

### Project Proposal

**Project idea:** For our class project, we propose replicating the analyses conducted in the paper titled “*Machine learning model for predicting Major Depressive Disorder using RNA-Seq data: optimization of classification approach*”. The study utilizes machine learning algorithms to classify RNA-Seq data into control, MDD, and MDD-S groups. We will replicate the methodology using random forest and KNN classifiers, testing the classification of gene expression data from the NCBI database. We aim to validate the results by analyzing the provided dataset and comparing our outcomes with the accuracy of the original paper. We have successfully located the data in the NCBI archive and have identified and sorted all the relevant raw data used in the study.

**Group members:**

Quoc Dung Van, V01073195

Umit Akirmak, V01063348

**Sample Data Access:** [NCBI Archive](#)

**Citation:** Verma, P., & Shakya, M. (2021). *Machine learning model for predicting Major Depressive Disorder using RNA-Seq data: optimization of classification approach*. Cognitive Neurodynamics, 16, 443-453. <https://doi.org/10.1007/s11571-021-09724-8>