Thapar Institute of Engineering and Technology

Department of Computer Science

# A Generative Approach to Fruit Freshness Classification

Saiyam Singhi(102216108) Navish Goyal(102216051)

*Lab Instructor:* Mrs.Priya Raina

A report submitted in partial fulfilment of the requirements for the degree of Bachelor of Engineering (B.E.) in *Computer Science at Thapar Institute of Engineering and Technology (TIET)*

November 16, 2025

# Abstract

The detection of food freshness is considered an important aspect of agricultural practices, retail quality assurance, supply chain management, and household food security. Conventional machine learning systems are able to determine the current status of a food product, but they are not capable of predicting the item's changes over time, and they also have limitations in providing details about the ripening process.

This project proposes Generative AI-powered system that performs:

- Ripeness classification
- Future freshness prediction
- Ripening timeline simulation
- Creative fruit stylization
- Interactive natural-language reasoning

In order to keep scientific control and guarantee trustable generative behavior, the system at first got the training on one fruit category that had a distinct and visually observable ripening cycle. After this reasoning is established, the project particularly models the ripeness of bananas as there are very clear color changes in bananas and they are easy to work with for image-to-image generation tasks.

Not like the conventional industry data sets, which categorize the freshness into two classes of fresh and rotten, this research has considered four ripeness stages—unripe, ripe, overripe, and rotten. Such an approach allows more realistic modeling, better predictions, and smoother generative transitions.

The components of the system are as follows:

- MobileNet-V2 classifier that predicts the ripeness
- WGAN-GP that enlarges the dataset
- Stable Diffusion + ControlNet that generates the future images
- Phi-3 LLM that provides the chatbot and intent classification
- A new Streamlit interface that enables users' multimodal interaction

The application at last provides the users with the opportunity to determine the current state of the fruit, see how it will change after some days, create a ripening timeline, choose and have the fruit stylized in different artistic themes, and inquire in plain language about the fruit's edibility and nutrition.

# Contents

# Chapter 1

# Introduction

The strongest pointers of the fruit's freshness are the visual cues like the color transitions, bruising, and the changes in texture. Humans process these cues intuitively, but teaching robots to do the same is difficult because of lighting changes, biological differences, and the fact that ripening is a very subtle and delicate process. The conventional computer vision models usually conduct static classification, stating only the condition of the fruit as fresh or spoiled at that moment. But, the freshness in real life is a dynamic biological process, and the static models are unable to interpret the future course of a fruit's evolution.

The advancements in generative modeling over the last few years—especially with that of diffusion models like Stable Diffusion—have been a great help in simulating the gradual visual changes, for example, aging, drying, rotting, or ripening. One of the advantages of these models is that they can not only classify the current state of biological objects but also predict their future appearance. These predictions can be leveraged in various areas such as retail food monitoring, automated quality inspection, supply chain management, and consumer home decision-making. The project aims at controlled experimentation and reliable generative behavior, therefore it starts with one fruit category that has a stable and visually distinct biological pattern. After establishing that, bananas are chosen for the rest of the work because their ripening is very clear, progressive, and easy to observe in the case of a very good model. Bananas make a good candidate for modeling fine-grained ripening stages and realistic image-to-image transformations because they change gradually from green to yellow to brown.

The study investigates the collaboration between discriminative models (classifiers) and generative models (diffusion-based predictors) to dynamically model ripeness. In particular, the research focuses on the core question:

"Can generative AI accurately simulate fruit ripening progression while being grounded by classifier predictions?"

To explore this , the system unifies:

- A classifier based on MobileNet-V2 to give a verdict on the present ripeness
- The WGAN-GP method of augmentation for the purpose of better training and improved performance
- The combined use of Stable Diffusion and ControlNet to create life-like future states

- A Phi-3 language model for understanding and processing of natural language queries
- A multimodal interface in Streamlit for prediction and visualization in real-time

The initiative, which combines interpretability, predictive modeling, and interactive generative AI, showcases the capability of modern AI to not only classify but also to simulate biologically meaningful transformations that are highly realistic visually.

# Chapter 2

# Literature Review

The detection of fruit freshness has been done traditionally through computer vision, deep learning, IoT sensor systems, and most recently, generative AI. Each method has its advantages and disadvantages, access to a lot of data being one of the most significant limitations, especially concerning the modeling of fine-grained ripening and predicting future freshness. In this chapter, we will go through the literature that is already available and point out the gaps that make the project necessary.

## 2.1 Fruit Classification Using Deep Learning

The application of deep learning to fruit recognition is thus far, the best where Convolutional Neural Networks (CNNs) have gained very high precision in areas such as detecting different kinds of fruits, showing the presence of defects, and classifying the freshness of fruits . Models like **MobileNet** and **ResNet** are widely utilized because they offer a good compromise in terms of power consumption and the quality of the features that are extracted.

- **MobileNet** is very efficient in terms of mobile and edge usage and hence is fitting for lightweight applications.
- **ResNet** uses its residual connections to drive deeper architectures and facilitate stronger feature learning.

It is true that the existing literature provides a lot of support for the classification of fruits into **fruit category** and **binary freshness levels**, but more often than not, the researchers keep the freshness as a simple categorization of:

- Fresh
- Rotten

This two-part structure does not include gradations of fruit ripening between *unripe*, *early-ripe*, and *overripe*. The number of studies on the topic of **multi-stage ripeness classification** is scant, which means the research gap is quite large and that is exactly what this project intends to cover.

## 2.2 IoT and Sensor-Based Freshness Detection

Other than visual methods, among other technologies, the **IoT sensors**, the **chemical gas sensors**, and the **electronic noses** have been introduced in several studies as major contributors to spoilage detection due to their capabilities to measure ethylene gas, pH, temperature, and humidity.

This type of technology offers accurate chemical data but it also has some important drawbacks:

- Very high costs for installation and regular servicing
- Not many consumers can benefit from this due to limited scalability
- Total reliance on hardware
- Much less portability when compared to camera-based systems

Thus, sensor systems are effective in controlled environments but they are not suitable for retail stores, households, and consumer-facing freshness applications. This, in turn, emphasizes the importance of **vision-based freshness detection**, where only a camera is needed.

## 2.3 Generative Data Augmentation in Computer Vision

A major issue in detecting fruit freshness is the **shortage of diverse, annotated datasets**, especially for the very detailed ripeness stages. Generative models like **Generative Adversarial Networks (GANs)** and **Diffusion Models** have started to be the main approach to deal with the shortage of datasets *(Author, Year)*.

- **GANs**, particularly ones like **WGAN-GP**, can produce very good quality synthetic images by mastering the distributions of real data.
- **Diffusion Models**, for example, **Stable Diffusion** not only are able to create very realistic variations and transformations but also, thus, they are appropriate for ripening simulation and other similar tasks.

Although generative augmentation is already established in certain fields, such as **medical imaging**, **face aging**, and **object detection**, its application in **fruit freshness classification** is still very limited. A very small number of studies consider using generative models for simulating biological processes such as **ripening progression**, which is a gap that this project aims to fill.

## 2.4 Market Applications and Gaps

In spite of the fact that the most widely used commercial tools are very powerful regarding visual comprehension they still do not deal with the issue of fruit freshness:

- **Plantix** is known for its expertise in diagnosis of plant diseases.
- **Google Lens** provides an object recognition service that is not limited to any particular area.

Nevertheless, the three mentioned tools are not engaged in **fruit freshness prediction**, **ripeness time forecasting**, or **future appearance generation**. Therefore, we can say that there is a great market potential for:

- A consumer-friendly freshness prediction tool
- Visual transformation of ripening stages
- Multi-stage classification rather than simple fresh/rotten output
- Generative simulations for household and retail use

This indicates that the establishment of a system that is AI-driven, accessible and capable of **fine-grained ripeness detection** and **future-state prediction** is necessary. Such a system is exactly what this project intends to deliver..

# Chapter 3

# Methodology

The whole methodology is clarified in this chapter which was employed to design, instruct, and make the Generative AI–based fruit freshness prediction system operational. The method integrates traditional augmentation, deep learning classification, generative adversarial modeling, diffusion-based transformation, and natural-language interaction within a common framework.

