



COSC2673/COSC2793 | SEMESTER 1 2020  
MACHINE LEARNING & COMPUTATIONAL MACHINE LEARNING

---

## Assignment 4 (v1.0)

# Machine Learning Project Proposal

---

**Weight:** 30% of the final course mark

**Due Date:** 5.00pm, Friday 12 June 2020 (Week 14)

**Learning Outcomes:** This assignment contributes to CLOs: 2, 5, 6

### Change Log

1.0

- Initial Release

# 1 Introduction

## 1.1 Summary

In this assignment you will develop a project proposal aimed at solving a challenging real-world machine learning problem. This assignment is designed for you to become more confident in applying machine learning to novel circumstances similar to those found in real-world. This is an individual project.

In this assignment you will:

- Explore a real word problem and identify how ML can be applied.
- Identify and Investigate techniques that are applicable to a given machine learning problem.
- Conduct a review of the state-of-the-art to identify the most suitable approach.
- Research into extending techniques that are taught in class.
- Produce a comprehensive plan to solve the problem.

This assignment has the following deliverables:

1. A report (of no more than 4 pages, plus up to 2 pages for appendices) that includes a literature review critically analysing the applicable techniques and a proposal describing how the problem can be solved (including a justified approach, evaluation plan and a plan to overcome possible challenges).

## 1.2 Learning Outcomes

This assignment contributes to the following course CLOs:

- **CLO 2:** understand a range of machine learning methods and the kinds of problem to which they are suited
- **CLO 5:** understand major application areas of machine learning
- **CLO 6:** understand the ethical considerations involved in the application of machine learning.

## 1.3 Plagiarism

**!** *Plagiarism is a very serious offence.*

The penalty for plagiarised assignments include zero marks for that assignment, or failure for this course. Please keep in mind that RMIT University uses plagiarism detection software to detect plagiarism and that all assignments will be tested using this software. See the RMIT website for more information about the university policies on Plagiarism and Academic Misconduct.

## 2 Task

The COVID-19 pandemic's rapid rise has increased interest in developing automated detection systems based on AI techniques. It has been identified that several chest imaging modalities like Computer Tomography (CT) and Chest X-Ray (CXR) contain information that can be used for diagnostic/prognostic purposes in relation to COVID-19. It is a challenging task to provide expert clinicians, that can interpret these images, to every hospital due to the limited number of radiologists. Therefore, simple, accurate, and fast AI models may be helpful to overcome this problem and provide timely assistance to patients.

In this assignment, you will explore how machine learning techniques that are covered in this course (and their extensions) can be applied to analysing COVID-19 in **Chest X-Ray images**. The machine learning task to be explored is to *classify between healthy patients, those who are ill with pneumonia and those who have COVID-19, from chest X-rays*.

In completing this assignment you should assume “Patterns of COVID-19 lung disease can be identified on conventional chest X-rays”. However, you may explore this hypothesis in your literature review and discuss your findings.

- You should also assume that you have access to a relatively small chest X-rays dataset provided to you by a public hospital in Victoria (around 200 X-rays images of patients with COVID-19, and 300 X-rays image from each other class).

You must conduct the following tasks:

1. Conduct a literature review on the above task (More information in Section 2.1).
2. Develop a proposal to solve the above problem (More information in Section 2.2).
3. Compile a report containing your findings and proposal (guidelines given in Section 2.4).

! Note that you are not required to perform any programming task in this assignment.

### 2.1 Literature Review

The literature review should explore the following:

1. The ML techniques that are used for diagnosing (classifying) medical images like x-ray.
2. Previous attempts at developing AI systems to diagnose COVID-19 and related respiratory diseases (e.g. influenza, pneumonia).
3. Common problems encountered in applying ML to medical images and the existing solutions.

You should not write your literature review in a Q&A format and provide a response to each of the above points one-by-one.

! The details in this spec are the *minimum requirements*. A thorough investigation *must* include evidence to justify your proposed methodology.

