



COMP3314 Tutorial 4

Tutorial for Assignment 3

TA for COMP3314

Assignment 3: image classification Kaggle challenge

The screenshot shows the Kaggle competition page for 'HKU COMP3314 Assignment 3: Image Classification'. The page is titled 'COMP3314 (2023-2024 Sem 2) Assignment 3'. It features a sidebar with navigation options like Home, Competitions, Datasets, Models, Code, Discussions, Learn, and More. The main content area includes an 'Overview' section with a description of the assignment, a 'Start' section with a progress bar, and a 'Requirements' section with two main tasks: '1. Register and Record Your Kaggle Account (5 Points)' and '2. PDF Report (Submit to Moodle, 30 Points)'. The 'Requirements' section also includes a list of tasks for the PDF report, such as 'Dataset analysis' and 'Classifier exploration'. The right sidebar shows 'Competition Host' (yxiao), 'Prizes & Awards', 'Participation' statistics, and a 'Table of Contents'.

`max_team_size == 3`

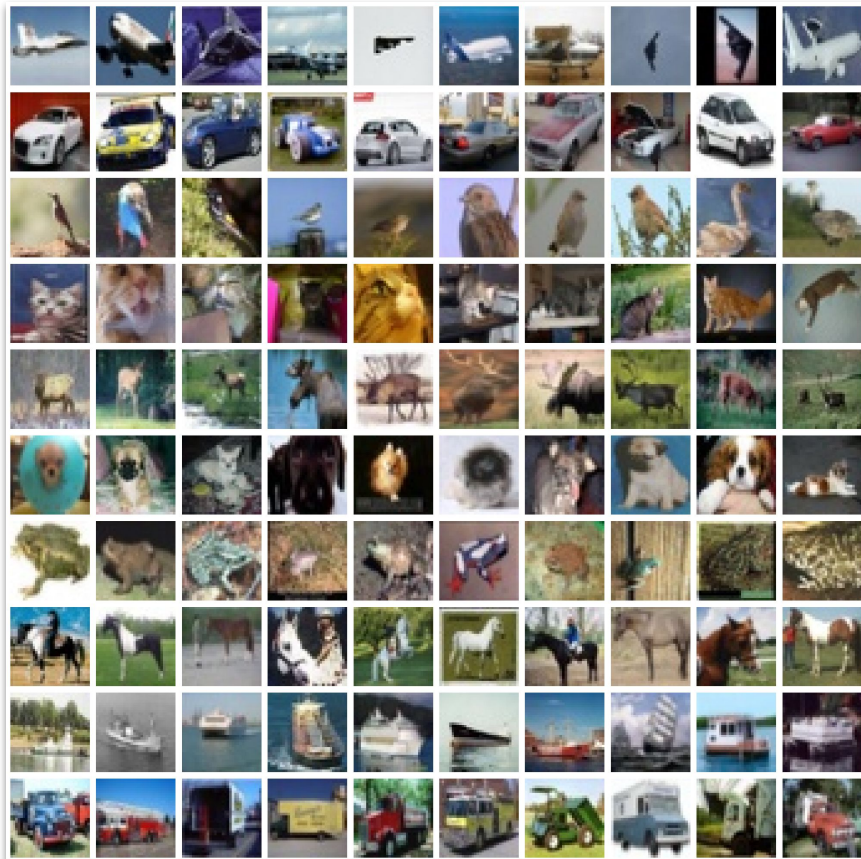
Kaggle invitation URL:

<https://www.kaggle.com/t/62283d169918407e85492b2591e1c5e1>

Kaggle competition URL:

<https://www.kaggle.com/competitions/comp3314-assignment-3-image-classification/>

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

HKU COMP3314 Assignment 3: Image Classification

Overview **Data** Discussion Leaderboard Rules Team Submissions

10000 unique values

d59d350.jpg	0
d59e3e9.jpg	0
d59e68e.jpg	0
d59ef00.jpg	0
d59f154.jpg	0
d59fa50.jpg	0
d5a1a79.jpg	0
d5a1c9a.jpg	0
d5a237e.jpg	0
d5a47bd.jpg	0
d5a49d9.jpg	0
d5a579a.jpg	0
d5a5b0e.jpg	0
d5a7656.jpg	0
d5a86ed.jpg	0
d5a9bfe.jpg	0
d5a9d67.jpg	0
d5ab460.jpg	0
d5ab77c.jpg	0
d5ad58f.jpg	0
d5b3770.jpg	0
d5b3826.jpg	0
d5b3c81.jpg	0
d5b6e23.jpg	0