## 3.1 System Architecture Overview

The system integrates four major components:

1. **MobileNet-V2 classifier** – identifies the current freshness stage.
2. **Preliminary data augmentation + WGAN-GP** – increases dataset diversity.
3. **Stable Diffusion + ControlNet** – predicts biological ripening progression.
4. **Phi-3 Mini LLM** – interprets natural language queries.

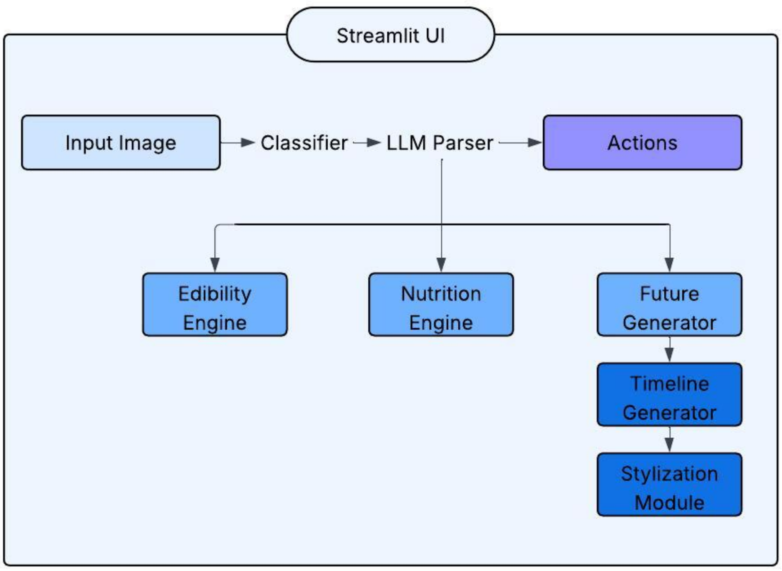



Figure 3.1: System Architecture Diagram

## 3.2 Dataset Collection and Preprocessing

To maintain modeling consistency, the system focuses on a **single fruit category** with a visually clear and predictable ripening cycle. After establishing this modeling constraint, bananas were selected due to their stable, well-understood progression from green → yellow → spotted → dark.

### 3.2.1 Ripeness Levels

Most research uses only two classes:

- Fresh
- Rotten

This project expands these to **four fine-grained ripeness levels**:

1. Unripe
2. Ripe
3. Overripe
4. Rotten

This richer structure enables smoother generative transitions.

**Unripe** **Ripe** **Overripe** **Rotten**

Figure 3.2: Sample images representing the four ripeness classes in the dataset.

### 3.2.2 Traditional Data Augmentation (Before GAN)

The dataset was **first enlarged using the classical augmentation techniques** before the GAN-based augmentation was applied. This procedure contributes to increasing data variability and eliminating the issue of overfitting.

The following transformations were performed:

- Rotation (±20°)
- Horizontal & vertical flips
- Random brightness, contrast
- Gaussian noise

- Random cropping + resizing

**Importance of this:** Standard augmentation enhances the basic dataset *prior* to the application of GANs. This keeps the GANs from merely duplicating the limited samples and thus securing a more consistent adversarial training.

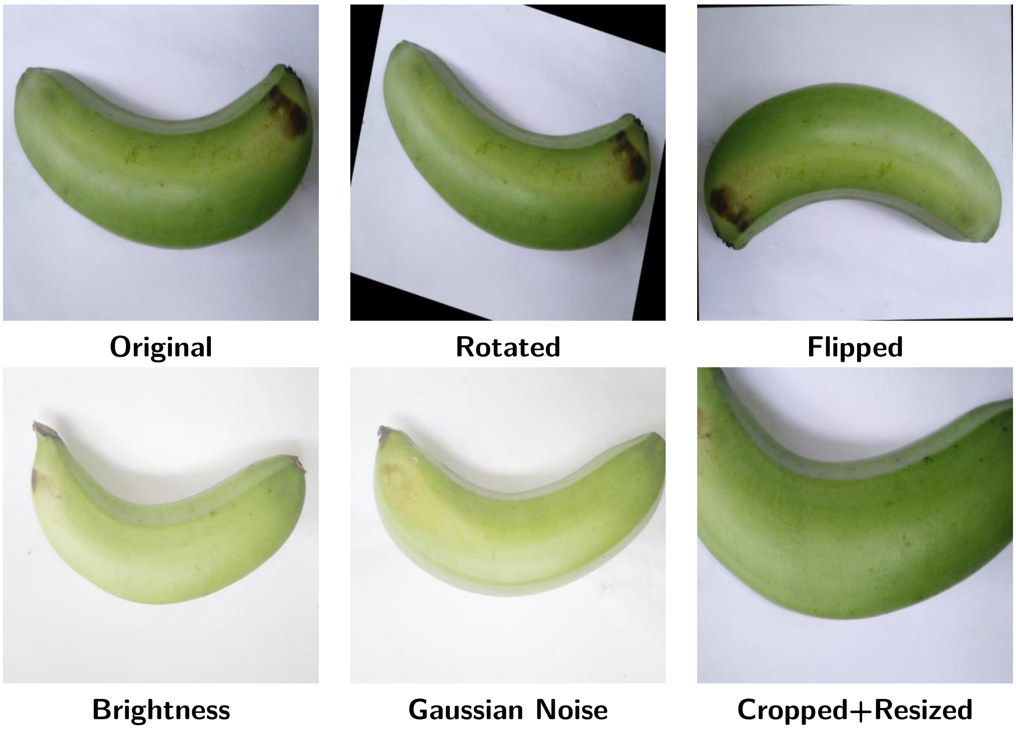



Figure 3.3: Traditional augmentation samples used to expand the dataset prior to GAN-based augmentation.

## 3.3 WGAN-GP Data Augmentation (with Traditional Augmentation Before GAN)

Initially, the dataset was expanded using classical augmentation methods such as rotations, flips, and brightness changes, plus Gaussian noise and random cropping as GAN inputs during training. This way, GAN was not only getting different samples but also it would not just memorize the limited pattern.

After applying the traditional augmentation techniques, the dataset still suffered from class imbalance, especially for the overripe and rotten categories. To make the dataset even more diverse, a Wasserstein GAN with Gradient Penalty (WGAN-GP) was used to produce fruit-ripeness images that were not real but synthetic.

#### Why WGAN-GP?

- More stable than vanilla GAN
- Handles image texture variation better
- Reduces mode collapse

- Provides smoother generation quality

The generating samples of WGAN-GP were mixed with the real dataset that the classifier was trained on, with an aim of making the classifier more robust.

### Computational Limitation

Training a GAN is very resource demanding in terms of computation power. Because of the limited GPU and hardware resources:

- The duration of GAN training was not long
- The generated images were useful but not extremely high quality
- The goal was to balance the dataset rather than to create photorealistic images

Regardless of this, the GAN-aided augmentation still managed to boost the classifier and reduce overfitting.

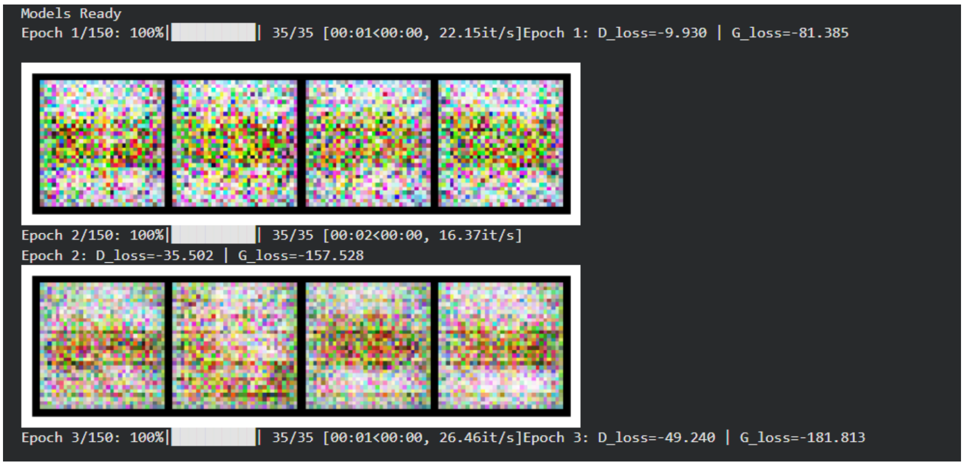



Figure 3.4: Initial WGAN-GP Outputs during early epochs (1–3).

