UNIVERSITÀ DEGLI STUDI DI MILANO

CULTIVATING INSIGHTS: PREDICTION MODELS FOR PLANT GROWTH

Candidate: Dzmitry Kurch 13652A
Supervisor: Professor Giancarlo Manzi
Co-supervisor: Professor Silvia Salini

Academic year 2023/2024

Motivation and goal

- Traditional methods of **manual phenotyping** (measuring plant traits by hand) are time-consuming and prone to human error. There's a growing need for **automated systems** that provide **accurate, real-time** plant growth monitoring for researchers and farmers alike.
- By applying **machine learning** and **computer vision** techniques, we can enhance the precision of growth predictions, leading to better crop management and ultimately higher yields.
- This research aims to bridge the gap by developing methods to automate the detection of plant growth stages through **image analysis** and leveraging **time-series models** to predict future growth trends based on environmental data and historical growth observations.

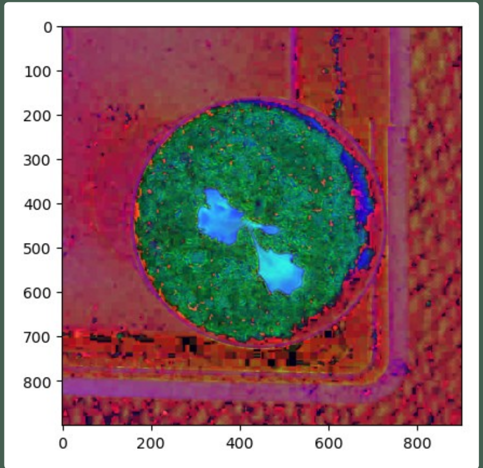
Literature background

- **Availability of Data Sources:** Although there are multiple plant image datasets available, the majority of these are either focused on a single plant identification or pacing of the aggregated crops.
- **Challenges in Leaf Counting & Surface Estimation:** Despite advances in image processing, leaf counting remains a challenging task due to issues like leaf overlap, occlusion, and changing lighting conditions. Inaccurate leaf detection can affect growth analysis and predictions. Surface estimation, while more stable, also faces challenges in maintaining accuracy across various growth stages.
- **Existing Models for Growth Prediction:** Traditional growth models such as the logistic growth model or exponential models have been widely used in agriculture but often lack the precision needed to handle complex environmental influences. Machine learning models, particularly time-series models like ARIMA, have gained popularity due to their ability to model dynamic growth patterns.

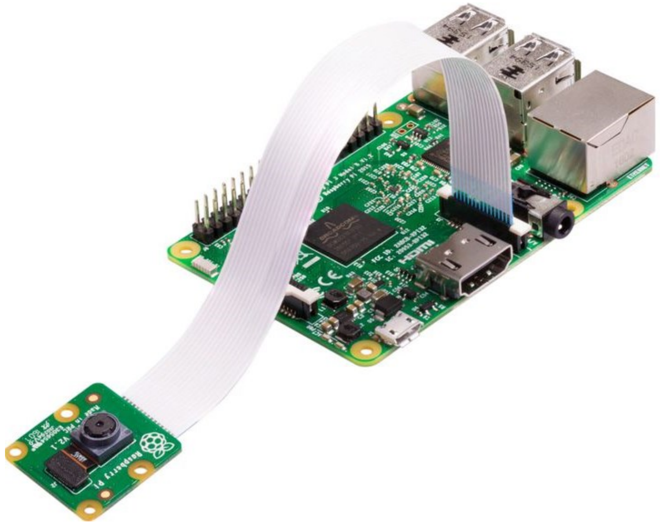
Project overview

a) In-house Data Collection
   a) Experiment hardware setup
   b) Custom software development
   c) Final measurements and data collection

b) Image analysis
   a) Raw data preprocessing
   b) Leaf count extraction
   c) Leaf surface extraction



c) Plant growth prediction
   a) Time-series modeling
   b) Bayesian framework



In-house data collection

Hardware setup:

- Raspberry Pi 3
- Pi3 camera
- DHT11 sensor

Software setup:

- Python 3.11
- Image processing libraries
- Script as a service

Data collection

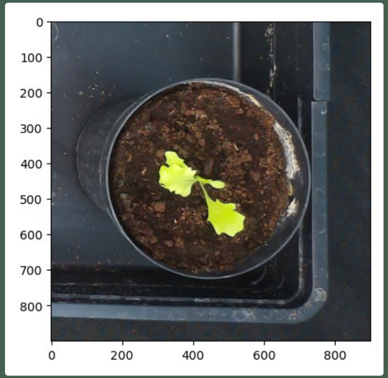
Image analysis

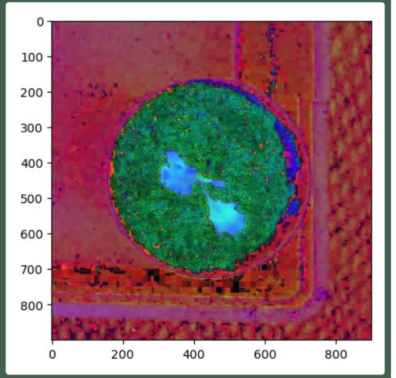
Leaf Detection:

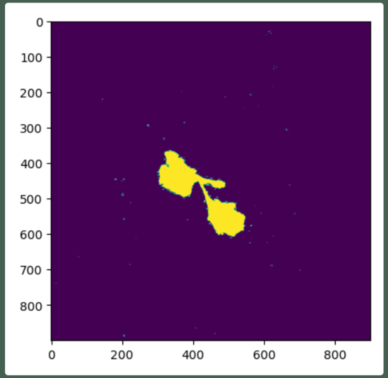
- **Raw Image Processing:** manual review and selection, detection of individual plants, tone normalization, color spaces.
- **Filtering and Morphology:** Uses morphological operations like opening and closing to refine leaf boundaries, removing noise and enhancing accuracy.
- **Challenges:** Difficulties arise in late growth stages due to overlapping leaves, making it harder to distinguish individual leaves.

Surface Area Calculation:

- **Pixel to Real-World Conversion:** Converts detected leaf areas from pixel-based measurements to real-world units (cm²), calibrated based on camera settings.
- **Advantages:** Unlike leaf counting, surface area estimation is more robust, especially when leaves overlap, as it measures overall plant coverage.
- **Applications:** Useful for tracking growth trends over time and provides a reliable metric for comparing plants in various stages of development.







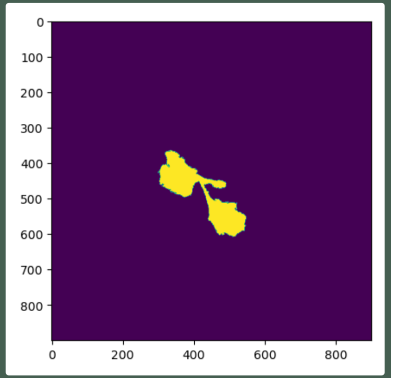
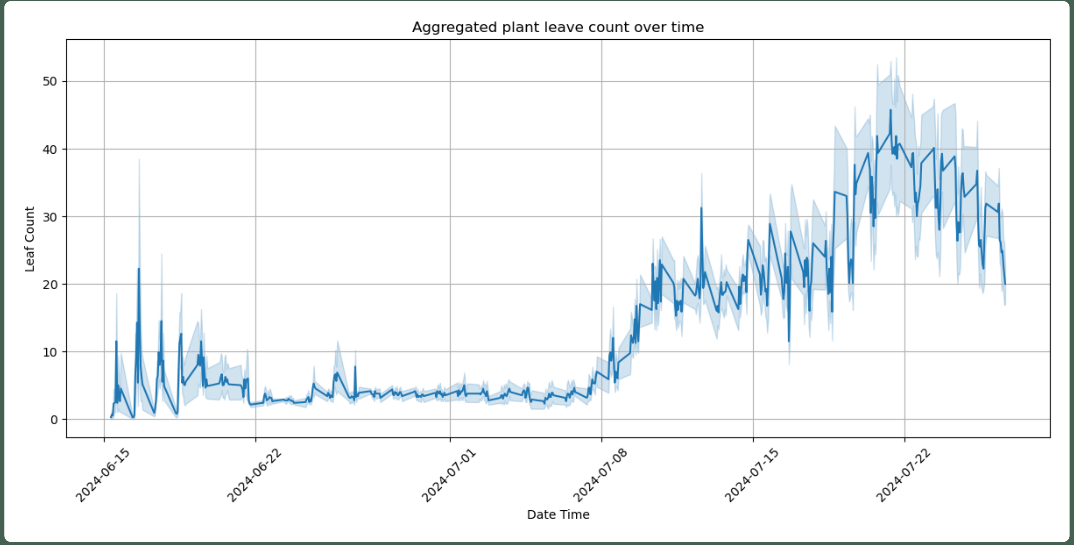


