COS40007 Design Project

COS40007 students are expected to undertake a design project on a focused topic of AI for Engineering. Students will get many sample data examples to work with on their projects. Students are also encouraged to collect similar data on their own if available. This document contains a summary description of each project. Week 4 and Week 5 Seminar will cover more detailed descriptions of the project topics.

A. Grouping Rules

- A group needs to be formed to complete the design project.
- The group should contain 4-5 students.
- The group must be formed within the same studio session. Formation of groups between students in multiple studio sessions will not be allowed unless there is special consideration. If you have changed your studio session to a different one you are registered for, please inform your tutor so you can attend the group of studio sessions you generally attend.
- The group are expected to produce the following outcome
  - ❖ Project Brief
  - ❖ Data Labelling
  - ❖ Data Exploration & Data Pre-processing
  - ❖ Training and validation of Machine Learning/Deep Learning models
  - ❖ Evaluation of Machine Learning/Deep Learning models
  - ❖ An AI Demonstrator of the final selected model
  - ❖ Project Presentation
  - ❖ A Final Report

B. Rubric, Report and Project Progress

- A rubric for the design project will be available around mid-semester
- An outline of the final report will be available after mid-semester
- Part of the studio session after mid-semester will be utilised to review and discuss design project progress

C. Project Topics

- The topics capture the overall project goals to solve AI problems in engineering but do not define specifications like portfolio tasks for how the project should be undertaken.
- Apply the knowledge and skills you acquired in this unit to complete your project.
- Project groups are expected to study the topic and develop a Project Brief and Plan.
- Project groups have the freedom to choose technology components.

D. Project Themes

Project Groups are expected to select from one of the following topics for a given theme. Please submit your group member's name, student numbers, studio session, and rank the project themes in this Microsoft Form or Scan below and submit by noon on 7 April. You will be assigned to one of your selected projects. Please note that the first preference will not be guaranteed. We will distribute the project so each studio session can have all five projects.

Theme 1: Smart City / Civil and Construction Engineering

AI areas: Deep Learning, Machine Learning, Object classification, Anomaly detection

Format of Data: Roadside Images of the city

Topic: Detecting roadside asset issues and identifying objects causing the issues using roadside image data.

Description: Using image data obtained from onboard cameras on vehicles, detect issues of roadside assets, such as damaged road signs, dumped rubbish, etc. The project group will get some existing image data but must label it. They are also encouraged to collect their data online or by taking photos of themselves.

Key Question to Answer

1. What roadside issue is detected by your model (e.g. damaged road sign, dumped rubbish)
2. What is the type of the detected issues (for a damaged sign, is it bent, cracked or graffiti, etc.; for dumped rubbish, what sort of rubbish (e.g., mattresses, couches, char tables, toys, etc.)

Input and Output of final AI model

Input: an image file

formating jpeg, jpg,... into unified format for consistency (color decoder, compression algorithms,..)

Output:

1. Identified issues along with confidence score
2. Identified objects/ detects along with confidence score

Data Label Annotation:
- Annotation Consistency
- Tools: LabelImg, CVAT,...

Data source

The data provided for this Theme is in the Design Project under the Theme1 data folder. You will find a data folder, and inside it, three folders: "rubbish", "not_rubbish", and "damaged-sign3". You can either choose the "rubbish"," not_rubbbish", or "damaged-sign3" dataset. This dataset contains roadside images of the city council. To develop an AI model, you first need to label data.

For "rubbish" data, you will need to annotate (using bounding box annotation) the location of the rubbish object in the image and also need to annotate at least 10 familiar objects within that rubbish, such as a mattress, electrical goods, chair, couch, trolley, toy, clothes, cartoon, rubbish bag, furniture so that your model can detect the location of the rubbish in the image and what are inside that rubbish. To train the model on what is not rubbish, some "not rubbish" images are also provided.

For "damage-sign3" data, you will need to annotate road signs (signs that are not damaged) and damaged signs in the data. For damaged signs, you must also annotate the type of damage (broken sheet, bent, crack, graffiti, rust/dust). So, your AI model will be capable of detecting damaged signs and the type of damage it was.

