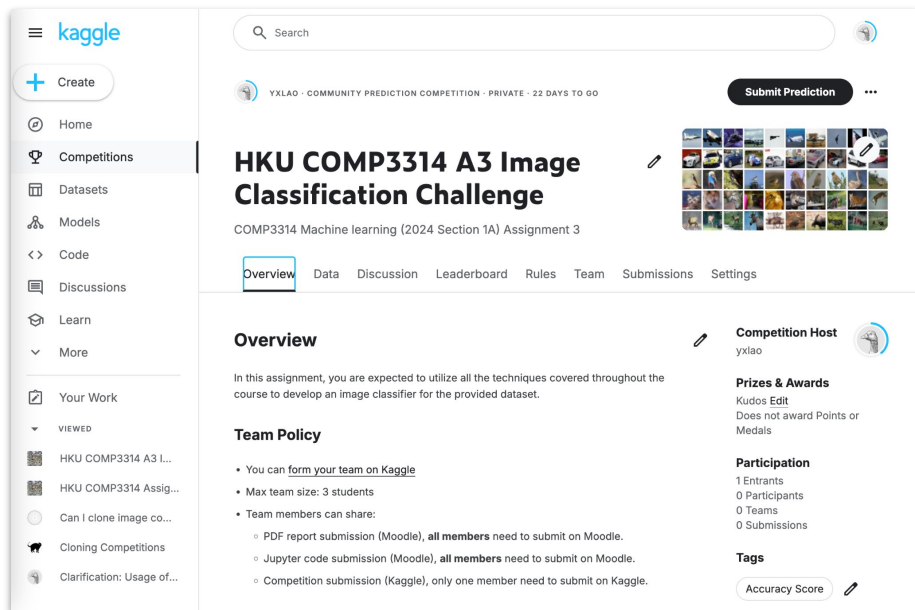# COMP3314 Tutorial 3

Assignment 3
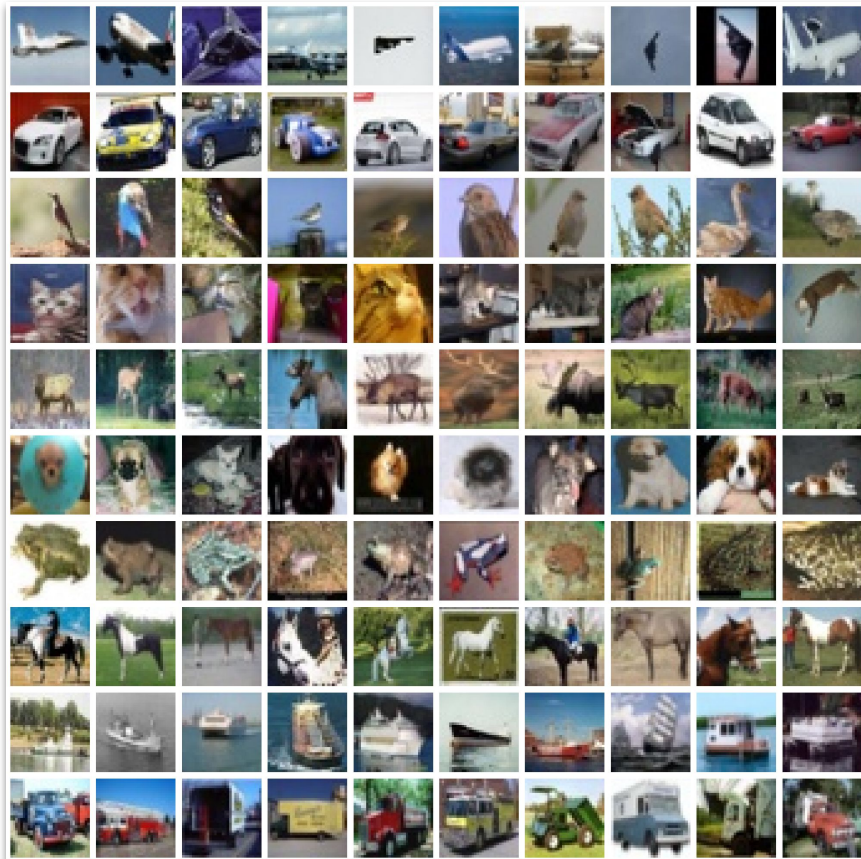
TAs for COMP3314

Assignment 3: image classification Kaggle challenge


```
max_team_size == 3
```

- Kaggle invitation URL (you must register with your @connect.hku.hk email):
  - https://www.kaggle.com/t/d0f3da431cee4a5f88d156576d99250f
- Kaggle competition URL:
  - https://www.kaggle.com/competitions/hku-comp3314-2024-1a-a3-image-classification-challenge
- Sign-up form:
  - https://forms.gle/KGjk5SN2pqWcCvhX7

Overview: image classification task

- In this assignment, you are expected to utilize all the techniques covered throughout the course to develop an image classifier for the provided dataset.
- We will be using a Kaggle **leaderboard** to manage your assignment submission.

