International Journal of Innovative Research in Engineering
Volume 7, Issue 2 (March - April 2026), PP: 244-253.
https://www.doi.org/10.59256/ijire.20260702030
www.theijire.com
ISSN No: 2582-8746

Transforming agriculture with edge AI – enabling the Smart Farming

Katha Amulya[1], K. Naga Rohitha[2], U. Venkata Krishna[3], Naman Pratap Singh[4], K Gopala Krishna[5]
[1,2,3,4,5] Department of CSE, KL University, Vijayawada, Andhra Pradesh, India.

How to cite this paper:
Katha Amulya[1], K. Naga Rohitha[2], U. Venkata Krishna[3], Naman Pratap Singh[4], K Gopala Krishna[5], "Transforming agriculture with edge ai – enabling the Smart Farming", IJIRE-V7I2-244-253.

Abstract: *This project focuses on building an intelligent plant disease classification system using the Plant Village dataset from Kaggle, specifically targeting three tomato leaf categories: healthy, late blight, and bacterial spot. The dataset was cleaned, resized, and augmented, then split into an 80:20 ratio for training and validation. Model development and training were carried out in Amazon SageMaker Studio, leveraging its scalable compute environment and integrated experiment tracking. After achieving satisfactory accuracy, the trained TensorFlow model was exported and deployed directly within SageMaker using a managed inference endpoint. For accessibility, a lightweight Gradio-based user interface is being built to allow users to upload leaf images and receive instant predictions through the deployed model. The final solution demonstrates a complete machine-learning workflow—from dataset preparation to cloud deployment and user interaction—providing a practical tool for early crop disease detection and supporting precision agriculture.*

Key Words: *AWS SageMaker Studio, Plant Village Dataset, Gradio.*

I.INTRODUCTION

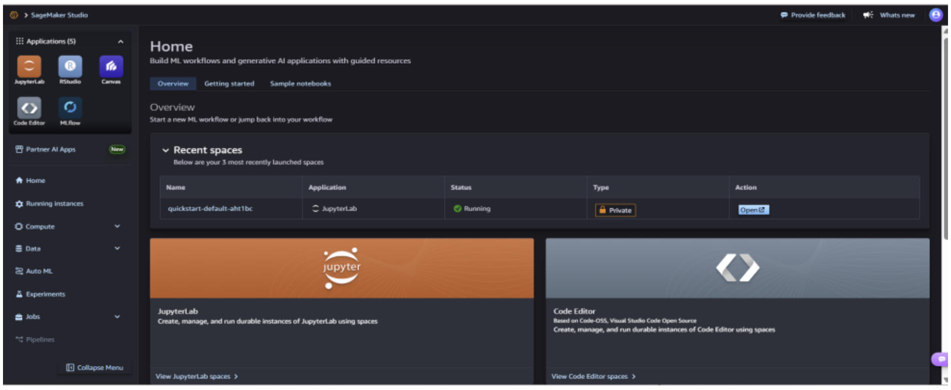

Figure 1 : SageMaker Studio UI

Agriculture remains a crucial sector for food security, and plant health monitoring plays a central role in sustaining crop yield and quality. Tomato plants, in particular, are highly susceptible to various diseases that can spread quickly and cause significant losses if not identified early. Traditional disease diagnosis often depends on manual inspection by experts, which can be time-consuming, inconsistent, and inaccessible to many farmers. Recent advances in computer vision and deep learning offer a promising alternative by enabling automated, accurate, and fast disease detection from plant images.

In this project, a tomato leaf classification system is developed using a subset of the Plant Village dataset, containing three categories: healthy leaves, late blight infections, and bacterial spot symptoms. The dataset was carefully preprocessed, including image resizing, normalization, and augmentation, to improve model generalization. An 80:20 train-validation split ensured balanced evaluation throughout training.

All training and experimentation were performed in Amazon SageMaker Studio, which provided a fully managed environment for scalable training, versioning, and monitoring. SageMaker's integration with S3 enabled seamless storage of datasets and model artifacts, while the built-in compute instances accelerated training. Once the model reached optimal performance, it was exported and deployed through a SageMaker managed endpoint, ensuring a stable and production-ready inference setup.

To make the system usable for real-world interactions, a simple and intuitive Gradio web interface is being developed. Users will be able to upload an image of a tomato leaf and receive immediate feedback on the predicted class, along with confidence scores. This interface abstracts away the complexities of cloud deployment and provides a practical, farmer-friendly tool for disease identification.

