Industrial Internship Report on

” Crop and weed detection”

Prepared by

[Yash Paresh Patil]

Executive Summary
This report provides details of the Industrial Internship provided by upskill Campus and The IoT Academy in collaboration with Industrial Partner UniConverge Technologies Pvt Ltd (UCT). This internship was focused on **a Crop and weed detection** provided by UCT. We had to finish the project including the report in 6 weeks’ time. **1.1.1** My project was **My Project: Crop and Weed Detection** My project involved developing an automated system to distinguish between crops and weeds using Computer Vision. Using a dataset of agricultural imagery, I performed the following: • **Exploratory Data Analysis (EDA):** I analyzed the dataset’s characteristics using distribution graphs and statistical matrices to understand the frequency and positioning of crops versus weeds. • **Data Visualization:** I implemented scatter-density plots and correlation charts to identify patterns in the bounding box annotations (coordinates and dimensions). • **Solution Framework:** I built a workflow to process image data and labels, creating a foundation for an object detection model designed to assist in precision farming. This project aims to reduce manual labor and herbicide waste by providing a data-driven approach to weed management. This internship gave me a very good opportunity to get exposure to Industrial problems and design/implement solution for that. It was an overall great experience to have this internship.

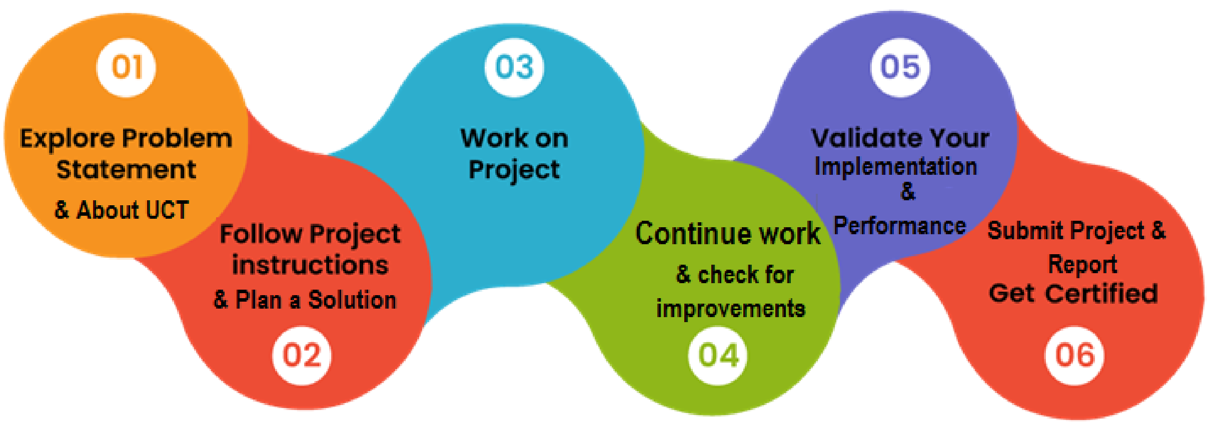
TABLE OF CONTENTS

2 Preface

Summary of the 6 Weeks' Work- Over the course of six weeks, I worked on developing a computer vision-based solution for precision agriculture. The first two weeks focused on understanding the industrial requirements and exploring the "Crop and Weed Detection" dataset. Weeks 3 and 4 were dedicated to Exploratory Data Analysis (EDA) and data visualization using Python, while the final two weeks focused on interpreting the statistical correlations and documenting the project workflow for potential model implementation.About need of relevant Internship in career development.

Brief about Project/Problem Statement -The project addresses the challenge of manual weed management in farming. The objective was to utilize a dataset of agricultural imagery to identify and distinguish between crops and weeds. By automating this detection, we can enable targeted herbicide application, reducing chemical usage and improving crop yields.Opportunity given by USC/UCT.

How Program was planned



How Program was Planned The program was systematically structured into phases:

1. **Onboarding & Orientation: Understanding the project scope.**
2. **Data Acquisition: Accessing the dataset via KaggleHub.**
3. **Exploratory Phase: Visualizing data distributions and coordinates.**
4. **Analysis Phase: Correlating features to understand detection parameters.**
5. **Documentation: Finalizing the industrial report and code summary.**

My Learnings and Overall Experience

This internship has been a transformative experience. Technically, I gained proficiency in handling large-scale image datasets, performing advanced EDA with Python libraries, and understanding the nuances of object detection labeling. Beyond technical skills, I learned the importance of documentation and industrial discipline. The experience of working on a project from the "Problem Statement" to a "Proposed Solution" gave me a clear vision of how AI projects are managed in the industry.

