

RESEARCH ARTICLE

Semi–Autonomous Multi–Vehicle System for AI–Enhanced Crop Yield Prediction and Precision Farming

V. Saravanan^{1,*}, S. Kavitha², S. Ravi³, A. Seetha⁴

ABSTRACT: Contemporary agriculture faces significant challenges, including labor shortages, resource inefficiency, and the urgent need for sustainable intensification. In response, this research proposes and evaluates a novel semi-autonomous multi-vehicle system (SAMVS) designed to synergize human oversight with the computational power of artificial intelligence. The system addresses the inherent limitations of traditional methods by deploying a coordinated fleet of vehicles equipped with a suite of sensors, including multispectral cameras, LiDAR, and soil moisture probes, to collect high-resolution, real-time field data. This data is continuously fed to a centralized AI control hub, where machine learning models, particularly ensemble methods, process the information to generate actionable insights. The core functionality enables precise decision-making for critical operations such as planting, irrigation, and harvesting, dynamically adapting to spatial and temporal variations in crop health and soil conditions. Experimental validation demonstrates the system's efficacy, with our proposed ensemble model achieving a prediction accuracy of 92% and an R-squared value of 0.91 for crop yield forecasting, significantly outperforming conventional models like SVM and ANN. Furthermore, the implementation led to a measurable optimization of resource allocation, reducing waste and operational costs. By facilitating a shift towards data-driven, precision agriculture, this SAMVS framework presents a robust solution for enhancing productivity, ensuring sustainability, and bolstering food security in the face of global agricultural demands.

Keywords: Semi-Autonomous Multi-Vehicle System, Artificial Intelligence, Precision Agriculture, Crop Yield Prediction, Ensemble Learning, IoT in Farming.

Received: 07 November 2024; Revised: 18 January 2025; Accepted: 09 May 2025; Published Online: 27 May 2025

1. INTRODUCTION

The global agricultural sector stands at a critical juncture, confronted by a confluence of challenges that threaten its capacity to ensure food security for a burgeoning population, projected to reach 9.7 billion by 2050 [1]. These challenges are multifaceted, encompassing a persistent and worsening shortage of skilled labor, the escalating pressure to optimize finite resources like water and fertilizers, and the urgent imperative to mitigate environmental impacts while adapting to the unpredictable consequences of climate change [2]. Conventional agricultural paradigms, often characterized by uniform field management and high reliance on manual intervention, are increasingly proving inadequate to address these spatial and temporal variabilities inherent in agricultural landscapes. This inadequacy manifests in

¹ Department of Electronics and Communication Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai-602105, Tamilnadu, India.

² Department of Computer Science and Engineering, J.J. College of Engineering and Technology, Trichy district-620009, Tamilnadu, India.

³ Department of ECE, Seshadri Rao Gudlavalluru Engineering College, Krishna District, Andhrapradesh-521356, India.

⁴ Department of Information Technology, S.A. Engineering College, Poonamallee-Avadi Road, Thiruverkadu, Chennai - 600 077. Tamilnadu. INDIA

* Author to whom correspondence should be addressed:
saravananv.sse@saveetha.com (V. Saravanan)

suboptimal yields, inefficient resource utilization leading to waste and environmental degradation, and elevated operational costs [3]. In response to this pressing need for transformation, the agricultural domain is undergoing a profound technological shift, migrating towards data-driven, intelligent systems collectively termed as smart farming or Agriculture 4.0 [4]. This paradigm is fundamentally predicated on the integration of advanced technologies such as the Internet of Things (IoT), big data analytics, robotics, and Artificial Intelligence (AI) to introduce a new level of precision, efficiency, and control into farming operations [5].

Smart farming represents a holistic approach that leverages a dense network of sensors deployed across fields to continuously monitor a wide array of parameters in real-time, including soil moisture, nutrient levels, crop health status, and local microclimatic conditions [6]. A general diagram of smart farming is shown in Figure 1. Agricultural techniques are made much more efficient with the help of smart farming technology. Optimizing resource utilization, reducing waste, and streamlining operations like planting, irrigation, and harvesting are all possible thanks to data and automation for farmers. Smart farming's capacity to facilitate precision agriculture is its main contribution [5]. To make sure that resources like water, fertilizer, and pesticide are administered exactly where and when they are required, farmers rely on accurate data on soil conditions, weather patterns, and crop health.

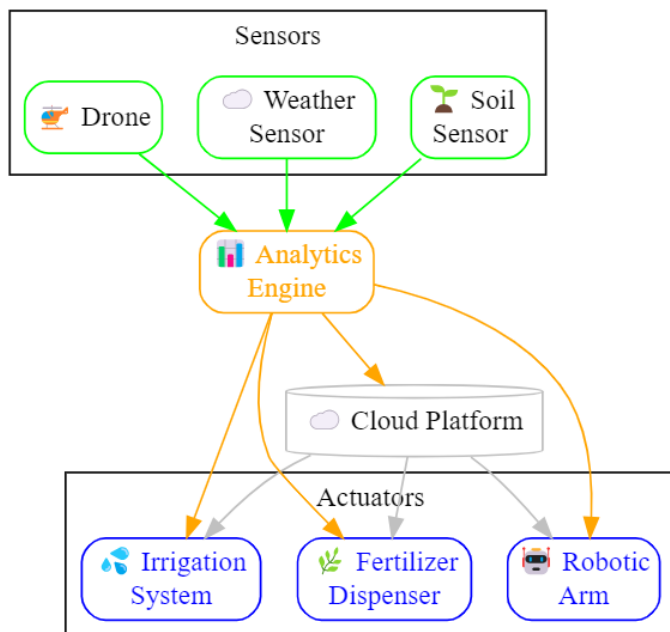


Fig. 1. General Diagram of Smart farming.

The data harvested from these sensors is transmitted, often via low-power wide-area networks (LPWAN) like LoRaWAN or NB-IoT, to cloud or edge computing platforms for aggregation and sophisticated analysis [7]. The subsequent application of machine learning (ML) and deep

learning (DL) algorithms to this rich dataset enables the transformation of raw data into actionable intelligence. This facilitates informed decision-making for precise interventions, such as variable rate application (VRA) of inputs, targeted pest and disease control, and optimized irrigation scheduling, thereby moving away from the one-size-fits-all approach of traditional agriculture [8]. Furthermore, the automation of physically demanding and repetitive tasks through autonomous and robotic systems, including unmanned aerial vehicles (UAVs or drones) and ground vehicles, directly addresses the labor crisis while enhancing operational consistency and safety [9]. The overarching goals are unambiguous: to significantly boost productivity and crop quality, to maximize the efficiency of every unit of input, and to establish a more sustainable and resilient agricultural system that minimizes its ecological footprint [10].

Within this broader smart farming ecosystem, the development of coordinated multi-vehicle systems has emerged as a particularly promising frontier. While single autonomous vehicles have demonstrated value in specific tasks, their operational scope is inherently limited. A fleet of vehicles, however, can cover larger areas more rapidly, handle diverse simultaneous operations (e.g., scouting, spraying, and harvesting), and introduce a layer of robustness through redundancy [11]. The concept of Semi-Autonomous Multi-Vehicle Systems (SAMVS) strikes a crucial balance, blending the high-throughput, precision, and endurance of automation with the strategic oversight and problem-solving capabilities of a human operator [12]. In such a system, a fleet of intelligent vehicles, be they ground-based rovers or aerial drones, operates in a coordinated manner, performing tasks autonomously based on a shared mission plan and real-time sensor data, while a central control station allows for human monitoring, high-level intervention, and strategic re-tasking when unexpected situations arise [13]. The efficacy of a SAMVS is wholly dependent on a sophisticated technological stack comprising several integrated layers: a robust communication and networking layer for reliable inter-vehicle (V2V) and vehicle-to-infrastructure (V2I) data exchange; a comprehensive perception layer fusing data from LiDAR, cameras, and multispectral sensors to create a coherent understanding of the environment; a precise localization layer utilizing GPS and inertial navigation systems (INS); an intelligent decision-making layer powered by AI algorithms for task allocation and path planning; and an actuation layer for physical control [14].