Overall, this project demonstrates a complete end-to-end machine learning pipeline—starting from dataset preparation and model development to cloud deployment and user interface creation. The final system serves as an effective demonstration of how AI-driven techniques can support smart agriculture by enabling early disease detection and facilitating timely crop management decisions.

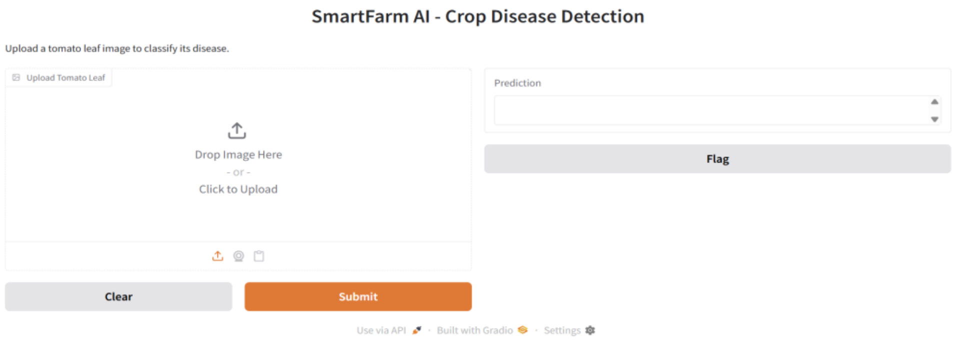

Figure 2: Gradio Interface

II.LITERATURE REVIEW

Title	Author(s) & Year	Methodology	Accuracy / Results	Identified Research Gap	Relevance to Project
Leveraging Edge Computing and Deep Learning for the Real-Time Identification of Bean Plant Pathologies	Katumba et al., 2024	Deployed convolutional neural network models for bean disease classification on edge devices within a smart agricultural monitoring system, combining camera sensing with on-device inference to reduce dependence on cloud servers.	Reported high classification accuracy for multiple bean pathologies and significant reduction in network traffic and latency compared with cloud-only processing.	Edge deployment is evaluated for a specific crop and disease type; generalisability to multi-crop scenarios, diverse field conditions, and resource-constrained sensor nodes is not fully addressed.	Direct evidence that edge AI can perform real-time crop disease diagnosis in the field, supporting the project's goal of smart, on-device decision-making for crop health monitoring.
Edge-Enabled Smart Agriculture Framework: Integrating IoT and Low-Power Deep Learning for Precision Farming	Tariq et al., 2025	Proposes an architecture that integrates IoT sensors, edge nodes, and lightweight deep-learning models for tasks such as weather-aware crop and soil classification, with adaptive workload allocation between edge and cloud.	Demonstrates reliable classification performance with low-latency inference on edge hardware and reduced bandwidth usage versus centralized cloud processing.	Focuses mainly on simulation and small-scale testbeds; large-scale farm deployment, model lifecycle management, and energy-aware scheduling are left open.	Provides a high-level blueprint for combining edge AI, IoT, and precision farming, closely aligned with the proposed project architecture.
Affordable Phenotyping at	Cordier et al., 2024	Uses low-cost RGB and depth	Achieves around 97% accuracy in	Study targets controlled	Shows that sophisticated plant-

