**Problem Statement:**

Drug research and development have a serious challenge in categorizing drug mechanisms of action (MoA), which is necessary to comprehend a drug's potential efficacy, safety, and adverse effects. In our effort, we use gene expression data from a cell-based assay to predict the MoA of medications. Because the data is multi-label, some rows may be connected to multiple targets. We will make accurate predictions about the MoA of medications using machine learning techniques to find trends in the data. We won't find any pre-existing external datasets because the assay that was used to gather the data is brand-new. Therefore, by using this new assay, our project will help to improve understanding of the MoA of drugs.

**Dataset:**

We will make use of the gene expression data for more than 20,000 chemicals evaluated on human cells from the "MoA Prediction" dataset offered by the Kaggle competition. This dataset includes gene expression data and cell viability data for a large number of compounds, along with labels for their MoA. Several machine learning algorithms, including neural networks, can be trained on the dataset's size to determine the MoA.

**Neural Network:**

We plan to use a standard form of the neural network, such as a deep feedforward neural network or a convolutional neural network, to classify the MoA of the compounds based on their gene expression profiles. We may also explore customizing the neural network to improve its performance. In addition, we may use other machine learning algorithms, such as random forests, gradient boosting, or support vector machines, to compare their performance with the neural network.

**Software:**

We will use PyTorch, a popular open-source machine learning library, to implement the neural network or other algorithms. PyTorch offers a flexible and efficient way to define, train, and evaluate deep learning models, and it supports both CPU and GPU acceleration. We may also use other Python libraries, such as scikit-learn, Pandas, and NumPy, for data preprocessing, visualization, and analysis.

**Reference Materials:**

We will refer to various research papers, articles, and online resources to obtain sufficient background on applying neural networks and other algorithms to drug MoA classification using gene expression data. Some of the references we may use include:

PRISM assay: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6708833/>

Gene Expression Omnibus (GEO): <https://www.ncbi.nlm.nih.gov/geo/>

<https://www.kaggle.com/competitions/lish-moa/overview>

**Performance Metrics:**

For our models, the log loss measure will serve as the performance metric. This metric, which evaluates the performance of our model in terms of its probability estimations, is frequently employed for classification issues.

**Project Schedule:**

This project should be finished in two to three weeks. The first week will be devoted to data discovery and preparation, and the second week will be used to create and train our models. Next, we will evaluate the model for two to three days, and the remaining time will be used to document and present the