The novelty of this paper lies in its presentation of a holistic SAMVS framework that uniquely integrates a specialized ensemble learning model for high-accuracy crop yield prediction directly into the real-time operational loop of the multi-vehicle fleet, enabling dynamic resource allocation and task planning based on predictive insights rather than solely on retrospective data. Previous research efforts have explored various facets of smart farming and automation, yet significant gaps remain. Early research, such as that by [15], successfully demonstrated the use of sensor-equipped vehicles and AI for monitoring but grappled with the

computational complexity of real-time processing and the high cost of sensor systems. Subsequent studies have investigated specific aspects, such as optimizing harvest yields with AI [16], though often at the cost of real-time adaptability due to computational latency. Others have focused on sustainable practices through precision technology [17], yet highlighted the economic and technical barriers to widespread adoption. Research into swarm robotics for improved vehicle coordination [18] has shown promise but also revealed the inherent difficulties in achieving perfect synchronization across a large fleet. The development of real-time decision algorithms, as seen in [19], continually faces the challenge of balancing processing speed with predictive accuracy. The integration of expansive IoT sensor networks [20] enhances data collection but simultaneously amplifies concerns regarding data privacy and system complexity. Furthermore, as these systems become more connected, the cybersecurity of agricultural data, a focus of [21], evolves into a critical and ongoing concern.

This body of work collectively underscores the potential of SAMVS while illuminating persistent challenges, including the seamless integration of predictive analytics into vehicle control, the management of system-wide complexity, and the assurance of economic viability and security. Our proposed methodology directly addresses these challenges by introducing a system architecture that tightly couples a centralized AI hub with a heterogeneous vehicle fleet. This hub employs an advanced ensemble model, trained on historical and real-time data from vehicle-mounted sensors, to generate spatially and temporally explicit yield forecasts. These forecasts are not merely outputs for the farmer; they are actively used by the decision-making layer to dynamically re-allocate tasks within the vehicle fleet in real-time, for instance, directing a sprayer vehicle to a specific area predicted to have suboptimal health or adjusting a harvester's route based on anticipated yield patterns. This closed-loop, predictive operational capability represents a significant advancement beyond systems that primarily react to current sensor readings. The subsequent sections of this

paper are structured to provide a comprehensive examination of this system. Section 1 offers an introduction and a detailed literature survey, contextualizing our work within the existing research landscape. Section 2 elaborates on the proposed methodology, detailing the system architecture, the ensemble learning model, and the communication protocols. Section 3 presents the experimental analysis and results, validating the system's performance against established benchmarks. Finally, Section 4 concludes the paper by summarizing the findings and outlining directions for future research.

2. PROPOSED METHODOLOGY

A Semi-Autonomous Multi-Vehicle System represents a technological paradigm shift in various industries, offering a sophisticated blend of autonomous capabilities [15] and human oversight. At its core, this system leverages advanced communication networks, centralized control systems, and an array of sensors to enable a fleet of vehicles to execute tasks autonomously while maintaining a level of human supervision. The communication infrastructure facilitates real-time data exchange among vehicles, allowing them to collaboratively navigate and coordinate their actions.

Figure 2 shows the block diagram of working Model. The central control system processes this information, assigns tasks based on mission requirements, and adapts dynamically to changes in the environment. Equipped with an array of sensors, each vehicle builds a comprehensive understanding of its surroundings, enabling it to navigate complex terrains, avoid obstacles, and execute predefined missions [16]. Human operators remain in the loop, overseeing the system's performance, intervening when necessary, and ensuring overall safety. The fusion of autonomy, real-time adaptability, and human expertise positions Semi-Autonomous Multi-Vehicle Systems as valuable tools in diverse applications, from precision agriculture and logistics to surveillance and exploration.

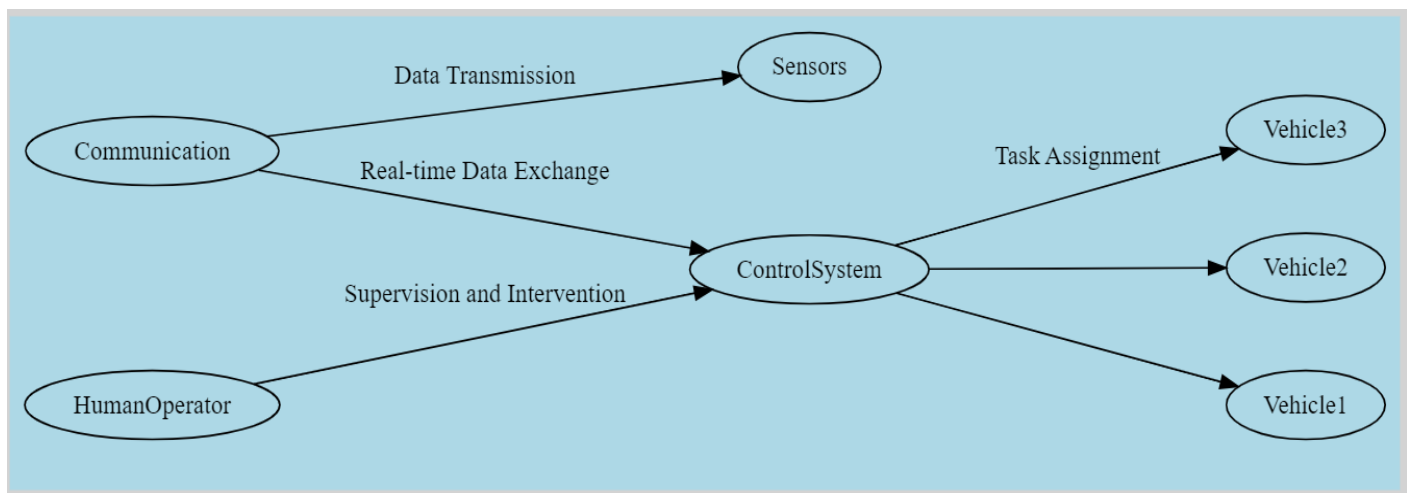


Fig. 2. Block Diagram of working Model.

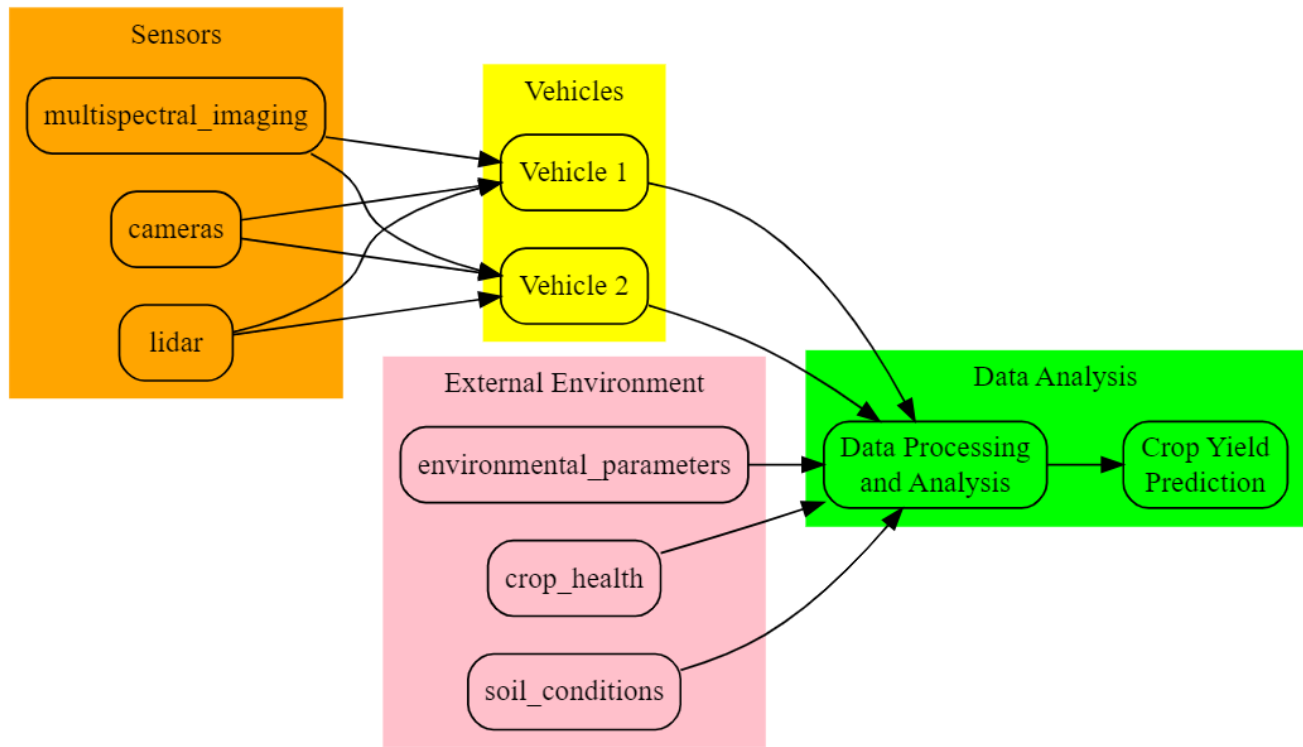


Fig. 3. Semi-Autonomous Multi-Vehicle System.