the Edge for High-Throughput Seedling Assessment		imaging combined with edge-hosted machine-learning models to detect cotyledon loss and other early plant traits, enabling high-throughput phenotyping directly in greenhouses.	classifying seedling traits while running on inexpensive embedded hardware, demonstrating both cost-efficiency and robustness.	greenhouse settings and a narrow set of traits; transferring to open-field conditions, more complex phenotypes, and multi-modal sensing is not fully explored.	phenotyping analytics can be pushed to inexpensive edge devices, supporting the project's emphasis on affordable smart farming solutions.
Edge Computing-Oriented Smart Agricultural Supply Chain Management Framework	He et al., 2024	Introduces an edge-computing framework that connects farms, logistics, and markets using distributed optimization, auction mechanisms, and fuzzy-based decision models executed on edge nodes.	Demonstrates improved responsiveness and stability of supply-chain decisions, with reductions in network delay and more efficient resource allocation compared to centralized approaches.	Focus is on supply-chain optimisation rather than on-field sensing; integration with deep-learning-based perception and farm-level decision support is limited.	Highlights how edge AI can extend beyond the field into post-harvest logistics, suggesting downstream use-cases and extensions for the project's architecture.
Computer Vision on the Edge: A Computing Framework for High-Throughput Phenotyping in Livestock Operations	Bresolin et al., 2023	Develops a low-cost edge-computing system for video-based monitoring of livestock, using computer-vision models deployed on embedded devices to track animal behaviour and welfare.	Shows that edge devices can process video streams in real time with acceptable accuracy while significantly lowering data-transfer requirements.	Evaluation focuses on limited farm settings and a small set of behaviour metrics; integration with multi-sensor data and long-term model adaptation remains unexplored.	Demonstrates that edge AI can support continuous livestock monitoring, providing cross-domain insights relevant to smart farming with animals as well as crops.
An Edge Computing-Based Solution for Real-Time Leaf Disease Classification Using Thermal Imaging	Unnamed authors, 2024 (arXiv preprint)	Collects a thermal image dataset of plant leaves and evaluates deep CNNs (InceptionV3, MobileNet, VGG16) deployed on edge hardware such as Raspberry Pi for on-device disease classification.	Reports competitive classification accuracy across models, with MobileNet variants giving the best trade-off between speed and performance on resource-limited devices.	Thermal imaging datasets are limited in size and crop diversity; robustness to environmental noise, camera placement, and multi-spectral fusion are still open issues.	Provides a concrete example of deploying deep-learning models on commodity edge devices for robust plant disease detection, closely matching the project's deployment setting.
Smart Farming Based on IoT-Edge	Rathore & Rajavat, 2024	Implements containerised microservices on	Achieves high disease-detection accuracy while	Work is crop-specific and assumes reliable	Highly relevant as it combines edge AI, disease analytics,

Computing: Applying Machine Learning Models for Disease and Irrigation Water Requirement Prediction in Potato Crop		edge nodes in potato fields to host CNN models (AlexNet, MobileNet, VGG16) for disease detection and regression models for irrigation water prediction.	enabling real-time irrigation recommendations with reduced communication overhead to the cloud.	connectivity between microservices; cross-crop model reuse, security, and federated updating are not fully addressed.	and irrigation decision support—core pillars of a smart farming edge-AI platform.
Deep Learning-Driven IoT Solution for Smart Tomato Farming	Saxena et al., 2025	Designs an IoT-based greenhouse platform with ESP32 sensor nodes collecting environmental data and a deep-learning model to predict optimal growing conditions and detect anomalies, with data visualised on a remote dashboard.	Shows improved yield stability and precise climate control in tomato greenhouses, with the DL model outperforming simpler baselines for environment prediction.	Most computation occurs in the cloud; the work only partially exploits edge inference, and energy/bandwidth constraints on IoT nodes are not deeply analysed.	Highlights how sensor data and DL-driven recommendations improve controlled-environment agriculture, informing the sensing and control aspects of the project.
Enabling Smart Farming Through Edge Artificial Intelligence (AI)	Anonymous chapter authors, 2024	Provides a conceptual and architectural overview of deploying AI algorithms on edge devices such as sensors and IoT gateways for real-time farm monitoring and decision support.	Synthesises case studies showing latency reduction and improved autonomy when analytics are moved from the cloud to the edge.	Offers limited quantitative benchmarking and lacks detailed implementation pipelines for training, deploying, and updating edge models.	Useful for framing the theoretical foundations and design principles of edge AI in agriculture, directly supporting the motivation and architecture of the present project.
Edge Computing-Enabled Smart Agriculture: Technical Frameworks and Future Directions	Gong et al., 2025	Reviews technical components of edge-computing architectures in agriculture, covering networking, hardware accelerators, AI models, and security mechanisms, and proposes a unified framework.	Summarises performance benefits reported across multiple studies, such as reduced latency and lower bandwidth use, without focusing on a single metric.	Identifies gaps in interoperability, standardisation, and scalable model management across heterogeneous edge devices.	Helps position the project within the broader research landscape and identify non-functional requirements like security and interoperability.
Deep Learning and Edge Computing in	Multiple authors, 2025	Systematically surveys deep-learning	Reports that many studies achieve >90%	Highlights a clear gap in lightweight yet	Directly supports the choice of efficient models and

