

CROP YIELD PREDICTION USING HISTORICAL DATA.docx

by Justin Gilbert Alexius Silvester .

Submission date: 09-Apr-2026 03:00PM (UTC+0800)

Submission ID: 2926561037

File name:

47529_Justin_Gilbert_Alexius_Silvester_-_CROP_YIELD_PREDICTION_USING_HISTORICAL_DATA_1170930_1555755923.docx
(146.04K)

Word count: 4138

Character count: 25766

CROP YIELD PREDICTION USING HISTORICAL DATA

JUSTIN GILBERT, ADELINE SNEHA JOHN, SIVAKUMAR VENGUSAMY,
MURUGANANTHAN VELAYUTHAM, PREETHI SUBRAMANIAM, THAVAMALINI
SIVASAMY, RAJA RAJESWARI, TULASI SATHYABAMA

Justin Gilbert, Asia Pacific University of Technology & Innovation

justin.gilbert@apu.edu.my

Adeline Sneha John, Asia Pacific University of Technology & Innovation

adeline.john@apu.edu.my

Sivakumar Vengusamy, Asia Pacific University of Technology & Innovation

dr.sivakumar@apu.edu.my

Murugananthan Velayutham, Asia Pacific University of Technology & Innovation

murugananthan@apu.edu.my

Preethi Subramaniam, Asia Pacific University of Technology & Innovation

preethi@apu.edu.my

Thavamalini Sivasamy, Asia Pacific University of Technology & Innovation

thavamalini@apu.edu.my

Raja Rajeswari, Asia Pacific University of Technology & Innovation

raja.rajeswari@apu.edu.my

Tulasi Sathyabama, Asia Pacific University of Technology & Innovation

tulasi.appalasamy@apu.edu.my

ABSTRACT:

Crop yield prediction is significant for reinforcing food security, as climate variability continues to disrupt agricultural production and planning. This study examines the possibility of combining historical weather data, soil conditions, and farming-related characteristics with machine learning techniques to predict crop yield more accurately. The complete modelling workflow is executed, including data understanding, preprocessing, feature engineering, and the training of several models such as Random Forest, XGBoost, MLP, and CatBoost. Based on the evaluation results, CatBoost is selected as the final model due to its strong predictive accuracy and stability when handling tabular agricultural data. The trained model is implemented in a complete web-based decision support system named *Smart Crop Manager* to demonstrate the practical application of the research. The system allows users to enter and manage crop information, automatically process weather data, generate yield forecasts, categorize risk levels, and provide scenario-based suggestions to support planting and harvesting decisions. In addition, an artificial intelligence crop assistant powered by Google Gemini provides readable summaries and responds to user queries to enhance the understanding of predictions and soil recommendations. Overall, this project demonstrates that artificial intelligence and web technologies can be integrated to support crop planning and promote more sustainable food production, thereby contributing to the advancement of Sustainable Development Goal 2: Zero Hunger.

Keywords: Crop yield, prediction, machine learning, historical data

1 Introduction:

Agricultural production has faced unprecedented challenges due to climatic variability, meteorological extremes, and soil degradation in recent years. These factors have led to uncertain crop yields, raising concerns about food security, economic efficiency, and ecological sustainability. Extreme weather events such as floods, droughts, and temperature fluctuations exacerbate the vulnerability of agricultural systems, resulting in reduced crop yields and disrupted food supply chains. Studies indicate that climate change will continue to pose a major threat to crop yields, particularly in regions reliant on rain-fed agriculture. As global temperatures rise, staple crops like wheat, rice, and maize are projected to experience reduced yields, which will not only impact food production but also increase food insecurity (Ray et al., 2021; Ashfaq et al., 2025).

Traditional crop yield forecasting methods rely on outdated models that fail to account for the complexities and dynamic nature of environmental factors affecting agricultural production. These methods typically depend on historical data, using simplistic assumptions that neglect the spatial and temporal variability critical for accurate crop yield predictions. Traditional models also focus on a limited number of variables, overlooking the interactions between key environmental factors such as soil conditions, weather patterns, and pest infestations. As a result, these approaches cannot adequately address the challenges posed by a rapidly changing climate, leading to unreliable predictions and suboptimal decision-making in agriculture (Mandal et al., 2023; Silva et al., 2024).

