

# Tests for Prospective GSoC 2022 Applicants

Below are the tests we will use to evaluate prospective GSoC students for the [DeepLense](#) projects. After completing the first common test, please thoroughly complete the specific second test for your project of interest and (optionally) other second tests if you would like to also be considered for additional DeepLense projects at the same time (this may increase your chances of success, but make sure you don't do it at the expense of the specific project you are interested in)

**Note:** please work in your own github branch (i.e. NO PRs should be made). Send us a link to your code when you are finished and we will evaluate it.

## Common Test I. Multi-Class Classification

**Task:** Build a model for classifying the images into lenses using **PyTorch** or **Keras**. Pick the most appropriate approach and discuss your strategy.

**Dataset:** [dataset.zip - Google Drive](#)

**Dataset Description:** The Dataset consists of three classes, strong lensing images with no substructure, subhalo substructure, and vortex substructure. The images have been normalized using min-max normalization, but you are free to use any normalization or data augmentation methods to improve your results.

**Evaluation Metrics:** ROC curve (Receiver Operating Characteristic curve) and AUC score (Area Under the ROC Curve)

- Those interested in the Lens Finding Project should complete Test II.
- Those interested in the Regression Project should complete Test III.
- Those interested in the Equivariant NN Project should complete Test IV.
- Those interested in the Transformer Project should complete Test V.
- Those interested in Extending the DeepLense Pipeline should complete Test III.

## Specific Test II. Lens Finding

**Task:** Build a model for classifying the images into lenses using **PyTorch** or **Keras**. Pick the most appropriate approach and discuss your strategy.

**Dataset (Updated as of 4/4/22):**

<https://drive.google.com/file/d/1eXmbZqUfpqhl-MWz2xkpK0dXQ4FttGQf/view?usp=sharing>

**Dataset Description:** A data set comprising images with and without strong lenses.

**Evaluation Metrics:** ROC curve (Receiver Operating Characteristic curve) and AUC score (Area Under the ROC Curve)

### **Specific Test III. Learning Mass of Dark Matter Halo**

**Task:** Using the provided dataset implement a regression algorithm to learn the mapping between lensing images and the lensing dark matter halo mass. You can use the machine learning algorithm of your choice. Please implement your approach in **PyTorch** or **Keras** and discuss your strategy.

**Dataset:**

<https://drive.google.com/file/d/1hu472ALwGPBcTCXSAM0VoCWmTktg9j-j/view>

**Dataset Description:** The data set consists of strong lensing images for cold dark matter with subhalo substructure. For each lensing image the corresponding fraction of mass in dark matter substructure is provided.

**Evaluation Metrics:** MSE (mean squared error)

### **Specific Test IV. Exploring Equivariant Neural Networks**

**Task:** Use an Equivariant Neural Network of your choice to build a robust and efficient model for binary classification or unsupervised anomaly detection on the provided dataset. In the case of unsupervised anomaly detection, train your model to learn the distribution of the provided strong lensing images with no substructure. Please implement your approach in **PyTorch** or **Keras** and discuss your strategy.

**Dataset:**

[https://drive.google.com/file/d/16Y1taQoTeUTP5rGpB0tuPZ\\_S30acvnqr/view?usp=sharing](https://drive.google.com/file/d/16Y1taQoTeUTP5rGpB0tuPZ_S30acvnqr/view?usp=sharing)

**Dataset Description:** A set of simulated strong gravitational lensing images with and without substructure.

**Evaluation Metrics:** ROC curve (Receiver Operating Characteristic curve) and AUC score (Area Under the ROC Curve)

**Specific Test V. Exploring Transformers**

**Task:** Use a vision transformer method of your choice to build a robust and efficient model for binary classification or unsupervised anomaly detection on the provided dataset. In the case of unsupervised anomaly detection, train your model to learn the distribution of the provided strong lensing images with no substructure. Please implement your approach in **PyTorch** or **Keras** and discuss your strategy.

**Dataset:**

[https://drive.google.com/file/d/16Y1taQoTeUTP5rGpB0tuPZ\\_S30acvnqr/view?usp=sharing](https://drive.google.com/file/d/16Y1taQoTeUTP5rGpB0tuPZ_S30acvnqr/view?usp=sharing)

**Dataset Description:** A set of simulated strong gravitational lensing images with and without substructure.

**Evaluation Metrics:** ROC curve (Receiver Operating Characteristic curve) and AUC score (Area Under the ROC Curve)

**Submission Guidelines**

- You are required to submit Jupyter Notebooks for each task clearly showing your implementation.
- Please also put your solution in a github repository and send us a link
- You must calculate and present the required evaluation metrics for the validation data (90:10 train-test split).

- When completed, please send your **CV**, a link to your **solution repository**, your **notebooks** and **trained model weights** to [ml4-sci@cern.ch](mailto:ml4-sci@cern.ch) with Evaluation Test: DeepLense in the title