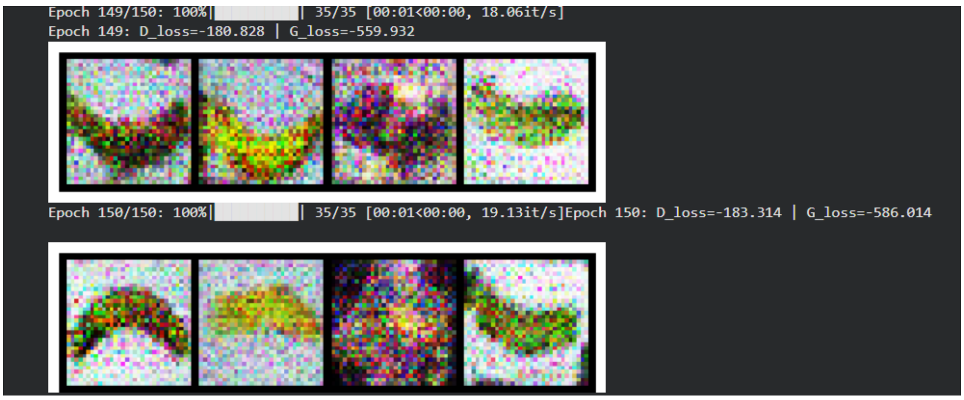



Figure 3.5: Final WGAN-GP Outputs after convergence (Epochs 149–150).

## 3.4 Classification Model – MobileNet-V2

The model MobileNet-V2 is chosen owing to the following reasons:

- It is very light and quick.
- It has great power in classifying images correctly.
- It is ready for Streamlit deployment.
- It withholds its strength even when the data is meager.

**Training Details**

- Epochs: 10–20
- Learning Rate: $1 \times 10^{-4}$
- Batch size: 32
- Optimizer: Adam
- Loss: Cross-entropy

The classifier achieved **87% accuracy**, which is very good for a **4-class, single-category fruit freshness** problem.

### 3.4.1 Classification Performance

The MobileNet-V2 classifier was evaluated on the test set using precision, recall, F1-score, and overall accuracy. The results for each ripeness class are shown in Table 3.1.

Table 3.1: Classification report for the four-class fruit ripeness classifier.

| Class | Precision | Recall | F1-score | Support |
|---|---|---|---|---|
| Overripe | 0.89 | 0.83 | 0.86 | 164 |
| Ripe | 0.87 | 0.88 | 0.87 | 184 |
| Rotten | 0.83 | 0.84 | 0.84 | 221 |
| Unripe | 0.91 | 0.94 | 0.92 | 133 |
| **Accuracy** | | 0.87 (702 samples) | | |
| **Macro Avg** | 0.87 | 0.87 | 0.87 | 702 |
| **Weighted Avg** | 0.87 | 0.87 | 0.87 | 702 |

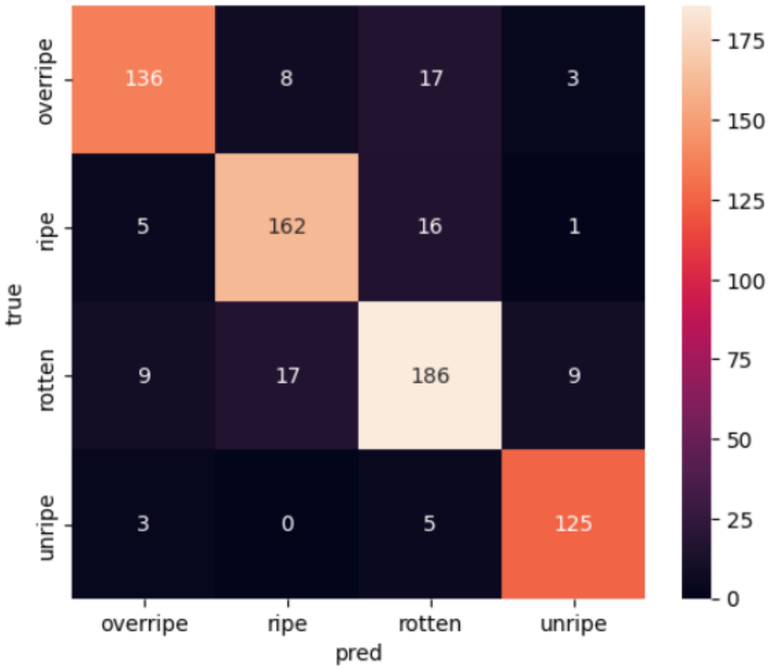
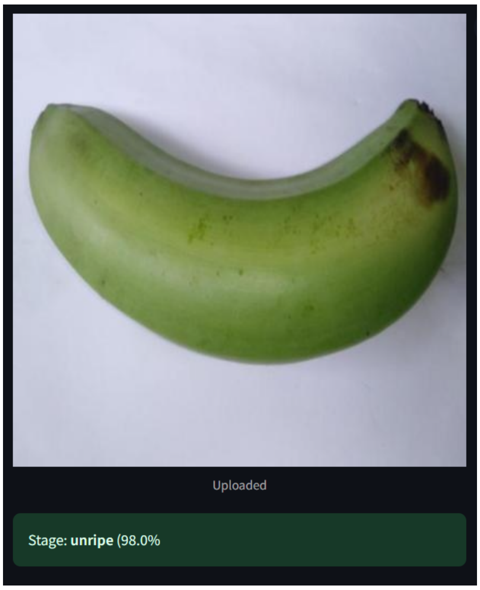
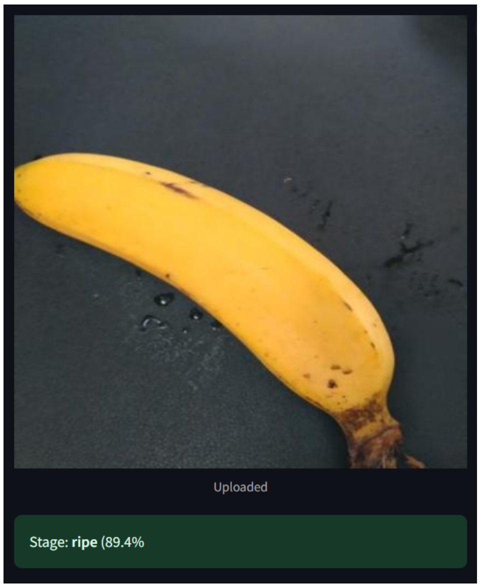



Figure 3.6: Confusion matrix showing classification performance across the four ripeness levels (Overripe, Ripe, Rotten, Unripe). The model performs strongest on **unripe** and **ripe** classes, while some confusion occurs between visually similar **overripe** and **rotten** stages.

