

**UNIVERSIDADE DE AVEIRO**  
**DEPARTAMENTO DE ELECTRÓNICA TELECOMUNICAÇÕES E INFORMÁTICA**

**CAA PROJECT 1 2024/2025- Instructions**

**Deadline for Project 1 submission 7/May 2025**

**Submit to elearning: report (pdf+ source files); presentation slides; implementation code**

**PROJECT PRESENTATION: 8/May 2025 (in class)**

Each group of two students is supposed to work on one project topic.

You are strongly encouraged to propose a deep learning (DL)-oriented problem you would prefer to work, not listed below, that may reflect better your interests. Please, discuss your idea with the instructor.

## **I. PROJECT GOALS**

The goal is to apply DL algorithms to solve a data science problem (classification, regression, clustering). Represent the results in graphical/table formats and make analysis and conclusions.

## **II. PROJECT PROPOSALS**

- 1) Heart Disease Cleveland UCI  
[https://www.kaggle.com/datasets/cherngs/heart-disease-cleveland-uci?fbclid=IwAR1Zs9rCsFKcC8-kOWTJf9sP5vB5ThC3pgcBjArKtHnt\\_uhLcXcv2petlS8](https://www.kaggle.com/datasets/cherngs/heart-disease-cleveland-uci?fbclid=IwAR1Zs9rCsFKcC8-kOWTJf9sP5vB5ThC3pgcBjArKtHnt_uhLcXcv2petlS8)
- 2) Human Stress Detection in and through Sleep  
<https://www.kaggle.com/datasets/laavanya/human-stress-detection-in-and-through-sleep?select=SaYoPillow.csv>
- 3) Thyroid Sickness Determination  
<https://www.kaggle.com/datasets/bidemiayinde/thyroid-sickness-determination>
- 4) Facebook metrics Data Set  
<https://archive.ics.uci.edu/ml/datasets/Facebook+metrics>
- 5) Water Quality<<https://www.kaggle.com/datasets/adityakadiwal/water-potability>> Drinking water potability
- 6) STROKE PREDICTION DATASET  
<https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset>
- 7) Wind Speed Prediction Dataset  
<https://www.kaggle.com/datasets/fedesoriano/wind-speed-prediction-dataset>
- 8) Top Hits Spotify from 2000-2019 Top songs spotify playlists. What is the Genre of the track.  
<https://www.kaggle.com/datasets/paradisejoy/top-hits-spotify-from-20002019>
- 9) GARBAGE CLASSIFICATION (IMAGE PROCESSING)  
6 classes: cardboard, glass, metal , paper, plastic and trash.  
<https://www.kaggle.com/datasets/asdasdasdasdas/garbage-classification>
- 10) Face Recognition With Olivetti Dataset (image processing)

**Data Repositories:**

- Kaggle Data Repository : <https://www.kaggle.com/datasets>
- UCI Machine Learning Repository : <https://archive.ics.uci.edu/ml/index.php>
- <https://www.isic-archive.com> (International Skin Imaging Collaboration )

**III. PROJECT ASSESMENT (25 % of the final grade)**

1. **Report.** The project is evaluated based on a submitted report (IEEE Latex format). The work done by each student has to be explicitly specified. All project's files (pdf and Latex files of the report, the presentation slides and the code implementing the algorithms) are submitted to the elearning.ua.pt page of the course in section SUBMISSION – PROJECT 1 in a compressed format having the following name: P1\_CAA2025\_XXXXX\_YYYYY (where XXXXX and YYYYY are substituted by the student number of each student).
2. **Oral presentation** of the report in class (about 10-15 min.).

**IV. Evaluation criteria** (total score 20)

1. *Report content (12):*
  - Data description and preprocessing (if necessary normalization, feature selection, transformation, etc.). Motivation for choosing the particular problem.
  - Data visualization (histograms, box plots, other plots).
  - Short description of the implemented ML models.
  - Model training (data splitting – train, validate, test, k-fold Cross validation). Visualize graphically the cost function trajectory over iterations. Training with regularized and non-regularized cost function.
  - Model hyper-parameter selection - regularization parameter  $\lambda$ , number of NN hidden layer units, number of hidden layers (if necessary),  $\sigma$ ,  $C$ ,  $k$ , etc.. Systematic approach instead of just one or several randomly chosen values.
  - For a classification problem, you need to present the confusion Matrix (accuracy, precision, recall, F1 score, etc.).
  - Performance comparison between the models.
  - Results in graphical or table formats.
  - Conclusions.
  - Problem complexity.
2. *Report formatting (2) :*
  - IEEE Latex format, affiliation (Department, University, subject, course instructor), abstract, keywords, work load per student.
  - Sufficiently detailed report.
  - References, reference citation in the report.
  - Clear figures (title, legends, axis labels) and tables referred in the text.
3. *Oral presentation (3)*
  - Slide Organization, slide numbers, affiliation.
  - Clear and convincing presentation by both students.
4. *Novelty and contributions (3)*
  - Compare your solution with the works of other authors (published references) , try to propose a better solution, e.g. improve the performance of the ML model in solving the problem you work with.