In Malaysia, there is a significant gap in the availability of data-driven forecasting tools that integrate key environmental factors, including soil health, weather conditions, and crop performance. Although some predictive tools are in use, they typically rely on outdated or generalized data, limiting their accuracy and utility for farmers. These tools also fail to provide real-time, region-specific insights that are necessary for informed decision-making. As a result, farmers struggle to optimize crop management practices, such as irrigation, fertilization, and pest control. The absence of robust, data-driven systems for agricultural forecasting hinders Malaysia's agricultural sector and prevents farmers from achieving optimal productivity (Saad et al., 2022).

Farmers and agricultural consultants in Malaysia also face challenges due to the lack of reliable and accessible prediction tools that are tailored to the needs of local farming practices. Existing tools, when available, are often too expensive, too complex, or too generalized for widespread adoption by smallholder farmers. Consequently, farmers rely on traditional methods that can be inefficient and prone to error, resulting in suboptimal use of resources. The inability to make timely, data-driven decisions about pest control, irrigation, and crop management further exacerbates crop losses and financial instability, especially during unpredictable weather events (Sharma et al., 2025).

While machine learning techniques, such as **Random Forest (RF)** and **Categorical Boosting (CatBoost)**, have shown promise in agricultural forecasting, existing research has not yet fully integrated temporal and multivariate environmental data into a single, unified model. Many forecasting models either focus on single-variable analysis or fail to account for the complex interactions between multiple environmental factors over time. However, the application of AI models like CatBoost, which combines various sources of data (e.g., weather, soil moisture, crop health), presents a valuable opportunity to provide more accurate, dynamic, and long-term predictions of crop yields. Despite the potential of these models, their application in agricultural forecasting remains underutilized, limiting their impact on decision-making and sustainable agricultural practices (Paudel et al., 2023; Ashfaq et al., 2025).

Problem Statement

Uncontrollable Weather and Climate Data Factors

Uncontrollable data refers to external factors that strongly influence crop yield but cannot be managed directly by farmers or the system, especially short-term environmental changes. In crop yield prediction, weather and climate variability are key uncontrollable inputs because they can change suddenly and affect plant growth conditions even when farming practices remain the same. Variations in temperature, rainfall, humidity, and extreme events such as droughts and floods introduce uncertainty into crop production, making yield outcomes harder to predict and increasing the risk of crop loss. Studies have reported that climate change has already reduced yields in multiple regions, and continued warming is expected to decrease yields for major staple crops such as wheat, rice, and maize, which can threaten food security and production stability (Ray et al., 2021). These climate-driven risks are particularly serious in regions that rely heavily on rain-fed agriculture and have fewer resources to adapt, where yield outcomes become more sensitive to uncontrollable environmental fluctuations (Ashfaq et al., 2025).

Limitations of Traditional Agricultural Models

Traditional agricultural forecasting models are increasingly inadequate in addressing the complex, non-linear interactions between climate variables and crop productivity. These models often rely on historical data and simplistic assumptions, such as constant weather patterns or fixed agronomic practices, which fail to capture the temporal and spatial variability that influences crop growth. Many traditional models also overlook the interactions between environmental factors, such as soil health, weather, and pest outbreaks, which are critical for accurate crop yield predictions. As a result, these outdated methods lead to errors and inaccuracies in predicting crop yields (Mandal et al., 2023; Silva et al., 2024).

Lack of Data-Driven Forecasting Tools in Malaysia

In Malaysia, the agricultural sector faces a significant gap in the availability of data-driven forecasting tools that incorporate multiple data sources, such as weather patterns, soil health, and crop performance. While some forecasting systems exist, they often rely on outdated or incomplete data, limiting their accuracy and utility for farmers. Furthermore, existing tools do not fully integrate real-time data or consider the unique environmental conditions in various regions of Malaysia. This lack of localized, accurate, and dynamic forecasting systems impedes farmers' ability to make informed decisions regarding irrigation, fertilizer application, pest

control, and crop rotation, ultimately hindering crop yield optimization (Saad et al., 2022; Taufik & Tharakan, 2021).