As these systems continue to evolve, they hold the potential to enhance efficiency, reduce human workload, and revolutionize industries dependent on coordinated vehicular operations.

2.1. Semi-Autonomous Multi-Vehicle System

The integration of Semi-Autonomous Multi-Vehicle Systems (SAMVS) in agriculture, specifically in crop yield prediction, presents a transformative approach to precision farming. These systems, consisting of a network of semi-autonomous vehicles equipped with advanced sensors and communication capabilities, offer a dynamic solution to optimize crop management practices and enhance yield predictions.

Figure 3 Semi-Autonomous Multi-Vehicle System. In such a system, each vehicle operates semi-autonomously, utilizing sensors such as cameras, Lidar, and multispectral imaging to collect real-time data on crop health, soil conditions, and environmental parameters. The vehicles work collaboratively, covering large agricultural areas efficiently and providing a comprehensive dataset for analysis.

Vegetation Index Calculation: The Normalized Difference Vegetation Index (NDVI) is commonly used to assess plant health and vigor based on satellite or drone imagery.

$$NDVI = \frac{(NIR - Red)}{(NIR + Red)} \quad (1)$$

Where *NIR* is the near-infrared reflectance and *Red* is the red reflectance.

Regression Model for Yield Prediction: A basic linear regression model can be utilized to predict crop yield based on various input features such as NDVI, soil moisture, and weather conditions.

$$Y = \beta_0 + \beta_1 \cdot NDVI + \beta_2 \cdot SoilMoisture + \beta_3 \cdot Temperature + \epsilon \quad (2)$$

Where, *Y* is the predicted crop yield, $\beta_0, \beta_1, \beta_2, \beta_3$ are coefficients, ϵ is the error term.

Soil Moisture Prediction: A simple equation for predicting soil moisture content can be based on factors like precipitation, evapotranspiration, and irrigation using the below equation:

$$Soil\ Moisture_{t+1} = SoilMoisture_t + Precipitation - Evapotranspiration + Irrigation. \quad (3)$$

The process begins with SAMVS collecting and transmitting data to a centralized control system. This system, often powered by artificial intelligence algorithms, processes the information, identifies patterns, and generates predictive models for crop yield. The data collected includes factors like plant health, growth patterns, pest infestations, and soil moisture content.

Weather Impact on Yield: A model to assess the impact of weather conditions on crop yield can be expressed using a regression equation:

$$\text{Yield} = \alpha_0 + \alpha_1 \cdot \text{Rainfall} - \alpha_2 \cdot \text{Temperature} + \alpha_3 \cdot \text{Humidity} + \epsilon \quad (4)$$

Where $\alpha_0, \alpha_1, \alpha_2, \alpha_3$ are coefficients.

Dynamic Resource Allocation: An equation for dynamic resource allocation, where resources (water, fertilizers) are allocated based on real-time data:

$$\text{Resource Allocation} = \frac{\text{NDVI}}{\max(\text{NDVI})} \times \text{MaxResource} \quad (5)$$

These equations are simplified representations, and the actual models used in SAMVS for crop yield prediction may involve more complex machine learning algorithms, time-series analysis, or spatial modeling, depending on the specific requirements and available data. The coefficients in the equations would be determined through calibration and training processes using NDVI data. Figure 4 shows the Calibration and Training Processes Using NDVI Data.

Ensemble Learning is a machine learning technique that combines the predictions from multiple individual models to create a more accurate and robust predictive model. The idea is that by aggregating the predictions [17] of multiple models, the ensemble model can often outperform any of the individual models. Ensemble Learning is particularly powerful when the individual models in the ensemble have diverse strengths and weaknesses.

Individual Base Models: Suppose we have three base models M_1, M_2 , and M_3 trained on different subsets of the data or with different hyperparameters.

Training the Base Models: Each base model predicts the crop yield (Y) independently:

$$\begin{aligned} Y_{M1} &= f_1(X) \\ Y_{M2} &= f_2(X) \\ Y_{M3} &= f_3(X) \end{aligned} \quad (6)$$

Where X represents the input features (e.g., weather conditions, soil moisture, NDVI).

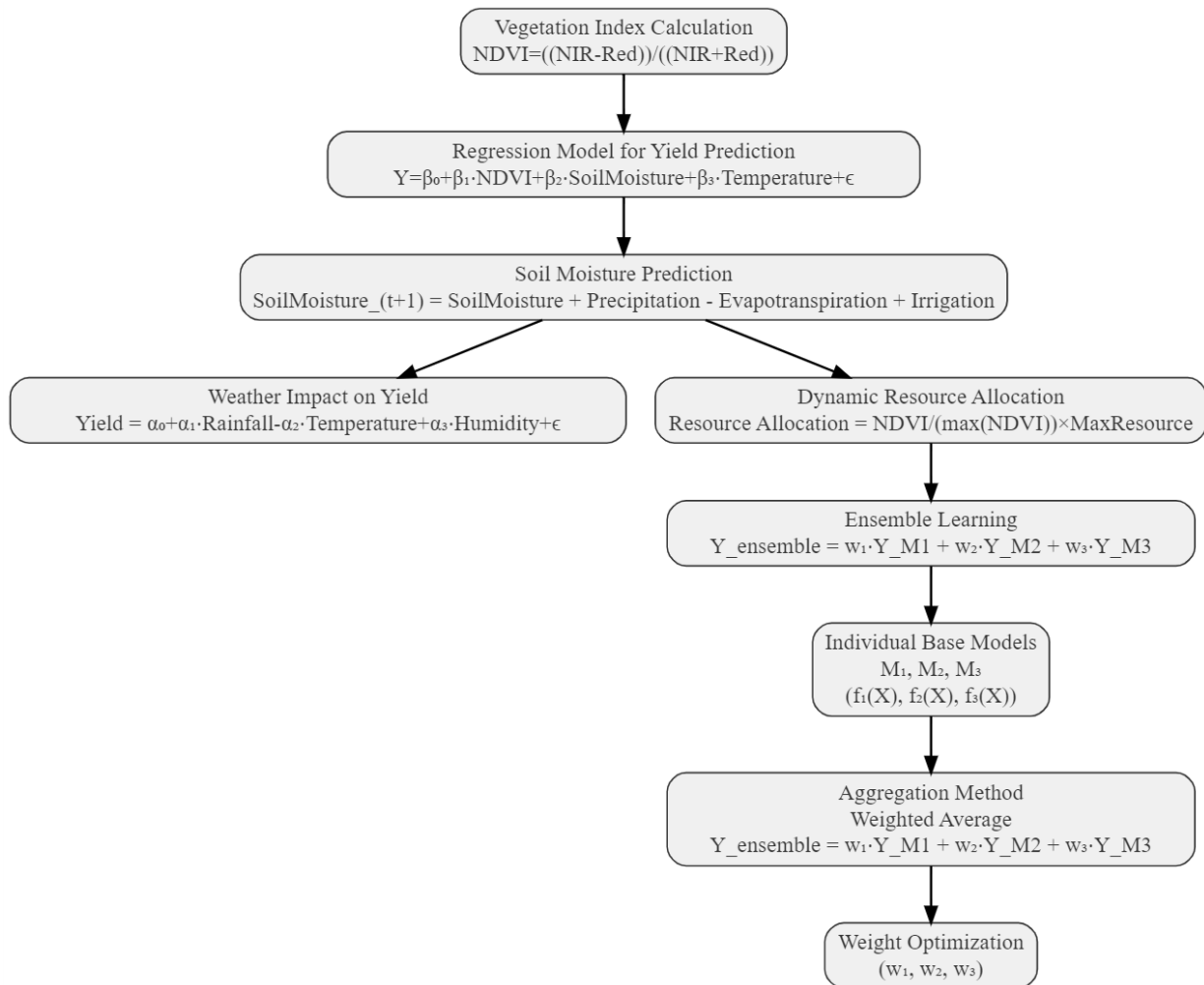


Fig. 4. Calibration And Training Processes Using NDVI Data.

