

**INDIVIDUAL ASSIGNMENT**

**TECHNOLOGY PARK MALAYSIA**

**CT098-3-2-RMCT**

**RESEARCH METHODS FOR COMPUTING AND TECHNOLOGY**

**APU2F2411CS(AI)**

**HANDOUT DATE:**

**HAND IN DATE:**

**LECTURER NAME:**

**STUDENT NAME:**

**TP NUMBER:**

**WEDNESDAY, 28TH MAY 2025**

**MONDAY, 9TH JUNE 2025**

**DR. MURUGANANTHAN VELAYUTHAM**

**MANREEN KAUR A/P JAGJIT SINGH**

**071290**

Table of Contents

[**Abstract** 3](#_Toc200311788)

[**Index Terms** 3](#_Toc200311789)

[**1.** **Introduction** 3](#_Toc200311790)

[**2.** **Problem Statement** 4](#_Toc200311791)

[**3.** **Research Aim** 4](#_Toc200311792)

[**4.** **Research Objectives** 4](#_Toc200311793)

[**5.** **Research Questions** 4](#_Toc200311794)

[**6.** **Research Significance** 5](#_Toc200311795)

[**7.** **Overview of the Proposed System/Research** 5](#_Toc200311796)

[**References** 7](#_Toc200311797)

**Development of an Agri Crop Yield Prediction Model**

Manreen Kaur A/P Jagjit Singh

TP071290@mail.apu.edu.my

# **Abstract**

The invention of an Agricultural Crop Yield Prediction Model provides a modern agricultural solution. The Agricultural Crop Yield Prediction Model forecasts crop productivity using soil nutrient values, environmental variables, and crop data. This project will develop a reliable and efficient model capable of predicting yields based on soil nitrogen, phosphorus, and potassium values; temperature conditions; and humidity conditions. The Agricultural Crop Yield Prediction Model will utilize machine learning algorithms to assess various inputs for predicting the yield possibility of different crops and empowering the farming community to make informed decisions based on maximum potential yield, informed by crop selection and management optimization, where efficiency and productivity could eventually enable farmers to make informed decisions regarding crop selection and crop management to maximize input. The aim is to utilize advanced analysis of data within farming, moving agriculture to sustainability to enable food security while helping limit environmental impact. The success of this project will depend on the ability to maximize crop management precision and crop management effectiveness, enabling the agricultural crop yield prediction model to provide farmers with a decision-making tool for crop and soil management planning, together with soil and climate conditions for maximizing yield potential.

# Index Terms

Agriculture, Agronomy, Climate adaptation, Crop yield prediction, Data-driven, Decision support, Deep learning, Environmental factors, eFarming systems, IoT, Machine learning, Optimization, Precision agriculture, Predictive modelling, Resource allocation, Sustainability

* + - 1. **Introduction**

The application of machine learning in agriculture has gained momentum as a solution for addressing the complexities of crop management and yield prediction. By utilizing data-driven approaches, machine learning algorithms can process vast amounts of agricultural data, offering predictions that surpass traditional methods in both accuracy and scalability (Seenu et al., 2024). These models utilize historical data on soil nutrients such as Nitrogen (N), Phosphorus (P), and Potassium (K), along with environmental parameters like temperature, humidity, and rainfall, to create precise yield forecasts (Ahmed et al., 2024). Farmers can optimize fertilization schedules, irrigation procedures, and crop rotation strategies by using machine learning to examine the intricate correlations between these variables (Islam et al., 2024).

Furthermore, predictive models based on soil health and environmental factors enable farmers to implement precision farming practices, lowering input costs and impact on the environment (Mgendi, 2024). These models assist in minimizing risks associated with market demand, price, and resource allocation by accurately predicting agricultural yields (Shahhosseini et al., 2021). Recent research has found that machine learning-based yield prediction models, particularly those that use deep learning techniques, outperform standard statistical models in terms of precision and adaptability (Jhajharia et al., 2023). This demonstrates machine learning's great potential for transforming agricultural methods and maintaining food security in the face of global issues (Islam et al., 2024). As the use of such models grows, further research is needed to improve their accuracy across various geographic regions and crop types, assuring broader application and long-term benefits for farmers (Weber et al., 2023). These developments highlight the growing importance of technology-based solutions for increasing agricultural productivity and sustainability.

In essence, the growing demand for accurate agricultural yield predictions, as well as the increased usage of machine learning technologies in agriculture, underline the importance of this research area and the necessity to investigate new and novel methods to optimize crop production. Integrating machine learning models with agricultural data offers a great opportunity to improve yield forecasts based on soil nutrient levels and environmental factors. This technology-driven strategy not only improves farming efficiency but also promotes sustainable agricultural practices, addressing the worldwide dilemma of food security in a changing environment.

* + - 1. **Problem Statement**

The research problem in this topic can be summarized as the need for a more effective and reliable solution to accurately predict crop yields, as traditional methods frequently fail to account for the complex interactions between soil nutrients, environmental conditions, and crop characteristics. While many existing models focus on isolated variables or specific crop types, there is a significant need for the development of integrated models that take into account soil nutrient levels (N, P, K), temperature, and humidity for more extensive and versatile yield prediction. As the demand for efficient resource management and sustainable farming practices grows, farmers require more precise tools to optimize crop productivity and make informed decisions. Farmers need more precise tools to maximize crop yield and make informed decisions as the need for effective resource management and sustainable farming practices develops.