Challenges Faced by Farmers and Consultants

Farmers and agricultural consultants in Malaysia face significant challenges due to the absence of reliable and user-friendly forecasting tools. Traditional methods, such as visual inspections or external consultations, are often labor-intensive, costly, and prone to errors. As a result, farmers lack the ability to make timely, informed decisions about when to plant, irrigate, fertilize, or manage pests. The lack of access to real-time, localized prediction tools prevents farmers from responding effectively to emerging threats such as climate shifts, pest infestations, or diseases. This inefficiency limits productivity and increases the vulnerability of farmers to economic losses (Sharma et al., 2025; Saad et al., 2022).

Project Objectives

- 1 To collect and preprocess historical weather and soil data relevant to the target crop's health, growth, and yield.
- To apply machine learning models to predict crop yields based on the historical weather and soil conditions.
- To assess and evaluate the accuracy of the model predictions using metrics like Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE).
- To visualize predicted crop yields and suggest optimization strategies for irrigation, fertilizer use, and crop management.

1 Traditional Methods

Traditional yield prediction methods have long been integral to agricultural planning, providing valuable insights based on experience, field sampling, and historical data. However, these methods are increasingly being challenged by the complexities of modern farming and changing environmental conditions. Below, we explore various traditional approaches to crop yield prediction, highlighting their strengths and limitations in the context of contemporary agricultural practices.

Farmer Experience and Heuristics

For centuries, farmers have relied on experiential knowledge passed down through generations to estimate crop yields. These estimations are often based on visual cues such as plant height,

leaf coloration, grain head density, and general field condition. In many rural areas, this heuristic approach still dominates because it requires no additional cost or technology (Bakar et al., 2021). However, the method's subjectivity means that two farmers observing the same field may arrive at vastly different yield estimates. Furthermore, changing climate patterns have reduced the reliability of these traditional cues, as crops may exhibit unfamiliar responses to extreme weather events. Without integration of objective measurements, farmer experience remains a valuable but incomplete tool in modern yield forecasting. AI-based tools can complement this knowledge by providing quantitative predictions that align with farmers' on-the-ground observations.

1 Manual Field Sampling

Manual field sampling is one of the earliest systematic approaches to yield estimation. It typically involves selecting representative plots within a field, harvesting the crops from these plots, and extrapolating the results to the entire field. While this method can provide more structure than pure observation, it is labour-intensive and time-consuming (Latif et al., 2022). Sampling bias is a common issue; for example, choosing healthier-looking areas can lead to overestimation, while sampling damaged sections can cause underestimation. Moreover, environmental variability within a field—such as differences in soil quality or microclimates—means that small samples may not accurately represent the whole crop. In large-scale operations, the time required to complete sampling can delay decision-making, reducing its practical value in fast-changing conditions.

1 Historical Averages

Using historical yield averages is a common practice in both farm-level planning and government forecasting. This method assumes that past production levels can predict future outcomes under similar conditions. While it provides a quick estimate, it fails to account for annual variations in weather, pest outbreaks, and management practices (Rahman et al., 2022). For example, a region that has maintained stable yields for a decade may suddenly face a drought or flood, rendering the historical average inaccurate. In the context of climate change, relying on static historical data becomes increasingly risky. Integrating historical averages with real-time environmental data can improve their utility, but on their own, they offer limited predictive power.

1 Government Statistical Models

Government agricultural departments often use statistical models—such as regression equations—to predict yields based on aggregated data like rainfall totals, temperature averages, and sowing dates. While these models can be transparent and easy to interpret, they are often constrained by the availability and quality of data (Nguyen et al., 2021). Aggregation at district or regional levels can mask significant variability at the farm scale. Additionally, traditional regression models are not well-suited to capturing complex, non-linear interactions among multiple variables, such as the combined effects of temperature spikes and water stress. As a result, their predictions may be less accurate compared to modern machine learning models that can process richer datasets.