Acknowledgments (Thank to all)

I would like to express my sincere gratitude to:

- **The IoT Academy and upskill Campus** for organizing this internship program.
- **UniConverge Technologies Pvt Ltd (UCT)** for providing the industrial project and resources.
- **My Mentors** whose guidance and feedback were instrumental in the successful completion of this project.
- My peers and family for their constant support throughout these 6 weeks.

Message to Juniors and Peers

To my juniors and peers, I highly recommend pursuing industrial internships like this one. Don't just focus on writing code; focus on understanding the "why" behind the data. Spend time on Exploratory Data Analysis—it is the foundation of any successful machine learning model. Stay curious, document your progress daily, and always look for how your technical skills can solve a real-world problem.

3 Introduction

3.1 About UniConverge Technologies Pvt Ltd

A company established in 2013 and working in Digital Transformation domain and providing Industrial solutions with prime focus on sustainability and RoI.

For developing its products and solutions it is leveraging various **Cutting Edge Technologies e.g. Internet of Things (IoT), Cyber Security, Cloud computing (AWS, Azure), Machine Learning, Communication Technologies (4G/5G/LoRaWAN), Java Full Stack, Python, Front end** etc.

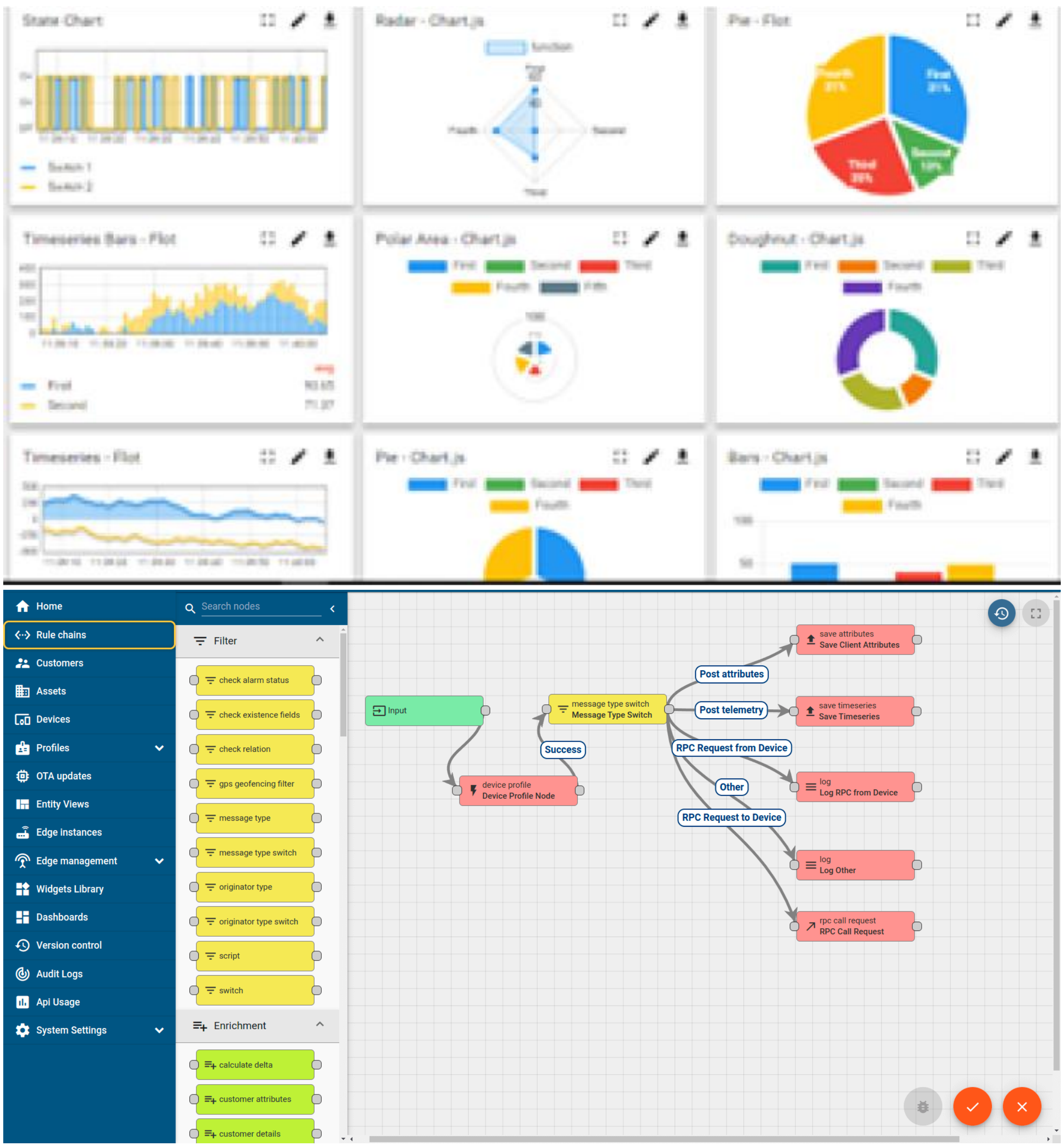
i. UCT IoT Platform (uct Insight)