Summary

- 60.0k files
- 4 columns

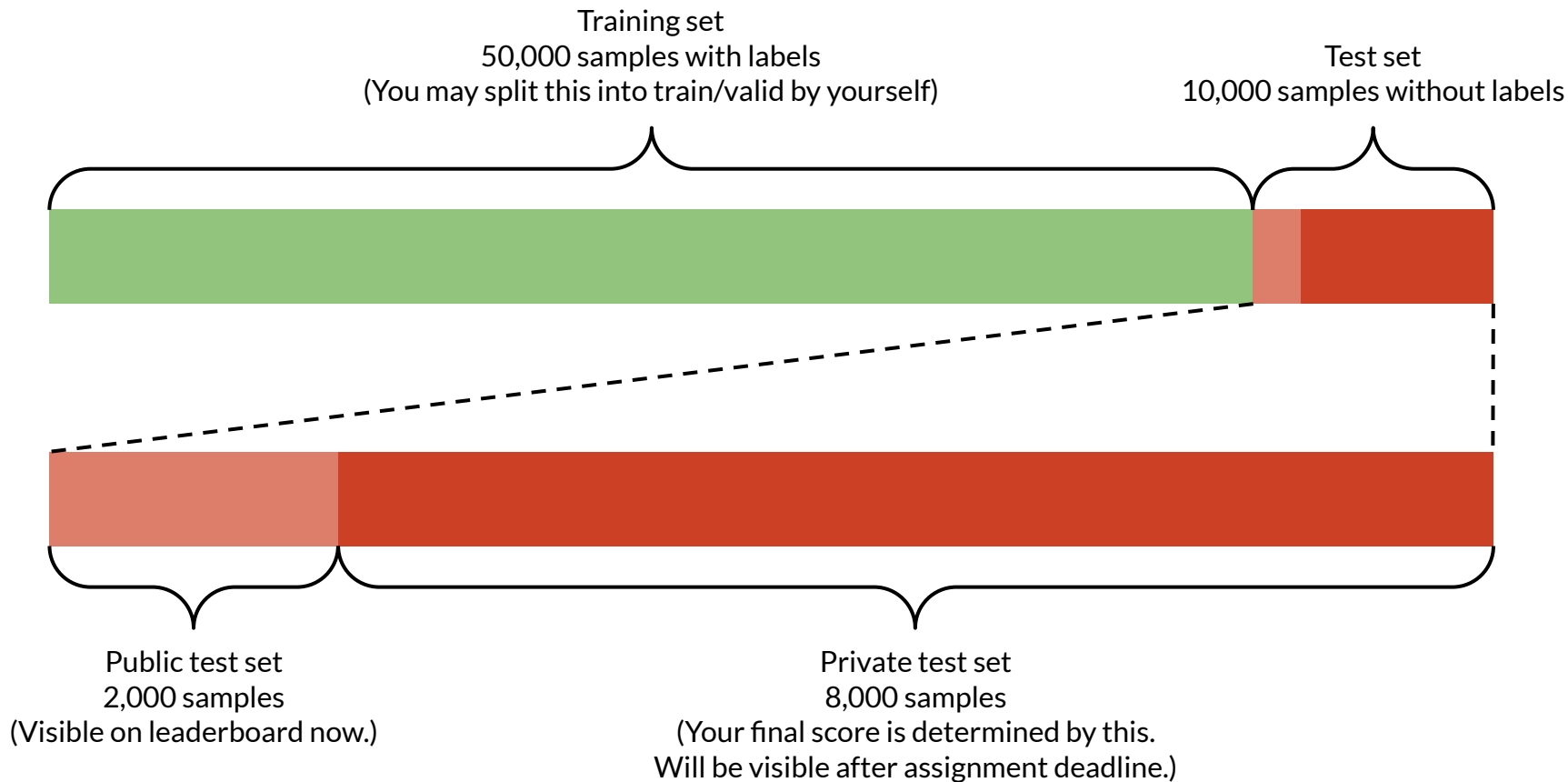
[Download All](#)

```
> kaggle competitions download -c comp3314-assignment-3-image-classification
```

data

```
├─ train.csv
├─ test.csv
├─ train_ims
│   ├── 00016cd.jpg
│   ├── 0001808.jpg
│   ├── 0002399.jpg
│   ├── 0003973.jpg
│   ├── 00061cc.jpg
│   └─ ...
├─ d59d147.jpg
└─ test_ims
    ├── d59d350.jpg
    ├── d59e3e9.jpg
    ├── d59e68e.jpg
    ├── d59ef00.jpg
    ├── d59f154.jpg
    └─ ...
    └─ fffe437.jpg
```