1 Limitations of Traditional Methods

The primary limitation of traditional yield prediction methods lies in their inability to handle the complexity and variability of modern agricultural systems. They often rely on a small number of variables, assume stable environmental conditions, and cannot quickly adapt to sudden changes like extreme weather events or pest outbreaks. Moreover, these methods typically provide predictions at fixed intervals, such as pre-harvest surveys, rather than offering continuous updates throughout the growing season. In a context where climate unpredictability is increasing and agricultural markets demand timely information; traditional methods alone are insufficient. The shift toward automated, data-driven models addresses these gaps by enabling real-time, high-resolution, and adaptive forecasting.

Automation in Crop Yield Prediction

Automation has transformed crop yield prediction by integrating advanced technologies such as remote sensing, machine learning, deep learning, and hybrid models. These technologies allow for continuous, real-time monitoring and processing of vast amounts of data, leading to more accurate and adaptive forecasting. Below, we delve into key methods and tools that are driving this transformation in crop yield prediction.

1 Remote Sensing Integration

Remote sensing has revolutionized crop yield prediction by enabling large-scale, continuous monitoring of crop growth and health. Through the use of satellite imagery, unmanned aerial vehicles (UAVs), and ground-based sensors, key vegetation indices such as the Normalized Difference Vegetation Index (NDVI) and Enhanced Vegetation Index (EVI) can be calculated

to assess canopy cover, chlorophyll content, and biomass accumulation (Nguyen et al., 2022). These indices provide early indicators of crop stress caused by drought, nutrient deficiency, or pest damage. Remote sensing data can be collected at regular intervals, allowing models to capture temporal growth patterns that are critical for accurate forecasting. The integration of remote sensing into automated systems also facilitates regional and national-scale predictions without the need for extensive on-the-ground surveys, significantly reducing costs and time requirements.

Machine Learning for Multivariate Data

Machine learning (ML) models, including Random Forest, XGBoost, and Gradient Boosting Machines, have become central to yield prediction due to their ability to process diverse datasets containing weather, soil, and management variables. These models excel at capturing non-linear relationships between variables, making them particularly effective for agricultural systems where yield is influenced by multiple interacting factors (Gilbert et al., 2023). For example, Random Forest models can determine the combined effect of temperature, rainfall, and soil pH on maize yield more accurately than traditional regression models. Machine learning algorithms also provide feature importance rankings, which help farmers and policymakers identify the most influential factors affecting yield in each season. This interpretability supports targeted interventions, such as optimizing fertilizer use or adjusting planting dates.

Model Building and Evaluation

Multiple machine learning models are trained using the same prepared dataset so their performance can be evaluated fairly. The goal is to identify the most suitable model for deployment based on prediction accuracy and reliability, instead of choosing a model based only on popularity.

Model Building Pipeline

The model building pipeline follows a structured workflow:

- 1) Load the processed dataset
- 2) Split the dataset into training and testing sets
- 3) Train selected models using the training set

4) Evaluate predictions using standard regression metrics.

This pipeline ensures each model is trained under the same conditions and the evaluation results can be compared directly to support final model selection.