One study found that "the use of advanced technologies such as machine learning in agriculture has shown great promise in improving crop yield predictions, with accuracy rates significantly higher than conventional methods" (Shahhosseini et al., 2021). Another study concluded that "accurate crop yield prediction remains a major challenge due to the complex nature of soil, environmental, and crop variables, requiring sophisticated data analysis techniques" (Khaki, 2019). However, these studies frequently fail to apply or validate their models across diverse geographic regions and crop conditions, limiting their practical applicability (Ishaq et al., 2025).

This study addresses these constraints by creating a machine learning model that includes critical soil nutrients and environmental parameters to improve the accuracy of crop yield predictions. It aims to assess the efficacy of such models across various datasets, resulting in a more scalable and generalizable solution for crop yield forecasting. These findings demonstrate the critical need for research in this area to develop models capable of accurately predicting yields and enhancing agricultural efficiency.

* + - 1. **Research Aim**

The project aims to develop an agri-based crop yield prediction model using IoT and Deep Learning based on soil parameters, environmental data, and crop features.

* + - 1. **Research Objectives**

1. To develop a Machine Learning-based crop yield prediction model that leverages soil nutrient data, environmental factors, and crop characteristics for accurate yield forecasting.
2. To employ advanced machine learning techniques, such as Decision Trees (DT), Random Forest (RF), K-Nearest Neighbors (KNN), XGBoost (XGB), and Support Vector Machines (SVM), for feature extraction and yield prediction across various crops.
3. To create a user-friendly Graphical User Interface (GUI) to visualize crop yield predictions and provide farmers with actionable insights to improve decision-making.
4. To evaluate the model’s performance by analysing accuracy, precision, and scalability using benchmark agricultural datasets and real-world scenarios in diverse geographic locations.
   * + 1. **Research Questions**
5. What is the best way to consider soil nutrients (N, P, K), temperature, and humidity in a machine learning model to generate better crop yield predictions?
6. Which machine learning methods (e.g., (DT), (RF), (KNN), (XGB), and (SVM)) offer the best trade-off of accuracy, interpretability, and computational efficiency for forecasting crop yields?
7. How can the performance and scalability of the crop yield forecasting model be assessed for predictions across multiple datasets, from distinct regions and crop types?
8. What design features are necessary for a user-friendly Graphical User Interface (GUI) to support farmers in interpreting predictions and help inform agricultural decisions?
   * + 1. **Research Significance**

This research is valuable because it provides a contemporary, data-driven approach to tackling one of agriculture's biggest problems, such as forecasting crop yield with the added complexity of changing environmental and soil conditions. Yield estimates are typically calculated based on historical data or single-variable analysis, neither of which is particularly adequate because agronomists often fail to address the inherent complexity posed by singular agronomical variables. In developing a more integrated machine learning model based on soil nutrients and climate data, this paper's objectives are designed to overcome these shortcomings and provide a more adaptive, scalable solution to yield forecasting.

The outcomes of the research include assisting farmers to make better crop planning decisions that may not only improve crop productivity but also reduce waste. Besides, contribute to more sustainable agricultural systems through optimized amounts and application rates of fertilizers and water resources. Other than that, assisting food security initiatives by providing more reliable yield forecasts regardless of local environmental conditions. Finally, act as a foundational model for other agricultural applications, including disease predictions, irrigation scheduling, and resource management.

* + - 1. **Overview of the Proposed System/Research**

The proposed system is an Agricultural Crop Yield Prediction Model that integrates Internet of Things (IoT) sensor data, environmental conditions, and soil nutrient levels into a machine learning framework. The primary aim is to provide farmers with accurate yield forecasts and practical agricultural recommendations to farmers by leveraging both real-time and historical data using a desktop-based application developed with Python's Tkinter library.

**System Functionality**

The Data Collection Layer involves the collection of critical information for crop yield prediction. Specifically, soil data such as (N), (P), and (K) content are taken into account. In addition to soil nutrients, environmental variables such as temperature, humidity, and possibly rainfall are incorporated. Furthermore, crop characteristics such as crop type, growth cycle, and historical yield data are included. To ensure relevance and comprehensiveness, these data can be acquired via simulated or real-world IoT sensors as well as public datasets.

In the Data Preprocessing step, the goal is to prepare the dataset for an effective machine learning application. This involves cleaning the dataset by removing inconsistencies and irrelevant records. Subsequently, normalization and transformation of data into a suitable format for machine learning are carried out. Additionally, this phase addresses handling missing values, detecting and treating outliers, and balancing skewed data distributions, thereby ensuring the data quality necessary for accurate predictions.