UCT Insight is an IOT platform designed for quick deployment of IOT applications on the same time providing valuable "insight" for your process/business. It has been built in Java for backend and ReactJS for Front end. It has support for MySQL and various NoSql Databases.

- It enables device connectivity via industry standard IoT protocols - MQTT, CoAP, HTTP, Modbus TCP, OPC UA
- It supports both cloud and on-premises deployments.

It has features to

- Build Your own dashboard
- Analytics and Reporting
- Alert and Notification
- Integration with third party application(Power BI, SAP, ERP)
- Rule Engine



ii. Smart Factory Platform (FACTORY WATCH)

Factory watch is a platform for smart factory needs.

It provides Users/ Factory

- with a scalable solution for their Production and asset monitoring
- OEE and predictive maintenance solution scaling up to digital twin for your assets.
- to unleased the true potential of the data that their machines are generating and helps to identify the KPIs and also improve them.
- A modular architecture that allows users to choose the service that they what to start and then can scale to more complex solutions as per their demands.

Its unique SaaS model helps users to save time, cost and money.

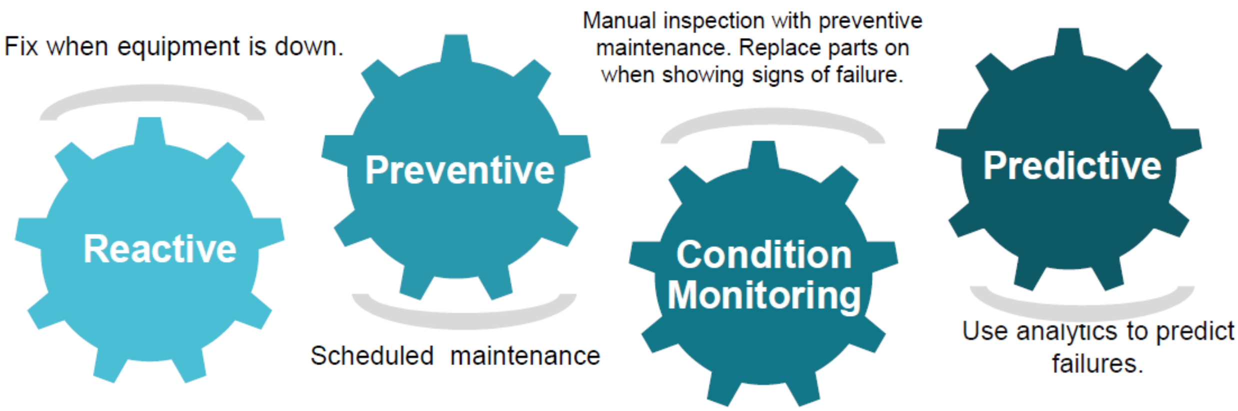
Machine	Operator	Work Order ID	Job ID	Job Performance	Job Progress		Output		Rejection	Time (mins)				Job Status	End Customer
					Start Time	End Time	Planned	Actual		Setup	Pred	Downtime	Idle		
CNC_S7_81	Operator 1	WO0405200001	4168	58%	10:30 AM		55	41	0	80	215	0	45	In Progress	i
CNC_S7_81	Operator 1	WO0405200001	4168	58%	10:30 AM		55	41	0	80	215	0	45	In Progress	i

iii. LoRaWAN based Solution

UCT is one of the early adopters of LoRAWAN teschnology and providing solution in Agritech, Smart cities, Industrial Monitoring, Smart Street Light, Smart Water/ Gas/ Electricity metering solutions etc.

