**Assignment: Protein 2D Structure Prediction Using Deep Learning.**

You will collect protein sequences, preprocess them, train a deep learning model, and evaluate the results. Submissions will include code, visualizations, and a detailed report.

**Instructions:**

1. **Data Collection:**
   * Collect protein sequences from **NCBI** (National Center for Biotechnology Information).
   * Choose a diverse set of sequences with varying lengths and characteristics.
   * Use a relevant format such as FASTA.
2. **Preprocessing:**
   * Tokenize the protein sequences into suitable input for your model.
   * Perform one-hot encoding, embedding, or any other representation technique as needed.
   * Divide the dataset into **training, validation, and test sets**.
3. **Model Development:**
   * Design and implement a deep learning model for protein 2D structure prediction.
   * Recommended architecture components include **ANNs**, **LSTMs, CNNs**, or **Transformers**.
   * Include multiple layers to capture complex sequence relationships.
4. **Training:**
   * Train the model for a sufficient number of epochs.
   * Use appropriate loss functions (e.g., categorical cross-entropy) and optimization algorithms (e.g., Adam or SGD).
   * Track the **weight updates** at each epoch and layer, and create **GIFs** illustrating the updates.
5. **Evaluation and Visualization:**
   * Plot the following graphs:
     + **Loss curve**: Training vs. validation loss over epochs.
     + **Accuracy curve**: Training vs. validation accuracy over epochs.
     + **Confusion matrix**: Visualize model performance on the test set.
   * Calculate evaluation metrics like precision, recall, F1-score, and accuracy for the test set.
6. **Submission Requirements:**
   * **Code:** Upload your entire project to **GitHub**, including scripts for data preprocessing, model training, and evaluation.
   * **Visualizations:**
     + Create **GIFs** showing the weight updates at each epoch and layer.
     + Include the graphs for loss, accuracy, and confusion matrix in your report.
   * **Report:**
     + Provide a detailed explanation of your work in the report.
     + The report should include:
       - **Introduction:** Brief overview of protein 2D structure prediction and its significance.
       - **Methodology:** Explain your data collection, preprocessing, and model architecture.
       - **Results:** Present evaluation metrics, result tables, and graphs.
       - **Model Architecture:** Draw a diagram of your model.
       - **Discussion:** Analyze your results and mention potential improvements.