Getting the data

Dataset overview

Training set
50,000 samples with labels
(You may split this into train/valid by yourself)

Test set
10,000 samples without labels

Public test set
2,000 samples
(Visible on leaderboard now.)

Private test set
8,000 samples
(Your final score is determined by this.
Will be visible after assignment deadline.)

Team policy

- Max team size: 3 students
- Team members can share:
  - PDF report submission (Moodle)
    - **All team members** need to submit on Moodle
  - Jupyter code submission (Moodle)
    - **All team members** need to submit on Moodle
  - Competition submission (Kaggle)
    - One of the team members needs to submit on Kaggle

Submission overview

1. Sign-up Form: https://forms.gle/KGjk5SN2pqWcCvhX7
   - Submission method: **Google Forms**
2. PDF Report: Document your process, findings, and methodology (30 pts)
   - Submission method: **Moodle**
3. Jupyter Notebook: Share your code in a runnable notebook (20 pts)
   - Submission method: **Moodle**
4. Prediction CSV: Submit your predictions on Kaggle for scoring (50 pts)
   - Submission method: **Kaggle**

Submission 1 of 4: Sign-Up

- https://forms.gle/KGjk5SN2pqWcCvhX7
- Yes, you can re-edit or re-submit this form if your team affiliation changes.
- If we do not receive this registration, you may receive 0 points for your Kaggle submission.
- **All members** of a team must submit the sign-up form.

COMP3314 Assignment 3 Sign-Up (2024 Sem 1)

All students (all members in a team) shall register in this sign-up form, in order to correctly assign your scores.

* Indicates required question

Your HKU Email *

Your answer

Your HKU UID *

Your answer

Your Kaggle Profile URL *
The profile page URL for your Kaggle account. It shall look like: https://www.kaggle.com/ikunkaggle

Your answer

Your Kaggle Team Name *

Your answer

Submit   Clear form

Never submit passwords through Google Forms.

This content is neither created nor endorsed by Google. Report Abuse - Terms of Service - Privacy Policy

Google Forms

Submission 2 of 4: PDF report (on Moodle)

The experiment report should cover the following aspects:

- Dataset analysis
  - Statistics on the number of categories
  - Visualizations of one example for each category
- Classifier exploration
  - Implement at least 2 different classifiers and compare their performance. Present the comparison and your analysis of your classifiers' performance. You may split the training set to "train" and "validation" based on your needs.
- Final solution description
  - Describe the core pipeline of your final solution, highlighting key components and methodologies utilized to achieve the desired classification results.
- **All members** of a team must submit the PDF report on Mooble

Submission 3 of 4: Jupyter Notebook (on Moodle)

You shall submit a Jupyter notebook to validate the reproducibility of the test results submitted on the test dataset.

- The Jupyter notebook **must contain the entire pipeline** of your final solution.
- Ensure the notebook is executable, with **pre-executed logs printed** for clarity.
- Upon running the notebook, it should generate a .csv file within the same directory as the notebook.
- Do not upload the dataset. Only the Jupyter Notebook shall be uploaded.
- TA may check the Jupyter Notebook to verify the results with your submitted results on Kaggle.
- **All members** of a team must submit the Jupyter Notebook on Mooble

Submission 4 of 4: prediction CSV file (on Kaggle)