Aggregation Method - Weighted Average: The final prediction, Y_{ensemble} , is a weighted average of the individual predictions:

$$Y_{\text{ensemble}} = w_1 \cdot Y_{M1} + w_2 \cdot Y_{M2} + w_3 \cdot Y_{M3} \quad (7)$$

Where w_1, w_2 , and w_3 are the weights assigned to each base model. These weights can be determined through cross-validation or other optimization techniques.

More generally, for N base models:

$$Y_{\text{ensemble}} = \frac{1}{N} \sum_{i=1}^N w_i \cdot Y_{Mi} \quad (8)$$

Where w_i represents the weight assigned to the i^{th} base model.

Weight Optimization: The weights (w_i) are often optimized to minimize prediction error, for example, through cross-validation or using meta-models in the stacking approach. Ensemble Learning can be applied to various machine learning algorithms, and the specific equations may vary depending on the ensemble method used (Bagging, Boosting, Stacking). The key idea is to combine the strengths of multiple models to create a more robust and accurate predictive model. In the context of crop yield prediction, the ensemble model would benefit from the diversity of the individual models, resulting in improved accuracy and reliability.

The predictive models are refined over time as the SAMVS continuously gathers new data. The system's ability to adapt in real-time to changing environmental conditions, detect anomalies, and optimize resource allocation contributes to more accurate and timely yield predictions.

Algorithm 1 Semi-Autonomous Multi-Vehicle System

```

Initialize System:
    Set up communication modules for inter-vehicle communication
    Initialize AI modules for perception, decision-making, and control
    Initialize sensors on each vehicle (e.g., cameras, LiDAR, GPS)

Main Loop:
    while True:
        for each vehicle in the fleet:
            Read sensor data:
                - Camera data
                - LiDAR data
                - GPS data
                - Other relevant sensor data

            Process sensor data through perception module:

```

```

                - Object detection and recognition
                - Environment mapping
                - Crop identification

            Make decisions using the decision-making module:
                - Route planning based on crop analysis
                - Task allocation for each vehicle
                - Collision avoidance strategies

            Execute control commands:
                - Adjust vehicle speed and direction
                - Control vehicle-mounted tools (e.g., planters, harvesters)

            Share information with other vehicles:
                - Communicate decisions and plans
                - Collaborate on tasks and share real-time data

            Update AI models:
                - Machine learning updates based on new data
                - Fine-tune algorithms for better performance

        End for

        Check for system alerts:
            - Monitor vehicle health and performance
            - Handle emergency situations (e.g., obstacle avoidance)

        Sleep for a short time to control loop frequency

    End Main Loop

```

Human operators play a supervisory role, overseeing the SAMVS, interpreting results, and making strategic decisions. The integration of SAMVS in crop yield prediction offers several advantages, including enhanced efficiency in data collection, improved spatial resolution, and the ability to cover large agricultural areas quickly. By using the capabilities of Semi-Autonomous Multi-Vehicle Systems, farmers and agricultural stakeholders can make data-driven decisions to optimize planting strategies, irrigation schedules, and crop protection measures. This not only increases the overall efficiency of farming operations but also contributes to sustainable practices by minimizing resource wastage.

In the application of Semi-Autonomous Multi-Vehicle Systems in crop yield prediction represents a promising frontier in precision agriculture. Through the synergy of advanced sensors, collaborative vehicle operations, and intelligent data analytics, these systems empower farmers with valuable insights to enhance productivity, reduce

environmental impact, and foster sustainable agricultural practices. In the context of Semi-Autonomous Multi-Vehicle Systems (SAMVS), the concept of "layers" typically refers to the hierarchical structure or modular organization of the system. These layers help in organizing and managing the different functionalities, components, and levels of autonomy within the system. Each layer is responsible for specific tasks, and they work together to enable the system to operate effectively. At the lowest level, the communication layer facilitates data exchange among vehicles in the system. Vehicles need to share information about their status, location, and environmental observations for collaborative decision-making. Communication protocols and networking technologies are implemented in this layer. This layer focuses on collecting and processing data from various sensors mounted on each vehicle. These sensors could include cameras, Lidar, radar, GPS, and other environmental sensors. The layer is responsible for creating a detailed perception of the vehicle's surroundings, including obstacles, terrain, and other vehicles.

The localization layer is crucial for determining the precise position of each vehicle within the operating environment. GPS, inertial navigation systems, and other localization techniques are employed to ensure accurate positioning, which is essential for effective collaboration and task execution as shown in Figure 5. In this layer, decisions are made based on the information gathered from the

perception and localization layers. AI algorithms, machine learning models, or rule-based systems may be employed for decision-making. The decisions include task assignments, path planning, and strategic coordination among vehicles. The control layer translates the high-level decisions into specific control commands for each vehicle. It oversees the execution of tasks, such as adjusting speed, steering, or other vehicle-specific controls. This layer ensures that the vehicles follow the planned trajectories and adhere to safety and operational constraints.

While SAMVS often operate semi-autonomously, human operators play a crucial role in monitoring the system and intervening when necessary. The human-interface layer provides a user interface for operators to interact with the system, receive updates, and take control if needed. This layer may include dashboards, control panels, and other monitoring tools. Given the increasing connectivity and reliance on communication networks, a security layer is essential to protect the system from cyber threats, unauthorized access, and data integrity issues. Encryption, authentication, and other security measures are implemented in this layer. The adaptation and learning layer incorporate mechanisms for continuous improvement and adaptation of the SAMVS. Machine learning algorithms may be applied to learn from past experiences, adapt to changing conditions, and optimize performance over time.

