# CS4442 - Final Project Report

# Fruit Classifier AI

Adrian Gramatke, 251069448

Mete Isiksalan, 250932899

Timofey Hartanovich, 251076987

Saif Uddin, 251105854

# Introduction

## Abstract

This project aims to create a machine learning model that can reliably categorize images of a fruit when given an image.

For our final project, we aim to develop a sophisticated Artificial Intelligence (AI) system that can analyze an image and accurately identify if there is a fruit present, and if so, determine the specific type of fruit. To accomplish this task, our AI will use a two-step approach:

- Firstly, we plan to extensively train our AI on a diverse collection of 131 distinct fruit and vegetable types. This training will enable the AI to recognize and classify a wide variety of fruits and vegetables accurately.
- Secondly, we will kernelize this model and apply it to a process to identify fruits and vegetables in other images.

With this two-pronged approach, we will be able to define different models to be used as kernels or pass multiple models that will be trained to detect specific features in images. By creating the model separately, we can test it against its own features. And attaching the model to another network becomes easy.

Our proposed AI system will require a significant number of computational resources, data, and extensive training to achieve optimal results. However, the potential applications of such a system are vast, including automation in agriculture, monitoring of fruit production, and quality control in the food industry. AI is always improving and by creating modular components that can perform specific object detection we will be able to chain the various classifiers together. Our project is meant to explore the application of creating filters from convolutional networks to apply to various images.

## Literary Review

A customized Convolutional Neural Network (CNN) architecture is the foundation of the approach presented for classifying fruit and vegetable images. This strategy is consistent with prior research, which has shown that adopting customized CNN architectures is successful for a range of classification applications, including fruits and vegetables.

The CNN architecture developed consists of two 3x3 Conv layers, followed by a 5x5 Conv layer, all with ReLu activation and normalization. In addition, the network uses pooling, two completely linked layers, and a third 3x3 Conv layer. A sigmoid function is used to normalize the output, and stochastic gradient descent with cross entropy loss as the loss function is used to train the model. The model was able to discriminate between 131 distinct fruits and vegetables with an accuracy of 92%.

The results section emphasizes two primary issues: a lack of data and problems applying the kernel to abnormalized images, which were experienced while trying to apply the model to additional picture sources. These difficulties highlight the requirement for more analysis and testing in the area.

For instance, a bigger and more varied data set, maybe comprising photos with different levels of noise and variation, could be used to overcome the problem of a lack of data. Moreover, investigating object identification methods, such as those based on edge detection, might be used to tackle the difficulty of applying the kernel and creating precise bounding boxes.

In conclusion, the suggested approach for classifying fruit and vegetable images utilizing a unique CNN architecture is consistent with previously published work in the area. The difficulties observed when using the model with different picture sources, however, point to interesting directions for further research, such as looking into object recognition strategies and creating more durable models for precise classification and localization.

## Methodology

To categorize photos of fruits and vegetables at the outset of the study, a convolutional network was created. For successful learning and feature extraction from the input photos, the network design has many layers. A 3x3 Conv layer and a 5x5 Conv layer were the initial two layers. The performance and generalization capacities of the model were enhanced by ReLu activation and normalizing in each of these three layers.

The 5x5 layer was pooled using a 2x2 matrix with a stride of 2 to lower the spatial dimensions and capture more abstract characteristics after the initial three convolutional layers. To further hone the learnt characteristics, a second 3x3 Conv layer was added. Two completely linked layers at the bottom of the network merged the retrieved characteristics to determine the final classification.

A sigmoid function was used in the output layer to normalize the output and show how definite the classification was. One of 131 potential types of fruits and vegetables was predicted by the program. To evaluate the model's performance during training, Stochastic Gradient Descent was utilized as the optimizer and Cross Entropy Loss as the loss function.

The model has a remarkable accuracy of 92% after 15 training iterations. The model was put to the test on a Google picture to gauge how well it performed on actual images. Unfortunately, max pooling early in the model did not offer much benefit because the pictures in the dataset were rescaled. Before using the model on fresh photos, scaling would be necessary to take into consideration size and aspect ratio variations.
(Scaling would have to happen before applying the model).

## Results

There were two major problems we encountered that prevented us from being able to apply our model to other sources:

1. A lack of data. Although the library was adequate for training a classifier for its own images with a high accuracy - Originally when creating the model, we neglected to add the sigmoid function for the outputs and the model was able to achieve 92% accuracy. However, currently with the sigmoid function we have been able to achieve 92%. The bigger issue is for the application of image detection, there is not enough noise or variance in the images to apply to other image sources. We also did not have a separate labeled library to test the images against.
2. Applying the kernel. To test the image detection, we passed the kernel over a couple of images found in google images and output labels that had over 92% confidence. The issue with the process was the lack of a bounding box detector for the images. This technology is still being researched and it will be interesting to further experiment with the project in the future by implementing max bounding boxes that scale the inputs to be passed to the model. By not rescaling the images, when passing the kernel over it sometimes a fruit may be bigger or smaller so results are not very accurate. To counteract this, another possibility is to scale the image into multiple different sizes by pooling and then pass the kernel over the various scaled versions. This, however, is computationally a lot more demanding than segmenting the image into various bounded boxes.

# Conclusion

Our model distinguished 131 distinct fruits and vegetables with a stunning 92% accuracy. Nevertheless, we found that the classification accuracy dramatically decreased when we tried to use the model on a variety of unlabeled photos. This restriction recalled early attempts at image identification when random elements of a picture may be incorrectly identified.

To solve this problem, a future project may construct a model that can recognize objects and generate bounding boxes using edge detection. This would provide the model with the ability to concentrate on particular interest areas, possibly producing more accurate classifications. The absence of an appropriate data set for evaluating this strategy was one of the difficulties encountered in the current endeavor.

We are certain that we can provide more solid and dependable results in subsequent iterations by creating a model that can automatically pre-process and scale photos before sending them to the classifier. As a result, our fruit classification model would be able to correctly identify fruits in a larger variety of photos, including those that are not pre-scaled and well defined.