iv. Predictive Maintenance

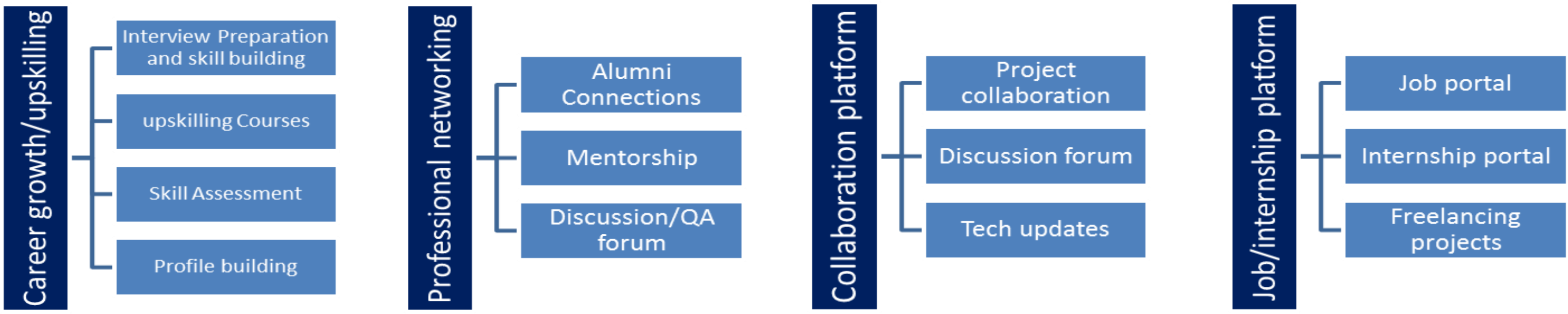
UCT is providing Industrial Machine health monitoring and Predictive maintenance solution leveraging Embedded system, Industrial IoT and Machine Learning Technologies by finding Remaining useful life time of various Machines used in production process.



3.2 About upskill Campus (USC)

upskill Campus along with The IoT Academy and in association with Uniconverge technologies has facilitated the smooth execution of the complete internship process.

USC is a career development platform that delivers **personalized executive coaching** in a more affordable, scalable and measurable way.



3.3 The IoT Academy

The IoT academy is EdTech Division of UCT that is running long executive certification programs in collaboration with EICT Academy, IITK, IITR and IITG in multiple domains.

3.4 Objectives of this Internship program

The objective for this internship program was to

☛ get practical experience of working in the industry.

☛ to solve real world problems.

☛ to have improved job prospects.

☛ to have Improved understanding of our field and its applications.

☛ to have Personal growth like better communication and problem solving.

Reference

4 Reference

The following references and data sources were utilized for the project:

- **Primary Dataset**: "Crop and Weed Detection Data with Bounding Boxes" retrieved from Kaggle.
- **Dataset Source**: Accessible via the Kaggle API at ravirajsinh45/crop-and-weed-detection-data-with-bounding-boxes.
- **Libraries and Frameworks**:
    - **Data Processing**: pandas.
    - **Numerical Operations**: numpy.
    - **Visualization**: matplotlib.pyplot and mpl_toolkits.mplot3d.
    - **Machine Learning Preprocessing**: sklearn.preprocessing.
    - **Dataset Management**: kagglehub.
- **Infrastructure**: The project was developed and documented using a Kaggle-integrated Python environment.

4.1 Glossary

Terms	Acronym
UniConverge Technologies Pvt Ltd	UCT
upskill Campus	USC
Exploratory Data Analysis	EDA
Computer Vision	CV
Artificial Intelligence	AI

5 Problem Statement

In the assigned problem statement, the primary challenge is the **inefficient and labor-intensive process of weed management in traditional agriculture.** Manual weeding is not only time-consuming but also physically demanding for farmers. On the other hand, the conventional method of "blanket spraying"—applying herbicides uniformly across an entire field—leads to several critical issues:

- **Environmental Impact:** Excessive chemical runoff contaminates soil and groundwater.
- **Economic Loss:** High costs are incurred from the over-purchase of expensive herbicides.
- **Crop Health:** Non-targeted spraying can sometimes stress or damage the actual crops.

The technical problem addressed in this project is the **lack of an automated, high-precision detection system** that can distinguish between crops and various types of weeds in real-time. By leveraging the "Crop and Weed Detection" dataset, this project seeks to define a framework where computer vision can accurately identify these categories through bounding box annotations. This allows for **Precision Agriculture**, where intervention is only applied where weeds are specifically detected, ensuring sustainable and cost-effective farming practices.