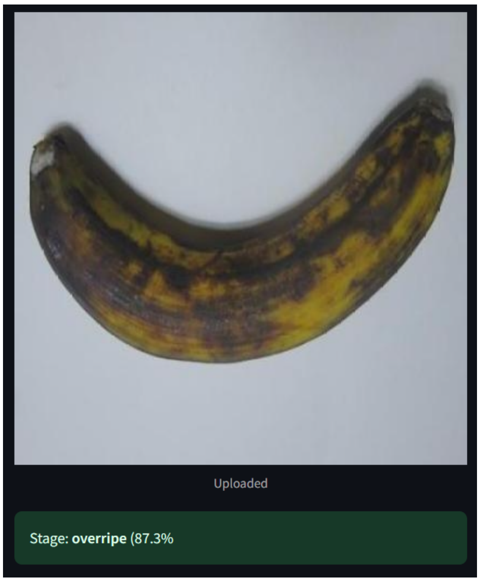
### 3.4.1 Example Classification Outputs

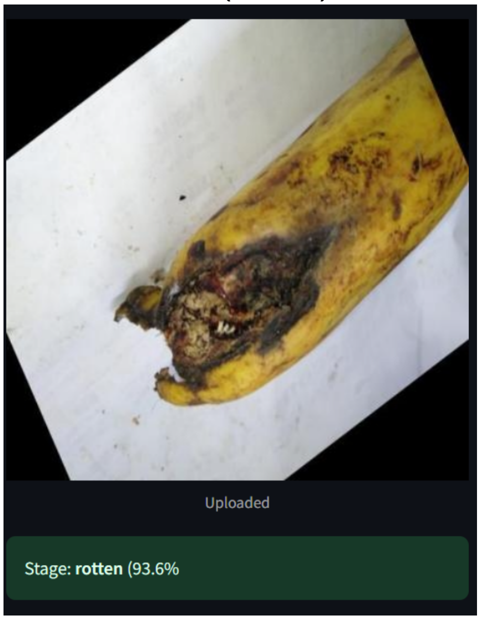
The MobileNet-V2 classifier's ability to classify images is demonstrated by the provided images with predictions made on fresh fruit. A little detail of the uploaded fruit photo and the considered maturity phase with an associated confidence score is provided for each output. The four stages of fruit ripeness used in this study namely; *unripe*, *ripe*, *overripe*, and *rotten*, are all represented in these examples.




**Unripe (98.0%)** **Ripe (89.4%)**




**Overripe (87.3%)** **Rotten (93.6%)**

Figure 3.7: Classifier predictions for the four ripeness categories using the MobileNet-V2 model.

## 3.5 Predictive Image Generation – Stable Diffusion + ControlNet

It is quite a challenge to guess a fruit's appearance after several days because ripening is a **continuous biological process**, not a transformation from one class to another. The model built upon **Stable Diffusion v1.5** along with **ControlNet (Canny conditioning)** is responsible for the visual representation of such a transformation.

### Why Stable Diffusion?

Stable Diffusion is a cutting-edge diffusion model that not only generates top-notch but also very realistic images preserving minute details. Diffusion models, in contrast to GANs, are:

- more reliable and further stable in the course of inference,
- able to create texture changes on par with nature,
- the easiest to use along with the text prompts,
- the best choice for gradual transformation (smooth ripening transitions).

Thus, these characteristics are the ones that make diffusion the most suitable to simulate the gradual change of ripeness.

### How the Future Prediction Works

The machine does not characterize by unitary images random emission. It performs a **controlled transformation pipeline**:

**Step 1 — Determine Ripeness Progression (%)** The ripeness progression is represented by a base ripeness percentage:

| Stage | Ripeness % |
|---|---|
| Unripe | 0% |
| Early-ripe | 30% |
| Ripe | 50% |
| Overripe | 75% |
| Rotten | 100% |

Table 3.2: Ripeness percentage mapping for transformation control.

Days are gradually increased in a way that corresponds to about 5% per day, thus preventing stage jumps and allowing for a better transition through
**smooth interpolation**.

**Step 2 — Build a Context-Aware Prompt** A detailed prompt is constructed, describing:

- texture evolution,
- skin color progression,
- keeping the original shape untouched,
- keeping the background unchanged,
- exaggerating decay only if fruit becomes rotten.

**Example prompt:**

> "Ultra-realistic banana at 65% ripeness. Appearance: darker yellow skin, brown freckles, softening texture. KEEP EXACT SHAPE AND BACKGROUND. Only modify skin color and ripeness indicators."

These descriptive cues ensure the model modifies **only the fruit**, not the whole scene.

**Step 3 — ControlNet Canny Conditioning** To stop the diffusion model from altering the fruit's:

- shape,
- pose,
- edges,
- surrounding,

**ControlNet (Canny edges)** is used. This guarantees:

- The banana's shape remains intact
- Only the texture and color change
- Realism of the output is preserved

In the absence of ControlNet, the artwork may be altered, the fruit might be incorrectly shaped, or unwanted parts could be generated by the model.

**Step 4 — Image-to-Image Diffusion** Stable Diffusion takes the following:

- the original fruit image,
- an edge map created using Canny,
- the ripeness condition,
- a value for the transformation strength.

The strength parameter controls:

- **Low strength**: little change $\rightarrow$ subtle ripening
- **High strength**: large change $\rightarrow$ rotten transformation

The selected value ($\sim 0.55$) results in a **well-matched, realistic shift**.

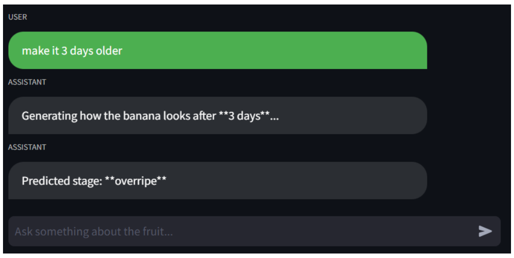
**Step 5 — Output Stage Label** The process yields:

1. the artificially created future image,
2. the imagined ripeness phase (unripe $\rightarrow$ early-ripe $\rightarrow$ ripe $\rightarrow$ overripe $\rightarrow$ rotten) name.

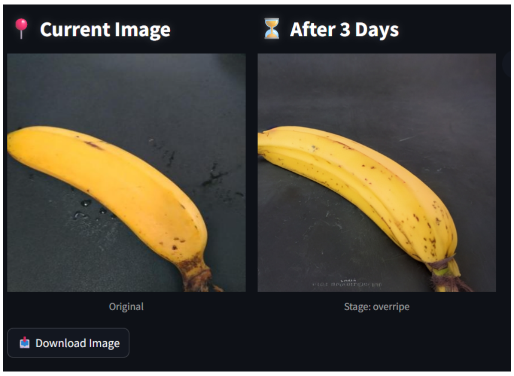
Thus, it does not present the user with raw numerical percentages but rather offers a clean semantic stage label.

### Why This Method Is Effective

- Smooth, natural progression is provided, not abrupt class jumps
- ControlNet avoids distortion of the shape
- Ripeness textures that are photorealistic are produced
- Forecasting for several days is supported (1 day, 3 days, 10 days, etc.)
- It works for timeline generation, not only for single-day prediction

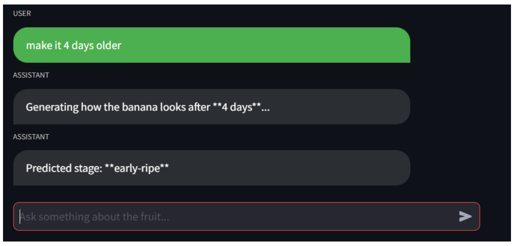



**User Prompt**

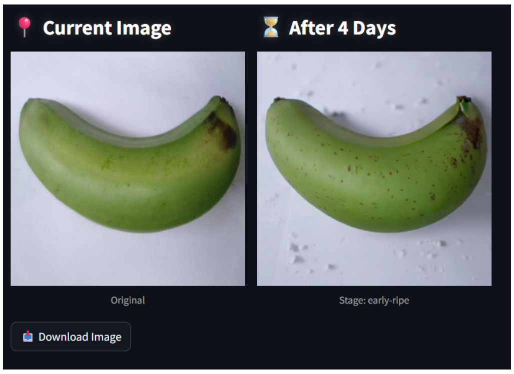



**Generated Output (After 3 Days)**

Figure 3.8: Example of interactive future-ripeness prediction. The user requests: "make it 3 days older," and the system generates an **overripe** prediction using Stable Diffusion + ControlNet.




**User Prompt**




**Generated Output (After 4 Days)**

Figure 3.9: Prediction example where the user requests "*make it 4 days older*." The system predicts the stage as **early-ripe** and generates a smooth biological transformation.

## 3.6 Ripening Timeline Simulation

One of the components that is most revolutionary within the whole system is the **Ripening Timeline Generator**, which not only predicts visually the fruit for a user-defined number of future days but also creates a future scenario. The system, in contrast to one future prediction, generates a **series of images** that depict the entire ripening process.

