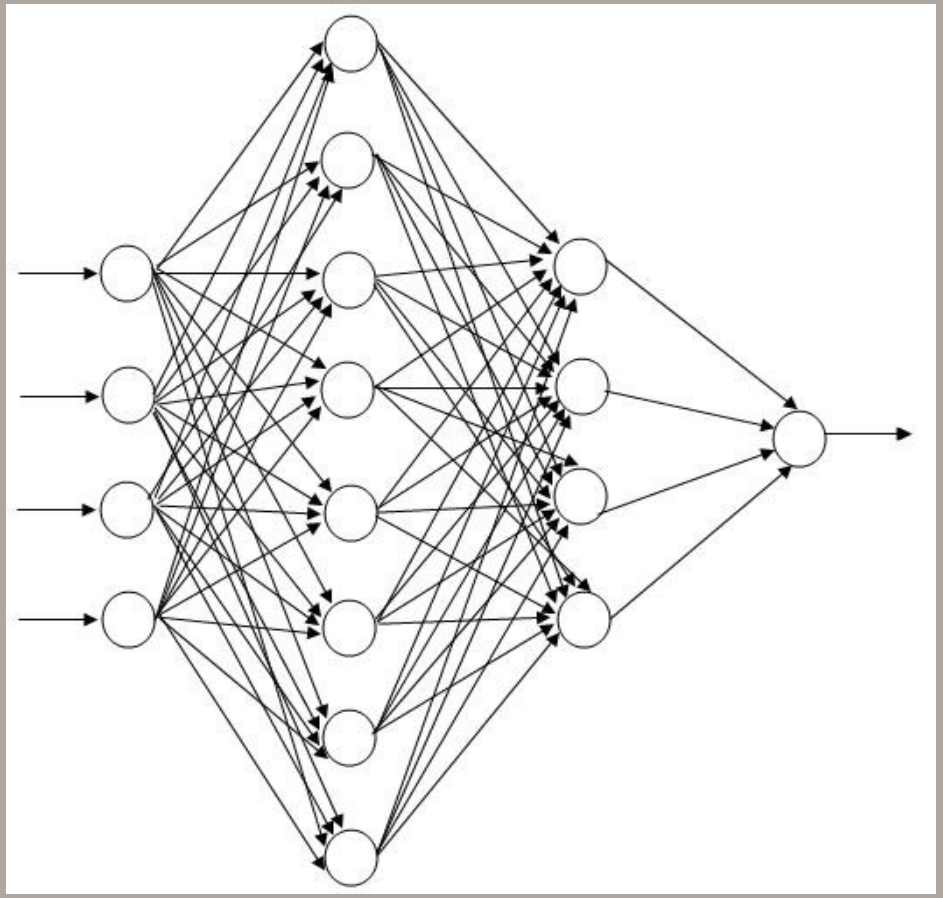
Bidirectional Recursive

Kohonen Self Organizing Maps (SOM)

Competitive Layers

Learning Vector Quantification

Hopfield Network

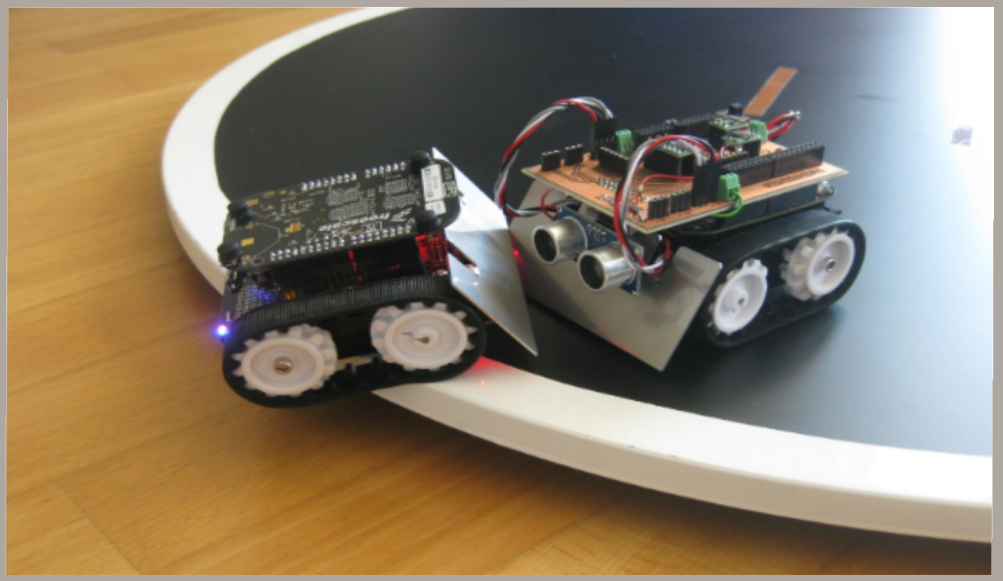


Advantages:

- Adaptive learning
- Self organization
- Real-time operation (parallel)
- Computationally powerful (complex patterns)

Disadvantages:

- ↑Number of neurons → ↑computation time
- Overfitting
- Difficult to implement
- Requires data normalization



Applying ANN to Bioinformatics

Extract features from molecular sequences

Select usable features and encode

Initialize neural network for training

- assign number of layers, amount of neurons per layer, and random weights for each neuron input

Train network

Encode output to protein secondary structure

Using ANN results in a 70 – 90% accuracy rate

Diseases incapacitate and sometimes kill their victims.

Artificial Neural Networks though not perfect provide 70 - 90% accuracy in predicting protein secondary structure.

Applying these results saves time and money when engineering new therapeutic drugs to fight pathogens.