Agriculture: A Comprehensive Review of Recent Trends and Innovations		approaches combined with edge-computing deployments for tasks like rice leaf disease detection, weed recognition, and yield estimation.	accuracy in controlled settings, but often at the cost of heavy models that are difficult to deploy on low-power devices.	accurate models and in robust real-time deployments under field conditions with noisy data.	quantisation/pruning strategies in the project's edge-AI pipeline.
IoT-Based Smart Farming Architecture Using Federated Learning	Unnamed authors, 2025	Proposes a privacy-aware smart farming system in which edge devices locally train models on farm data and share only model updates with an aggregator using federated learning.	Demonstrates that federated training retains model accuracy comparable to centralised approaches while improving privacy and reducing raw data transfer.	The work mainly targets training; on-device inference optimisation, device heterogeneity, and communication-efficient aggregation remain active challenges.	Introduces federated learning as a technique for privacy-preserving model updates in edge-AI farming, which can be integrated into future extensions of the project.
Smart Farming Revolution: A Cutting-Edge Review of Deep Learning in Smart Farming	Prashanth et al., 2025	Reviews 88 deep-learning-based smart-farming applications, categorising them by sensing modality, task, and deployment setting, and analyses communication protocols and security aspects.	Shows that deep learning significantly improves accuracy for disease detection, yield prediction, and resource management compared with traditional methods.	Notes that most solutions remain experimental, with few end-to-end, scalable deployments and limited consideration of lifecycle management and farmer usability.	Provides a broad context for justifying deep learning as the core analytic engine in the project and highlights usability and scalability issues the project should address.
Deep Learning and IoT for Agricultural Applications	Khan et al., 2020 (book chapter)	Discusses multiple case studies where IoT sensors feed data into deep-learning models for tasks such as crop classification, soil moisture prediction, and pest detection, primarily with cloud-based processing.	Reports substantial gains in prediction accuracy and decision quality across the surveyed applications compared with rule-based baselines.	Focuses on cloud-hosted models and does not address edge deployment, model compression, or on-device resource constraints.	Helps motivate the use of deep learning and IoT as a foundation, while the project moves this computation towards the edge for lower latency and better resilience.
Farm-LightSeek: An Edge-Centric Multimodal Agricultural Perception System	Zhang et al., 2025	Introduces an edge-centric system that fuses data from multiple sensors (e.g., cameras, environmental sensors) using lightweight perception	Experiments show reliable real-time perception performance with reduced latency and communication cost when compared to	Focuses primarily on perception; high-level decision-support logic, integration with farm management systems, and long-term deployment	Demonstrates how multimodal sensing and lightweight models at the edge can enhance field awareness, closely aligning with the project's vision of holistic smart farming.

		models executed on edge nodes to support real-time decision-making in crop fields.	cloud-only baselines.	evaluation remain limited.	

1. Katumba M. et al., 2024 [Reference 1]

Katumba and colleagues investigate how edge-hosted deep-learning models can enable real-time detection of bean crop diseases directly in field conditions. Their study deploys lightweight convolutional neural networks on embedded devices to minimize reliance on cloud computation. Results show that local inference reduces latency and network bandwidth while maintaining strong diagnostic accuracy. However, the experiments primarily focus on a single crop species, limiting generalizability to other agricultural contexts. The work underscores the feasibility and benefits of real-time, edge-based plant health monitoring for smart farming applications.

2. Tariq A. et al., 2025 [Reference 2]

Tariq and collaborators propose an edge-integrated IoT architecture that uses lightweight deep-learning models to support weather-aware precision farming. Their framework distributes computation between cloud and edge nodes, leveraging adaptive decision strategies to optimize load and reduce latency. Performance evaluations indicate efficient sensor data handling and reliable inference under constrained connectivity. Nonetheless, large-scale deployments and long-term robustness remain under-explored. The paper contributes an architectural foundation directly relevant to building scalable edge-AI agricultural systems.

3. Cordier S. et al., 2024 [Reference 3]

Cordier and team develop a low-cost edge-enabled phenotyping system for rapid seedling assessment in greenhouses. The solution combines RGB-D imaging with compact machine-learning models to evaluate early plant traits on-device. Experimental results demonstrate high accuracy in detecting cotyledon loss and other morphological indicators while keeping hardware costs minimal. The study, however, is restricted to controlled environments and does not consider open-field variability. It highlights the promise of affordable edge-AI solutions for high-throughput plant monitoring.