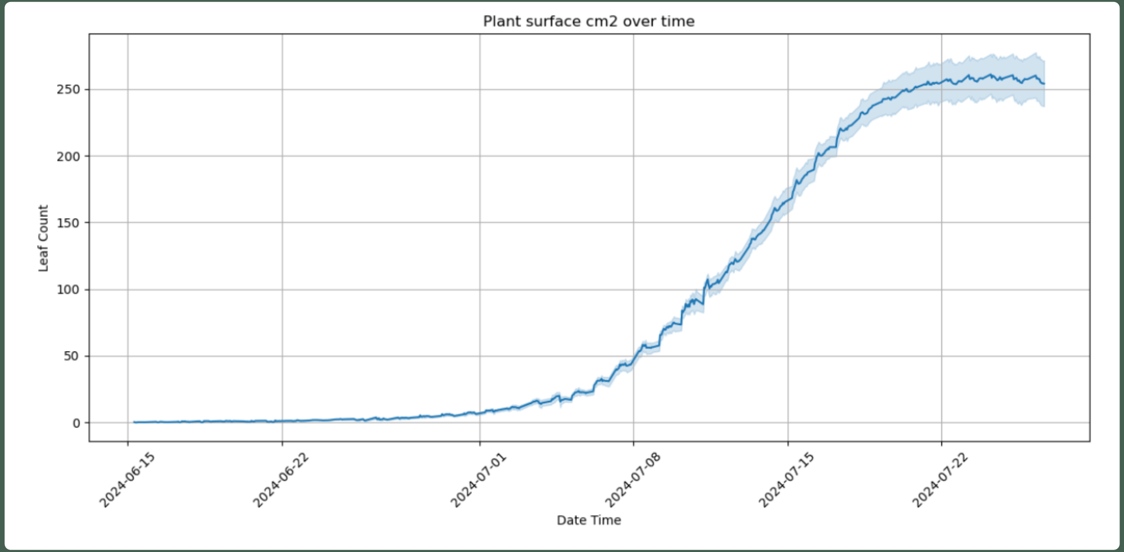
Image analysis

Leaf count calculation



- Big variance during germination
- Accurate in early stages
- Struggles with overlapping leaves in harvesting stage

Surface area calculation



- Robust in early to mid stages
- Requires calibration for accurate leave overlapping in harvesting stages

ARIMA for plant growth

Train-Test Split:

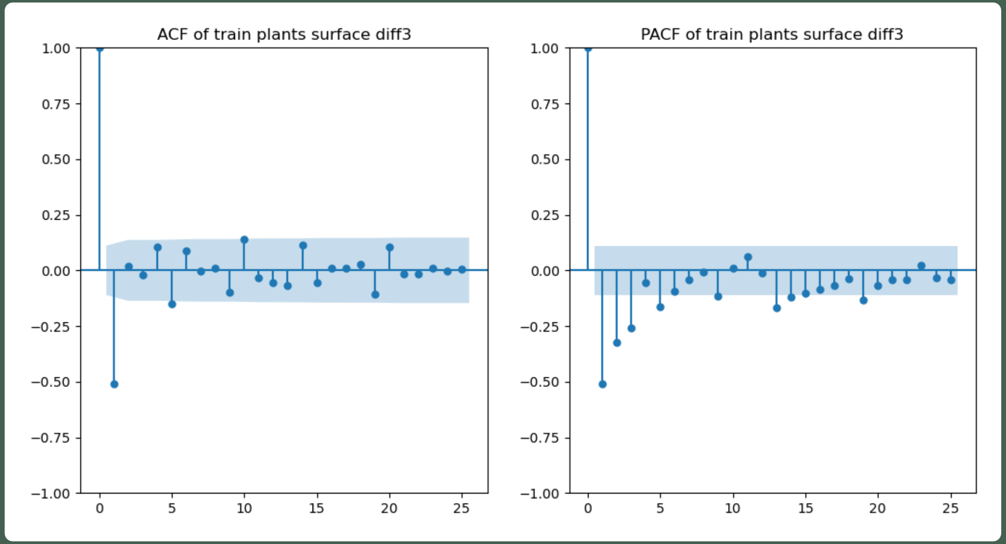
- **Train Data**: Used 6 plants to train the model by joining their time-series data with gaps to avoid abrupt transitions.
- **Test Data**: Kept 2 plants for testing and made sequential predictions by expanding the training set with test data day-by-day.