Results and Discussion

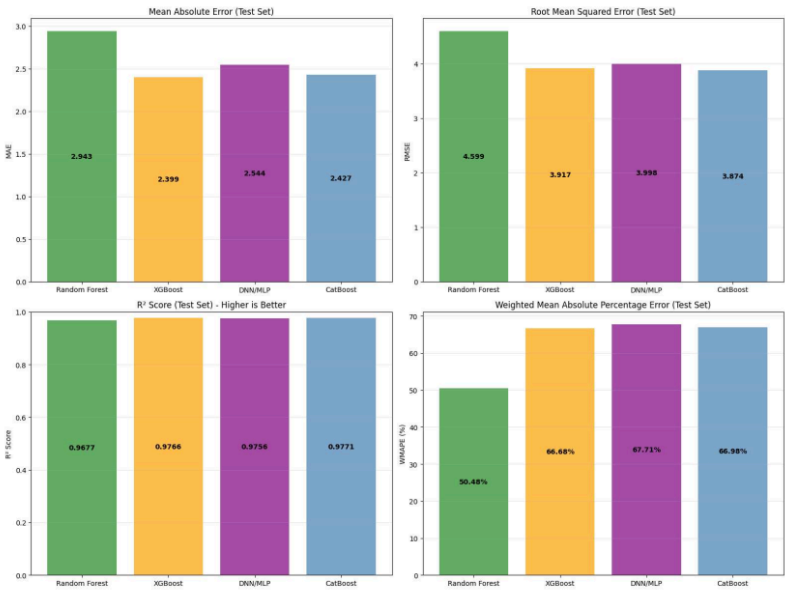


Fig 1: Chart for Model Comparison

The MAE chart shows XGBoost has the lowest MAE while CatBoost is very close, and Random Forest is the highest. The RMSE chart shows CatBoost has the lowest RMSE, followed by XGBoost and DNN or MLP, with Random Forest highest. The R2 chart shows CatBoost is the highest, then XGBoost, then DNN or MLP, and Random Forest last. The WMAPE chart shows Random Forest has a much lower WMAPE than the other three models, while XGBoost, DNN or MLP, and CatBoost are clustered around similar higher WMAPE values.

Model	MAE	RMSE	R2	WMAPE
Random Forest	2.9428	4.5992	0.9677	50.4813
XGBoost	2.3994	3.9173	0.9766	66.6770
MLP	2.5444	3.9978	0.9756	67.7105
CatBoost	2.4266	3.8742	0.9771	66.9764

Table 1 Model Comparison Table

Based on Table 5, all four trained models (Random Forest, XGBoost, MLP, and CatBoost) demonstrate strong predictive capability, with high R^2 values above 0.96, showing that the selected features are effective for explaining yield variation. However, differences in error metrics indicate that some models are more reliable than others for deployment. Random Forest produces the weakest results overall, with the highest MAE (2.9428) and RMSE (4.5992), meaning it makes larger average prediction errors. XGBoost and MLP improve the performance substantially, with lower MAE/RMSE and higher R^2 , but CatBoost provides the best overall balance, achieving the highest R^2 (0.9771) and the lowest RMSE (3.8742), which is important because RMSE penalizes large mistakes more heavily and better reflects prediction stability in real decision-making. Therefore, after comparing the models, CatBoost is selected as the final model to be used in the system, and it is exported and saved as best_crop_yield_model.cbm so the platform consistently applies the best-performing trained model during yield prediction.

Comparison with Similar Works

Source	Data Scope	Model	Deployment	Performance
Proposed Smart Crop Manager	Integrated crop + weather + soil-related features	CatBoost (tuned)	Web-based decision support	Test: MAE 2.4266, RMSE 3.8742, R^2 0.9771
Kalmani et al., 2024	Wheat + multi-factor features; intended for institutional decision support	DNN / MLP	PC/Workstation; portal integration	R^2 0.912, RMSE 146.2 kg/ha, MAE 112.5 kg/ha

Lu et al., 2024	Field yield (2014–2020) + Sentinel-2 indices + ECMWF weather	CNN-LSTM-Attention	GPU training; cloud/edge intention	Maize: R ² 0.80, RMSE 375 kg/ha, MAPE 4.21% (Rice 0.76, Soybean 0.78)
Ashfaq et al., 2025	Wheat across growth stages (2017–2022) + climate + soil nutrients	Multi-phase stage-wise model	PC/Workstation which is exportable to advisory apps	RMSE 0.128 t/ha, R ² 0.905

Table 2 Similar Work Comparison

Across the selected works, performance is commonly reported using R² and error metrics, but direct comparison of RMSE/MAE across papers is naturally limited because units and yield ranges differ by crop, region, and dataset scale (e.g., kg/ha vs t/ha). For that reason, the most interpretable high-level comparison is that the proposed system’s tuned CatBoost model reports R² = 0.9771 on the test set, indicating a strong ability to explain variance within the study’s dataset.

A key technical difference is the optimization strategy. CatBoost in this study was selected through a structured grid search with 768 combinations × 3 folds and stored as the best-performing configuration, which strengthens the credibility of the final model choice compared with using a default configuration. The reported evaluation also shows close Train/Validation/Test metrics, supporting the claim that the selected model generalizes well within the dataset split rather than only fitting training data.

