Prof. Adil Khan (Module Leader)
a.m.khan@hull.ac.uk
02 September 2025

# Assignment – Machine Learning

662086 - 2025/2026

## Description:

The assessment for this module will consist of this coursework that accounts for 100% of your final mark. As part of this coursework, you will use Python to compare a set of machine learning approaches for solving a multi-label real-world classification problem. The results of your work should be summarised in a single written report of no more than 3000 words submitted as a PDF file. You are required to submit your code alongside your report, but it will not be separately graded.

## Problem Statement:

The Triple-MNIST dataset is an extended version of the traditional MNIST dataset (which you have used in your labs), which has become a standard benchmark in machine learning and computer vision. The traditional MNIST dataset contains 28x28 grayscale images of handwritten digits (0 through 9), with 60,000 training images and 10,000 test images.

In contrast, the Triple-MNIST dataset contains images of size 84x84 pixels, where each picture contains 3 digits.  The goal is to identify each of the three digits present in the image, meaning labels are now combinations of three digits (e.g., '059', '348', etc.).

Your primary challenge here is: Given an image from the Triple-MNIST dataset, build a model to predict the sequence of the three handwritten digits it contains. This means the output space is no longer 10 possibilities (one for each digit), but is now significantly expanded to accommodate every combination of three digits.

The dataset is already pre-divided into "train", "validation", and "test" directories for ease of use. Your journey will be structured as a series of tasks, progressing from basic exploratory data analysis to more complex modelling. For each task, document your work/findings/results/outcomes/analysis in your written report.

## Task 1:

Read the dataset in Python. Visualize random samples from the dataset to gain a visual understanding. Using these insights to identify and discuss how this problem differs from a typical classification task, considering the nature of the dataset.

## Task 2:

Implement a basic approach to solve the problem and set a performance baseline.

- Pre-process the image data by flattening it into a single vector.
- Train two models:
    - Logistic Regression or Decision Tree
    - Convolutional Neural Network (CNN)
- Utilize the validation set for hyperparameter tuning for both models.
- Evaluate both models on the test data and compare their performance using suitable metrics. Suggested metrics include accuracy, F1 score, and Confusion Matrix.
- Document the performance of the models, highlighting which model delivered the best results and which one was the least effective.
- Dive deep into and explain why a certain model outperformed the other.

## Task 3:

Analyse the current results and find ways to further improve the model's performance.

- Examine the training and validation loss curves from your previous best model to determine if there's underfitting or overfitting.
- Based on your analysis, consider if the model's performance can be enhanced.
- Implement **two** techniques from the following to improve performance:
    - increasing model complexity,
    - changing model's architecture,
    - using a different more advanced model,
    - basic data augmentation,
    - dropout layers, or other regularization methods to bolster the  performance.

Document the impact of these enhancements on model performance, using the test set for evaluation.

## Task 4:

Dive deeper into the problem's nuances and develop a more apt solution.

- Pre-process the image data by splitting each image into three distinct pieces. Ensure that the splitting technique is generalisable.
- Train a CNN-based model tailored to predict a single label for each of the three image pieces.
- Concatenate the results of the three predictions to derive a final three-digit prediction.
- As before, make full use of the validation set for hyperparameter tuning.
- Evaluate this new model's performance on the test set and compare it to the models from Task 2. Note the differences and improvements.
- Document the performance of the models, highlighting which model delivered the best results and which one was the least effective.
- Dive deep into and explain why a certain model outperformed the other.

## Task 5:

(a) Analyse the multi-label structure of the problem and identify and implement an alternative CNN model that is able to solve the given classification problem, with an appropriate accuracy, without requiring you to split the image first.

(b) Utilise GANs to generate synthetic images to augment the training data and evaluate the impact of this augmentation on model's performance.
- Implement a GAN, using an appropriate architecture such as DCGAN or a similar variant) and training it on existing training dataset to produce realistic images of same format.
- Generate a set of synthetic images, visualise them and comment on their quality.
- Combine the synthetic images with original training data, retrain the model on the augmented data, and evaluate the performance on the test set, and discuss the final results.

**Please note** that we will not go into your Python code to find answers to the questions asked – anything that is not provided in the report will be considered as absent and will not incur points. However, anything included in the report which is not backed up in the attached code (i.e. is missing) will also be considered as absent.

## Code Submission:

You will need to submit your code alongside your report. As stated earlier, it will not be marked separately but can be checked to ensure that it supports the functionality described in the report and is not plagiarism. As before, *please note that anything you want us to see is in the report as we are not awarding marks for the code separately.*

## Hand-in deadline:

The report and the code is due via canvas on Thursday, 21 December 2023.