Dataset overview



Submission overview

1. Kaggle Account: Record your Kaggle username via the Google form (5 pts)
 - Submission method: **Google form**
2. PDF Report: Document your process, findings, and methodology (30 pts)
 - Submission method: **Moodle**
3. Jupyter Notebook: Share your code in a runnable notebook (15 pts)
 - Submission method: **Moodle**
4. Prediction CSV: Submit your predictions on Kaggle for scoring (50 pts)
 - Submission method: **Kaggle**

Register Kaggle competition

Kaggle invitation URL:

<https://www.kaggle.com/t/62283d169918407e85492b2591e1c5e1>

Kaggle competition URL:

<https://www.kaggle.com/competitions/comp3314-assignment-3-image-classification/>

Team policy

- Max team size: 3 students
- You shall form your team on Kaggle.
- Every team member shall:
 - Register Kaggle
 - Record the account on Google Form
- Team members can share:
 - PDF report submission (Moodle)
 - Jupyter code submission (Moodle)
 - Competition submission (Kaggle)
- Only one team member need to submit on Moodle for the PDF report and Jupyter code.

The screenshot shows the Kaggle interface for the 'HKU COMP3314 Assignment 3: Image Classification' competition. The left sidebar contains navigation links: Home, Competitions, Datasets, Models, Code, Discussions, Learn, More, Your Work, and VIEWED. The main content area is titled 'HKU COMP3314 Assignment 3: Image Classification' and includes a 'Submit Prediction' button. The 'Team' tab is selected in the navigation bar. The 'Your Team' section shows a team named 'Demo' with a description: 'Everyone that competes in a Competition does so as a team - even if you're competing by yourself. [Learn more](#).' The 'General' section includes a 'TEAM NAME' field with the value 'Demo' and a note: 'This name will appear on your team's leaderboard position.' The 'Host Contact' section has a toggle for 'Share your email address with the host' and a note: 'When you share your email address with the host, they may contact you for issues related to the competition (e.g. a professor administering a classroom competition). If you have teammates, this will not share their email addresses.' The 'Let others know you're looking for teammates' section includes a note: 'To use this feature, you need to reach the Contributor level on Kaggle. [Learn more](#)' and a toggle for 'Display your status with [icon] on the leaderboard'. The 'Team Members' section shows a team leader 'demo (You)' and a note: 'Your team can have a maximum of 2 members.' The 'Send Invitation' section includes a note: 'You can invite 1 more person to join your team. They will need to have joined the competition.' and a search bar for the team name. The 'Received Invitations' section shows a note: 'You haven't received invitations from other teams.'

Submission 1 of 4: Record your Kaggle username

Register a Kaggle account and record your Kaggle account using this form.

You may also form a team on Kaggle afterwards. Every team member must record their Kaggle account using the Google Form.

URL:

https://docs.google.com/forms/d/e/1FAIpQLSdKXKj5-MuCmn1nL3GeRM_jlQr67kaqhtds24Ls_nVd84NBw/viewform

HKU COMP3314 Assignment 3 Sign-Up


In-class competition for HKU COMP3314 (2023-2024 Sem 2) Assignment 3 Sign-Up

yixing.lao@gmail.com [Switch account](#)

🔒 Not shared

* Indicates required question

Your HKU Email *

Your answer 

Your HKU UID *

Your answer

Your Kaggle user name *

If your Kaggle profile is on <https://www.kaggle.com/example-name>, your Kaggle user name is "example-name".

Your answer

Submit

Clear form

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.

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.