Conclusion

In conclusion, the project manages to provide a complete end-to-end Smart Crop Manager decision-support system enabling the relationship of data inputs, yield prediction, scenario comparison, and an AI explanation layer into a single workflow. Technically, the model analysis indicates that CatBoost had the greatest generalization of the models that were tested with XGBoost and the MLP following it very closely with slightly lower Test R. This facilitates

the research objective of the project to compare models and choose an ultimate approach not on presumption but on evidence.

Limitations

The greatest weakness is data representativeness and external validity. Even though the data set after the filtering is large and balanced, it is a Kaggle-based dataset and, therefore, may not represent the actual variability of Malaysian farms, microclimates, and soil lab accuracy and real management practices. This implies that the high R2 result is just an evidence that it will perform in other regions, seasons, or in introduced farming conditions, but not in Malaysia.

There were also limitations exist in the input quality of the real world and weather realism. It is based on an external weather API to get historical conditions and falls on a seasonal fallback when dates are ahead of time or an API retrieve failure, resulting in averages that are, at farm scale, reasonable but not local. This creates a question mark in predictions, particularly in decision making with the slightest variation in yield altering the recommendation of the best date to pick. Besides this, the optimization method samples combinations within discrete ranges though grid search, thus the best possible recommendation is limited in discrete candidate values as opposed to the optimal continuous values.

Future Enhancements

First, enhance real-world validity through the addition of local/field-aligned datasets and the enhancement of the data pipeline. Build on Kaggle, incorporating farm-record data, Malaysian weather station feeds and optional IoT inputs like soil moisture, rainfall gauges so that the model trains on conditions more relevant to the desired context. Such additional measures as periodic retraining and drift monitoring can be also applied, as the current dataset changes through 2014-2023 and weather patterns may change over time.

Second, improve quality of decision support through explainability and uncertainty, not just predictions. CatBoost is robust, however, users must have reasons and confidence levels, SHAP-based explanations and uncertainty represent the best addition to reducing over-trust and allow farmers to make the decision on when to take advice (Gilbert & Hwei, 2026). As your system already does structured feature ordering and scenario outputs it is well placed to add explanations consistently at the inference time.

Third, enhance readiness of the system by deploying it and increasing its usage. One of the options to consider is hybrid deployment or deploying into cloud to ensure that the tool would

be supporting multiple users, farmer organizations, and extension officers and still retain control over privacy. Simultaneously, increase the number of stakeholders involved in UAT and transform feedback into quantifiable usability KPIs like task completion time, error rate, perceived trust. Provided that the project is framed with SDG 2: Zero Hunger, also suggest monitoring outcome-oriented measures like the decrease in cases of late planting, better targeting of fertilizers and decrease in the cases of crop-loss in order to have contribution to be demonstrated outside of the model performance.

References:

- Ashfaq, M., Khan, I., Afzal, R. F., Shah, D., Ali, S., & Tahir, M. (2025). Enhanced wheat yield prediction through integrated climate and satellite data using advanced AI techniques. *Scientific Reports*, 15(1). <https://doi.org/10.1038/s41598-025-02700-w>
- Bakar, A. S., Rahman, M. T., & Mollah, M. M. (2021). Traditional yield prediction methods and modern challenges in climate-smart agriculture. *Sustainability*, 13(5), 2191. <https://doi.org/10.3390/su13052191>
- Gilbert, J., Hwei, C.M. (2026). Deep Learning Models for Predictive Maintenance in Aviation with Explainable AI Using SHAP. In: Bhateja, V., Hoong, A.L.S., William, Y.G.S., Omar, Z. (eds) Smart Computing Paradigms: Human-Centric Systems for Sustainable Development. SCI 2025. Lecture Notes in Networks and Systems, vol 1685. Springer, Cham. https://doi.org/10.1007/978-3-032-08253-4_2
- Gilbert, J., Hamid, S., Hashem, I.A.T. *et al.* The rise of user profiling in social media: review, challenges and future direction. *Soc. Netw. Anal. Min.* **13**, 137 (2023). <https://doi.org/10.1007/s13278-023-01146-0>
- Kalmani, V. H., Dharwadkar, N. V., & Thapa, V. (2024). Crop Yield Prediction using Deep Learning Algorithm based on CNN-LSTM with Attention Layer and Skip Connection. *Indian Journal of Agricultural Research, Of.* <https://doi.org/10.18805/ijare.a-6300>
- Latif, S., Aslam, M., & Fayyaz, H. (2022). Evaluation of traditional yield prediction methods in Pakistan's agriculture: A case study of maize and wheat. *Journal of Agricultural Studies*, 10(1), 66-75. <https://doi.org/10.5296/jas.v10i1.18859>
- Lu, J., Li, J., Fu, H., Tang, X., Liu, Z., Chen, H., Sun, Y., & Ning, X. (2024). Deep Learning for Multi-Source Data-Driven crop yield prediction in Northeast China. *Agriculture*, 14(6), 794. <https://doi.org/10.3390/agriculture14060794>

