

**Name That Mushroom!**

*The world of mycology is a veritable universe unto itself, in the soil around all of us. Forest floors invite exploration and discovery for young and old, novices and experts. And the world of mushrooms is diverse enough that even a seasoned mushroom hunter will find mushrooms she can’t always identify. Mushrooms are at once alluring and anxiety-causing; the right mushroom can be delicious, but the wrong mushroom can be a toxic experience. Imagine if a computer could help you identify which mushroom you’re looking at? In this project, we will train a machine learning model on images of different common mushroom genus, to see if it can be like having a mushroom expert in your pocket!*

**1. Data**

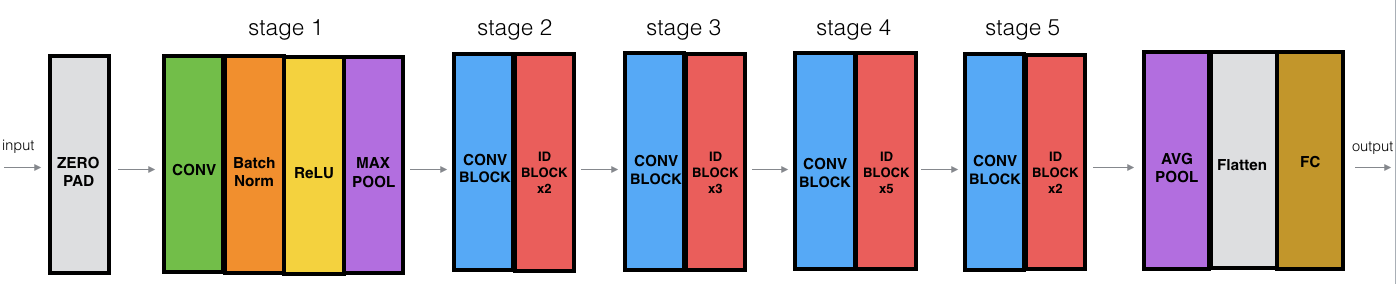
The series of mushroom images comes from the Kaggle website, link below:

* [Kaggle Dataset](https://www.kaggle.com/maysee/mushrooms-classification-common-genuss-images)

**2. Method**

Neural networks, with their ability to identify and look for patterns in visual data, are particularly useful for image classification. Two neural network types are explored:

1. **Convolutional Neural Network (CNN):** Convolutional methods “convolve” the individual pixel data, aggregating them in useful ways for pattern recognition.
2. **Residual Neural Network (ResNet):** Extremely deep learning (i.e., many layers) suffer from issues with vanishing gradients. ResNet algorithms have been in use for a while, but in the last several years have gained prominence for their ability to deal with vanishing gradients with techniques such as skip connections, allowing deep learning and greater classification accuracy.



The ResNet50 Model

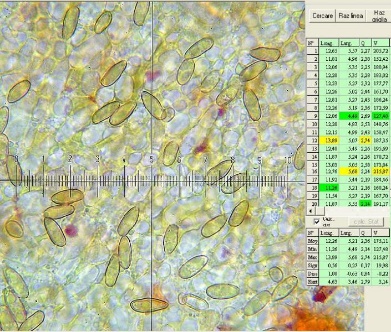
**3. Data Cleaning**

The Kaggle dataset is pre-organized into subfolders, each representing a common mushroom genus. While the vast majority of the images are perfectly fine for use, there are a couple data cleaning opportunities:

* **Problem 1:** The original dataset contains images with writing, highlights, arrows and circles, probably gleaned from books and websites about the subject of mushroom classification. While a human can ignore these image features as clearly non-natural, a computer may not be able to.



* **Problem 2:** The original dataset contains images, some microscopic, of spores of the genus types.



* **Problem 3:** Several images which show sliced mushrooms. While these are not a natural form for mushrooms in the forest, different mushrooms can have different discolorations when cut and exposed to air. Ultimately, I decided to leave these images in the dataset to preserve this useful, if potentially confusing, information.



**4. EDA**

**5. Algorithms & Machine Learning**

**6. Predictions**

**7. Future Improvements**

**8. Credits**