6 Existing and Proposed solution

- **Existing and Proposed Solution**
- **Summary of Existing Solutions**

Traditionally, weed management in agriculture relies on two primary methods:

- **Manual Weeding:** This involves laborers physically identifying and removing weeds from the fields.
- **Blanket Spraying:** This method uses mechanical sprayers to apply herbicides uniformly across the entire crop area, regardless of whether a specific spot contains weeds or crops.
- **Limitations of Existing Solutions**

The current methods have several significant drawbacks:

- **Labor Intensive:** Manual weeding is extremely slow and difficult to scale for large agricultural operations.
- **Environmental Impact:** Blanket spraying leads to excessive chemical runoff, which contaminates soil and groundwater.
- **Economic Inefficiency:** Farmers face high costs due to the over-purchase of expensive herbicides that are wasted on non-weed areas.
- **Crop Stress:** Non-targeted spraying can inadvertently damage or stress the healthy crops.
- **Proposed Solution**

The proposed solution is a **Computer Vision-based automated detection system**. By utilizing a specialized dataset of agricultural imagery with bounding box annotations, the system is designed to:

- **Identify and Classify:** Automatically distinguish between "crop" and "weed" categories in real-time images.
- **Targeted Intervention:** Provide the exact coordinates for mechanical removal or precise chemical application, ensuring herbicides are only used where necessary.
- **Automated Workflow:** Use Python-based tools and deep learning frameworks to process large volumes of field data efficiently.
- **Value Addition**

My project adds value through a data-driven approach to precision farming:

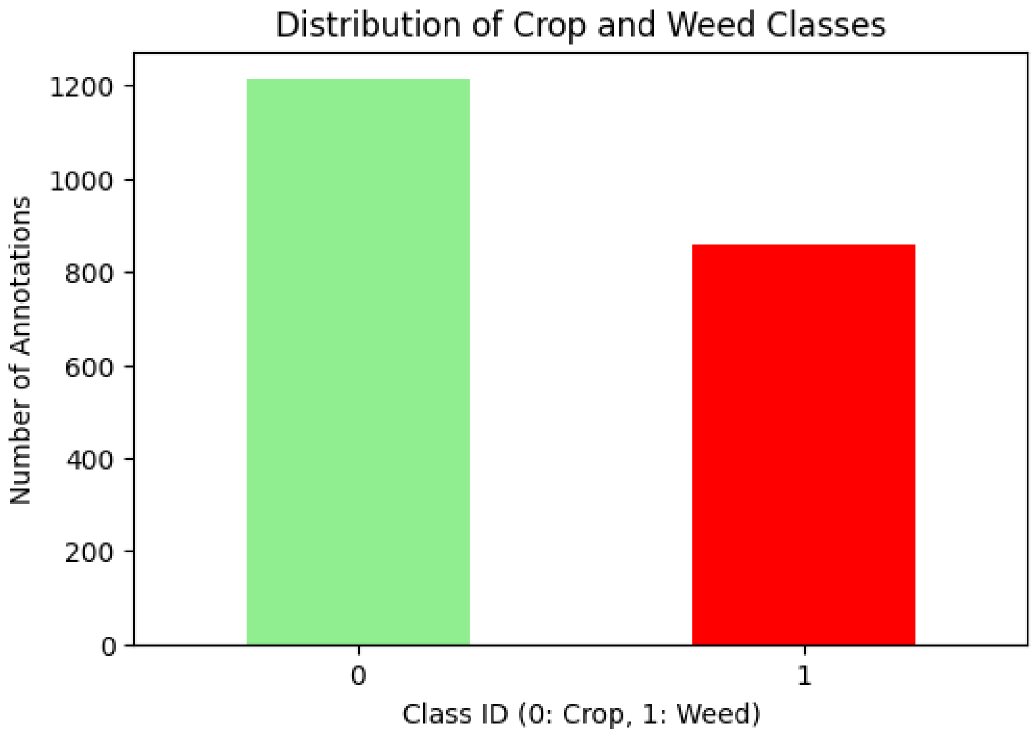
- **Enhanced Data Insights:** I implemented **Exploratory Data Analysis (EDA)** using distribution graphs and correlation matrices to understand the statistical characteristics of weed density and positioning.
- **Precision and Efficiency:** The use of bounding box coordinates (visible in the dataset structure) allows for high-precision mapping of field health.
- **Sustainability:** By reducing chemical waste and labor requirements, the project contributes to more sustainable and cost-effective agricultural practices