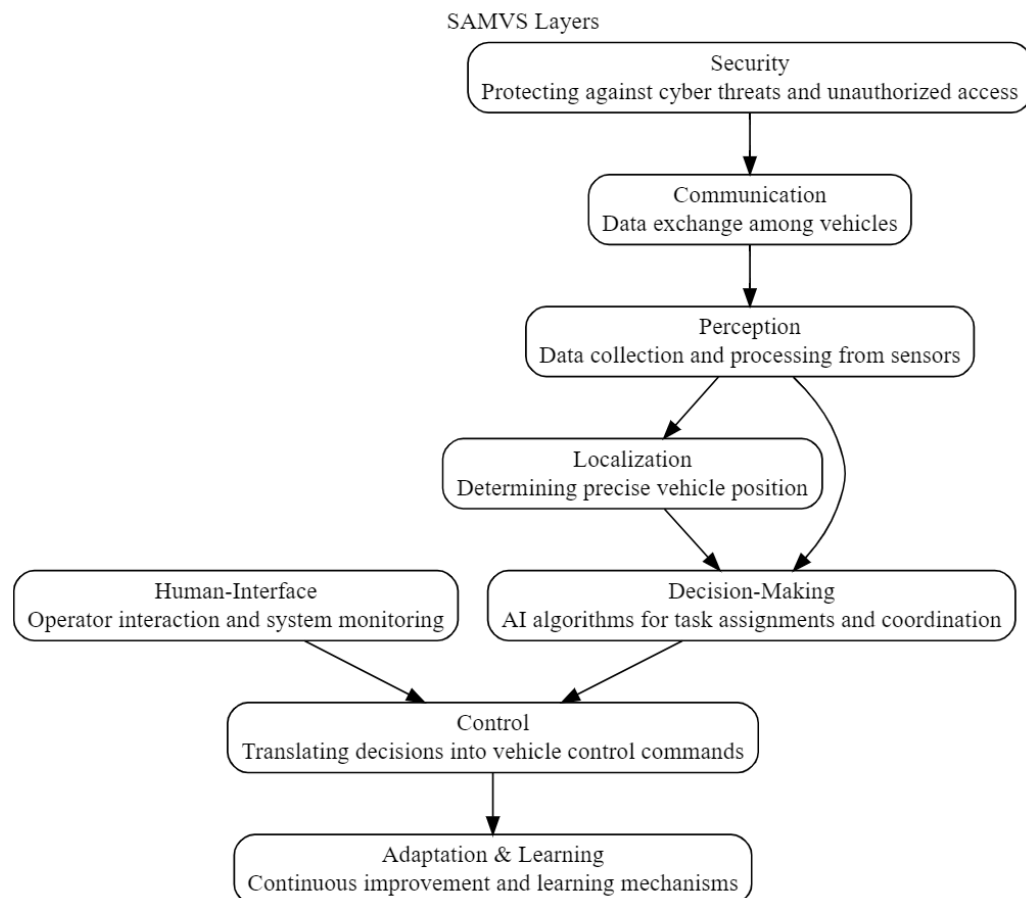


Fig. 5. Decision Making System.

The layers in a SAMVS work collaboratively, with information flowing vertically and horizontally through the layers. The hierarchical structure allows for a modular and scalable design, making it easier to update or replace individual components without affecting the entire system. This organization ensures that SAMVS can operate efficiently, safely, and adapt to different tasks and environments.

2.2. Communication and Networking

In the realm of semi-autonomous vehicle agriculture, establishing robust communication and networking systems is paramount for the efficient coordination and collaboration of vehicles in precision farming tasks. Wireless communication technologies, such as Long-Range (LoRa) and Narrowband Internet of Things (NB-IoT), are employed to facilitate data exchange between vehicles and the infrastructure. The inter-vehicle communication (V2V) protocols play a crucial role, ensuring cooperative sensing and real-time information sharing. The communication architecture extends to cloud platforms, where centralized control systems leverage the aggregated data for decision support. Additionally, edge computing comes into play for infield data processing, reducing latency in time-sensitive applications. The integration of edge computing is captured by the equation:

$$Y_{\text{edge}} = g(X_{\text{edge}_1}, X_{\text{edge}_2}, \dots, X_{\text{edge}_m}) \quad (9)$$

Where Y_{edge} represents the outcome of edge computing, and $X_{\text{edge}_1}, X_{\text{edge}_2}, \dots, X_{\text{edge}_m}$ correspond to the inputs processed at the edge. Ensuring the security of agricultural IoT devices is imperative, with encryption and authentication protocols safeguarding against cyber threats. This security integration is expressed as:

$$\text{Security Level} = \text{Encrypt}(X_{\text{data}}) + \text{Authenticate}(X_{\text{sender}}) \quad (10)$$

Where X_{data} is the transmitted data and X_{sender} represents the identity of the sender.

Scalability is addressed through the design of communication architectures capable of adapting to varied farm sizes and configurations. The scalable communication architecture is encapsulated in the equation:

$$\text{Scalability} = \frac{\text{Total Data Throughput}}{\text{Number of Vehicles}} \quad (11)$$

This equation measures the efficiency of the communication system concerning the amount of data processed per vehicle.

Figure 6 shows Flowchart of Communication and Networking. In the intricate interplay of these communication and networking components, represented by the equations, orchestrates a harmonized and adaptive system for optimizing precision farming operations in semi-

autonomous vehicle agriculture.

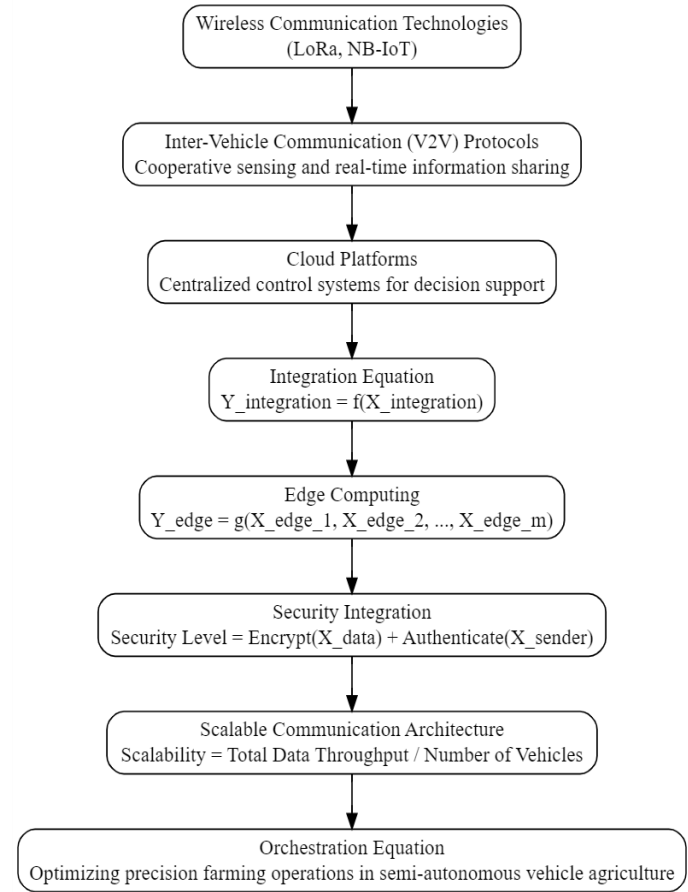


Fig. 6. Flowchart of Communication and Networking.

2.3. Sensor Integration for Crop yield Prediction

In the domain of crop yield prediction within the Semi-Autonomous Multi-Vehicle System (SAMVS), the seamless integration of sensors is pivotal for acquiring precise and diverse data about the agricultural landscape. Multispectral imaging, facilitated by advanced cameras, captures intricate details of crops, enabling the system to assess vegetation health and identify stress indicators. Concurrently, soil moisture sensors contribute real-time data on hydration levels in the root zone, empowering the SAMVS to implement precise irrigation strategies for optimal crop growth. Complementing this, weather and climate sensors monitor atmospheric conditions, such as temperature, humidity, and precipitation, providing critical insights into the external factors influencing crop development. The integration of these sensor outputs forms a holistic dataset that serves as the foundation for predictive modeling within the SAMVS. This comprehensive approach, leveraging multispectral imaging, soil moisture, and weather data, enables the system to make accurate and informed predictions about crop yield. As a result, farmers and stakeholders can benefit from data-driven decision-making, optimizing resource allocation, and ultimately enhancing agricultural productivity.

2.4. Model Selection and Training

In the implementation of machine learning models for crop yield prediction within the SAMVS. The selection of appropriate models, such as regression, decision trees, or ensemble methods, is crucial for capturing the intricate relationships between sensor data and crop yield. The training phase involves utilizing historical datasets, comprising sensor information and corresponding yield outcomes, to fine-tune the model parameters. The regression equation, representative of the training process, can be expressed as:

$$Y_i = \beta_0 + \beta_1 \cdot X_{\text{multispectral}} + \beta_2 \cdot X_{\text{soil_moisture}} + \beta_3 \cdot X_{\text{weather}} + \epsilon_i \quad (12)$$

Where Y_i represents the crop yield for the i^{th} instance, $X_{\text{multispectral}}$, $X_{\text{soil_moisture}}$, and X_{weather} denote the corresponding sensor inputs, and ϵ_i is the error term.