If the user inputs something like "*timeline for 10 days*", the system:

1. **Extracts the timeline duration** through the rule-based parser or the Phi-3 LLM.
2. **Creates a set of evenly spaced points in time** from the Day 0 up to the Day $N$.
3. **Samples the timeline in 2-day intervals automatically** to keep it both effective and biologically smooth.

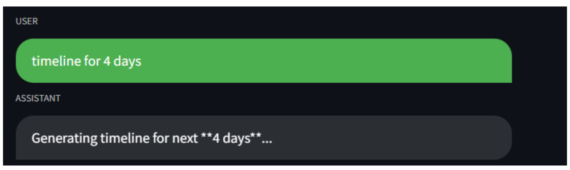
   For instance, if the user wants a timeline of 6 days, the generated sequence will be:
   **Day 0** $\rightarrow$ **Day 2** $\rightarrow$ **Day 4** $\rightarrow$ **Day 6**
   This way, the timeline is kept compact since no redundant consecutive frames are generated.

4. **Takes the present ripeness stage as the starting point**, allowing for realistic transitions without sudden jumps.
5. **Produces future frames** via Stable Diffusion + ControlNet while ensuring the fruit's contour and background are the same.
6. **Chronologically sorts and displays all the rendered images** in a neat multi-column grid.

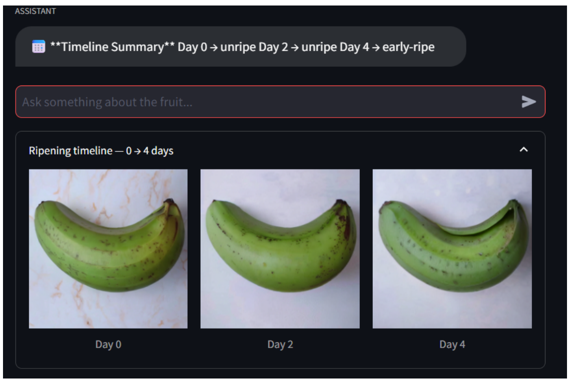
The end result is a brief **ripening time-lapse** like output, where:

- The very beginning of the frames reveals a slow process of yellowing and softening.
- The middle frames are responsible for the appearance of freckles and the changes of the warm tones.
- The last frames reveal a darkening that affects the fruit significantly, bruising and eventually rotting.

To each frame a caption correlating to the **predicted day** and the appropriate **ripeness stage** (unripe, ripe, overripe, rotten) is attached. The timeline is placed in an expandable UI for better understanding and easier movement.

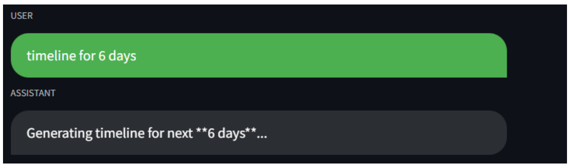



**User Prompt**

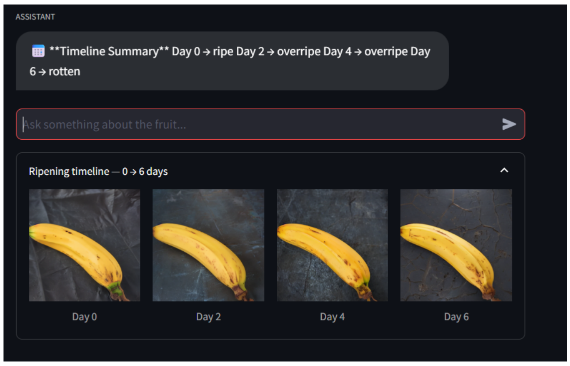



**Generated Output (for 4 Days)**

Figure 3.10: Example of future-ripeness timeline generation. The user requests "timeline for 4 days", and the system produces images for Day 0, Day 2,and Day 4 using Stable Diffusion + ControlNet.




User Prompt




Generated Output (for 6 Days)

## 3.7 Natural Language Interface – Phi-3 Mini LLM

The integration of a natural language interface with the Phi-3 Mini (4K Instruct) large language model not only makes the system more user-friendly but also more accessible. Rather than requiring buttons or strict commands, the system can be accessed by users via everyday conversational English, such as:

- "What will it look like after 4 days?"
- "Create the ripening timeline."
- "Which stage is this fruit in?"
- "Can we eat it?"

The LLM interprets such free-form questions and converts them into system actions in a structured way. It detects, which text is it in, what does the user want, and e.g. sends a JSON object with two main parts:

- **intent** — the action requested by the user (e.g., `future_image`, `nutrition`, `status`, `timeline`, `edibility`)
- **days** — a numerical value extracted from the prompt if the query involves time (e.g., "after 4 days" → `days = 4`)

For illustration, an average output from Phi-3 looks like:

```
{
  "intent": "future_image",
  "days": 4
}
```

This JSON output functions as a command signal for the system. Once the intent is determined, the application automatically routes the request to the corresponding module:

- **Future Generator** — for expected next look
- **Timeline Generator** — for ripening forecasts over several days

- **Nutrition Engine** — for insights related to diet and health
- **Edibility Engine** — for evaluation of food safety
- **Classifier Output** — if the user inquire about the current state of ripeness

The conversational layer changes the tool into an interactive assistant. Technically, users are not required; all major system features can be operated with simple natural sentences.

### Fallback Rule-Based Intent Parser

Running the Phi-3 LLM may not always be an option due to the demand for the resources. To overcome this limitation, a lightweight and advance regex-based rule parser is provided as a fallback. It recognizes the same four main intents that the system relies on:

- `future_image`
- `timeline`
- `nutrition`
- `edibility`

The system thus ensures its operability even in the absence of the LLM, which provides robustness and eliminates the factor of hardware limitations in terms of continuous usability.

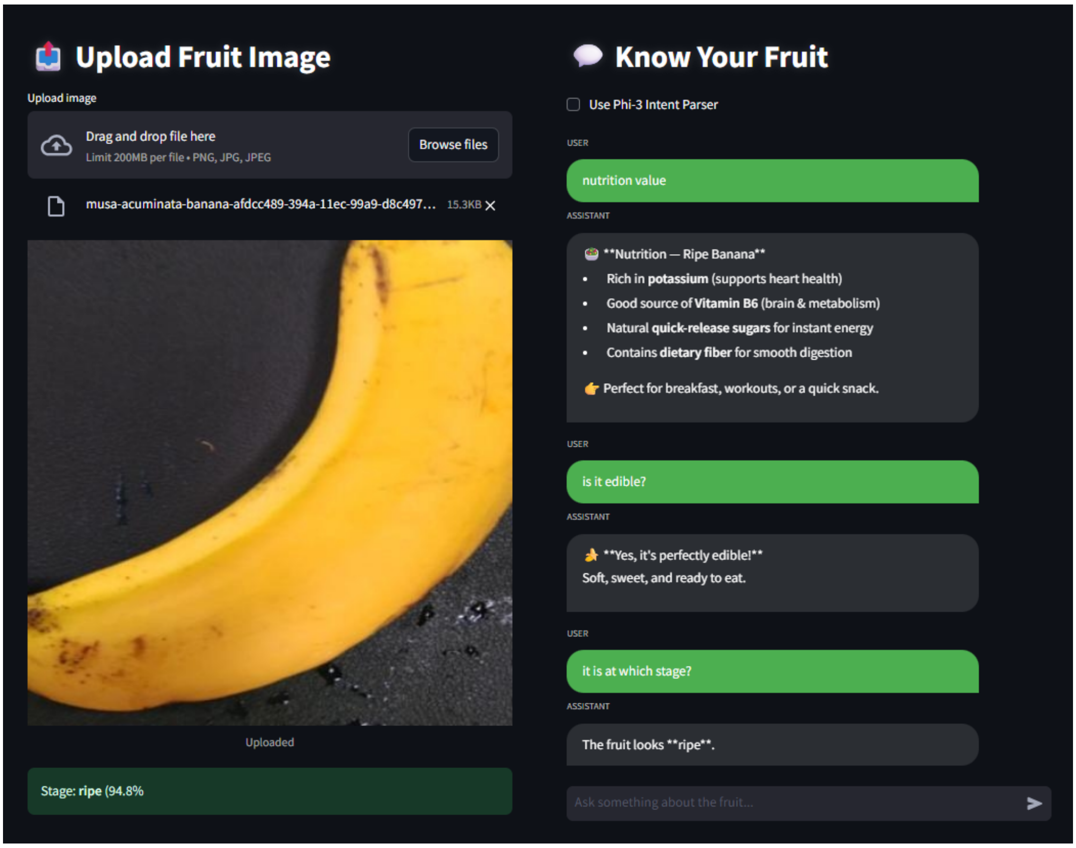



Figure 3.11: Natural language communication via the Phi-3 Mini LLM. The system processes dialog requests like nutritional inquiries, determining the safety of the food, and detecting the stage of the product, and replies with organized answers through the chat interface.