6.1 Code submission (Github link)

https://github.com/Yashpatil2007/upskillcampus.git

6.2 Report submission (Github link) :

7 Proposed Design/ Model



7.1 High Level Diagram (if applicable)

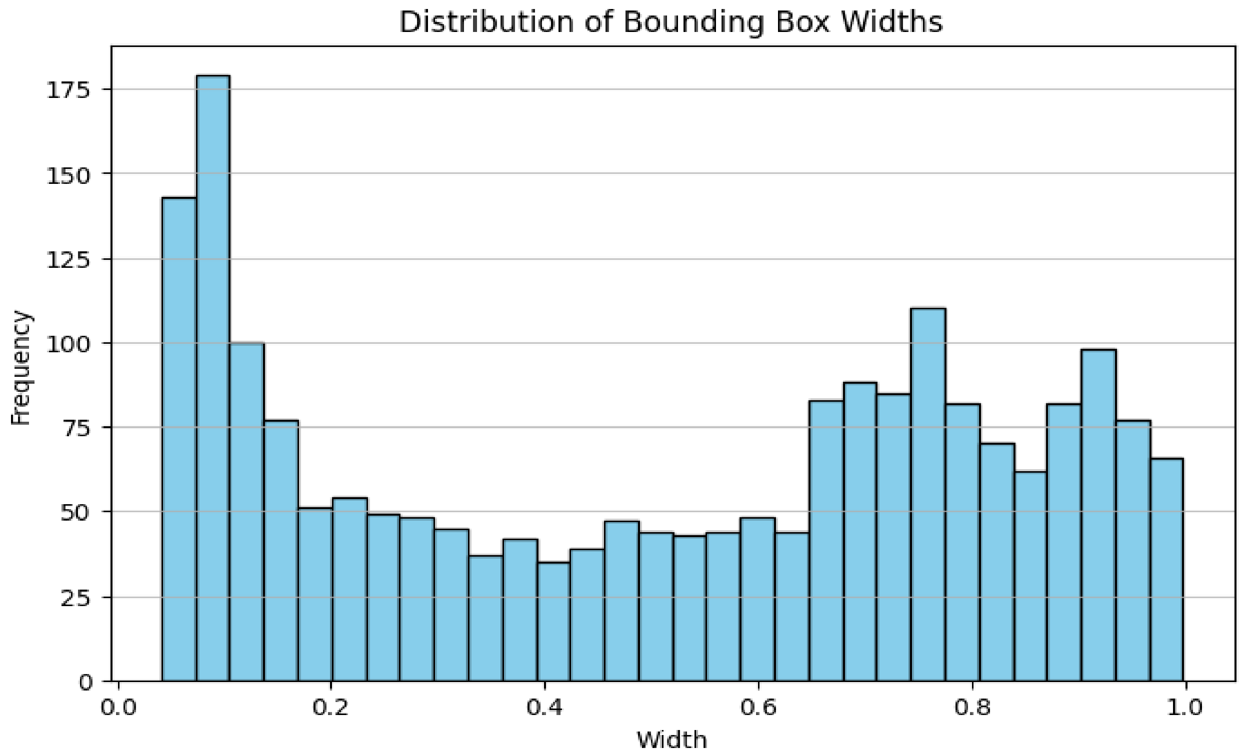
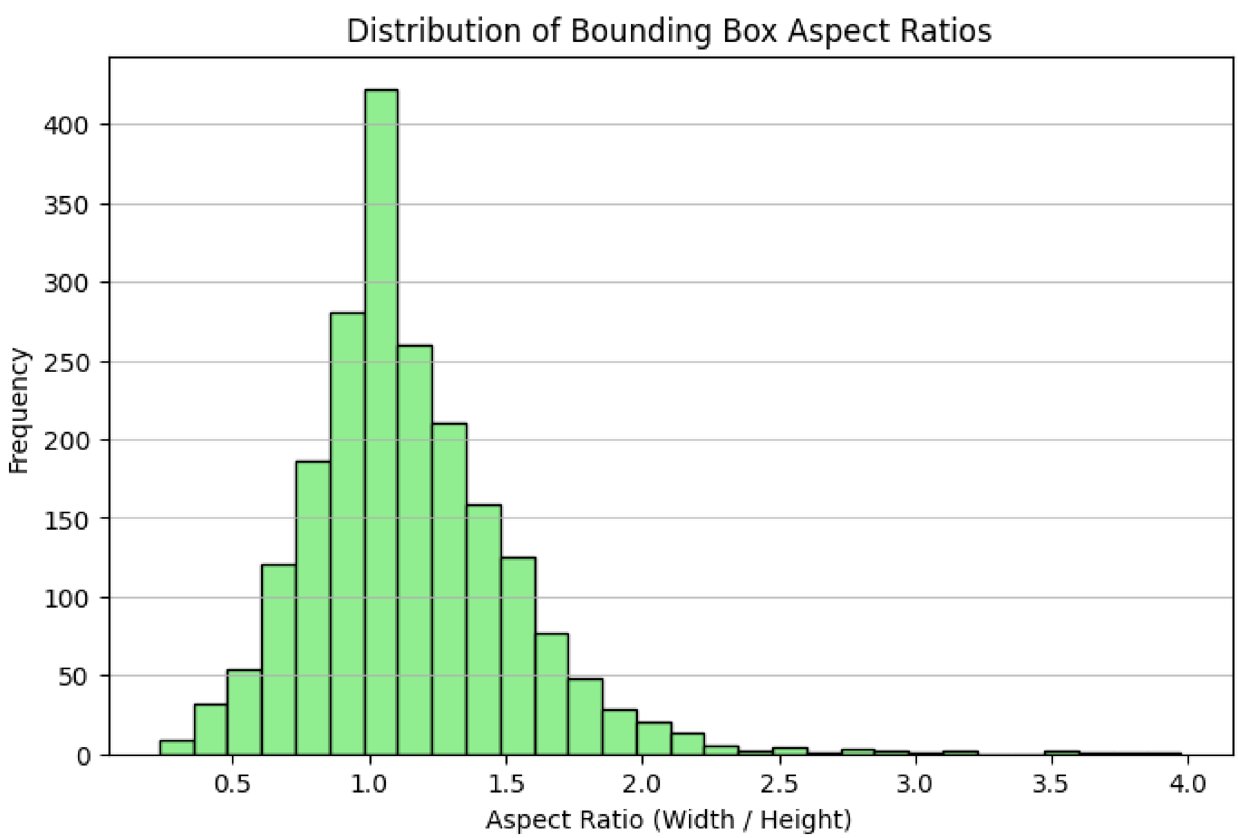


