Thesis Proposal for CSE 4000

Roll: 1807082 **Name:** Md. Shoriful Islam

Thesis Title:

Machine learning techniques for protein structure prediction.

Problem Statement:

Due to the notion that sequence dictates structure and structure determines function, protein structure prediction (PSP) has long been a major issue in bioinformatics. A protein's three-dimensional shape can be calculated from its amino acid sequence using a technique called protein structure prediction. This issue is crucial since a protein's shape greatly affects how it functions, but determining protein structures experimentally can be challenging.

Objectives:

Determine an efficient machine learning system for predicting protein sequences.

Method:

Use deep neural networks, natural language processing-based sequence embedding, and a wide variety of predictions, including those for protein function, tertiary structure, secondary structure, residue interactions, intrinsic disorder, and hotspots that bind proteins, peptides, and nucleic acids.

Expected Outcomes:

Protein three-dimensional structures and end-to-end differentiability.

References:

1. https://link.springer.com/chapter/10.1007/3-540-26888-X_5
2. https://www.sciencedirect.com/science/article/pii/S1367593121000508