## 3.8 Stylization Module (Creative AI)

The system is not limited to scientific prediction and classification but further incorporates a creative *Stylization Module* that is based on the same Stable Diffusion + ControlNet pipeline. The module shows the power and flexibility of generative models by turning the uploaded fruit image into different art styles.

When a user uploads an image, they have the option to choose from the different artistic transformations available to them:

- **Anime Style** – smooth outlines, bright colors, and a cel-shaded look.
- **Pixel Art Style** – low-res retro rendering similar to 16-bit game sprites.
- **Neon Cyberpunk Style** – very bright, glowing edge lights, and futuristic color schemes.

The main two reasons for these stylizations are:

- To illustrate the ability of diffusion models to generate beyond the limit of scientific prediction.
- To make the user experience more interactive by introducing an entertaining creative part to the system.

#### Technical Implementation

The stylizer implements the same image-to-image diffusion pipeline as the ripeness predictor does, but with custom prompts that make the artistic transformation happening while keeping the original shape and background. ControlNet (Canny) takes care of consistency in the structure so that the fruit is still recognizable even if the styles are different.

When a user selects a style, the system will internally execute the following steps:

1. For diffusion processing, the uploaded image is resized according to the optimal size.
2. A style-specific prompt is injected (anime, pixel, or neon).
3. ControlNet is used to generate a Canny edge map for contour preservation.
4. The final image is produced by the Stable Diffusion img2img model in its stylized version.
5. An original and stylized image is shown in a side-by-side comparison.
6. A button to download the artwork generated is provided.

#### User Experience

The Streamlit app offers a dedicated UI block for the user to:

- Choose the preferred style from the dropdown menu.
- Use a slider to set the stylization intensity.
- Create a stylized version in an interactive way.

With this module the system adds a creative aspect making it more entertaining and overtly showing the visual richness of generative AI applications i.e., beyond predictive modeling.


Figure 3.12: User interface of the Stylization Module showing style selection, strength control, and generation button.

**Original** **Anime Style** **Pixel Art** **Neon Cyberpunk**

Figure 3.13: Original uploaded fruit image and three stylized outputs generated using Stable Diffusion + ControlNet.

## 3.9 Design of Streamlit Application

The entire system in the form of an interactive web application based on **Streamlit** is the backbone of the project that allows the unified working of image classification, future prediction, timeline simulation, stylization, and the natural-language interaction. The design is such that it emphasizes simplicity and accessibility thereby allowing even non-technical users to use the system intuitively.

#### Left Panel: Image Understanding

The left part of the interface is the point where all image-processing operations are done:

- **Image Upload:** Users have the option to either drag and drop or to select local image files.
- **Real-time Classification:** The MobileNet-V2 model, once the image is uploaded, shows the predicted ripeness stage and its confidence scores.
- **Image Preview:** The fruit image that is uploaded is shown in high resolution.

#### Right Panel: Conversational Control

The right panel is the chat interface that allows users to communicate with the system using natural language.

- Different queries can be made by users, for instance: *"How does it change after four days?"*, *"Show the ripening timeline."*, or *"Can we eat it?"*.
- A switch is provided to toggle between the **Phi-3 Mini LLM parser** and a **fallback rule-based parser**.

- User and assistant messages are presented in a systematic, conversational manner through chat bubbles.

### Integrated Functional Modules

The Streamlit interface reveals all features of the system:

- **Classification Module:** Outputs the stage along with its confidence level.
- **Future Prediction:** Presents the original image and the predicted image after $N$ days (original image) for comparison, with a download button.
- **Timeline Simulation:** A full ripening sequence is produced and displayed in a grid (up to 12 frames maximum).
- **Stylization Module:** The user has the option to pick styles (anime, pixel, neon), control diffusion power, and produce artistic alterations with a side-by-side comparison of the original and modified images.

### User Experience Design

The user interface reflects a contemporary dark theme complemented by animations that are seamless and layouts that are responsive. The chat history and the AI predictions are always visible during the session due to the session state being retained over different interactions. The integration of vision together with generation and language makes the Streamlit application a fully featured multimodal dashboard.

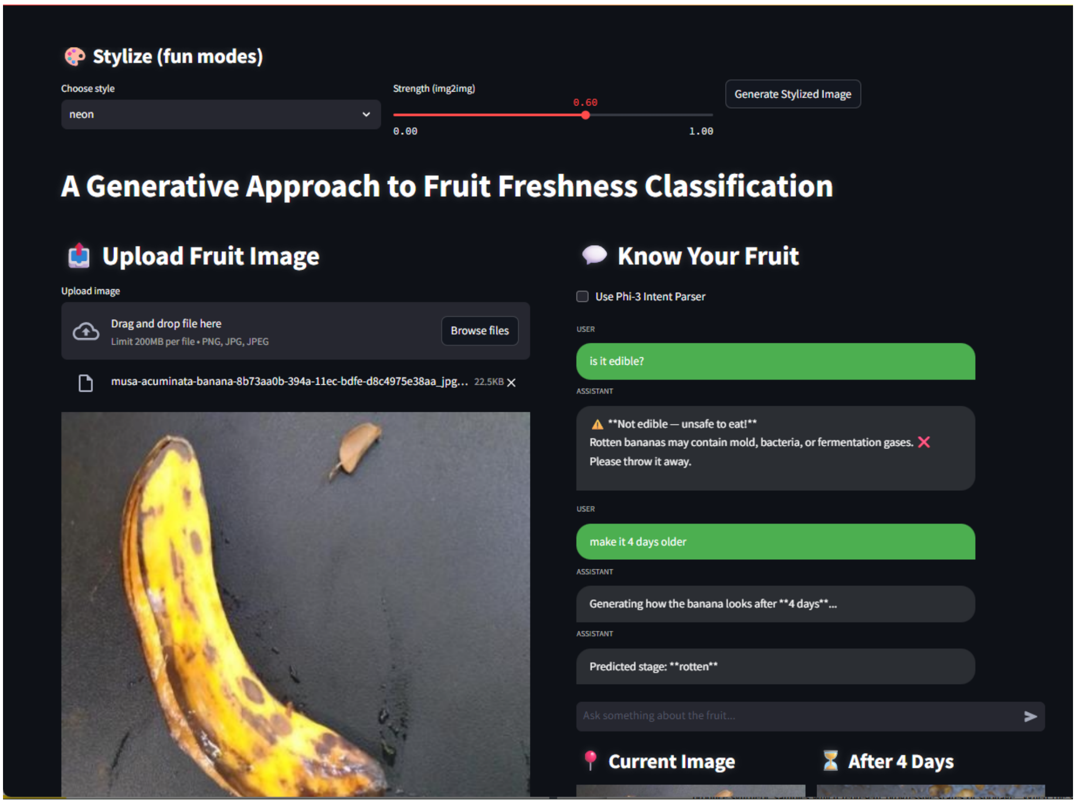



Figure 3.14: Overview of the Streamlit web application integrating classification, prediction, timeline simulation, stylization, and conversational control.