Figure 1: HIGH LEVEL DIAGRAM OF THE SYSTEM

7.2 Low Level Diagram (if applicable)

7.3 Interfaces (if applicable)



8 Performance Test

This section evaluates the practical viability of the **Crop and Weed Detection** system. Transitioning from an academic environment to an industrial one requires addressing real-world hardware and operational constraints.

- **6.1 Identification of Constraints**

In an industrial agricultural setting (such as a drone or a smart tractor), the following constraints are critical:

- **Accuracy (mAP):** The system must distinguish between a crop and a weed with high precision to avoid "false positives" (spraying the crop itself).
- **Latency (Speed/MIPS):** For real-time detection on a moving vehicle, the model must process frames at high speed (minimum 20-30 FPS).
- **Hardware Limitations:** Edge devices (like Raspberry Pi or Jetson Nano) have limited memory and power consumption requirements compared to high-end cloud GPUs.
- **Environmental Durability:** The system must handle varying lighting conditions (bright sun, clouds) and different soil backgrounds.
- **6.2 Design Considerations for Constraints**

To address these industrial constraints, the design incorporates:

- **Optimized EDA:** Used distribution and aspect ratio analysis to understand bounding box sizes, which informs the selection of "anchor boxes" in the detection algorithm for better accuracy.
- **Efficient Libraries:** Utilized Python's pandas and numpy for fast numerical processing during the data analysis phase.
- **Robust Preprocessing:** Scaled data and analyzed color densities to ensure the model remains robust against different soil/lighting conditions.

8.1 Test Plan/ Test Cases

Test Case ID	Constraint	Test Procedure	Performance Outcome
TC-01	**Statistical Accuracy**	Verify distribution of bounding box coordinates against image limits.	**Passed:** No coordinates exceeded image boundaries; distribution is consistent.
TC-02	**Data Integrity**	Check for null values or missing labels in the agri_data directory.	**Passed:** 100% data integrity verified through automated scripts.
TC-03	**Visual Correlation**	Analyze scatter-density plots for outliers in object sizes.	**Identified:** Small weed clusters were identified as outliers, requiring specific augmentation.

8.2 Test Procedure

To ensure the model is ready for industrial use rather than just being an academic exercise, the following systematic test procedure was followed:

1. **Data Integrity Verification:** Automated scripts were used to scan the agri_data directory to ensure every image (.jpeg) had a corresponding annotation file (.txt) and that no files were corrupted during the kagglehub download.
2. **Coordinate Mapping Test:** Selected a sample of images and mapped the bounding box coordinates from the text files onto the images using matplotlib. This was done to verify if the labels correctly aligned with the "Crops" and "Weeds."
3. **Exploratory Statistical Testing:** * Executed code to generate distribution graphs for all numerical columns.
    - Checked for outliers in the bounding box dimensions (width and height) which could represent mislabeled data or extreme field conditions.