- Mandal, M., Prakash, A., & Sharma, S. (2023). *Machine learning for precision agriculture: Current challenges and future trends. Computers and Electronics in Agriculture*, 216, 107283. <https://doi.org/10.1016/j.compag.2023.107283>
- Nguyen, P. T., Phan, H. L., & Tan, T. D. (2021). Rice yield forecasting models based on environmental data and remote sensing. *Field Crops Research*, 259, 107998. <https://doi.org/10.1016/j.fcr.2021.107998>
- Paudel, D., De Wit, A., Boogaard, H., Marcos, D., Osinga, S., & Athanasiadis, I. N. (2023). Interpretability of deep learning models for crop yield forecasting. *Computers and Electronics in Agriculture*, 206, 107663. <https://doi.org/10.1016/j.compag.2023.107663>
- Rahman, M., Ali, M., & Singh, A. (2022). Machine learning for crop yield prediction: A review of current models and techniques. *Agricultural Systems*, 196, 103275. <https://doi.org/10.1016/j.agsy.2022.103275>
- Ray, D. K., et al. (2021). *Climate change impacts on global food production and security: A systematic review. Nature Sustainability*, 4(12), 1044–1056. <https://doi.org/10.1038/s41599-021-00631-7>
- Saad, R., Zainuddin, Z. M., & Arshad, R. (2022). *Data-driven decision-making in Malaysian agriculture. Environmental Science & Policy*, 112, 1-12. <https://www.sciencedirect.com/science/article/pii/S1462901121000576>
- Sharma, S., Rai, S., & Krishnan, N. C. (2020, November 3). *Wheat Crop Yield Prediction Using Deep LSTM Model*. arXiv.org. https://arxiv.org/abs/2011.01498?utm_source=chatgpt.com
- Silva, J. V., Van Heerwaarden, J., Reidsma, P., Laborte, A. G., Tesfaye, K., & Van Ittersum, M. K. (2023). Big data, small explanatory and predictive power: Lessons from random forest modeling of on-farm yield variability and implications for data-driven agronomy. *Field Crops Research*, 302, 109063. <https://doi.org/10.1016/j.fcr.2023.109063>
- Taufik, S., & Tharakan, P. (2021). *Towards data-driven agriculture in Malaysia: Enhancing food security through technology adoption. Journal of Agricultural Informatics*, 14(3), 50-60. <https://doi.org/10.24915/jai.v14i3.625>

CROP YIELD PREDICTION USING HISTORICAL DATA.docx

ORIGINALITY REPORT

55%	5%	7%	54%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Asia Pacific University College of Technology and Innovation (UCTI) Student Paper	53%
2	Submitted to Universiti Teknologi Malaysia Student Paper	1%
3	Submitted to University of Liverpool Student Paper	<1%
4	Abhilasha Sharma, Vishwas Rathi, Anupam Biswas, Anil Singh, Omer Rana. "Multimodal Artificial Intelligence in Precision Agriculture - Practices, Challenges, and Applications", CRC Press, 2026 Publication	<1%
5	Almarar, Maryam. "Detecting Fraud in Police Reports Using Machine Learning and Natural Language Processing", Rochester Institute of Technology, 2025 Publication	<1%
6	uu.diva-portal.org Internet Source	<1%
7	"Smart Computing Paradigms: Human-Centric Systems for Sustainable Development", Springer Science and Business Media LLC, 2026 Publication	<1%
8	managementjournal.usamv.ro Internet Source	<1%

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		