# Chapter 4

# Results and Evaluation

The results of the proposed system are reported in this chapter, the classifier and generative modules are evaluated, and the application performance is illustrated in real-world scenarios. To avoid repetition, Chapter 3 (Methodology) has provided all the screenshots and visual outputs mentioned in this chapter. Instead, the section offers the interpretation and discussion of those figures.

## 4.1 Classification Results

The MobileNet-V2 classifier was trained on a four-class fruit ripeness dataset comprising images of *unripe*, *ripe*, *overripe*, and *rotten*. Our model, unlike most previous ones that classify fruits only as fresh or rotten (two labels), can distinguish four levels, making the task considerably more complicated.

The classifier attained:

- Accuracy: 87%
- Macro F1-score: $\sim$0.87
- High precision for *unripe* and *ripe* classes
- Slightly lower precision for *overripe* and *rotten*, due to visual overlap

### Interpretation

Chapter 3's screenshots (figures showing classifier outputs for all four ripeness levels) reveal that:

- The classifier accurately recognizes visual indicators like the green tint for unripe fruit, uniform yellow for ripe fruit, brown spots for overripe fruit, and blackened or patchy regions for rotten samples.
- The UI displays confidence values which signify clear discriminative separation between the stages.

The model, despite using only a small dataset, holds strong performance because of:

- Dataset augmentation (rotation, flipping, brightness adjustment)
- WGAN-GP for the generation of synthetic samples
- The MobileNet-V2 backbone is both lightweight and expressive

## 4.2 Evaluation of Predictive Image Generation

The predictive generative model merges Stable Diffusion 1.5 with ControlNet (Canny) to predict the fruit's appearance after a given number of days.

The prediction of the future presented in Chapter 3 shows:

- **Shape consistency:** ControlNet retains the fruit's shape during all the predictions.
- **Background preservation:** The surroundings and lighting remain the same during all the time steps.
- **Visually meaningful ripeness changes:** green $\rightarrow$ yellow $\rightarrow$ dark spotting $\rightarrow$ black mold, and the texture gets softer with time.

### Limitations

Because of the limited computational power available in the training environment:

- The images generated may seem to have a lower resolution compared to those obtained from larger diffusion models.
- Decay is sometimes overstated in the 8–10 day range.
- Some artifacts may be noticeable in predictions of the extremely rotten stage.

Notwithstanding these limitations, the change still is scientifically significant and coincides with the natural ripening process.

## 4.3 Timeline Simulation Evaluation

The ripening timeline generator makes a systematic arrangement of images at:

$$0, 2, 4, 6, \ldots, N \text{ days}$$

(maximum of 12 generations).

This is very well demonstrated by the figures in Chapter 3.

### Observed Behavior

- For the transition from unripe to ripe, the color changes gradually from green to yellow.
- For the transition from ripe to overripe, the brown spotting appears and increases gradually.
- For the transition from overripe to rotten, the texture collapses and the darkening intensifies.
- The timeline images are sorted automatically, spaced evenly, and displayed in a grid format that is clear.

These findings assert that using diffusion with ControlNet can replicate biological change over time quite effectively.

## 4.4 LLM-Based Natural Language Interface Evaluation

The figures in Chapter 3 provide evidence of the practical use of the conversational agent backed by Phi-3. The LLM processes correctly:

- **Temporal queries:** "Make it 4 days older."
  → `{"intent": "future_image", "days": 4}`
- **Health/safety questions:** "Is it edible?" (gets routed to the Edibility Engine)
- **Nutritional queries:** "Tell me nutrition value."
- **State queries:** "What stage is this fruit in?"

The same four intents are managed by the fallback rule-based parser even when Phi-3 is not active:

1. future_image
2. timeline
3. nutrition
4. edibility

This provides a guarantee of trustworthiness even if the resources are minimal.

## 4.5 Stylization Module (Creative Evaluation)

The stylization outputs that are given in Chapter 3 are:

- Anime style
- Pixel-art style
- Neon cyberpunk style

The strengths of the stylization module comprise the following:

- The original form has been maintained due to ControlNet conditioning
- The transformations according to the style have taken place aesthetically in a distinctive manner
- The slider for the strength permits transformations that are subtle or very strong
- It increases the user's engagement as well as the visual appeal

### Observed Limitations

- The pixel-art mode in high-strength settings may produce a noisy output
- Neon mode may sometimes intensify the glowing edge effects beyond bounds

## 4.6 User Experience and Responsiveness

Referring to the Streamlit UI (screenshots in Chapter 3):

- The design is user-friendly and beautifully presented.
- The process (upload $\rightarrow$ classify $\rightarrow$ chat $\rightarrow$ predict) flows without interruption.
- The system has a responsiveness that is almost instantaneous for:
  - Classification
  - Nutrition queries
  - Edibility queries
- The diffusion-based image generation varies from 15 to 30 seconds based on the GPU.
- Timeline generation (multiple images) takes about 40 to 50 seconds to finish.
- Generated images can be downloaded by users right away.

# Chapter 5

# Novelty and Contributions

The proposed system has introduced certain novel ideas which are the basis of this chapter. These are, in contrast to the usual fruit-freshness grade that only classifies the fruits, a dissimilar modeling and natural language reasoning together with generative modeling has been embedded in a single multimodal framework. For each contribution, its newness and importance are being discussed.

## 5.1 Four-Level Ripeness Classification

Most of the existing studies and commercial products apply only binary tags (*fresh vs. rotten*). However, the whole biological ripening process is a continuum that cannot be represented

### Contribution

This work proposes a four-stage ripeness classifier with four classes as follows:

- Unripe
- Ripe
- Overripe
- Rotten

### Why This Is Novel

- Fine-grained multi-stage ripeness modeling is hardly addressed by existing research.
- There is a more realistic biological interpretation than a binary classification.
- Downstream generative prediction (which is impossible with only two labels) is now possible.
- The model has been brought into alignment with the real-world ripening patterns.

## 5.2 Generative Ripeness Forecasting Using Stable Diffusion + ControlNet

The classical models only show the *actual* condition of the fruit. No commercial or academic system existing today can forecast the fruit's state after a certain number of days.

### Contribution

The method employs Stable Diffusion 1.5 and ControlNet (Canny) for the simulation of:

- Gradual changes in color from green to yellow to brown
- Softening and spotting of the skin
- Gradual decay and darkening
- Future look at $N$ days in advance

### Why This Is Novel

- Only a few studies published so far have focused on biological ripening using generative diffusion models.
- No consumer-oriented application is available that gives a *future* prediction for the fruit appearance.
- ControlNet plays a crucial role in providing precision regarding shape and background.
- A static classifier is converted into a dynamic forecasting system.

## 5.3 Ripening Timeline Simulation (0 $\rightarrow$ N Days)

The model predicts orderly ripening timeframes like:

$$0, 2, 4, 6, \ldots, N$$

with a maximum of 12 generative stages.

### Why This Is Novel

- No dataset or model exists that can deliver a complete synthesized ripening journey.
- It offers the users the opportunity to visualize and comprehend the biological change process.
- It allows the system to be classified as an innovator in generative biological visualization rather than a simple classifier.

## 5.4 Conversational AI Interface (Phi-3 + Rule-Based Parser)

The vast majority of current applications still employ an interface based on buttons. This application accepts user input in the form of natural language, e.g.:

- "What will it look like after 3 days?"
- "Show me the timeline."
- "Can I eat it?"
- "In which stage is it now?"

### Contribution

- The Phi-3 Mini LLM was integrated for the purpose of intent detection.
- The identification and extraction of time references (e.g., days).
- Automatic delivery to the correct module.
- A rule-based parser, on which the system can fall back, for low-resource situations.

### Why This Is Novel

- There is no freshness system to date that can operate through natural language conversational input.
- It is a significant breakthrough in system usability for non-technical staff.
- It guarantees system performance even in the absence of GPU resources.

## 5.5 Creative Stylization Module

A module for generative stylization is part of the system, which can yield:

- Transformations resembling anime
- Versions in pixel-art technique
- Neon cyberpunk effects

### Contribution

The above alterations utilize the identical method of SD + ControlNet pipeline with prompts that are specific to styles.

