Alexander Munoz
Atousa Nourmahnad
Tina Zeina

Machine Learning for Cervical Cancer Risk

For our project, we will analyze a dataset from the UCI Machine Learning Repository: “Cervical cancer (Risk Factors) Data Set.” This dataset contains 36 features on 858 patients scraped from electronic medical records. The features include data such as age, number of sexual partners, IUD, STD status, and previous diagnoses. Using these features, we will predict Cervical Cancer scores: Hinselmann, Schiller, cytology score, and biopsy score. We will first attempt this model using a simple linear regression. We will compare this baseline model to a neural network perceptron model with one hidden layer. We will additionally perform an L2-regularized lasso regression model. We will compare the errors of these models and analyze why these differences occurred. Finally, we will state conclusions which are consistent across all models.

Link: https://archive.ics.uci.edu/ml/datasets/Cervical+cancer+%28Risk+Factors%29