Key Steps:

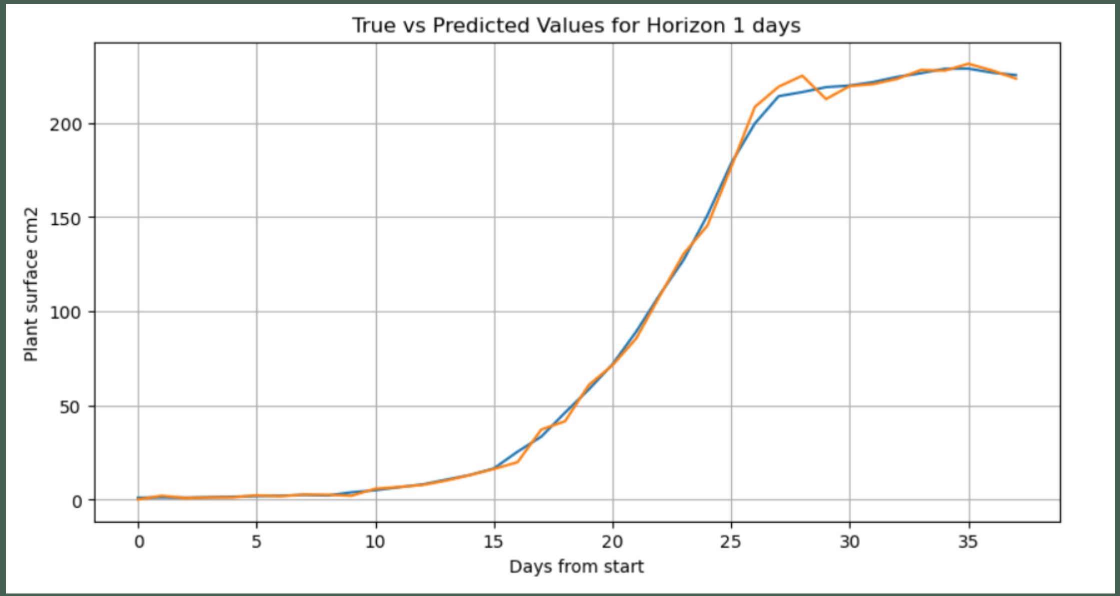
- Time-Series Decomposition
- Differencing for Stationarity
- ACF and PACF for model parameters
- Model fit and quality measurement

Results

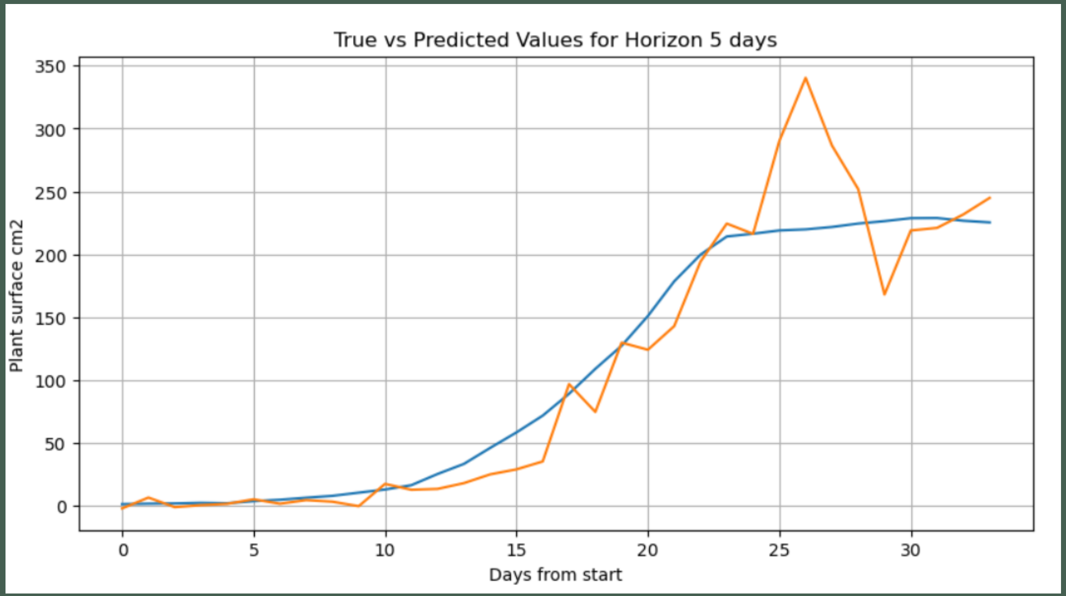
- Good for **short-term predictions** (next couple of days)
- Inappropriate for **long-term forecast**
- Struggles with **non-linearity** in later growth stages



