COS40007 Design project

COS40007 students are expected to undertake a design project on a focused topic of AI for Engineering. Students will get a lot of example sample data to work with on their projects. Students are also encouraged to collect similar data by their own if available. This document contains a summary description of each project. More detailed description related to the project topics will be covered in Week 4 and Week 5 Seminar.

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 same studio session. Formation of group between students of multiple studio session will be not allowed unless there is a special consideration. If you have changed your studio session to a different one you are registered please inform this to your tutor so you can attend in the group of studio session you generally attend
- Group are expected to produce the following outcome
    - project brief
    - data labelling
    - data exploration and data pre-processing
    - training and validation of machine learning/deep learning models
    - evaluation of machine learning/deep learning models
    - an AI demonstrator of 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 final report will be available after mid-semester
- Part of 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 problem in Engineering, but do not define specification like portfolio tasks on how the project should be undertaken.
- You should apply your knowledge and skills you acquired in this unit to complete your project
- Project groups are expected to study on the topic and come up with a project brief and project plan
- Project groups have their own freedom to choose technology components

D. Project Themes

Project Group are expected to select from one of the following topics for a given theme. Please submit your group members name, Student numbers, studio session and your choice of 3 project theme in this google form by 26 August 12 PM. You will be assigned in one of your selected projects. Please note that, the first preference will not be guaranteed. We will distribute the project such a way that each studio session can have all 5 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 city

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

Description: Using image data obtained from on board 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, however, will need to label it. They are also encouraged to collect own data from online or by taking photos of their own.

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 damaged sign is it bent, cracked or graffiti etc.; for dumped rubbish, what sort of rubbish (e.g., mattresses, couch, char table, toy etc.)

Input and Output of final AI model

Input: an image file

Output:

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

Data source

The provided data for this Theme is located in Design Project under Theme1 folder. You will find a data folder and inside it 3 folders “rubbish”, “not_rubbish” and “damaged-sign3”. You can either choose “rubbish”,”not_rubbbish” or “damaged-sign3” dataset. This dataset contains roadside images of city council. For developing AI model, you first need to label data

For “rubbish” data you will need annotate (using bounding box annotation) the location of rubbish object in the image and also need to annotate at least 10 familiar objects within that rubbish such as 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 provided also.

For “damage-sign3” data you will need to annotate road sign (sign that are not damaged) and damaged sign in the data. For damaged sign you will also need to annotate type of the

damage (broken sheet, bent, crack, graffiti, rust/dust). So, your AI model will be capable of detecting damaged sign and what type of damage it is

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 along with knife sharpness using body worn sensors to understand worker's productivity and safety in manufacturing plant.

Description: Using motion data obtained from body worn sensors in different body position, detect worker's activity such as cutting, slicing along with knife sharpness quality. The project group will get existing data but need to label 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 next state of the sharpness

Data Source

The provided data for this Theme is located in Design Project under Theme2 folder. There are two folders named "P1" and "P2" which refers to Person1 and Person2. For each person there are 2 cutting type activity data: boning and slicing, Under "boning" and "slicing" you will find data file with the name format: MVN-J-abc-xyz-pqr where abc refer to one of the cutting type(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 different shift for different knife sharpness. Each data file is in .xlsx format which has 18 tabs. The first tab "General Information" contains some metadata about the file. The "Markers" tab contains labelling, the type of activity the worker was doing during the duration of the marked frame in the list. The activities are labelled as categorical value (0,1,2,3,4,5,..). The remaining tabs contain sensor fusion data of 16 different fusion 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 x,y,z position data of 23 body positions. So, this data is already labelled. 1 frame corresponds to 1 second data.

You will need to convert knife sharpness to 3 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 interval 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 worker was cutting (activity 4) with a medium knife.

So you need to develop 3 Machine learning models and combine there outcomes together to provide the answer. You will need to do 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 production run of product

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

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

Key Question to Answer

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

Input and Output of 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 provided data for this Theme is located in Design Project under 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 type (part). The data contains machine settings (SPs) and Machine sensor values during the day time. There are 3 CSV files -> good, low_bad and high_bad which refers to 3 type of solid consistency of the final product. Each of these files has 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 the settings that can be controlled by human operators

The remaining columns: machine sensors

Here, FFTE, TFE and Extract Tank refer to 3 different running system

There is another folder called “Downtime” that contains information of production shutdown events due to some anomalies. This downtime data is only for 2 months (May-June 2020).

To solve the problem you need to pre-process data and separate them for different yeast time. Using 1 year production run data that resulted 3 different quality of solid you will need to build Machine Learning model which can tell what are the recommended SP values in current situation based on SP and PV data to get good solid. Also using 2 months data it 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 detections

Format of Data: Images containing structural defects

Topic: Structural defects detection

Description: Using image data of structural defects (such as corrosion in bridge, crack in solar panel) identify defects and classify them. The project group will get some exiting image data, however will need to label it. They are also encouraged to collect own data from online or by taking photos of their own.

Key Question to Answer

1. Is there any structural defect detected by your model
2. What type of defect it is (e.g.: corrosion, crack)

Input and Output of final AI model

Input: An image file

Output:

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

Data source

The provided data for this Theme is located in Design Project under Theme4 folder. You will find "tower" folder and inside this there are some high resolution tower images captured by drones. Some of the parts of those towers has corrosion you will need to annotate the tower using polygonal annotation and corrosion in the tower using bounding box annotation. Now you will need to develop 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 with 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. Latency) identify geographical zone (from longitude and latitude) with different performance level. Also predict network performance of zone 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. 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 provided data for this Theme is located in Design Project under Theme5 folder. You will find "data" folder and inside this there data files of 5G network performance of Brimbank city council area. 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 you can ignore those data points

Speed: speed of the truck

Truck: truck id

Svr1-svr4: latency value using 4 different servers

Value and unit of different upload and download stream measures (column R to AD), Note that here RX refers to receive or download

Send_data: data uploaded by application running in the truck

Square_id: a 1 kilometer square area zone defined by network operator (e.g., optus)

You will need to pre-process the data by day and combine the knowledge together 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 number of clusters using latency, data upload and download properties measure and their locations using gps coordinates. Then you will need o label them based on the aggregated values of network performance.

You can use the data as time series forecasting to recommend values for next hour 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