The Machine Learning Engine is responsible for implementing various algorithms to perform the prediction task. It includes (DT), (RF), (KNN), (XGB), and (SVM). Moreover, there is a potential use of a Deep Learning model for comparison or future scalability. Cross-validation approaches are used to optimize model performance using hyperparameter tuning and feature selection. These models are evaluated using metrics such as accuracy, precision, recall, and F1-score to discover which model performs best.

Tkinter is a Python library used to create the Graphical User Interface (GUI). It is designed to be lightweight, responsive, and easy to navigate. This interface allows users to provide soil and weather parameters, choose crop type, and view predicted yield output. Furthermore, the results are provided in a user-friendly format, with graphs and summarized suggestions. Importantly, the GUI is developed for offline use, making it ideal for farmers who have only limited internet access or no internet connectivity.

In the Deployment and Scalability stage, the system is deployed as a desktop application, which users can install and operate on their local computers (Windows/Linux). Although the system is not web-based, its modular design allows for future migration to web platforms (e.g., Flask or Django) to increase accessibility as desired. Scalability is measured in terms of model generalizability by training and testing on several datasets from various geographic regions and types of crops. Additionally, the application is designed to handle an increasing amount of data without sacrificing performance through efficient code and data handling. The overall architecture of the proposed system is illustrated in Figure 1 Architectural Design.

**Architectural Design**

A diagram of a machine learning process

AI-generated content may be incorrect.

Figure 1 Architectural Design

**Research Overview**

This research aims to investigate the viability of machine learning algorithms for predicting agricultural output by combining data from a variety of sources, including IoT sensor data, environmental variables, and soil nutrients. To achieve this, the study involves collecting and preprocessing diverse agricultural datasets to prepare them for analysis. Moreover, the research includes implementing and comparing multiple machine learning models to identify the most accurate and robust predictor.

Moreover, the study stresses assessing models using standard performance metrics (accuracy, precision, recall, and F1-score) to ensure reliability. Plus, the research focuses on creating a desktop application interface that allows end-users to interact with the system intuitively, even when offline. Ultimately, the research also involves assessing the usability and scalability of the system, including potential future migration to web-based platforms. The step-by-step research methodology is summarized in Figure 2 Research Flowchart.

**Research Flowchart**

A close-up of a blue screen

AI-generated content may be incorrect.

Figure 2 Research Flowchart

# References

1. Ahmed, F.U, Das, A., & Zubair, M. (2024, March 28). A machine learning approach for crop yield and disease prediction, integrating soil nutrition and weather factors.

<https://arxiv.org/pdf/2403.19273>

1. Seenu, A., Kumar, P. R., Uma Rani, B., Nuthalapati, C., Pagidimalla, P., & M, S. (2024). Enhancing Crop Yield Prediction and Management in eFarming Systems through Machine Learning.

<https://www.researchgate.net/publication/380115352_Enhancing_Crop_Yield_Prediction_and_Management_in_eFarming_Systems_through_Machine_Learning#fullTextFileContent>

1. Islam, M. M., Alharthi, M., Alkadi, R. S., Islam, R., & Masum, A. K. M. (2024). Crop yield prediction through machine learning: A path towards sustainable agriculture and climate resilience in Saudi Arabia. AIMS Agriculture and Food, 9(4), 980–1003.

<https://www.aimspress.com/article/doi/10.3934/agrfood.2024053>

1. Mgendi, G. (2024). Unlocking the potential of precision agriculture for sustainable farming. Discover Agriculture, 2(1).

<https://doi.org/10.1007/s44279-024-00078-3>

1. Shahhosseini, M., Hu, G., Huber, I., & Archontoulis, S. V. (2021). Coupling machine learning and crop modeling improves crop yield prediction in the US Corn Belt. Scientific Reports, 11(1).

<https://www.nature.com/articles/s41598-020-80820-1>

1. Jhajharia, K., Mathur, P., Jain, S., & Nijhawan, S. (2023). Crop Yield Prediction using Machine Learning and Deep Learning Techniques. Procedia Computer Science, 218, 406–417.

<https://www.sciencedirect.com/science/article/pii/S1877050923000236?via%3Dihub>

1. Weber, M. N., Hohmann, P., Demyan, M. S., & Reineking, B. (2023, August 16). Predicting crop yield with machine learning: An extensive analysis of input modalities and models on a field and sub-field level.

<https://arxiv.org/pdf/2308.08948>

1. Sajid, S. S., Shahhosseini, M., Huber, I., Hu, G., & Archontoulis, S. V. (2022). County-scale crop yield prediction by integrating crop simulation with machine learning models. Frontiers in Plant Science, 13.

<https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2022.1000224/full>

1. Khaki, S., Wang, L., & Archontoulis, S. V. (2019). A CNN-RNN framework for crop yield prediction.

<https://arxiv.org/pdf/1911.09045>

1. Ishaq, R. A. F., Zhou, G., Jing, G., Shah, S. R. A., Ali, A., Imran, M., Jiang, H., & Obaid-ur-Rehman. (2025). Geospatial Robust Wheat Yield Prediction Using Machine Learning and Integrated Crop Growth Model and Time-Series Satellite Data. Remote Sensing, 17(7), 1140.

<https://doi.org/10.3390/rs17071140>