FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE D
COIMBRA

Deep and Reinforcement Learning
2025/2026 - 1st Semester

Discriminate blood with ML!

Jorge Henriques
João Correia

1 Introduction

We present here the first practical project, part of the students' evaluation process of the Deep and Reinforcement Machine Learning course of the Master in Informatics Engineering of the University of Coimbra. This work is to be done autonomously by a group of **two students**. The **deadline** for delivering the work is **31th of October of 2025** via Inforestudante. The quality of your work will be judged as a function of the value of the technical work, the written description, and the public discussion. All sources used to perform the work (including the code) must be clearly identified. If you use AI tools in the production of this work (e.g. ChatGPT), you must clearly identify all the parts in which the tool was involved. Please note that, during the defence, you will be required to demonstrate a deep understanding of the content generated by the tool, and this knowledge will be subject to evaluation. The document may be written in Portuguese or in English, using a word processor of your choice[1]. The **written report** is limited to **10 pages long**.

The document should be well structured, including a general **introduction**, **a description of the problem**, the **approach**, the **experimental setup**, an **analysis of the results**, and a **conclusion**. The report should follow the **Springer LNCS format**[2]. The final mark will be given to each member of the group individually. To do the work the student may consult any source he/she wants. Nevertheless, plagiarism will not be allowed and, if detected, it will imply failing the course. While doing the work and when submitting it, you should pay particular attention to the following aspects (whose relative importance depends on the type of work done):

- description of the approach to the problem
- description of the general architecture of the methods used;
- description of the experiment, including a table with the parameters used which should allow full replication;
- description of the evaluation metrics used for the validation: quality of the final result, efficacy, efficiency, diversity, or any other most appropriate;

[1] Latex is preferred

[2] Template available for word and latex under the "Important downloads for authors" at https://www.springer.com/gp/computer-science/lncs/conference-proceedings-guidelines

Do not forget, besides what was just said, that it is fundamental: (1) to do a correct experimental analysis; (2) to do an informed discussion about the results obtained; (3) to put in evidence the advantages of the chosen alternative.

2 Problem Statement

Image classification is a fundamental computer vision problem where machine learning has achieved impressive results. Traditionally, tackling this task required expert knowledge and feature engineering, where practitioners carefully designed and extracted visual features for machine learning models to utilize. After this step, classical models such as *Adaboost, Random Forests, Support Vector Machines (SVMs)*, and *Multi-layer Perceptrons (MLPs)* were typically employed.

With the rise of deep learning, particularly *Convolutional Neural Networks (CNNs)*, image classification has seen significant advancements. Although CNNs first emerged in the late 1980s and early 1990s, they gained widespread adoption after 2010 due to advances in *GPU computing*. CNNs have demonstrated superior performance in image classification tasks by integrating *feature extraction and learning* into the model itself. Unlike traditional approaches, CNNs automatically extract hierarchical features from raw images, feeding them into dense layers to make predictions, similar to fully connected Multi-Layer Perceptrons.

One area where CNNs have been particularly impactful is *medical image analysis*, including applications to datasets such as *MedMNIST* [3]. The *MedMNIST dataset* is a benchmark dataset specifically designed for lightweight medical image classification. It consists of multiple subsets covering a variety of medical imaging modalities, including *X-rays, CT scans, MRIs, and histopathology images*, across different medical classification tasks. Unlike general-purpose image classification datasets, *MedMNIST focuses on biomedical applications*, making it a valuable resource for evaluating machine learning models in healthcare. Applying CNNs to *MedMNIST* allows for automated feature extraction from medical images, reducing reliance on handcrafted features and domain expertise.

As the field of medical AI continues to grow, *MedMNIST* serves as an important benchmark for developing *efficient and interpretable deep learning models*, ensuring progress in *computer-aided diagnosis* while addressing challenges such as *data scarcity, model interpretability, and ethical considerations* in medical imaging.

[3]MedMNIST: https://www.nature.com/articles/s41597-022-01721-8

3 Objective

The main objective is to analyse the dataset and create a machine learning approach with neural networks as the models that can perform supervised image classification of the dataset that is address in Section 3.1. To do that you should attend to the following objectives:

- Prepare the machine learning pipeline for the image classification dataset.
- Explore solutions and compare the results using at least, the following ANNs models:
  - Multi-Layer Perceptrons (MLP)
  - Convolutional Neural Networks (CNNs)
- And for each model, experimentation with different layer architectures, loss functions, optimisers and hyperparameters. Explore at least:
  - 2 different Loss functions
  - 2 different Optimizers

Exploring other solutions than the listed ones that are suitable for the problem at hand and considered as extra work, can be as compensation points to cover problems in the listed ones above.

3.1 Dataset and Evaluation

BloodMNIST is derived from the *BCCD (Blood Cell Count and Detection)* dataset, containing microscopic *peripheral blood smear images* annotated by experts. It comprises **17,092 color images** categorized into **eight blood cell types**, organized as a **multi-class classification task**. The dataset is divided into **training, validation, and test sets** according to the official **MedMNIST v2 split** (**11,959 / 1,721 / 3,412** samples, respectively). To ensure consistency and computational efficiency, all images are resized from their original resolution to $3 \times 28 \times 28$.

Each image corresponds to one of the following 8 classes:

0: Basophil

1: Eosinophil

2: Erythroblast

3: Immature Granulocyte

4: Lymphocyte

5: Monocyte

6: Neutrophil

7: Platelet

Tabela 1: Class distribution in the BloodMNIST dataset (MedMNIST v2).

