



*AI Project: BiteCheck Food Classification Dataset Documentation*

**Group 12:**

Eyra Inez Anne-Marie Agbenu (47152026), PraiseGod Ukwuoma Osiagor (63962026),

Rose Carlene Mpawenayo (49262027), Kelly Kethia Gacuti (32742027)

CS254\_B: Introduction to Artificial Intelligence

Dennis Asamoah Owusu

April 29, 2025

## **BiteCheck Food Classification Dataset Documentation**

For our BiteCheck project, we developed a reliable machine learning model that classifies food images as healthy or unhealthy.

### **Data Source and Collection Methods:**

We used the Food101[\[2\]](#) dataset from Kaggle, containing 101 food categories with 1,000 images each. Then, we selected 15 categories most common at Ashesi University, maintaining the 1,000 images per class. Then we organised these selected folders into a new dataset directory which served as our working dataset for the model development. To explore and validate the directory structure, we used Python's `os.walk()` function to traverse all directories and subdirectories.

### **Preprocessing Steps:**

To prepare the dataset for training we applied several key preprocessing techniques which were directory structuring and splitting. Here, the dataset was organised into three subsets: training (70%), validation (15%) and testing (15%). The next technique was image validation and cleaning, in which we used `Image.open()` from the Pillow library [\[1\]](#) to help identify and exclude corrupted files (images), if any, from the dataset. Lastly, we performed Image standardization where all the images were resized to 224x224 pixels, which is the input requirement for ResNet50 architecture.

### **Visualizations:**

To communicate the dataset properly, we created a training history plot which showcased the training vs validation accuracy, training vs validation loss and a clear marking of when the fine-tuning began. Furthermore, we generated a confusion matrix to visualize class-wise performance and, lastly, we used the seaborn module for better visualization of our images.

## References:

- [1] Alex Clark, et al. 2023. Pillow: Python Imaging Library (PIL Fork). <https://python-pillow.org>
  
- [2] Lukas Bossard, Matthieu Guillaumin, and Luc Van Gool. 2014. Food-101 – Mining discriminative components with random forests. In European Conference on Computer Vision (ECCV '14). Springer, Cham, 446-461. Retrieved from <https://www.kaggle.com/datasets/dansbecker/food-101>