Arima for plant growth



Actual (blue) VS predicted (orange) value for test plant with prediction horizon **1 day**



Actual (blue) VS predicted (orange) value for test plant with prediction horizon **5 days**

Bayesian framework

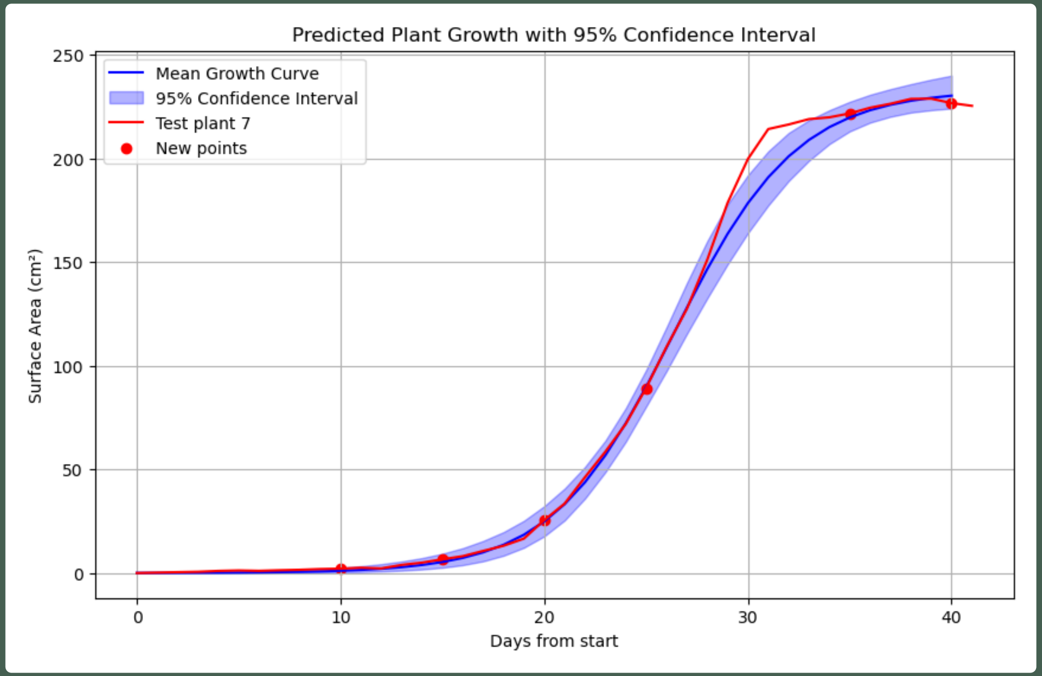
Key concepts

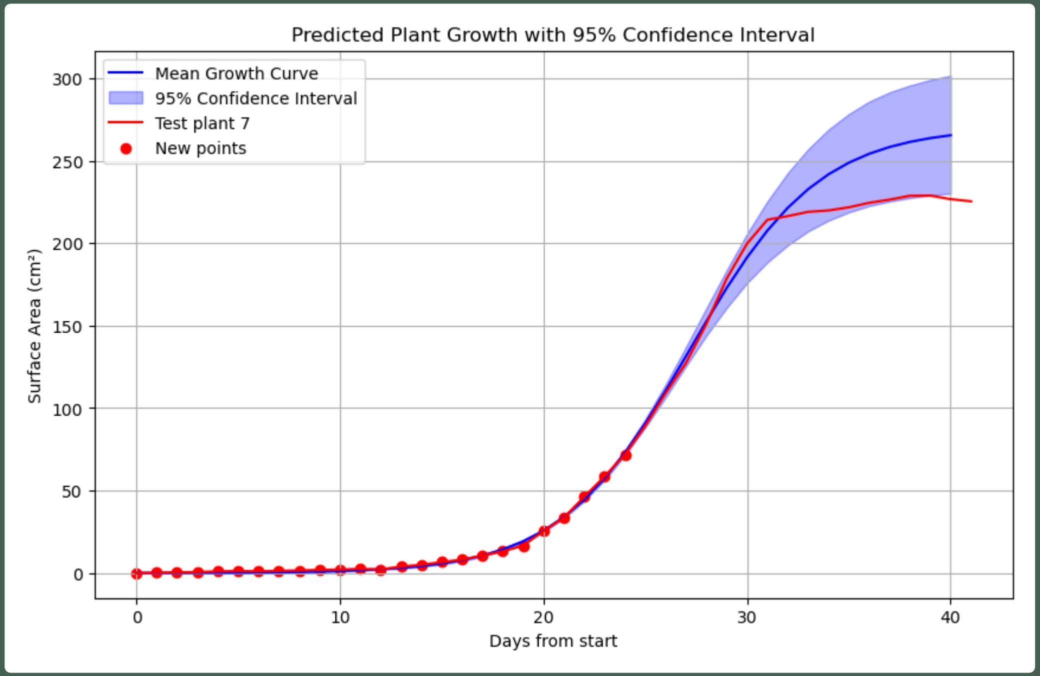
- **Bayes' Theorem**: Updates prior beliefs with new data to create posterior distributions, capturing **uncertainty** in predictions.
- **Sigmoid Growth Curve**: Utilized a sigmoid function to model plant growth over time, reflecting the three phases of growth (slow start, rapid boost, and plateau).

