

# Projects for ML course

## 1 Description

For the last weeks of the semester, you will have to implement a mini-project.

The task of the project is to **do a classification and/or regression task on a dataset, using one of the usual machine learning algorithms.**

Examples:

- image classification with CNN networks
- classification or prediction on other data, using MLP networks or logistic regression
- any other idea, with any other algorithm you may know or want to use

A list of datasets to choose from is provided at the end of this file.

You can use Matlab or other programming languages you may know (e.g. Python).

### 1.1 Deliverables

The project should be finalized with a **written report**, consisting of a few pages, which should discuss:

- The dataset used
  - What sort of data it contains
  - Examples of the images from each class
  - Is the dataset balanced or unbalanced (roughly same amount of data in each category, or not)
  - Anything else of interest
- What is desired to obtain
- What algorithm is used, and how you used it (parameter values etc.)
- Results: include classification results, training results (if relevant), any other graphical or numerical result relevant for the problem
- Any personal comments or conclusions

The project must be presented in 5-10 minutes presentation, either in the last week project sessions or on some agreed date before the exam.

## 1.2 Tips'n Tricks and other instructions

- If the dataset is too large, you can use only a smaller part of it (i.e. classify between 2 or 3 classes instead of 100)
- Images can be resampled to a desired resolution (e.g. 224 x 224 for AlexNet)
- Small MNIST like images (28x28) can be used with MLP networks. The images can also be resized to larger values such for CNNs to be used (e.g. 224 x 224 for AlexNet)
- This list contains only **suggested** datasets and tasks. You can use a custom dataset (for safety, check with me first). The only requirement is that it is interesting.

## 2 List of Project Topics

1. Brain tumor classification from MRI images with CNN networks

[Data source](#)

2. Credit card fraud detection with MLP networks

[Data source](#)

3. Fruit and vegetables image detection with CNN networks

[Data source](#)

4. Image detection with MLP networks on the Fashion-MNIST dataset

[Data source](#)

5. Recognize dog vs cat image with CNN networks

[Data source](#)

6. Detect handwritten letters with MLP networks

[Data source](#)

7. Detect heart disease based on health measurements with logistic regression or MLP networks

[Data source](#)

8. Gastrointestinal Cancer MSI MSS Prediction with CNN networks

[Data source](#)

9. Predict mushroom toxicity with linear regression

[Data source](#)

10. Detect pneumonia from chest X-Rays using CNN networks  
[Data source](#)
11. Detect diabetes from diagnostic measurements with logistic regression or MLP (csv file)  
[\[Data source\]](#)
12. Detect dog breed from images using CNN networks  
[Data source](#) (only the dog breed part of the dataset)
13. Recognize sign language using MLP networks on the Sign Language MNIST  
[Data source](#)
14. Recognize Chinese handwritten numbers using MLP networks  
[Data source](#)
15. Bird species classification with CNN networks  
[Data source](#)
16. Bus vs car image detection with CNN networks  
[Data source](#)

## 2.1 MedMNIST datasets

All data for the next project topics are available here: [13579/:https://wwwaggle.com/datasets/niccleju/medmnist-v1mat-files](https://wwwaggle.com/datasets/niccleju/medmnist-v1mat-files)

For a detailed description, see the website [13579/:https://medmnistom/v1](https://medmnistom/v1) or the paper [13579/:https://arxivrg/pdf/11df](https://arxivrg/pdf/11df)

Most images are resized to size 28x28 from larger images available elsewhere.

17. Cancer pathology detection from the PathMNIST dataset with MLP networks  
[Data source](#), file `pathmnist.mat`
18. Chest Xray classification on ChestMNIST with MLP networks  
[Data source](#), file `chestmnist.mat`
19. Skin lesion classification on DermaMNIST with MLP networks  
[Data source](#), file `dermamnist.mat`
20. Retina disease classification on OCTMNIST with MLP networks  
[Data source](#), file `octmnist.mat`

21. Pneumonia detection on PneumoniaMNIST using MLP networks (or logistic regression)

[Data source](#), file `pneumoniamnist.mat`

22. Diabetic retinopathy severity classification on RetinaMNIST with MLP networks

[Data source](#), file `retinamnist.mat`

23. Breast cancer detection on BreastMNIST with MLP networks (or logistic regression)

[Data source](#), file `breastmnist.mat`

24. Organ identification from CT image slices (OrganMNIST\_Axial or Coronal or Sagittal) with MLP networks

[Data source](#), file `organmnist_axial.mat`, or `organmnist_coronal.mat`, or `organmnist_sagittal.mat`

## 2.2 Other datasets

25. **Or any other** interesting dataset found on the Internet, but only if you contact me first and I agree to it.

Popular dataset sources:

- [Kaggle](#)
- [Zenodo](#)
- [UCI archive](#)