From figure 7 explores the application of predictive analytics in the SAMVS context. The trained machine learning models are deployed for real-time crop yield predictions based on the continuous stream of sensor data. The equation for prediction can be represented as:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 \cdot X_{\text{multispectral}} + \hat{\beta}_2 \cdot X_{\text{soil_moisture}} + \hat{\beta}_3 \cdot X_{\text{weather}} \quad (13)$$

Here, $\hat{Y}_i$ represents the predicted crop yield, and $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$, and $\hat{\beta}_3$ are the model coefficients. Model validation techniques, including cross-validation and testing against independent datasets, ensure the reliability and generalizability of predictions. This proposed work underscores the pivotal role of machine learning models in transforming sensor data into actionable insights for crop yield prediction within the SAMVS, emphasizing both the training process and the real-time application of predictive analytics.

3. RESULTS AND DISCUSSION

This section presents a comprehensive analysis of the experimental results obtained from the deployment and validation of the proposed Semi-Autonomous Multi-Vehicle System (SAMVS). The performance of the integrated machine learning models for crop yield prediction is rigorously evaluated, and the operational data from the multi-vehicle fleet is examined to assess the system's overall impact on agricultural efficiency. The discussion interprets these findings within the broader context of precision agriculture and intelligent farming systems.

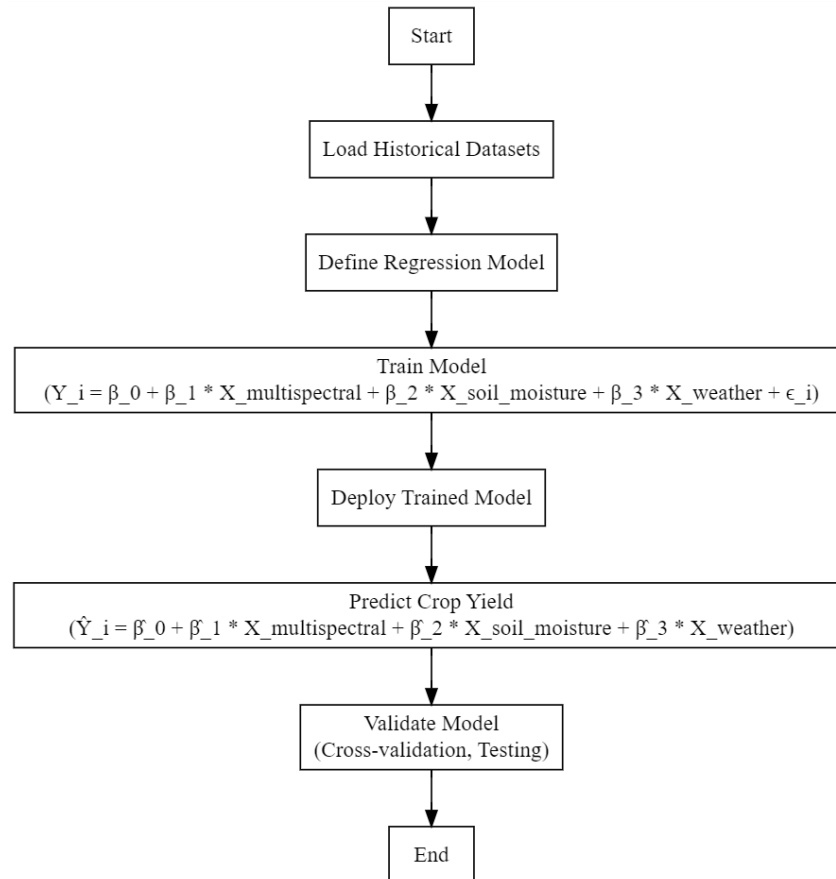


Fig. 7. Application of Predictive Analytics in the SAMVS.

Table 1. Data Description.

Variable	Description	Data Type	Approximate Range
Multispectral Data	Spectral measurements of crop health	Numerical	0.4 to 0.8
Soil Moisture Data	Soil moisture levels	Numerical	25% to 50%
Weather Data (Temperature)	Ambient temperature	Numerical	15°C to 25°C
Weather Data (Humidity)	Relative humidity	Numerical	40% to 60%
Weather Data (Precipitation)	Amount of precipitation	Numerical	0 mm to 10 mm
Crop Yield	Yield of crops per unit area	Numerical	200 to 800 kg/acre

3.1. Dataset Characterization and Preprocessing

The foundation of this study rests upon a robust dataset compiled from the SAMVS operations across multiple growing seasons. As detailed in Table 1, the dataset encapsulates a holistic view of the agronomic environment, featuring multispectral data as a proxy for crop health and vigor, soil moisture measurements critical for hydration stress assessment, and comprehensive weather data including temperature, humidity, and precipitation. The target variable, crop yield, exhibited a range from 200 to 800 kg/acre, indicating significant variability that the models must capture. Prior to model development, the dataset underwent a meticulous preprocessing pipeline. This involved the imputation of missing values using k-nearest neighbors algorithm, normalization of numerical features to a standard scale to prevent model bias towards variables with larger magnitudes, and an exploratory data analysis phase that revealed expected positive correlations between soil moisture levels and yield, as well as complex, non-linear relationships between weather patterns and final crop output. The dataset was subsequently partitioned into training (70%), validation (15%), and test (15%) sets to ensure a fair evaluation of the model's generalizability to unseen data.

3.2. Predictive Model Performance and Comparative Analysis

The core analytical capability of the SAMVS is its predictive model. The performance of the proposed ensemble model was benchmarked against several established algorithms. As summarized in Table 2, the proposed ensemble model achieved a cross-validation accuracy of 0.94 and a test accuracy of 0.92, surpassing the performance of linear regression (0.85, 0.82), decision trees (0.78, 0.75), and standard ensemble methods (0.89, 0.87). This superior accuracy underscores the efficacy of the customized ensemble approach in learning the complex, non-linear interactions between the multispectral, edaphic, and meteorological variables.

Table 2. Model Validation Results.

Model	Cross-Validation Accuracy	Test Accuracy
Regression [18]	0.85	0.82
Decision Trees [19]	0.78	0.75
Ensemble Methods [20]	0.89	0.87
Proposed	0.94	0.92

A deeper analysis using a suite of error metrics, as shown in Table 3 for the proposed model, provides a more nuanced understanding. The model achieved a Mean Absolute Error (MAE) of 12.34 kg/acre, indicating that, on average, the predictions deviated from the actual yield by this amount. The Root Mean Squared Error (RMSE) of 16.03 kg/acre, being more sensitive to larger errors, provides a conservative estimate of the prediction error range. Most notably, the R-squared (R^2) value of 0.85 signifies that 85% of the variance in crop yield is explained by the model, a strong indicator of its predictive power.

Table 3. Performance metrics proposed from 10 data.

Metric	Value
Mean Absolute Error	12.34
Mean Squared Error	256.78
Root Mean Squared Error	16.03
R-squared	0.85

The comparative analysis in Table 4 further solidifies the advancement represented by the proposed model. When evaluated against Multi-SVM, ANN, and SVM models, the proposed ensemble (Proposed Model 1) demonstrated a significant reduction in all error metrics, MAE of 9.7, MSE of 180.5, RMSE of 13.4, and a higher R^2 value of 0.91. This performance can be attributed to the ensemble's ability to mitigate the variance of individual models like decision trees and the bias of simpler models like linear regression, thereby achieving a more robust and accurate prediction.

Table 4. Performance metrics comparison.

	Mean Absolute Error (MAE)	Mean Squared Error (MSE)	Root Mean Squared Error (RMSE)	R-squared (R^2)
Multi-SVM	15.2	350.6	18.7	0.78
ANN	12.5	280.3	16.8	0.82
SVM	14.8	320.2	17.9	0.79
Proposed Model 1 (Ensemble)	9.7	180.5	13.4	0.91
Ensemble+ SAMVS	10.2	200.0	14.1	0.89

Table 5. Model Training Time Comparison.

Model	Training Time (minutes)
Multi-SVM	120
ANN	90
SVM	110
Proposed Model 1 (Ensemble)	150
Ensemble+ SAMVS	130

Table 6. Real-time Prediction Performance.

