Case Study Rubric

DS 4002 - April 28th 2025 - Bridget Holt
Due: TBD
Submission Format: Upload link to GitHub repository on UVA Canvas page

Individual Assignment

Why am I doing this?
This case study gives you the opportunity to apply deep learning techniques to a real-world image classification problem. By training and evaluating a CNN model, you'll gain experience with tools and workflows commonly used in computer vision and data science. This project challenges you to think critically about model design, data quality, and evaluation, skills that are essential in both academic and industry settings.

What am I going to do?

In this case study, you will use a convolutional neural network (CNN) model, specifically VGG16 with transfer learning, to classify flower species based on images. You'll begin by cleaning and merging datasets from the Oxford Flowers collection, perform exploratory data analysis, and prepare the data for modeling. You will train and evaluate your model on UVA's Rivanna high-performance computing cluster.

Your task includes selecting model parameters, interpreting training and validation accuracy over multiple epochs, and analyzing results using confusion matrices. You will also reflect on potential bias, image quality challenges, and how background features may influence model performance.

The final deliverable includes:

- a written report
- modeling code
- GitHub repository containing all files and documentation

How will I know I have succeeded?
You will meet expectations on the case study when you successfully follow the criteria on the rubric below:

Spec Categories	Spec details
Formating	● Written Portion ○ Submit a PDF that includes the case study write-up (problem, model, results, and reflection). ○ Include a title page with your name, project title, and date. ● Data & Code ○ Submit a GitHub repository containing all relevant code and any preprocessed or merged data files. ○ The repository should be named using the format: CS3-[YourFirstNameLastName]. ○ Use folders to organize code, images, data, and documentation clearly. ● README.md ○ Include a high-level summary of the project, modeling approach, and instructions for how to run the code. ○ Add links to key files or subfolders within the repo. ● References ○ Include references in IEEE format in a REFERENCES.md file or on a final page of the written PDF. ○ Provide a short annotation under each citation explaining how it was used in your project.
README.md	Include a brief summary of your case study including the research question, modeling approach, and dataset. ● Provide clear instructions on how to navigate the repository and run your code.
Source Code File	Your code submission should be a well-documented Jupyter Notebook

	or Python script that contains the full modeling workflow. It must include: ● Data cleaning and merging across the three Oxford datasets (images, index, and flower names) ● Image preprocessing steps (e.g., resizing, normalization, and augmentation if used) ● Setup and training of the CNN model using VGG16 with transfer learning ● Implementation of model evaluation tools, including: ○ Confusion matrix ○ Training and validation accuracy and loss over epochs ● Adjustments made based on training performance (e.g., increasing epochs, tuning batch size or learning rate) ● Code used to submit and run training jobs on Rivanna ● Clear and consistent comments throughout to explain your logic and decisions
References	Include a list of references in a separate section of the written PDF or in a REFERENCES.md file in IEEE format. Mention any sources used for tutorials, documentation, or datasets, and add brief notes explaining their relevance to your work.