Diego Silva
Kayla Malloy

CSEN 140 Machine Learning
and Data Mining

December 6, 2025

Assessing Fruit Quality with Computer Vision

Background and Motivation

Problem Domain

Fruit Quality Classification

Supply Chain Quality Control

Image Classification

Convolutional Neural Networks

Significance

Automated Quality Control in Supply Chain

Applicable for Multiple Portions of Supply Chain

Manual Sorting/Checking is Slow, Inconsistent, and Costly

Public Health, Manufacturing Costs, and Agricultural Productivity Benefits

Project Objective and Research Question

Primary Goal

Build an efficient and accurate image classification model capable of identifying fruit types and conditions (fresh vs. rotten vs. formalin-mixed) from images.

Research Question

Can we correctly identify the quality of a fruit based off an image in order to efficiently run quality control in the supply chain?

How accurately can pretrained image models classify fruit images across 15 categories?

Data Source and Dataset Details

"FruitVision: A Benchmark Dataset for Fresh, Rotten, and Formalin-mixed Fruit Detection", Mendeley Data

Originally contained:

10,154

images

→

Augmented to:

73,407

images

70/15/15

split for train/validate/test

- 15 classes:
  - 5 fruit types: Apple, Banana, Mango, Orange, Grape
  - 3 categories: Fresh, Rotten, Formalin-Mixed
- Augmentation Techniques: rotation (i.e., 45, 60, 90 degree), horizontal flipping, zooming, brightness adjustment, shearing, and Gaussian noise addition

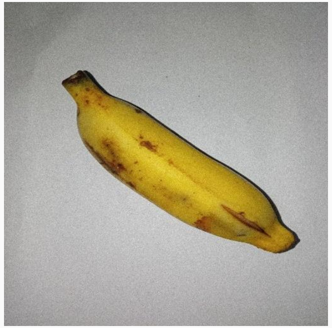
Sample Images From Our Dataset

Fresh Banana

Formalin-Mixed Mango

Fresh Apple

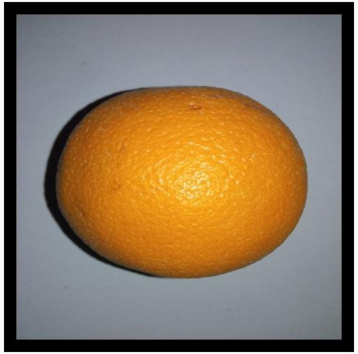
Fresh Grape



Formalin-Mixed Banana

Rotten Mango

Rotten Apple



Formalin-Mixed Orange

Method and Procedure

Algorithms and Techniques	• Pre-trained CNNs: ResNet50 and EfficientNetB0 • Transfer Learning and further Data Augmentation • Hyperparameter tuning
Experimental Design	• As mentioned, 70/15/15 train/validation/test ratio • Validation monitoring • Early stopping • Freezing layers
Tools, Libraries, and Frameworks	• TensorFlow • Matplotlib & Scikit-learn • Keras Tuner & Hyperband • NumPy
Computational Resources	• We used the T4 High-RAM GPU • Reduced image size, fixed epoch size • Data prefetch and pipelining

Experimentation and Implementation

Baseline Model (no pretrained weights)

Efficiently Loading Image Data (image_dataset_from_directory)

Normalizing, Prefetching & Caching

ResNet50: Global Average Pooling, Dropout, Dense

Transfer Learning

Efficiently Loading Image Data 224×224 images

EfficientNetB0 with Pre-Trained Weights

Freezing/Unfreezing Layers

Hyperparameter Tuning

Utilized Keras Tuner and Hyperband for optimization

Early stopping to stabilize hyperband, reduce runtime, and prevent overfitting

Tuning learning rate, dropout, and freezed layers

Final Model

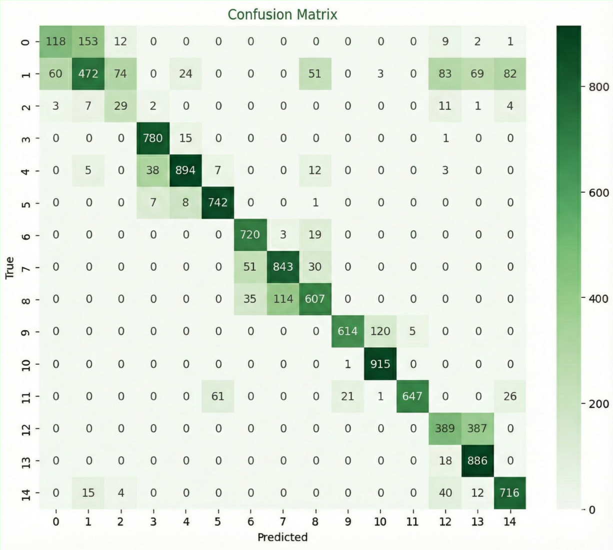
Efficiently Loading Image Data

Prefetching, Image Augmentation

EfficientNetB0: Using Parameters from Hyperparameter Tuning Model

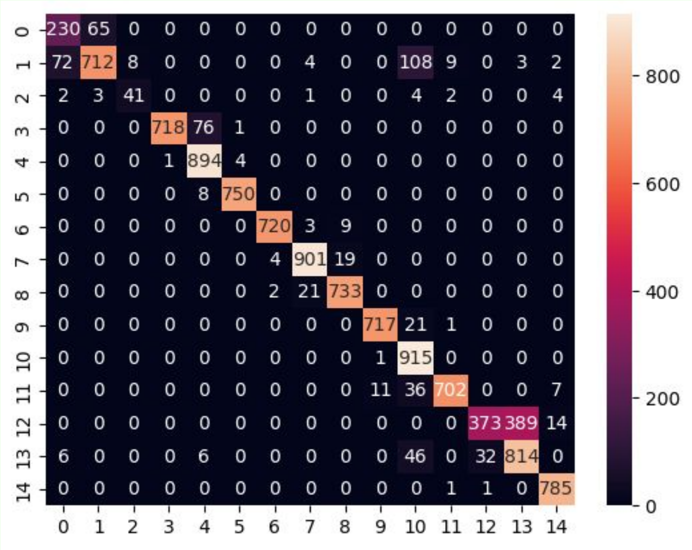
Results and Evaluation

	Accuracy	Precision	Recall	F1-Score
Baseline Model	0.84	0.81	0.80	0.79
Transfer Learning	0.83	0.84	0.79	0.78
Final Model	0.91	0.91	0.89	0.89



Baseline Model

Final Model



Misclassified Images

Conclusion and Reflection

Key Findings:
- CNN creation process
- The Power of Fine-Tuning

What Worked Well:
- Data Pipeline Optimization
- Error Analysis

Limitations:
- Class Confusion Persistence
- Time & Memory Constraints

Future Work:
- Grad-CAM Explainability
- Targeted Data Augmentation

Implications:
- Automated Quality Control
- Scalability

91% Accuracy

Automated Supply Chain Quality Control

References

→ Bijoy, et al. (2025). *FruitVision: A Benchmark Dataset for Fresh, Rotten, and Formalin-mixed Fruit Detection.* Science Direct

→ Mehta et al. (2021). *Fruit Quality Analysis using modern Computer Vision Methodologies.* IEEE

→ FruitVision: A Benchmark Dataset for Fresh, Rotten, and Formalin-mixed Fruit Detection, Mendeley Data
https://data.mendeley.com/datasets/xkbjx8959c/2

→ TensorFlow Documentation

→ Scikit-learn Documentation

→ Chat GPT & Gemini used for planning and memory-usage guidance

Thank you!