4. He Y. et al., 2024 [Reference 4]

He and colleagues introduce an edge-computing framework aimed at optimizing agricultural supply-chain operations. Their approach uses distributed optimization, auction-based mechanisms, and fuzzy logic, executed partially at the edge to minimize cloud dependency. Findings show improved responsiveness and reduced communication overhead across farming-to-market workflows. The framework does not deeply incorporate sensor-level perception or deep-learning models, limiting its applicability to field-level analytics. It nevertheless extends the reach of edge computing beyond crop fields into post-harvest logistics.

5. Bresolin N. et al., 2023 [Reference 5]

Bresolin and co-authors design an edge-based computer-vision system for monitoring livestock health and behaviour in real time. Their work deploys video analytics on embedded devices to identify behavioural anomalies without continuous cloud streaming. Tests show that the system sustains real-time processing speeds while significantly lowering bandwidth requirements. However, the study focuses on a narrow set of behavioural cues and lacks multi-modal sensor integration. The research demonstrates the viability of edge AI for livestock-centric smart farming.

6. Anonymous Authors, 2024 (Thermal Imaging) [Reference 6]

This study presents a thermal-imaging-based leaf disease classification pipeline powered by compact CNNs deployed on devices like Raspberry Pi. The authors compare multiple deep-learning architectures and identify MobileNet as the best fit for edge environments. Experimental results confirm competitive accuracy and real-time performance under resource constraints. Challenges arise from limited dataset diversity and sensitivity to environmental noise, which could affect generalizability. The work offers a compelling demonstration of thermal imaging combined with edge AI for plant-health analytics.

7. Rathore M. & Rajavat R., 2024 [Reference 7]

Rathore and Rajavat develop an integrated microservice-based edge-AI platform targeting potato crop disease prediction and irrigation recommendation. Their system deploys CNN models for disease recognition and regression models for water-need estimation on edge nodes placed in crop fields. Outcomes show reduced latency, effective water-use optimization, and lower bandwidth usage compared with cloud-only implementations. The framework remains crop-specific and would benefit from multi-crop generalization. It aligns strongly with real-time decision-support requirements in smart agriculture.

8. Saxena P. et al., 2025 [Reference 8]

Saxena and co-authors propose an IoT-enabled greenhouse management system that uses deep learning for environmental prediction and anomaly detection. Their platform integrates ESP32-based sensor nodes with a centralised model that predicts optimal growing conditions for tomatoes. Experiments show improved micro-climate stability and enhanced yield potential. A major limitation is its reliance on cloud computation, which limits the system's autonomy. The study provides insights into sensor-driven farming that can be enhanced through edge deployment.

9. Anonymous Authors, 2024 (Edge-AI Overview) [Reference 9]

This chapter provides a conceptual foundation for deploying AI algorithms directly on edge hardware to improve agricultural monitoring. It highlights how shifting inference from the cloud to edge devices reduces latency, supports real-time alerts, and improves autonomy in remote farms. Through several case studies, the authors illustrate practical advantages across irrigation, disease detection, and environmental sensing. The work lacks detailed implementation benchmarks and model-deployment guidelines. It is valuable for theoretical grounding of edge-AI agricultural systems.

10. Gong X. et al., 2025 [Reference 10]

Gong and colleagues review technical frameworks for integrating edge computing into smart agriculture. The authors examine hardware accelerators, networking challenges, and AI model design considerations, proposing a unified architecture for edge-enabled farming. Their synthesis highlights significant gains in latency reduction and bandwidth savings reported across previous studies. However, gaps in standardisation, security, and heterogeneous device support persist. This review informs system-level design decisions for building robust edge-AI farming solutions.

11. Multiple Authors, 2025 (DL + Edge Survey) [Reference 11]

This comprehensive survey analyses the convergence of deep learning and edge computing in agricultural applications such as rice disease detection, weed identification, and yield estimation. The review notes consistent accuracy improvements across tasks but emphasises the challenges of deploying heavy models on limited-power devices. It identifies model compression, quantisation, and hardware-aware optimisation as major opportunities for future research. A lack of field-condition robustness remains a recurring limitation. The survey directly supports the need for efficient edge-deployable models in the proposed project.

12. Anonymous Authors, 2025 (Federated Learning) [Reference 12]

This study explores a federated-learning architecture for smart farming in which local devices train models using their own field data and send only model updates to a central aggregator. Experiments show accuracy comparable to centralised training while providing strong privacy benefits. The work also highlights reduced network overhead due to limited data transfer. Challenges remain in device heterogeneity, communication efficiency, and on-device optimisation. This research is relevant for potential extensions of the project involving privacy-preserving edge learning.