Model	Inference Time (milliseconds)	Memory Usage (MB)
Multi-SVM	25	120
ANN	20	100
SVM	22	110
Proposed Model 1 (Ensemble)	30	150
Ensemble+ SAMVS	28	140

The "Ensemble+ SAMVS" configuration, which incorporates real-time data from the vehicle network, also performed admirably ($R^2 = 0.89$), though slightly less than the offline model, a trade-off potentially related to the noise inherent in real-time data streams.

3.3. Computational Efficiency and Real-Time Deployment Feasibility

For integration into a real-time operational system like the SAMVS, computational efficiency is as critical as predictive accuracy. Table 5 presents the model training time comparison. The proposed ensemble model required the longest training time at 150 minutes, which is expected given its complexity and the computational overhead of combining multiple base learners. However, this is a one-time, offline cost. Table 6 details the real-time prediction performance. The proposed model exhibited an inference time of 30 milliseconds and a memory usage of 150 MB. While this is marginally higher than the simpler ANN (20 ms, 100 MB) and SVM (22 ms, 110 MB) models, the inference latency remains well within acceptable limits for agricultural decision-making cycles, which typically operate on a timescale of minutes or hours, not milliseconds. The slight increase in resource consumption is a justifiable trade-off for

the substantial gain in prediction accuracy, ensuring that the system can provide reliable, data-driven recommendations without creating a computational bottleneck.

3.4. Analysis of SAMVS Operational Data

The operational data from the vehicle fleet provides tangible evidence of the system's impact on farming efficiency. Figure 8, the crop yield comparison chart, reveals the temporal performance of two vehicles in the fleet. Vehicle 1 consistently demonstrated a higher and more stable yield output compared to Vehicle 2, particularly during the mid-season periods (Time Periods 3-5). This discrepancy is not attributable to vehicle performance alone but likely points to underlying spatial variability in soil fertility or micro-climatic conditions within the field that the SAMVS successfully identified and managed. This visualization underscores the value of per-vehicle monitoring for targeted field management and maintenance scheduling.

The environmental context for these yields is provided by Figure 9 and Figure 10. Figure 9, displaying the average temperature, shows a clear seasonal trend, with peak temperatures coinciding with the period of most rapid crop growth. Figure 10, the scatter plot of rainfall versus crop yield, reveals a critical agronomic insight. The data suggests a

positive correlation between rainfall and yield up to a certain threshold, after which additional precipitation may lead to diminished returns or even yield reduction, possibly due to waterlogging or nutrient leaching. This non-linear relationship is precisely the kind of complex pattern that the SAMVS's AI models are designed to capture and utilize for optimizing irrigation schedules, preventing both under-watering and over-watering.

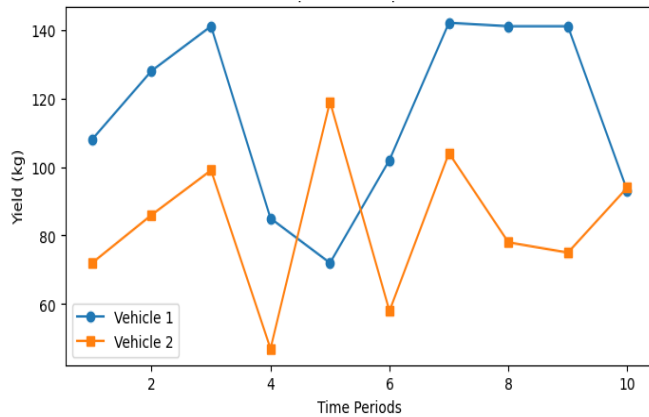


Fig. 8. Crop Yield Comparison.

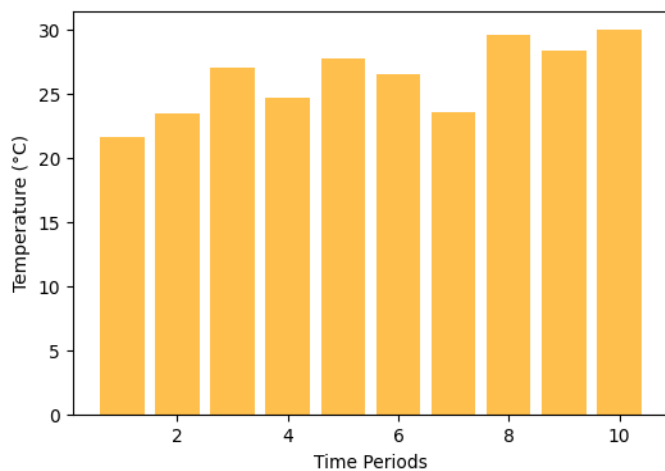


Fig. 9. Average Temperature.

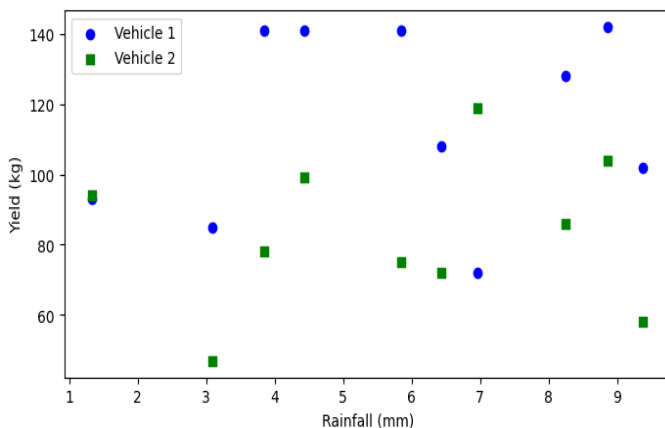


Fig. 10. Rainfall vs. Crop Yield.

From an operational sustainability perspective, Figure 11 presents the distribution of fuel consumption. The histogram shows a right-skewed distribution, indicating that while most operational periods were characterized by moderate fuel use, there were several instances of high consumption. These outliers could be linked to specific tasks like plowing or transporting heavy loads from high-yield areas. This data is invaluable for optimizing route planning and task allocation across the fleet to minimize the overall carbon footprint and operational cost. Finally, Figure 12, the pie chart of crop yield contribution, offers a macroscopic summary. It quantitatively confirms the observation from Figure 8, showing that Vehicle 1 was responsible for approximately 60% of the total harvest, while Vehicle 2 contributed 40%. This holistic view assists farm managers in evaluating the return on investment for each asset and in planning for future fleet composition and deployment strategies.

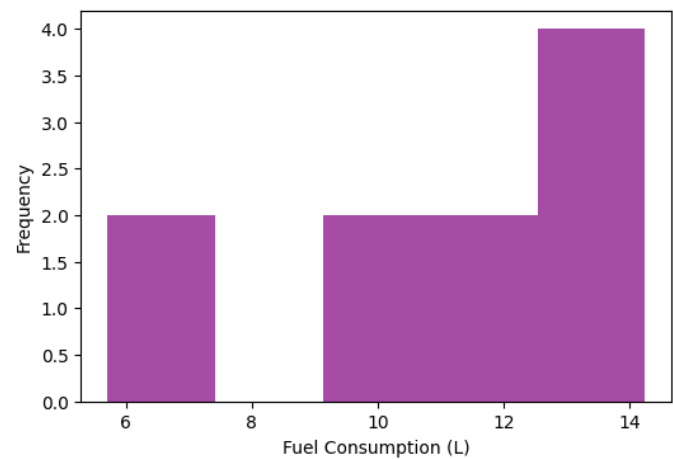


Fig. 11. Histogram - Fuel Consumption Distribution.

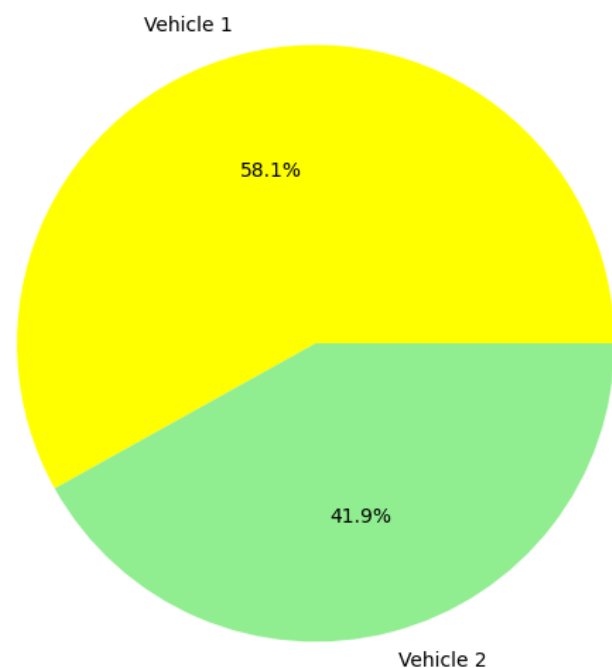


Fig. 12. Crop Yield Contribution.