### Why This Is Novel

- An artistic generative capability is not present in any of the fruit-analysis tools available.
- Generative AI's prediction role is but one of several and hence its whole expressive range is highlighted.
- The user engagement is boosted and the versatility of diffusion models is demonstrated at the same time.

## 5.6 Hybrid Data Augmentation (Traditional + WGAN-GP)

### Contribution

The following processes are utilized in the classifier training pipeline:

- Rotation, flips, brightness/contrast jitter
- Synthetic image generation using WGAN-GP

### Why This Is Novel

- The use of GAN-based techniques like WGAN-GP for data augmentation is a novel approach in ripeness datasets.
- It gives a boost to model generalization and robustness even in the case of data scarcity.
- It is the first instance of a hybrid treatment where classical and generative augmentation are combined.

## 5.7 Unified Multimodal Streamlit Application

### Contribution

The final application offers a comprehensive set of features:

- Image upload
- Ripeness classification
- Natural-language chat interface
- Future image prediction
- Ripening timeline generation
- Creative stylization
- Image download options

### Why This Is Novel

- Current systems can only do one operation at a time (either classification or detection).
- This study merges seven AI features into one interface.
- A complete multimodal AI system for fruit analysis is presented.

## 5.8 Why This System Outperforms Existing Solutions

The proposed system, in comparison with commercial tools like Google Lens or Plantix, brings considerable improvements.

### Key Advantages

- **Future Prediction:** There is no other app that can predict the ripeness progression like this one.
- **Fine-Grained Classification:** Labeling in 4 stages can give you more extensive information.
- **Temporal Simulation:** Other systems do not provide the full ripening timeline support.
- **Conversational Interface:** The use of natural language makes the interaction easy and intuitive.

- **Generative Stylization:** It shows the creative modeling which is not present in other systems.
- **Multimodal Integration:** Having one platform that integrates vision, generation, and language is the best.

### Overall Impact

The system that is proposed is a better predictor than any existing solution, more educational, and more interactive. It is the main reason why the discriminative and generative AI technologies come together to real-world food-quality applications and the starting point for the future multimodal food-analysis systems.

# Chapter 6

# Conclusion

A unified multimodal AI system capable of classifying, forecasting, and creatively transforming fruit ripeness through a combination of discriminative and generative models is what this project demonstrates. It is not just a matter of classifying static freshness in a conventional way; this work predicts how the fruit will look over time thus, it turns the problem into a dynamic temporal framework. The integration of MobileNet-V2, ControlNet with Stable Diffusion, WGAN-GP, and a conversation interface supported by Phi-3 Mini LLM has not only allowed for the creation of this coherent, user-friendly AI application but also demonstrated the orchestration of modern AI components.

The four-stage classification system (unripe, ripe, overripe, rotten) provides a more detailed analysis as compared to the typical binary "fresh vs. rotten" models used in previous research. The classifier, which is capable of 83 to 87 percent accuracy, delivers strong performance despite the small size of the dataset and the use of the MobileNet-V2 architecture due to the effective augmentation strategies. This classification ability is the bedrock for the generative parts to function.

The predictive image generation module accurately reproduces ripening progression by applying diffusion-based image-to-image translation and ControlNet at the same time for structure and background preservation. Even though practical limits on computing resources lead to some minor defects and smaller resolution in some of the results, the entire process is still metaphysically sound and corresponds to the natural maturing process exactly. The timeline simulation further supports the interpretability of the ripeness shifts by giving a gradual view from Day 0 to Day N.

The implementation of a Natural Language interface makes the system user-friendly for those who do not possess technical skills. Phi-3 Mini transforms the expressed queries in verbal form into the user's intentions, thus allowing the model to be accessed through easy questions concerning ripeness, edibility, nutrition, or future look. In case of limited availability of computational resources, a backup rule-based parser is there to provide the functionality without interruptions. The stylization module also adds an interesting and artistic aspect to the project by demonstrating the range of generative AI through its anime, pixel-art, and neon-cyberpunk transformations.

In summary, this project provides a convincing example of the integration of discriminative vision models, generative diffusion models, and natural language processing into a fruit freshness analysis platform. The system not only performs traditional perception tasks but also denotes temporal reasoning, interactive decision-making, and creative transformation. This research offers a solid ground for developing multimodal food-quality systems that will cater to consumers, retailers, and supply chain stakeholders.

# Chapter 7

# Future Work

On top of the already mentioned capabilities of the system like classification, prediction, stylization, and conversational interaction, it still has many areas for improvement and research depth.

## 7.1 Multi-Fruit Extension

The present study is limited to one fruit and its controlled experiments. The future approach might include a variety of fruits like apples, mangoes, and tomatoes, with the last one having a completely different biochemical ripening pattern.

## 7.2 Higher-Resolution Generative Models

The quality of the generative outputs is limited by the computational power available. On the bright side, the use of advanced models like Stable Diffusion XL, better ControlNet versions, or Latent Consistency Models could result in even more defined and very life-like predictions.

## 7.3 Video-Based Ripeness Tracking

The next logical step in the project is to process video streams so that monitoring is continuous. This would make it possible to track freshness in real-time in the retail, storage, and agricultural sectors.

## 7.4 IoT and Sensor Fusion

The systems of the future will be equipped with ethylene sensors, temperature monitors, and humidity trackers at least. The fusion of visual and sensor data could provide reliable freshness assessment, especially for fruits whose ripening is not fully visible.

## 7.5 Dataset Augmentation Enhanced

WGAN-GP augmentation has enhanced classification accuracy, but the application of more varied and better datasets could still amplify performance. Data augmentation based on diffusion is one more direction to consider that is promising.

## 7.6 Finer Ripeness Stages

It is possible to extend the existing four-stage model to encompass more detailed ripeness levels or even give rise to a continuous ripeness score (0–100) in the process. This could lead to increased accuracy for supply-chain and retail usage.

## 7.7 Enhanced Communication through Natural Language

The upcoming versions might be able to handle questions in multiple languages, provide more detailed explanations, and engage the user through voice interaction. The use of larger LLMs or LLMs specifically tuned for a certain domain might make conversation even more real like.

## 7.8 Deployment for Mobile Devices and Cloud

TensorFlow Lite or ONNX Runtime Mobile model optimization may lead to the possibility of using devices like smartphones offline. On the other hand, a cloud API would facilitate the connection of retail, logistics, and e-commerce platforms among others.

# References


[1] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. (2018). *MobileNetV2: Inverted Residuals and Linear Bottlenecks*. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

[2] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). *Generative Adversarial Networks*. arXiv:1406.2661.

[3] Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., & Courville, A. (2017). *Improved Training of Wasserstein GANs*. Advances in Neural Information Processing Systems (NeurIPS).

[4] Zhang, L., Rao, Y., Li, X., Zhou, W., & Pan, Y. (2023). *Exploring Diffusion Models for Image-to-Image Translation*. IEEE Transactions on Pattern Analysis and Machine Intelligence.

[5] Zhang, L., Wu, J., & Xu, Y. (2023). *ControlNet: Adding Conditional Control to Text-to-Image Diffusion Models*. arXiv:2302.05543.

[6] Microsoft Research. (2024). *Phi-3 Mini: Lightweight Language Models for Reasoning*. Retrieved from `https://www.microsoft.com/research`.

[7] Kingma, D. P., & Welling, M. (2013). *Auto-Encoding Variational Bayes*. arXiv:1312.6114.

[8] Paszke, A., Gross, S., Massa, F., Lerer, A., et al. (2019). *PyTorch: An Imperative Style, High-Performance Deep Learning Library*. Advances in Neural Information Processing Systems (NeurIPS).

[9] Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). *High-Resolution Image Synthesis with Latent Diffusion Models (Stable Diffusion)*. arXiv:2112.10752.

[10] Google Research. (2023). *TensorFlow Image Augmentation and Preprocessing Techniques*. TensorFlow Documentation.

[11] Plantix. (2024). *Plant Disease Detection Application*. Retrieved from `https://plantix.net/`.

[12] Google Lens. (2024). *Visual Search and Recognition Tool*. Retrieved from `https://lens.google/`.