

# Dermatological Image Classification Analysis -- Rubric

**Due:** TBD

**Submission format:** Upload a PDF document and a link to the Github repository to Canvas

## Individual Assignment

**General Description:** Analyze and classify dermatological image lesions as benign or malignant using a CNN model, studying the key features that lead to these classifications. Submit your PDF document and a link to your Github repository for this project.

**Why am I doing this?** This case study is an opportunity for you to become familiar with convolutional neural network (CNN) models, which are becoming increasingly popular for the analysis of various impactful data. Additionally, you will be introduced to the basics of working with image data and are responsible for applying your knowledge to a case study representing a real-world situation.

- Course Learning Objective: Learn the basics of working with CNN models through the analysis of image data, applying concepts in the context of a real-world situation.

**What am I going to do?** Through previous assignments and coursework, you will have accumulated technical and conceptual skills in data science. This is your opportunity to apply these skills in an independent case study. You will conduct research on CNN models and image data preprocessing, and create a model that can be used to classify dermatological lesions as Benign, Benign\* (benign, but at risk of becoming malignant), or Malignant. Deliverables include:

- Written portion of the submission as a PDF document – including references
- Github repository – houses code, data, and supporting documents

### Tips for success:

- **Focus is key!** – Avoid distractions, it is much easier to complete this analysis in a timely manner with your full focus.
- **Research is the key to knowledge** – More likely than not, you will be largely unfamiliar with working with CNN models and image data. Be sure to spend time on researching these unfamiliar topics so you can determine the best way to proceed.
- **Don't be afraid to ask for help** – Talk to your professor and TA for guidance or to discuss your thoughts. They are there to make sure you succeed, so take advantage of the resources they are willing to provide you!

**How will I know I have Succeeded?** You will meet expectations on this case study when you follow the criteria in the rubric below.

Spec Category	Spec Details
Formatting	<p>Submit each of the following components listed in this rubric as advised below:</p> <ul style="list-style-type: none"> <li>○ Written Portion</li> <li>○ Data and Code</li> <li>○ References</li> </ul> <p>Include all in one Github repository (submitted via link on Canvas)</p>
Written Documentation	<ul style="list-style-type: none"> <li>● <u>Goal</u>: This folder contains all of the written portions of this case study.</li> <li>● Background Investigation <ul style="list-style-type: none"> <li>○ In your own words, discuss what convolutional neural network (CNN) models are, what they can be used for, and why they are important. Additionally, summarize the real-world problem presented in this case study and its significance.</li> </ul> </li> <li>● Project Outline <ul style="list-style-type: none"> <li>○ Create a plan for the completion of this case study. Discuss each step at a high-level, and include what the task is, how you plan to approach it, how you actually approached it, and any other relevant information.</li> <li>○ Include a simple graphic that summarizes your analysis plan for this case study.</li> </ul> </li> <li>● Reflection <ul style="list-style-type: none"> <li>○ Discuss the results of your case study. Be sure to include any challenges you faced and the solutions you came up with for them. Highlight the significance of your results, as well as potential areas of improvement of your analysis.</li> </ul> </li> </ul>
Data and Code	<ul style="list-style-type: none"> <li>● <u>Goal</u>: This folder contains all of the relevant data and code for this case study.</li> <li>● Raw Image Data <ul style="list-style-type: none"> <li>○ Initial image data set, with no preprocessing.</li> </ul> </li> <li>● Merged, Normalized Image Data <ul style="list-style-type: none"> <li>○ Final image data set, post-merging and normalization</li> </ul> </li> <li>● Exploratory Data Analysis (EDA) <ul style="list-style-type: none"> <li>○ Include impactful figures highlighting image characteristics and distribution of lesion classifications in the dataset.</li> </ul> </li> <li>● Image Merging <ul style="list-style-type: none"> <li>○ Merge lesion outline image with full image to highlight a concentrated area of the lesion, then normalize.</li> </ul> </li> <li>● CNN Model <ul style="list-style-type: none"> <li>○ HINT: Start with ResNet-50 as your CNN model.</li> <li>○ Train and test model using merged image data, and generate confusion matrix and other model performance scores.</li> </ul> </li> </ul>
References	<ul style="list-style-type: none"> <li>● All references should be listed at the end of the written documentation</li> <li>● Use IEEE Documentation style (<a href="#">link</a>)</li> </ul>