**Final Project**

The objective of the final project will be to apply what you have learned in the second half of the course, in particular neural networks, to datasets of your choice. The datasets should be related to education. The analysis should involve the use of images along with other data.

Some sources of data are below. There are many more you can find by googling. You can also create your own image dataset by taking photos. The dataset does not need to be very big.

Educational datasets

<https://pslcdatashop.web.cmu.edu/>

<http://learnsphere.org/index.html>

<https://eedi.com/projects/neurips-education-challenge>

<https://osf.io/59shv/>

<https://nces.ed.gov/>

Image datasets

<https://www.kaggle.com/datasets?sortBy=relevance&group=featured&search=image>

<https://registry.opendata.aws/>

<https://datasetsearch.research.google.com/>

<https://www.visualdata.io/>

<https://ieee-dataport.org/topic-tags/computer-vision>

<https://www.taqadam.io/open-sourse-datasets/>

<https://github.com/TIBHannover/SlideImages>

Please pay attention to the grading rubric below and make sure you do everything that is asked for in the rubric.

**Grading rubric**

|  |  |
| --- | --- |
| Item | Points |
| **Description of dataset and objective of analysis** - Clear description of the dataset you are using, where it was sourced from, and any wrangling/cleaning you may have performed. What is the objective of your analysis? What questions are you trying to answer? | 5 |
| **Preprocessing/EDA**:  Perform any necessary preprocessing on the images  Perform EDA  The points you get will depend on the rigor of the EDA you have performed | 10 |
| **Use of different types of neural networks**:  Regular multilayer fully connected (dense) networks  Convolutional Neural Networks | 5 |
| **Performance Tuning:**  Tune various hyperparameters – batch size, epochs, learning rate etc. Use learning rate schedules, data augmentation etc.  Tune weight initialization, different types of gradient descent, different activation functions, different cost functions etc.  Try automating tuning using keras tuner  The points you get will depend on the rigor of the tuning you have performed. | 25 |
| **Try various network architectures:**  Several combinations of different types of layers, Regularization – L1/L2 regularization, dropouts etc.  The points you get will depend on the extent to which you have tried various architectures to improve performance. | 20 |
| **Transfer learning:**  Use pre-trained CNNs like ResNet etc. Try several different pre-trained models. Minimum of two models should be used. A list of available models in keras can be found at <https://keras.io/api/applications/>. Comment on the accuracy difference between the pre-trained model and the model you built from scratch.  The points you get will depend on how many pre-trained models you tried and how you have tuned it further on your dataset. | 15 |
| **Presentation** – A professional presentation that highlights key aspects of all the topics mentioned above. The presentations should be max 10 to 12 minutes. No need to present code (you can pull up the code if you want to highlight something specific that others could learn from). Spend some time addressing difficulties encountered and how rectified, so others can learn from how you solved the issues you faced. The presentation has to be complete in that it addresses everything mentioned above. The points you get will depend upon the completeness and quality of the presentation. | 5 |
| **Report** - A professional and succinct report of the project that should contain the code and output, but it should NOT just be code and output without any written commentary.  Please make sure there are comments in the code or text paragraphs explaining in detail in your own words what each section of your code does. Please write a paragraph or paragraphs on what were the things that you tried that led to improvement in accuracy and what did not.  You can turn in a Jupyter notebook as your report or a document generated via R Markdown containing all of the above.  The points you get will depend upon the quality and completeness of the report. | 5 |
| **Peer evaluation** | 10 |