**Task 4: Active Learning for GSNR Prediction**

**Objective:** In this assignment, you will explore and implement active learning techniques to improve the prediction of General Signal-to-Noise Ratio (GSNR). You will apply active learning to iteratively select the most informative data points from the European or USA dataset to train a model. The goal is to reduce the amount of labeled data needed to achieve high performance.

1. **Understanding Active Learning:**
   * Research and read about active learning, focusing on its principles, benefits, and common techniques. Look for reliable sources such as academic papers, textbooks, and reputable online articles.
   * Understand the concepts of uncertainty sampling, query by committee, and other active learning strategies.
2. **Data Preparation:**
   * Use the European and USA datasets provided.
   * Preprocess the datasets as necessary (e.g., handling missing values, normalization).
3. **Model Training on European Dataset:**
   * Choose a suitable model architecture for GSNR prediction (you may use the same model from the previous task).
   * Train the model on the European or USA dataset.
   * Evaluate the model's performance on a validation set.
4. **Active Learning Implementation:**
   * **Initial Model Training:**
     + Train the initial model on a small subset of the EU or USA dataset.
   * **Active Learning Loop:**
     + Use an active learning strategy (e.g., uncertainty sampling) to select the most informative data points from the EU or USA dataset.
     + Label these selected data points (assume labeling is done by an oracle).
     + Add the newly labeled data points to the training set and retrain the model.
     + Repeat the process for a predefined number of iterations or until performance plateaus.
   * Evaluate the model's performance on the EU or USA validation set after each iteration.
5. **Comparison and Analysis:**
   * Compare the performance of the model using active learning with the performance from the previous task (transfer learning with feature extraction and fine-tuning).
   * Analyze the performance improvements and discuss the benefits of active learning.
   * Provide visualizations (e.g., training/validation loss curves, GSNR prediction accuracy) to support your analysis.
   * Analyze the amount of labeled data needed to achieve similar performance with and without active learning.
6. **Report:**
   * Include code snippets and relevant plots in your report.
   * Discuss the active learning strategy used and the rationale behind its selection.
   * Provide insights and recommendations based on your findings.

**Deadline: 12th July 2024**