Class	**Cell Type**	**Number of Samples (%)**
0	Basophil	404 (2.36%)
1	Eosinophil	1,048 (6.13%)
2	Erythroblast	1,924 (11.26%)
3	Immature Granulocyte	1,826 (10.68%)
4	Lymphocyte	3,769 (22.06%)
5	Monocyte	1,659 (9.71%)
6	Neutrophil	5,812 (34.01%)
7	Platelet	650 (3.80%)
Total		**17,092 (100%)**

Figure 1 shows representative examples from the dataset.

Use the provided script to retrieve and handle the dataset. The main goal is to use the training data to design, implement and validate your approaches, while the test will be used to evaluate the generalisation ability of your models.

Given the splits you should see how to fit the models that you are training/creating and select the appropriated metrics to evaluate performance under this context. Thus, the validation part of this work is crucial and you should select the most appropriate set of metrics and justify them.

4 Conclusion

A few short comments. First, the control of the progression of your work will be done during the classes (T and PL). Moreover, you can discuss eventual problems by presenting yourself during office hours. Second, the projects reflect for the most part your current knowledge. The rest will be the object of lecturing soon. Third, we try to balance the difficulty of all the work, but

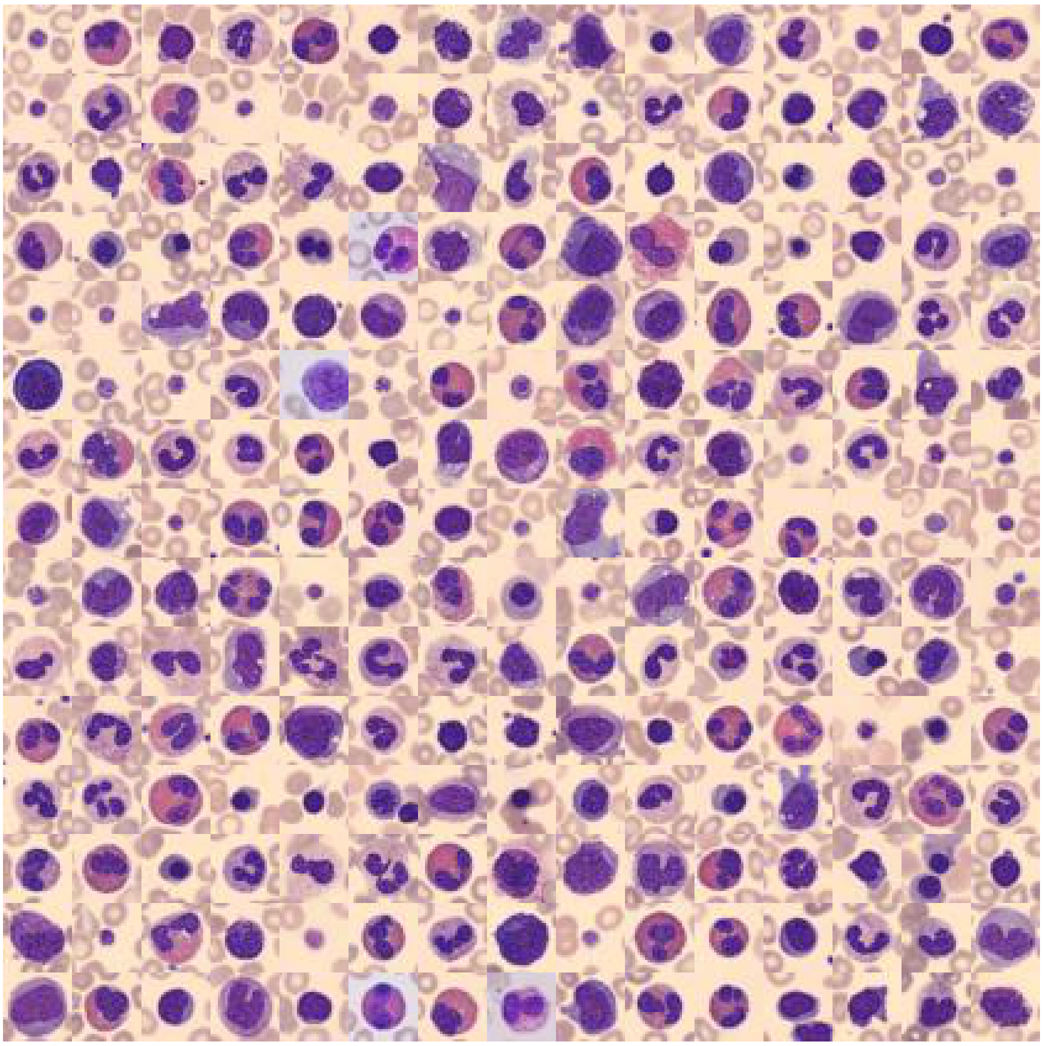

Figura 1: Example of the images from the training dataset.

we are aware that this is not an easy task and it is somehow a subjective matter. Fourth, we try to ask for a workload compatible with the value of the work for the final mark.

Methodological issues, like the statistical background, were elucidated during the previous lectures. You may use the statistical tool you feel at ease with, including the Python code that was provided. Finally, even if this is a work that asks you to do simulations and analyse the results, i.e., it has a practical flavour, there is however a theory behind the work, and you are advised to consult the necessary literature.