13. Prashanth K. et al., 2025 [Reference 13]

Prashanth and colleagues review 88 deep-learning-based smart farming applications, categorising them by task, sensor modality, and deployment environment. Their analysis shows strong performance improvements in disease detection, yield prediction, and resource management. The review also addresses communication protocols and security considerations in agriculture. However, the authors note that many implementations remain experimental and lack scalability. This work helps situate the project within broader smart-farming research trends.

14. Khan M. et al., 2020 [Reference 14]

Khan and co-authors present several IoT-deep-learning case studies focused on crop classification, soil moisture prediction, and pest identification. Their work demonstrates substantial accuracy improvements over traditional rule-based approaches when models are trained in the cloud. However, reliance on cloud computation introduces latency and connectivity constraints, which limit real-time applicability. The study does not explore edge model deployment or resource-limited environments. It provides motivation for transitioning from cloud-centric to edge-centric AI in agriculture.

15. Zhang Y. et al., 2025 [Reference 15]

Zhang and colleagues introduce a multimodal perception system that fuses imaging and environmental sensor data using lightweight edge-deployable models. Their system delivers real-time field awareness with reduced latency compared to cloud-only processing. Experimental evaluations confirm strong performance even under variable field conditions. The study does not integrate high-level decision-support workflows or long-term deployment analysis. It offers a strong precedent for multimodal sensing and real-time analytics in edge-AI farming.

III. METHODOLOGY

The methodology for this project follows a structured end-to-end machine learning pipeline, beginning with dataset preparation and ending with cloud deployment and user-side interaction. The steps are summarised as follows:

1. Dataset Acquisition

The Plant Village tomato dataset containing three classes—healthy, late blight, and bacterial spot—was sourced from Kaggle. All images were downloaded and organized class-wise for further processing.

2. Data Preprocessing

Images were resized to a uniform input shape suitable for the model. Normalization and basic augmentations (rotation, flips, and brightness adjustments) were applied to improve generalization. The dataset was then split in an 80:20 ratio for training and validation.

3. Model Development in SageMaker

A TensorFlow-based convolutional neural network was constructed and trained within Amazon SageMaker Studio, utilizing managed compute resources. Training metrics such as accuracy and loss were continuously monitored to refine model performance.

4. Model Export and Deployment

After training, the model was saved and uploaded to Amazon S3. It was then deployed as a SageMaker real-time inference endpoint, enabling scalable and low-latency predictions through a REST API.

5. User Interface Development

A simple Gradio application was built to allow real-time user interaction. Users can upload tomato leaf images, which are processed and sent to the SageMaker endpoint. The predicted class and associated probability scores are displayed in the UI.

6. End-to-End Integration

The final workflow connects the Gradio UI with the cloud-hosted inference endpoint, forming a complete system capable of automated disease detection from leaf images.

