

Project Overview

We identified 26 proteins whose tyrosine phosphorylation relieves autoinhibition, based on an extensive literature search. You can find this data in the "Auto_inhibited Protein's Data" folder.

Next, we computed the CT and CO values for these proteins using the provided CT calculation code. After that, we implemented a Gradient Boosting Machine (GBM) classification and conducted a greedy search to optimize hyperparameters for classifying two groups of datasets:

- **Oncogene vs. Non-Oncogene**
- **Autoinhibited vs. Non-Oncogene**

The results of these classifications, based on the CT and CO features, can be found in the "Results" folder.

Additionally, we ran a MATLAB script to generate box plots and perform a statistical test (p-value calculation) for the XGB energy feature, comparing oncogenes with non-oncogenes. The results, also stored in the "Results" folder, indicate that the XGB energy feature alone could not distinguish between the two groups (oncogene vs. non-oncogene) in the pancancer dataset. It requires the inclusion of CT and CO features for accurate classification.