Image preprocessing (resizing, normalization, augmentation)

Model development

Model Evaluation and Issue Detection

User Interface

Marking criteria distribution

Task	**% weights in marking**
Data labelling and image processing	40
Training and validation	20
Detection of issues and objects in unseen data	15
Evaluation metrics	15
User Interface	10

Theme 2: Electronics Engineering / Biomedical Engineering

AI areas: Machine Learning, Activity Recognition, Feature Engineering, Predictive analytics

Format of Data: Raw motion data from sensors

Topic: Detecting Worker's activity and knife sharpness using body-worn sensors to understand Worker's productivity and safety in a manufacturing plant.

Description: Using motion data obtained from body-worn sensors in different body positions, detect the Worker's activity, such as cutting, slicing, and knife sharpness. The
The project group will get existing data but must label it by themselves.

Key Question to Answer

1. What activity is detected by your model (e.g., cutting, slicing, idle)
2. What is the quality of the knife (blunt, medium, sharp) and predict when to sharpen the knife

Input and Output of final AI model Input: raw sensor data of 1 minute
Output:

1. Identified Worker's activity
2. Identified knife sharpness and recommendation for the next stage of the sharpness

Data Source

The data provided for this Theme is in the Design Project under the Theme2 folder. There are two folders, "P1" and "P2", which refer to Person1 and Person2. For each person, there are two cutting-type activity data: boning and slicing. Under "boning" and "slicing", you will find a data file with the name format: MVN-J-abc-xyz-pqr where abc refers to one of the cutting types (boning or slicing), xyz is the knife sharpness factor (the more this factor is close to 100 means the current knife is sharper), pqr is the data collected in a different shift for different knife sharpness. Each data file is in .xlsx format, which has 18 tabs. The first tab, "General Information", contains metadata about the file. The "Markers" tab contains labelling, the type of activity the Worker was doing during the marked frame in the list. The activities are labelled as categorical values (0,1,2,3,4,5,..). The remaining tabs contain sensor fusion data of 16 different fusions for 17 body-worn sensors in 23 body positions. For this project, you can only focus on "Segment velocity" and "segment.

Acceleration" tab data. Each tab contains the frame value, the activity class label of that frame and the x,y, and z position data of 23 body positions. So, this data is already labelled. One frame corresponds to 1 second of data.
You must convert knife sharpness into three categories:

- 85 and above: Sharp
- 70 to 84: Medium
- Below 70: Blunt

To solve the problem, you will need to pre-process and extract features of 23 body position data in per-minute intervals in such a way that your model can say which activity the Worker is doing and using what type of knife if new raw data of 60 frames are provided

For example, This data is boning, and the Worker cut (activity 4) with a medium knife.

So, you need to develop 3 Machine learning models and combine their outcomes to provide the answer. You must do the necessary sampling on data to balance the classes.

Marking criteria distribution

Task	**% weights in marking**
Data pre-processing and feature extraction	40
Training and validation using different ML models	20
Classifications on unseen data	15
Evaluation metrics and model comparisons	15
User Interface	10

Theme 3: Product Manufacturing / Mechanical Engineering

AI areas: Machine Learning, Prescriptive analytics

Format of Data: Machine sensors and machine settings data during a production run of the product

Topic: Recommend machine settings values that can produce desired product consistency during production.

Description: Using machine sensors and machine settings data during the production run of manufacturing vegemite, develop ML models that can recommend machine settings values to get the desired product quality during production. You must also detect anomalies (production downtime) during the production run. The project group will get existing data however, they need to label the data by themselves.

Key Question to Answer

1. What are the recommended values of machine settings during the production process for different classes of product quality
2. What anomalies can occur in the production process for which a production run can fail

Input and Output of the final AI model

Input: current machine sensor and machine settings

Output:

1. Recommended machine settings values to get desired product quality
2. Detected anomalies that can happen with current settings

Data Source

The data provided for this Theme is in the Design Project under the Theme3 folder. The folder "data_02_07_2019-26-06-2020" contains almost 1 (July 2019 to June 2020) year data of production batches for different yeast types (part). The data contains machine settings. (SPs) and machine sensor values during the day. There are 3 CSV files -> good, low_bad and high_bad, which refers to 3 types of solid consistency of the final product. Each of these files has the following columns:

- VYP batch: batch_id with date
- Part: raw material/type of yeast used
- Set time: the date and time of the observed values from machine settings and sensors.
- Columns suffixed with SP: Set points to the settings that human operators can control.
- The remaining columns: machine sensors

The columns FFTE, TFE and Extract Tank refer to 3 different running systems.