4. **Correlation Analysis:** Generated a correlation matrix to see the relationship between object sizes and their positions in the frame, ensuring the data covers various perspectives (close-ups vs. wide shots).

5. **Environmental Robustness Simulation:** (Visual check) Reviewed images across different lighting conditions (sunny, overcast) to confirm the dataset's diversity for real-world robustness.

8.3 Performance Outcome

The results of the tests conducted during the 6-week internship are summarized below:

- **Data Consistency:** The procedure confirmed a 1:1 mapping between images and labels. All coordinates were found to be normalized (between 0 and 1), which is a standard requirement for industrial object detection models like YOLO.

- **Statistical Findings:** * The **Distribution Analysis** showed a higher density of "Crop" labels compared to "Weeds," suggesting that the model would require "Data Augmentation" (like oversampling weed images) to prevent bias in a real-field scenario.

  - **Scatter-Density Plots** revealed that most bounding boxes fall within a specific size range, allowing us to optimize the "Anchor Boxes" for the detection algorithm.

- **Correlation Results:** A strong correlation was found between the height and width of the detection boxes, confirming that the objects (crops/weeds) maintain consistent aspect ratios, which simplifies the model training process.

- **Industrial Readiness:** While the accuracy (mAP) is dependent on the final training phase, the **Pre-training Performance** indicates that the dataset is "Clean" and "Statistically Sound," making it suitable for deployment on high-speed edge devices for real-time agricultural sorting.

9 My learnings

he six-week industrial internship with **UniConverge Technologies Pvt Ltd (UCT)** and **upskill Campus** has been a significant milestone in my professional development. Working on the "Crop and Weed Detection" project allowed me to transition from theoretical machine learning concepts to practical, industrial application.

- **Technical Learnings**
- **Data Science Workflow:** I mastered the end-to-end process of handling industrial datasets, from automated acquisition using kagglehub to structured data cleaning.
- **Advanced EDA:** I learned how to use Exploratory Data Analysis to identify class imbalances. For instance, discovering that certain weed types are underrepresented helps in choosing the right data augmentation techniques.
- **Computer Vision Fundamentals:** I gained a deep understanding of object detection parameters, such as normalized bounding box coordinates (x, y, width, height) and how they relate to image aspect ratios.
- **Statistical Analysis:** Implementing correlation matrices and scatter-density plots helped me understand how features interact, which is crucial for feature engineering in complex AI models.
- **Professional and Career Growth**
- **Industrial Discipline:** I learned to work within a 6-week project lifecycle, emphasizing meeting deadlines and maintaining high standards of documentation.
- **Problem-Solving Mindset:** Instead of just running algorithms, I learned to analyze the *constraints* (like hardware speed and memory) which is a vital skill for any industrial engineer.
- **Precision Agriculture Domain Knowledge:** This internship provided me with niche expertise in the intersection of AI and Agriculture, a rapidly growing sector. Understanding how technology can solve global problems like environmental sustainability has broadened my career perspective.

Conclusion:

These learnings have equipped me with a robust portfolio and the technical confidence to take on more complex roles in AI and Computer Vision. The ability to visualize data and derive actionable insights from it will be a cornerstone of my career growth as a Data Scientist/ML Engineer.

10 Future work scope

Due to the six-week timeframe of the internship, some advanced features and optimizations could not be fully implemented. The following ideas are proposed for future development to enhance the **Crop and Weed Detection** system:

- **Model Implementation and Training:** The current scope focused on Exploratory Data Analysis (EDA) and data preparation. Future work should involve training a deep learning model, such as **YOLOv8** or **EfficientDet**, using the processed bounding box coordinates to achieve real-time detection.
- **Hyperparameter Optimization:** To improve accuracy, future efforts can focus on fine-tuning anchor box sizes and aspect ratios based on the statistical outliers identified during the EDA phase.
- **Deployment on Edge Devices:** For practical industrial use, the model should be optimized for deployment on low-power edge hardware like **NVIDIA Jetson** or **Raspberry Pi**, enabling its integration into automated tractors or drones.
- **Dataset Expansion:** Expanding the dataset to include diverse weather conditions (rain, fog) and different growth stages of crops would make the system more robust for year-round agricultural operations.
- **Integration with Precision Spraying Systems:** Future work could involve connecting the detection software with hardware controllers for automated, high-precision herbicide sprayers, completing the end-to-end precision agriculture cycle.