$$S(t) = \frac{A}{1 + e^{-k(t-t_0)}}$$

  - **A** represents the **maximum surface area** (upper limit of plant growth).
  - **k** is the **growth rate**, determining how fast the plant grows.
  - **t0** is the **inflection point**, indicating when growth is fastest.
- Prior distribution parameters for **A**, **k**, and **t0** are chosen using observations from the training set
- As more observations (test set) are collected, the model updates the priors and creates posterior distributions for **A, K,** and **t0**. The more data points seen, the narrower the posterior becomes

Bayesian framework





- 6 random measurements from test plant to adjust the prior
- Get the growth curve for all period with uncertainty interval

- First 25 days of growth from test plant to adjust the prior
- Get the future growth pace with uncertainty interval

Plant growth prediction

ARIMA

- **Limited Handling of Non-Linearity**: Effective for **short-term predictions** but struggled with the non-linear plant growth patterns, especially as growth slowed during the plateau phase.
- **Requires Data Regularity**: The ARIMA model worked best when provided with **regular time intervals** of data.
- **Best for Short-Term Predictions**: While ARIMA produced accurate results in **early growth stages**, as growth followed more predictable trends, it became less reliable as the forecast horizon extended, especially when new, unseen data were introduced.
- **Simpler Setup**: ARIMA was **easier to set up** compared to the Bayesian model.
- **Increasing Error Over Time**: As the forecast horizon increased, ARIMA's error grew significantly, making it less suitable for long-term predictions compared to the Bayesian model, which maintained more accurate estimates with wider uncertainty bounds.

BAYESIAN APPROACH

- **Uncertainty Modeling**: Provided **credible intervals** for plant growth predictions, offering a range of possible outcomes.
- **Flexibility**: Used a non-linear **sigmoid growth curve** to model plant growth, which captures the **S-shaped** pattern typical in plant development (slow start, rapid growth, and plateau).
- **Use of Priors:** Incorporated **prior knowledge** from the training. These priors were updated dynamically with new observations, refining the model over time.
- **Handling different use-cases**: You may use only several observations at any stage to get predicted growing pace during all the time
- **Real-World Adaptation**: Adapted well to the complexity of plant growth, incorporating environmental variability and non-linearity, making it more practical for biological data like plant growth.

Future extensions

- **Expanding to Different Plant Species**: The models can be adapted to other plant species with varied growth patterns, such as crops with longer growth cycles or complex structures (e.g., fruit-bearing plants).
- **Integrating Additional Environmental Variables**: Future work can incorporate real-time data from soil moisture sensors, light intensity measurements, and nutrient levels to enhance prediction accuracy.
- **Improving Image Processing Techniques**: Applying **deep learning** (e.g., CNNs) could enhance leaf detection and segmentation, especially in cases with heavy occlusion or complex leaf shapes.
- **Developing Real-Time Monitoring Systems**: The research could be extended into an automated system for **real-time plant monitoring**, providing farmers with continuous growth predictions and alerts.
- **Scaling for Precision Agriculture**: Incorporating these methods into **precision agriculture** applications, enabling large-scale, data-driven crop management to optimize yield and resource use.