CSC3009/3109 MACHINE LEARNING
AY2023/2024 Tri 3

Group Project Specifications

1. Objectives

The overall objective of the group project is to put the concepts of machine learning, in particular deep learning, into practice. The students have already learnt the basic concepts and techniques in class, so it is time to put them in use to solve real-world challenges. It empowers the machine learning team with comprehensive understandings of the advanced methods to excel in solving open AI problems in computer vision for health screening purpose. This group project makes students aware of various aspects of machine learning, including data processing, modelling, learning approaches and decision making, as well as it instils in the students a constructive attitude towards large-scale machine learning projects.

2. Key Concepts

Machine learning, Real-world challenges, Advanced techniques, Reporting

3. Project Description

Today, Machine Learning (ML) continues to make technological advancements in solving real-world challenges in many fields, such as health care, climate change and energy. ML solutions are growing exponentially and becoming more and more prevalent in our daily lives. With the ability to "learn", machines are equipped with intelligence to deliver insights from data and to make better decisions for complex tasks.

As a machine learning team, the aim is to develop solutions for real-world challenges. In this project, you will learn various new knowledge and skills of machine learning such as data processing, modelling, analysis, benchmarking, and presentation, which are highly demanded by the industry. The team will also utilise modern deep learning frameworks such as Pytorch, Tensorflow, MxNet, DeepLearning4J or Microsoft Cognitive Toolkit.

The key objective of the project is to learn advanced machine learning techniques, understand real-world problems, and implement own solutions. The teams will use one of the popular dataset and submit own deep learning solutions for performance evaluation. The evaluation of the performance of your solutions should be clearly recorded in the teams' final reports.

There are thousands of open source machine learning libraries; nevertheless, students will need to register the use of open source libraries with valid justifications. In addition, the team has to report own contributions/enhancements/customisations to existing open source libraries. For example, you should **not** download and simply run existing open sources to solve the challenge without appropriate understandings and enhancements.

subject to change

In details, the project work shall fulfil multiple requirements as shown below:

3.1) You will be given a dataset with images different from other team. The task is to correctly classify 4 categories of brain MRI images based on the given dataset.

3.2) Based on comprehensive understandings of this real-world problem, you will need to provide clear motivations, exploratory data analysis, data preparation, model training and tuning, prediction, and performance evaluation.

3.3) The team will need to propose and develop at least 3 deep learning models in the project. Strengths and weaknesses of the models should be discussed extensively.

3.4) The team will also be required to deliver key insights from the analysis of their results. The performance of the proposed solutions of the individual members should be documented in the submitted final report.

4. Assessment criteria

The group assignment will be assessed according to the criteria listed in the marking scheme in Table 1. Note that the criteria are subject to change depending on the progress and situation throughout the trimester.

Table 1. Marking Scheme

Group assessment will be weighted by peer review, if needed

Criteria	Weight
Understanding and solving a real-world challenge: • Background and motivations • Problem statement and project objectives • Existing approaches to solve the problem	20

Machine Learning solutions:
• Exploratory data analysis
• Data preparation and pre-processing
• Investigation of machine learning approaches (the number of approaches should be the same or more than the number of team members)
    ○ Model training and tuning
    ○ Machine learning inference
    ○ Performance evaluation
    ○ Strengths and weaknesses
• Results, insights, and discussions
	60

Report: • Professional writing with a clear logic and structure • Reflection what you have learnt from the study and how to apply them into your project • Future directions of work	10
Video:	10

| • Make one 3-5 minutes video to demonstrate the performance of the proposed approaches. | |

NOTE: Bonus marks will be awarded to those who can demonstrate extensive understandings of the knowledge domain and integrate more innovative approaches, results and discussions.

5. Timeline and Deliverables

Final submission: **Friday, 12 July 2023, 12:00 PM**.

Each group is to submit assignment files to their group's xSITE Dropbox. One submission per team only.

The submission should contain the PDF document of the final report and the compressed file (including the video and supplementary materials) using the following conventions:

- T<Team Number>.pdf (e.g., T01.pdf)
- T<Team Number>-file.zip (e.g., T01.zip)

The final report should be around 30 pages (and should not be more than 50 pages) contain the following:

- Overall description of the machine learning project (3-6 pages)
    - Background and motivations
    - Problem statement
    - Project objectives
    - Existing approaches to solve the problem
- Machine Learning solutions (15 – 18 pages)
    - Exploratory data analysis
    - Data preparation and pre-processing
    - Investigation of deep learning approaches
        - Model training and tuning
        - Machine learning inference
        - Performance evaluation
        - Strengths and weaknesses
    - Results, insights, and discussions
- Clear and concise description of each team member's contribution (1 – 2 pages)
- Reflection on what you have learnt from the study and how to apply them into your project (2 – 3 pages)
- Future directions of work (1-2 pages)

6. Teams

Students will be grouped in a team of 5-6 members on Week 2. Each team will work on the machine learning project based on the dataset given, to be able to understand and solve real-world challenges.

7. Late Submission

A penalty of 20% per day for each deliverable will be imposed for late submission unless extension has been granted by the lecturers prior to the submission date. Request for an extension will be granted on a case-by-case basis. Any work submitted more than 4 days after the submission date will not be accepted and no mark will be awarded.

8. Plagiarism

The University's policy on copying does not allow you to copy software as well as your assessment solutions from another person. Copying another person's work is unacceptable. It is the responsibility of all students that their assessment solutions are their own work. You must also ensure that others do not obtain access to your solutions for the purpose of copying a part of them. Where such plagiarism is detected, both of the assessments involved will receive ZERO mark.