

Final Project

Forensics investigation from fingerprint microbes

Recently it was discovered that DNA samples from human fingerprints are unique to individuals. Therefore, it is possible to get samples from computer keyboards and identify who is using the computer. This task provides very strong patterns and the recognition rate is high. However, the harder task is to detect which hand (left or right) the samples are gathered from. In this project, the task is to identify which hand of an individual touched the computer keyboard. This project involves assessing classification performance of clinical data gathered from DNA data on computer keyboards. The task is to perform supervised learning on the dataset and report the classification performance.



Data

There are 271 samples (first 136 left, second 135 right hands). Each sample contains 3302 features. Therefore each file contains a table of 3302 x 271 entries. 136 of the samples are gathered from right-hand, and 135 are from left-hand. The datasets *otu.xlsx*, and *otu.csv* are provided. The first row of the files is the sample names, and the second row indicates if they are collected from left or right hands. They all have the same data in different formats.

Goal

With this project it is expected to have the highest possible correct classification percentage. In order to achieve that you are expected to perform classification with these attributes.

Classification Algorithms

The projects will be performed individually. **ANY** classification algorithm/technology covered in the lectures can be used. **ANY** programming language and platforms including machine learning packages (e.g. WEKA, Python Sklearn, etc.) can be used. If you program the project, the scripts/source code is requested. If a platform like WEKA is used, the program parameters are requested. Present your results in a short report.

Performance Measures

Sensitivity, *specificity* is requested as the output of the program performance.

$$\text{Sensitivity:} = \frac{\text{correct number of prediction of the first class}}{\text{total number of elements in the first class}}$$

$$\text{Specificity:} = \frac{\text{correct number of prediction of the second class}}{\text{total number of elements in the second class}}$$

NOTE: THE PERFORMANCES WILL BE MEASURED EITHER WITH SPLITTING THE DATA INTO TEST-TRAIN DATASETS, OR USING CROSS-VALIDATION.

DEADLINE IS JANUARY 26TH, 2022.

The result reports will be sent to e-mail: nalbantoglu@odev.erciyes.edu.tr

Good Luck!