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

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

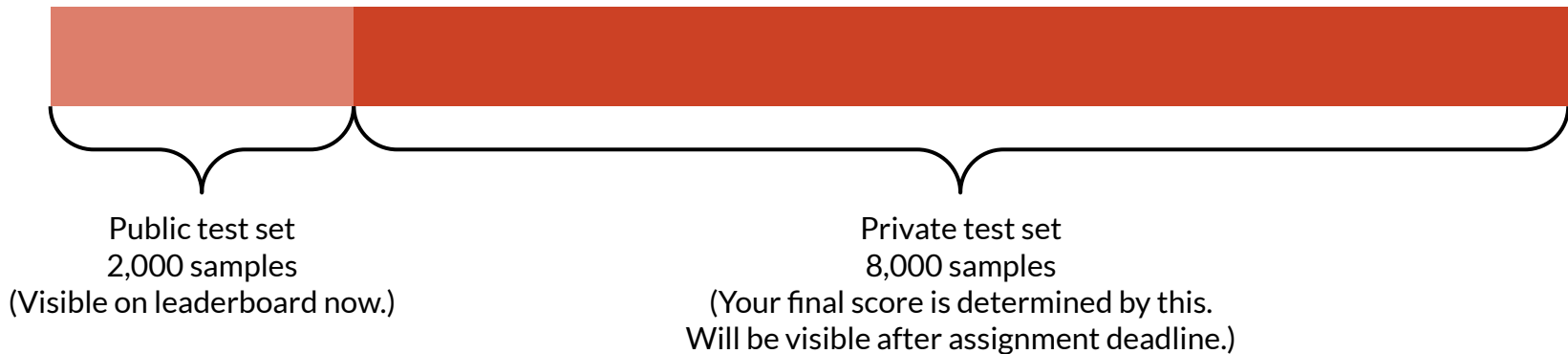
```
im_name,label  
d59d350.jpg,0  
d59e3e9.jpg,0  
d59e68e.jpg,0  
d59ef00.jpg,0  
d59f154.jpg,0  
d59fa50.jpg,0  
d5a1a79.jpg,0  
d5a1c9a.jpg,0  
...
```

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



- **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 policy based on your ranking

- 50 points if you rank within the top 0% - 10%.
- 40 points if you rank within the top 10% - 20%.
- 30 points if you rank within the top 20% - 50%.
- 20 points if you rank within the top 50% - 70%.
- 10 points if you rank within the top 70% - 100%.

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.

Discussion policy (bonus!)

For questions and discussions, please post on **Kaggle discussion board** so that your classmates can also benefit. Unless there are privacy concerns, public discussions are preferred over emails to encourage collaboration.

To encourage class collaboration, **the top three most active student respondents (who reply to questions) will receive a bonus of 5 points** (the maximum score remains 100 points). You may only discuss topics related to this assignment. Please note that directly posting answers is not allowed.

Leaderboard submission

The screenshot displays the Kaggle interface for a competition titled "HKU COMP3314 Assignment 3: Image Classification". The left sidebar contains navigation links: Home, Competitions (selected), Datasets, Models, Code, Discussions, Learn, More, Your Work, and a list of viewed competitions including "HKU COMP3314 Assig...". The main content area shows the competition details, including the title, subtitle "COMP3314 (2023-2024 Sem 2) Assignment 3", and a grid of image samples. A red circle highlights the "Submit Prediction" button in the top right corner. Another red circle highlights the "Submissions" tab in the navigation bar. Below the tabs, the "Submissions" section is visible, showing instructions for selecting up to 100 submissions and a progress indicator "0/100". The "Submissions" section also includes a filter for "Auto-selection candidates" and a list of submission statuses: All, Successful, Selected, and Errors. The bottom of the page shows a table header with columns for "Submission and Description", "Public Score", and "Select".

Kaggle

Search

YXLAO · COMMUNITY PREDICTION COMPETITION · PRIVATE · 17 DAYS TO GO

HKU COMP3314 Assignment 3: Image Classification

COMP3314 (2023-2024 Sem 2) Assignment 3

Overview Data Discussion Leaderboard Rules Team **Submissions**

Submissions

Select up to 100 submissions that will count towards your final leaderboard score. If less than 100 are selected, Kaggle will automatically select from your best scoring submissions. [Learn More](#)

Auto-selection candidates ?

All Successful Selected Errors

Recent ▼

Submission and Description Public Score Select

Timeline

- Moodle submission: Apr 21, 23:59
- Kaggle submission: Apr 22, 00:05
- Private leaderboard result will be released afterwards.

Recommendations

Form Your Team
Start Early
Have Fun

Q&A