A good literature review:

- Should follow a logical structure.
- Should **not** just provide a list of related papers. Papers should be discussed in relation to why are they relevant for this problem, what is good about them and what are the limitations.
- Papers should be summarised in your own words. **Directly copying sentences from the reviewed paper is considered as plagiarism and significant penalties will apply** (see Section 1.3).
- Will discuss literature from peer-reviewed sources. *Wikipedia* or web discussion forums are **not** considered as peer-reviewed sources.

Remember that good literature review provides factual statements that summaries the work in literature in a way that is useful for the reader to understand the context and follow your rational of the proposed methodology. Statements such as:

“Ruwan did <xyz>. Andrew did <xyz>. Arian did <xyz>”

is **not** a literature review. This is an annotated bibliography. Instead, you should aim for statements such as:

*“To solve the problem <xyz> Ruwan did <xyz>, while this method can handle <xyz> it has the limitations <xyz>. To overcome this limitation, Andrew proposed <xyz> ...”*

## 2.2 Project Proposal

The project proposal should:

1. Clearly identify all the elements of the proposed end-to-end machine learning solution.
  - Data preparation and pre-processing (e.g. feature extraction, feature construction - if relevant, data normalization, data augmentations, etc. )
  - Proposed machine learning model and learning framework.
  - Hyper parameters of the model and how they will be tuned.
2. Clearly identify the evaluation framework and mention how your evaluations will result in unbiased conclusions.
3. Identify *technical* issues that you may face in completing the project and propose possible solutions.
4. Discuss the biases that may be present in your final trained model due to the nature of data presented to it and the nature of the ML algorithm you selected. Also include any ethical considerations that might be relevant.

## 2.3 Sample Dataset

To gain an idea of the type of images that are involved in detecting COVID-19 in x-ray images, you may refer the dataset found at: <https://github.com/ieee8023/covid-chestxray-dataset>. Note that you are not required to perform any programming task in this assignment.

A starting point for your review might be:

*Ozturk, T., Talo, M., Yildirim, E. A., Baloglu, U. B., Yildirim, O., & Rajendra Acharya, U. (2020). Automated detection of COVID-19 cases using deep neural networks with X-ray images. Computers in Biology and Medicine, 103792. Advance online publication.*

The above is for guidance only and it should **not be** assumed that the method proposed in the above paper is the best/adequate approach.

## 2.4 Report

Your report:

- Must be compiled using Sample template file provided on canvas.
- Must be no longer than 4 pages of text - maximum two pages for literature review and maximum two pages for proposal.
- May contain ***an additional*** 2 pages for appendices.
- The appendices may only contain citations, figures, diagrams, or data tables that provide evidence to support the statements in your report.
- Include the name(s) and student id's of the student(s) who wrote the report.

! Any over length content, or content outside of these requirements will not be marked. For example, if you report is too long, ONLY the first 4 pages of text will be read and marked.

## 3 Additional Information

### 3.1 Sources of Help

Most questions should be asked on Canvas, however, please do not post any code. There is a FAQ, and anything in the FAQ will override what is specified in this specifications, if there is ambiguity.

Your lecturer is happy to discuss questions and your results with you. Please feel free to come talk to us during consultation, or even a quick question, during lecture break.

### 3.2 Marking Rubric

The rubric is attached on Canvas.

### 3.3 Submission Instructions

Submission instructions will be placed on Canvas.

### 3.4 Late Assessment Policy

A penalty of 10% of the maximum mark per day (including weekends) will apply to late assignments up to a maximum of five days or the end of the eligible period for this assignment, whichever occurs first.

Assignments will not be marked after this time.

#### 3.4.1 Assessment Eligible Period

The eligible period for this assignment is from release on Canvas until 5.00pm Friday 19 June 2020. Late submissions for extensions through ELP, assessment adjustments or special consideration will not be permitted after the end of this eligible period. Extensions through ELP or assessment adjustment that would extend the submission date beyond the eligible period will not be granted. Any special consideration that would extend the submission date beyond the eligible period will result in an equivalent assessment.

Assignments submitted after the end of the eligible period will not be marked.