Another folder called "Downtime" contains information about production shutdown events due to some anomalies. This downtime data is only for 2 months (May- June 2020).

You need to pre-process data and separate them for different yeast times to solve the problem. Using 1 year of production run data that resulted in 3 different quality of solid, you will need to

build a Machine Learning model which can tell the recommended SP values in the current situation based on SP and PV data to get good solid. Also, 2 months of data can tell what downtime may happen with the current data settings and in what machine.

Marking criteria distribution

Task	**% weights in marking**
Data pre-processing and feature extraction	35
Training and validation	20
SP recommendations and downtime predictions	25
Evaluation metrics	10
User Interface	10

Theme 4: Structural Engineering / Chemical Engineering

AI areas: Deep Learning, Defect Detection

Format of Data: Images containing structural defects

Topic: Structural defects detection

Description: Using image data of structural defects (such as corrosion in bridges and cracks in solar panels), identify defects and classify them. The project group will get exciting image data but needs to label it. They are also encouraged to collect their data online or by taking photos of themselves.

Key Question to Answer

1. Does your model detect any structural defect
2. What type of defect is it (e.g., corrosion, crack)

Input and Output of final AI model

Input: An image file

Output:

1. Detection of objects containing defects along with confidence score
2. Identified type of defects along with confidence score

Data source

The data provided for this Theme is in the Design Project under the Theme4 folder. You will find the "tower" folder; drones capture some high-resolution tower images inside this folder. Some parts of those towers have corrosion; you will need to annotate the tower using polygonal annotation, and corrosion in the tower will be done using bounding box annotation. Now, you will need to develop an AI model that can show the tower in the image the location of corrosions / detects in the tower. You can also use any public dataset of a similar nature for this Theme.

Marking criteria distribution

Task	% weights in marking
Data labelling and image processing	40
Training and validation	20
Object and issue detections	15
Evaluation metrics	15
User Interface	10

Theme 5: Electrical Engineering / Telecom Engineering

AI areas: Machine Learning, Clustering, Predictive analytics

Format of Data: 5G network performance data in CSV format

Topic: Grouping of zone based on 5G network performance and prediction of 5G network performance

Description: Using 5G network performance data (such as throughput and latency), identify geographical zones (from longitude and latitude) with different performance levels. Also, the network performance of the zone can be predicted using time-series data. The project group will get some existing network performance data.

Key Question to Answer

1. How many groups can be categorised using 5G network performance and their location? What are the key properties of each group?
2. The prediction value for a given network performance data.

Input and Output of final AI model

Input: 5g network performance values

Output:

1. The group and zone the data belongs to
2. The prediction of network performance for the next period

Data source

The data provided for this Theme is in the Design Project under the Theme5 folder. The "data" folder contains files on the Brimbank city council area's 5G network performance. The naming of the data has the following format

[Date_of_data_collection]-[truck_number]-combined-kml.csv

Each CSV file has the following columns:

- time: timestamp in unix format (you may not need this column), date and time information columns
- GPS coordinate (latitude and longitude): 99,999 is used for invalid detection so that you can ignore those data points
- Speed: speed of the truck
- Truck: truck id
- Svr1-svr4: latency value using four different servers
- Value and unit of different upload and download stream measures (column R to AD). Note that here, RX refers to receiving or downloading.
- Send_data: data uploaded by the application running in the truck
- Square_id: a 1-kilometre square area zone defined by the network operator (e.g., Optus)

You will need to pre-process the data by day and combine the knowledge to recommend the desired requirement using this data.

You can either do clustering or forecasting using this data.

For example, you will need to find a number of clusters using latency, data upload and download properties measure and their locations using GPS coordinates. Then, you must label them based on the aggregated network performance values.

You can use the data as time series forecasting to recommend next-hour values based on historical observations.

Marking criteria distribution

Task	**% weights in marking**
Data pre-processing and feature extraction	40
Training and model development	20
Clustering/ forecasting	15
Evaluation metrics	15
User Interface	10