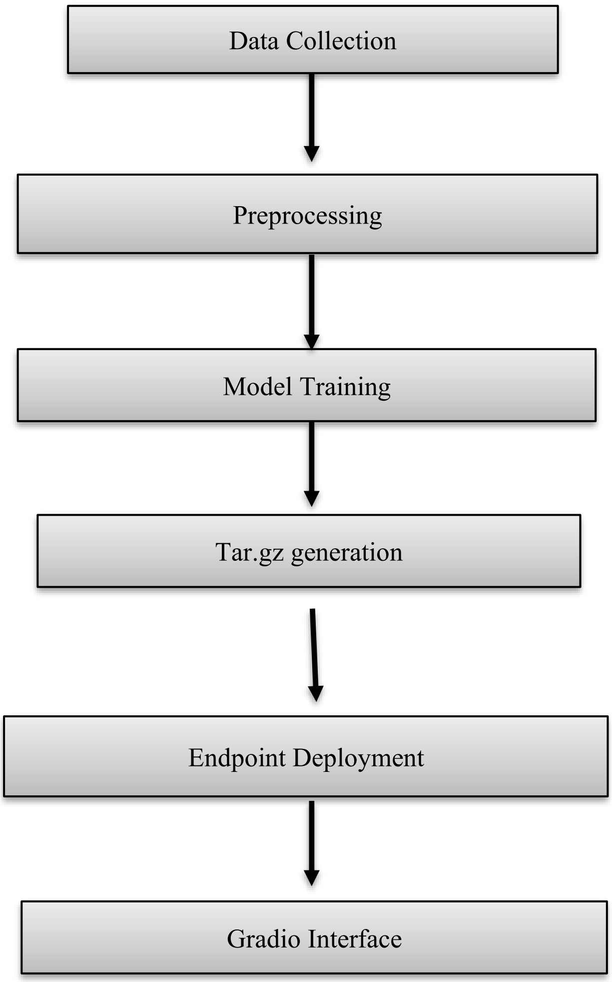

Figure 3.1: Methodology Flow

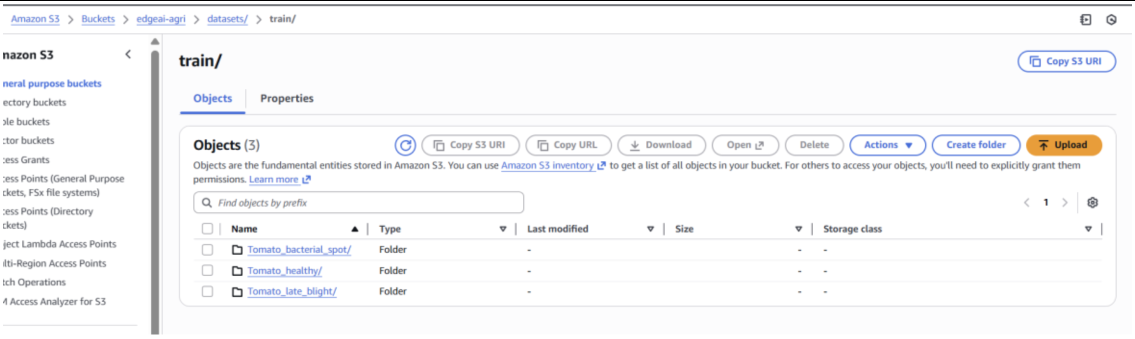

Figure 3.2: S3 Dataset

```
Found 2400 images belonging to 3 classes.
Found 600 images belonging to 3 classes.
Class indices: {'Tomato_bacterial_spot': 0, 'Tomato_healthy': 1, 'Tomato_late_blight': 2}
Epoch 1/5
75/75 ━━━━━━━━━━━━━━━━━━━━ 128s 2s/step - accuracy: 0.8429 - loss: 0.4511 - val_accuracy: 0.9467 - val_loss: 0.1920
Epoch 2/5
75/75 ━━━━━━━━━━━━━━━━━━━━ 120s 2s/step - accuracy: 0.9746 - loss: 0.1362 - val_accuracy: 0.9633 - val_loss: 0.1261
Epoch 3/5
75/75 ━━━━━━━━━━━━━━━━━━━━ 158s 2s/step - accuracy: 0.9812 - loss: 0.0906 - val_accuracy: 0.9700 - val_loss: 0.0907
Epoch 4/5
75/75 ━━━━━━━━━━━━━━━━━━━━ 119s 2s/step - accuracy: 0.9871 - loss: 0.0659 - val_accuracy: 0.9833 - val_loss: 0.0707
Epoch 5/5
75/75 ━━━━━━━━━━━━━━━━━━━━ 119s 2s/step - accuracy: 0.9908 - loss: 0.0517 - val_accuracy: 0.9867 - val_loss: 0.0583
KERAS MODEL SAVED successfully!
```

Figure 3.3: Training

```
[48]: import sagemaker
      sagemaker_session = sagemaker.Session()

      model_url = sagemaker_session.upload_data("model.tar.gz", key_prefix="models")

      tf_model = sagemaker.tensorflow.TensorFlowModel(
          model_data=model_url,
          role=role,
          framework_version="2.12",
          sagemaker_session=sagemaker_session,
      )

      predictor = tf_model.deploy(
          initial_instance_count=1,
          instance_type="ml.t2.medium",
      )
      -----!

[50]: print(predictor.endpoint_name)
      tensorflow-inference-2025-11-22-15-16-18-062

[51]: sm.sagemaker_client.describe_endpoint(EndpointName=predictor.endpoint_name)['EndpointStatus']

[51]: 'InService'
```