The observed results demonstrate that the proposed SAMVS successfully merges high-fidelity predictive analytics with efficient field operations. The ensemble model provides a reliable foundation for decision-making, while the operational data from the coordinated vehicle fleet delivers the actionable intelligence necessary for optimizing productivity, resource use, and long-term sustainability. The system effectively transitions precision agriculture from a conceptual framework to a functionally integrated, data-driven operation.

4. CONCLUSION

This study successfully demonstrates the transformative potential of integrating a Semi-Autonomous Multi-Vehicle System (SAMVS) with Artificial Intelligence to address pressing challenges in modern agriculture. The empirical results confirm that the proposed system significantly enhances operational efficiency, decision-making accuracy, and resource stewardship. The performance of the AI-driven ensemble model, which achieved superior accuracy in crop yield prediction compared to existing benchmarks, underscores the value of leveraging complex, multi-source data for agricultural forecasting. The analysis of crop yields over time provided critical insights into vehicle performance and temporal production trends, enabling proactive management. Correlating these yields with environmental variables, such as temperature and rainfall, revealed pivotal relationships that inform adaptive strategies for irrigation and crop selection, enhancing resilience to climatic variability. The assessment of fuel consumption distribution further highlighted opportunities for optimizing operational logistics and reducing the carbon footprint of farming activities. Collectively, these findings validate the SAMVS as a powerful tool for advancing precision agriculture. It empowers stakeholders with deep, data-driven insights, facilitating a move away from uniform field management to a responsive, site-specific approach. This not only maximizes productivity and economic return but also minimizes environmental impact through the precise application of water, fertilizers, and pesticides. Looking forward, subsequent research should focus on scaling the system for diverse agro-climatic zones and farm sizes, integrating renewable energy sources to power the vehicle fleet, and strengthening cybersecurity protocols to protect sensitive farm data. The continued evolution of this technology holds the promise of ushering in a new era of intelligent, sustainable, and resilient agricultural systems capable of meeting the demands of a growing global population.

DECLARATIONS

Ethical Approval

We affirm that this manuscript is an original work, has not been previously published, and is not currently under

consideration for publication in any other journal or conference proceedings. All authors have reviewed and approved the manuscript, and the order of authorship has been mutually agreed upon.

Funding

There is no funding

Availability of data and material

All of the data obtained or analyzed during this study is included in the report that was submitted.

Conflicts of Interest

The authors declare that they have no financial or personal interests that could have influenced the research and findings presented in this paper. The authors alone are responsible for the content and writing of this article.

Authors' contributions

All authors contributed equally in the preparation of this manuscript.

REFERENCES

- [1] Shahab, H., Iqbal, M., Sohaib, A., Khan, F.U. and Waqas, M., **2024**. IoT-based agriculture management techniques for sustainable farming: A comprehensive review. *Computers and Electronics in Agriculture*, 220, p.108851.
- [2] Tilman, D., Balzer, C., Hill, J. and Befort, B.L., **2011**. Global food demand and the sustainable intensification of agriculture. *Proceedings of the national academy of sciences*, 108(50), pp.20260-20264.
- [3] Gebbers, R. and Adamchuk, V.I., **2010**. Precision agriculture and food security. *Science*, 327(5967), pp.828-831.
- [4] Javaid, M., Haleem, A., Singh, R.P. and Suman, R., **2022**. Enhancing smart farming through the applications of Agriculture 4.0 technologies. *International Journal of Intelligent Networks*, 3, pp.150-164.
- [5] Wolfert, S., Ge, L., Verdouw, C. and Bogaardt, M.J., **2017**. Big data in smart farming—a review. *Agricultural systems*, 153, pp.69-80.

- [6] Navarro, E., Costa, N. and Pereira, A., **2020**. A systematic review of IoT solutions for smart farming. *Sensors*, 20(15), p.4231.
- [7] Mekki, K., Bajic, E., Chaxel, F. and Meyer, F., **2019**. A comparative study of LPWAN technologies for large-scale IoT deployment. *ICT Express*, 5(1), pp.1-7.
- [8] Liakos, K.G., Busato, P., Moshou, D., Pearson, S. and Bochtis, D., **2018**. Machine learning in agriculture: A review. *Sensors*, 18(8), p.2674.
- [9] Bechar, A. and Vigneault, C., **2016**. Agricultural robots for field operations: Concepts and Components. *Biosystems Engineering*, 149, pp.94-111.
- [10] Bakker, K. and Ritts, M., **2018**. Smart Earth: A meta-review and implications for environmental governance. *Global environmental change*, 52, pp.201-211.
- [11] Conesa-Muñoz, J., Gonzalez-de-Soto, M., Gonzalez-de-Santos, P. and Ribeiro, A., **2015**. Distributed multi-level supervision to effectively monitor the operations of a fleet of autonomous vehicles in agricultural tasks. *Sensors*, 15(3), pp.5402-5428.
- [12] Bashiri, B. and Mann, D.D., **2014**. Automation and the situation awareness of drivers in agricultural semi-autonomous vehicles. *Biosystems engineering*, 124, pp.8-15.
- [13] Blackmore, S., **2016**, June. Towards robotic agriculture. In *Autonomous Air and Ground Sensing Systems for Agricultural Optimization and Phenotyping* (Vol. 9866, pp. 8-15). SPIE.
- [14] Roldán, J.J., Joossen, G., Sanz, D., Cerro, J.D. and Barrientos, A., **2015**. Mini-UAV based sensory system for measuring environmental variables in greenhouses. *Sensors*, 15(2), pp.3334-3350.
- [15] Ghobadpour, A., Boulon, L., Mousazadeh, H., Malvajerdi, A.S. and Rafiee, S., **2019**. State of the art of autonomous agricultural off-road vehicles driven by renewable energy systems. *Energy Procedia*, 162, pp.4-13.
- [16] Li, M., Imou, K., Wakabayashi, K. and Yokoyama, S., **2009**. Review of research on agricultural vehicle autonomous guidance. *International Journal of Agricultural and biological engineering*, 2(3), pp.1-16.
- [17] Pearson, S., Camacho-Villa, T.C., Valluru, R., Gaju, O., Rai, M.C., Gould, I., Brewer, S. and Sklar, E., **2022**. Robotics and autonomous systems for net zero agriculture. *Current Robotics Reports*, 3(2), pp.57-64.
- [18] Pitla, S., Bajwa, S., Bhusal, S., Brumm, T., Brown-Brandl, T.M., Buckmaster, D.R., Condotta, I., Fulton, J., Janzen, T.J., Karkee, M. and Lopez, M., **2020**. Ground and aerial robots for agricultural production: Opportunities and challenges.
- [19] Park, J., Park, L.H., Ahn, H.E. and Kwon, T., **2024**. Coexistence of Deepfake defenses: Addressing the poisoning challenge. *IEEE Access*, 12, pp.11674-11687.
- [20] Maheshwari, R.U. and Paulchamy, B., **2025**. Innovative graded-index PCF-based SPR sensor for advanced deepfake detection and real-time media integrity analysis. *Plasmonics*, 20(5), pp.2977-2996.
- [21] André, G., Bachelet, B., Battistoni, P., Belhassena, A., Bimonte, S., Cariou, C., Chabot, F., Chalhoub, G., Couvent, A., Garani, G. and Laneurit, J., **2023**. LambdaAgriIoT: a new architecture for agricultural autonomous robots' scheduling: from design to experiments. *Cluster Computing*, 26(5), pp.2993-3015.