## Marking Criteria:

Task 1: 10%

Task 2: 20%

Task 3: 20%

Task 4: 20%

Task 5: 20%

Quality of writing and presentation: 10%

## Detailed Grading Criteria Table:

| Criteria | First | 2:1 | 2:2 | Third | Poor |
|---|---|---|---|---|---|
| Task 1 (10%) | Comprehensive exploration of various facets of the dataset. Multiple visualization tools used to derive deep insights. Clear identification of how the problem stands distinct from regular classification, with implications for modelling. Deep understanding of dataset intricacies and potential challenges. | Good coverage of the dataset's essential facets, with minor misses. Utilizes visualization but might not fully leverage a wide variety. Good identification of primary differences from regular classification. Overall good understanding of the dataset with minor misses. | Surface-level exploration of the dataset. Limited use of visualization tools, with mostly generic insights. Recognizes obvious differences from typical classification but misses depth. Surface understanding of dataset nature. | Very limited exploration, missing major dataset facets. Minimal to no meaningful visualizations. Struggles to identify how the task differs from typical classification. Clear gaps in understanding the dataset. | Inadequate or incorrect exploration of the dataset. No meaningful use of visualization tools. Fails to identify or incorrectly identifies the nature of the problem. Fundamental misunderstanding or lack of engagement with the dataset. |
| Task 2 (20%) | Both models excellently implemented, tuned, & compared with a deep analytical approach. | Both models implemented & compared. Minor errors or shallow analysis. | One model well-implemented. Limited comparison & analysis. | Incomplete implementation of models. No valid comparison. | Models not implemented or completely incorrect approach. |
| Task 3 (20%) | Multiple advanced techniques excellently implemented & successful improvement in model | Some advanced techniques used & minor improvement shown. | Attempted only basic techniques with little to no improvement. | Little attempt to use improvement techniques. | No use of any improvement techniques. |

| | | | | | |
|---|---|---|---|---|---|
| | performance with detailed analysis. | | | | |
| Task 4 (20%) | Superior and generalisable solution tailored & executed with precision. Deep analysis of improvements. | Good solution with some tuning and generalisation. Analysis of the improvements is done. | Basic solution implemented. Limited insights on improvements. | Implementation has major flaws. No insights on improvements. | No implementation or completely incorrect approach. |
| Task 5 (20%) | A comprehensive attempt at implementing an alternative advanced CNN model justified with references from the literature and better performance.<br><br>Well implemented GAN generating high quality images, augmented data well integrated with performance improvement. Thorough evaluation and deep analysis of the results | An adequate attempt at implementing an alternative advanced CNN model with some improvement and basic discussion.<br><br>Adequate implementation of GAN with good quality images, good integration into the dataset with minor improvement, Adequate evaluation and discussion, and evidence of literature exploration. | Basic attempt at implementing an advanced alternative CNN model.<br><br>GAN implemented with good quality images, but no augmentation for performance improvement, and basic evaluation and discussion, with evidence of literature exploration. | Poor attempt at both the advanced model and the GANs with evidence of literature exploration. | No implementation but there is at least an attempt at explaining a possible approaches with evidence of literature exploration. |

| | | | | | |
|---|---|---|---|---|---|
| | and the literature. | | | | |
| **Quality of Writing & Presentation (10%)** | Report is eloquent, well-structured, with excellent use of references. | Report is well-structured with minor errors. Good referencing. | Report has a clear structure. Some referencing issues. | Report lacks a clear structure. Many referencing issues. | Report is incoherent with no structure or references. |

# Learning Objectives:

- **Data Exploration and Visualization**:
  - Ability to load, inspect, and understand the basic structure of a dataset.
  - Skilfully use visualization techniques to derive meaningful insights from image datasets.
  - Understand the intricacies and potential challenges of a dataset.

- **Problem Understanding and Classification Differentiation**:
  - Grasp the nature of a multi-label classification problem and how it stands distinct from traditional classification problems.
  - Develop an understanding of the implications of different problem types on the modelling approach.

- **Model Implementation and Evaluation**:
  - Understand and apply pre-processing techniques suitable for image datasets.
  - Implement basic machine learning models and neural networks for classification tasks.
  - Gain proficiency in model evaluation techniques, understanding the importance and use of metrics such as accuracy, F1 score, and confusion matrices.
  - Implement hyperparameter tuning effectively using validation datasets.

- **Advanced Neural Network Modelling**:
  - Develop the ability to design and implement complex CNN architectures tailored for specific problems.
  - Understand and apply strategies to process parts of an image individually and collate results.

- **Model Improvement and Iteration**:
  - Analyse training and validation curves to diagnose issues such as underfitting or overfitting.
  - Understand and apply advanced techniques, like data augmentation, dropout layers, and other regularization methods, to enhance model performance.

- **Critical Thinking and Analysis**:
  - Interpret model results and deduce reasons behind a model's performance.

- Draw logical conclusions from performance metrics and suggest potential improvements.

- **Effective Communication**:
    - Develop the ability to document processes, results, and conclusions effectively.
    - Improve writing skills to communicate technical content clearly and concisely.
    - Learn to organize and present information in a manner that's accessible to the intended audience.