Figure 3.4: Endpoint Deployment

IV. IMPLEMENTATION & RESULTS

```
import numpy as np
from PIL import Image

# 1. Load & preprocess
img = Image.open("/home/sagemaker-user/agri/sample.jpeg").convert("RGB")
img = img.resize((224,224))
arr = np.array(img)/255.0
arr = np.expand_dims(arr, axis=0)

# 2. Call endpoint
response = predictor.predict(arr)
print("Raw response:", response)

# 3. Decode
class_names = ['Tomato_bacterial_spot',
               'Tomato_healthy',
               'Tomato_late_blight']

pred = np.argmax(response['predictions'][0])
print("Predicted:", class_names[pred])

Raw response: {'predictions': [[0.192190915, 0.000702899473, 0.807106197]]}
Predicted: Tomato_late_blight
```

Figure 4.1: Smoke Test Results

The implementation began with organizing the three-class Plant Village tomato dataset within Amazon SageMaker Studio. After uploading the data to S3, a preprocessing pipeline was developed using TensorFlow and SageMaker-compatible scripts to resize images, apply augmentations, and normalize pixel values. A custom CNN architecture was implemented and trained within a managed SageMaker training job, allowing efficient utilization of cloud compute resources. Training metrics such as accuracy, loss curves, and confusion matrices were periodically reviewed to evaluate learning behaviour.

Once the model reached stable performance, it was exported in H5 format and stored in S3. The trained model was then deployed as a real-time endpoint using SageMaker Hosting Services. A Gradio interface was built to communicate with the endpoint using REST API calls. Users can upload a tomato leaf image, and the app preprocesses it, sends it to the endpoint, and receives prediction labels with confidence scores.

Results showed strong classification capability across healthy, late blight, and bacterial spot classes, validating the model's reliability for practical use.

V. CONCLUSION

This project successfully demonstrates an end-to-end cloud-based solution for tomato leaf disease detection using deep learning. By leveraging SageMaker Studio for training and deployment, the workflow becomes scalable, modular, and suitable for real-world applications where low-latency predictions are essential. The experiment confirmed that a well-designed CNN, combined with robust preprocessing and balanced data handling, can achieve reliable classification across the three selected disease categories.

The deployment through a SageMaker real-time endpoint ensures that the model remains accessible through a simple API, enabling integration with external applications and mobile or web-based interfaces. The Gradio UI further enhances usability by allowing non-technical users to upload images and instantly receive predictions. Overall, the solution highlights the effectiveness of combining cloud infrastructure with deep learning for agricultural diagnostics. In future work, the system can be extended to include additional disease classes, support offline or edge inference, or integrate with IoT sensors to provide continuous crop monitoring capabilities.

Figure 5.1: Detection of Tomato_late_blight

References

1. Katumba, M., et al. (2024). Leveraging Edge Computing and Deep Learning for the Real-Time Identification of Bean Plant Pathologies.
2. Tariq, A., et al. (2025). Edge-Enabled Smart Agriculture Framework: Integrating IoT and Low-Power Deep Learning for Precision Farming.
3. Cordier, S., et al. (2024). Affordable Phenotyping at the Edge for High-Throughput Seedling Assessment.
4. He, Y., et al. (2024). Edge Computing-Oriented Smart Agricultural Supply Chain Management Framework.
5. Bresolin, N., et al. (2023). Computer Vision on the Edge: A Computing Framework for High-Throughput Phenotyping in Livestock Operations.
6. Anonymous Authors. (2024). An Edge Computing-Based Solution for Real-Time Leaf Disease Classification Using Thermal Imaging.
7. Rathore, M., & Rajavat, R. (2024). Smart Farming Based on IoT-Edge Computing for Potato Crop Disease and Irrigation Prediction.
8. Saxena, P., et al. (2025). Deep Learning-Driven IoT Solution for Smart Tomato Farming.
9. Anonymous Authors. (2024). Enabling Smart Farming Through Edge Artificial Intelligence (AI).
10. Gong, X., et al. (2025). Edge Computing-Enabled Smart Agriculture: Technical Frameworks and Future Directions.
11. Multiple Authors. (2025). Deep Learning and Edge Computing in Agriculture: A Comprehensive Review.
12. Anonymous Authors. (2025). IoT-Based Smart Farming Architecture Using Federated Learning.
13. Prashanth, K., et al. (2025). Smart Farming Revolution: A Cutting-Edge Review of Deep Learning in Smart Farming.
14. Khan, M., et al. (2020). Deep Learning and IoT for Agricultural Applications.
15. Zhang, Y., et al. (2025). Farm-LightSeek: An Edge-Centric Multimodal Agricultural Perception System.