Example **test.csv**
(filled with dummy labels)

```
im_name,label
d59d350.jpg,0
d59e3e9.jpg,0
d59e68e.jpg,0
d59ef00.jpg,0
d59f154.jpg,0
d59fa50.jpg,0
d5a1a79.jpg,0
d5a1c9a.jpg,0
...
```
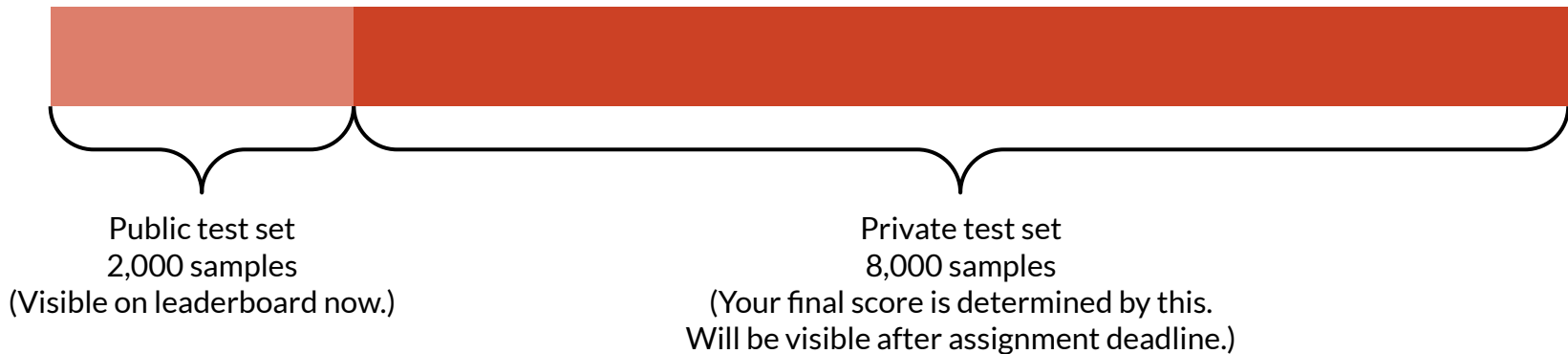
Example **submission.csv**
(filled with predicted labels)

```
im_name,label
d59d350.jpg,3
d59e3e9.jpg,1
d59e68e.jpg,4
d59ef00.jpg,1
d59f154.jpg,5
d59fa50.jpg,9
d5a1a79.jpg,2
d5a1c9a.jpg,6
...
```

Note: pay attention to the image name and label mapping.

Leaderboard

Public test set
2,000 samples
(Visible on leaderboard now.)

Private test set
8,000 samples
(Your final score is determined by this.
Will be visible after assignment deadline.)

- **Evaluation metric**: Accuracy score
- **Public test set**: During the competition, your submission's accuracy on the public test set will be visible on the leaderboard.
- **Private test set**: However, final rankings and the determination of award points will be based on your submission's performance on the private test set, which will be revealed at the competition's conclusion.

Grading policies

- PDF report (total 30 points)
- Jupyter notebook (total 20 points)
- Kaggle competition (total 50 points)
  - 50 points if accuracy score >= 0.70
  - 45 points if accuracy score >= 0.60
  - 40 points if accuracy score >= 0.50
  - 35 points if accuracy score >= 0.40
  - 30 points if accuracy score >= 0.30
  - 25 points if accuracy score >= 0.20
  - Bonus: top 0% - 10% teams get 5 bonus points
    - (Max 105 points)

Important Rules

- Neural networks (CNNs, RNNs, Transformers, etc.) are not allowed.
- Additional datasets or pre-trained models can not be utilized.
- Plagiarism of code or prediction results from external sources is strictly prohibited.
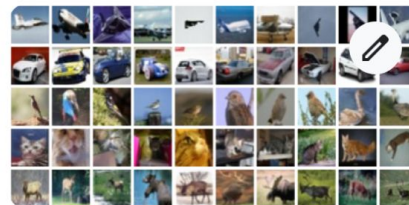
Leaderboard submission

YXLAO · COMMUNITY PREDICTION COMPETITION · PRIVATE · 22 DAYS TO GO

Submit Prediction

HKU COMP3314 A3 Image Classification Challenge

COMP3314 Machine learning (2024 Section 1A) Assignment 3

Overview   Data   Discussion   Leaderboard   Rules   Team   **Submissions**   Settings

Submissions

Select up to 5 submissions that will count towards your final leaderboard score. If less than 5 are selected, Kaggle will automatically select from your best scoring submissions. Learn More

🟦 Auto-selection candidates ⍰

0/5

Timeline

- Moodle submission: 11/24/2024 23:59 PM
- Kaggle submission: 11/25/2024 00:05 AM
  - Private leaderboard result will be released after the deadline

Recommendations

Form Your Team
Start Early
Have Fun

Q&A