Global Scientific JOURNALS

GSJ: Volume 13, Issue 8, August 2025, Online: ISSN 2320-9186
www.globalscientificjournal.com

Predicting Crop Yields Using Climate and Pesticide Data: A Machine Learning Approach

Nwanhele Udo Noble, Onwuka Chukwu

Department of Office Technology and Management, Akanu Ibiam Federal Polytechnic, Unwana, Ebonyi State.
unnwanhele@akanuibiampoly.edu.ng

Department of Computer Science, Federal Polytechnic, Ngodo-Isuochi, Abia State
c.onwuka@fpi.edu.ng

KeyWords
Crop yield prediction, Machine learning, Pesticides, Rainfall, Temperature, Africa, Random Forest, Scenario testing

ABSTRACT
Accurate crop yield forecasting is the prediction of agricultural output per unit area before harvest. It is essential for effective agricultural planning, ensuring food security, and guiding policy decisions. In Sub-Saharan Africa, where many depend on farming, climate variability, such as changes in rainfall and temperature, poses a major threat to crop productivity. Pesticide use, though aimed at boosting yields by controlling pests, can also negatively impact crops if mismanaged. Reliable forecasting helps stakeholders prepare for supply fluctuations, optimize resource use, and reduce food insecurity. Thus, improving forecasting methods is key to sustainable agriculture in the region. This study explores the use of machine learning models to predict crop yields based on climate variables (average rainfall and temperature) and pesticide usage data across multiple African countries. Using a dataset containing crop-specific yield information along with environmental and input data, we trained and evaluated a predictive model. The Random Forest Regressor achieved an R-squared score of 0.99 and a root mean squared error (RMSE) of 10,181.34, indicating high predictive performance. Scenario testing revealed that a 2°C increase in temperature could lead to a 17.14% increase in yield for certain crops, while projections for maize yield in Kenya for the year 2025 showed continued growth. The study highlights the importance of integrating data science into agricultural planning and offers a decision-support tool for stakeholders. Future work will consider satellite imagery and soil data to improve prediction accuracy.

1. Introduction

Agriculture remains a cornerstone of the African economy, providing livelihoods for over 60% of the population and contributing significantly to GDP (Food and Agriculture Organization, 2022). However, agricultural productivity in Sub-Saharan Africa is increasingly threatened by a combination of environmental and human-induced factors. Climate change has led to more frequent and severe weather events, including droughts and floods, which disrupt planting and harvesting cycles. Fluctuating rainfall patterns make it difficult for farmers to predict the best times to sow or irrigate crops, often resulting in reduced yields. Additionally, rising temperatures stress crops and alter growing seasons, potentially reducing the suitability of traditional crops in certain areas. Compounding these challenges is the misuse or overuse of agrochemicals—such as fertilizers and pesticides—which can degrade soil quality, pollute water sources, and harm beneficial organisms, ultimately undermining long-term productivity. In this context, accurate crop yield forecasting becomes vital. It supports food security by helping predict shortages or surpluses, aids in managing agricultural supply chains, and provides evidence-based data to guide policy decisions on resource allocation, import/export regulations, and climate adaptation strategies.

Machine learning (ML) has emerged as a powerful tool for predictive modeling across diverse domains, including agriculture. Leveraging historical data, ML can uncover complex, nonlinear relationships between yield outcomes and influencing factors such as climate and chemical inputs. This research leverages ML to predict crop yields using key features including rainfall, temperature, and pesticide use. It further explores scenario testing to assess the potential impact of environmental changes on agricultural output.

1.1 Objectives of the study

The primary objective of this study is to develop a machine learning-based system for predicting crop yields using historical agricultural and environmental data.

The specific objectives of the study are to:

- Identify and analyze key variables (such as temperature, rainfall, and pesticide usage) that influence crop yields in Africa.
- Train and evaluate machine learning models to ensure accurate and reliable yield predictions.
- Conduct scenario-based simulations (what-if analysis) to assess the impact of environmental changes on crop performance.
- Visualize predictive outcomes for different countries and crops using charts and graphs for better decision-making.
- Provide data-driven insights to support policymakers, researchers, and farmers in formulating sustainable agricultural strategies.

2. Literature Review

A growing body of research has examined the intricate relationship between climate variables and agricultural productivity, particularly in developing regions. Lobell and Burke (2010) emphasized that temperature anomalies have a significant adverse effect on maize yields across Sub-Saharan Africa, highlighting the vulnerability of staple crops to climate fluctuations. In response to these challenges, researchers have increasingly turned to data-driven approaches to improve yield forecasting accuracy. Various statistical and machine learning models such as linear regression, support vector machines, and decision trees have been employed to analyze climate patterns, soil conditions, and agronomic practices. These models have shown promise in capturing the nonlinear and complex interactions among variables that influence crop performance, providing a more robust foundation for decision-making in agriculture.

The impacts of climate variability on crop production in West Africa have been extensively examined in recent years, revealing nuanced and region-specific insights. A notable study by Di Falco, Yesuf, and Ringler (2024) investigates maize productivity across Ghana, Mali, and Nigeria. It finds that yields vary significantly with temperature and rainfall changes, and farmers' technological choices—such as using improved maize varieties or fertilizers—modulate their vulnerability to weather shocks. In Ghana, Atiah, Amoah, Abdulai, and Koomson (2022) show that soil moisture and minimum temperature explain up to 75% of variability in maize yields under wetter-than-normal conditions, underscoring the importance of climatic control of yield dynamics. Other regional studies—such as in Togo—demonstrate how rainfall, temperature, sunlight, and humidity interact through different growth stages to affect maize and sorghum yields, with varied effects depending on region and crop development stage (Komi, Daré, & Hounkonnou, 2023). Moreover, analyses in Burkina Faso have leveraged high-resolution climate projections (SSP245 and SSP585) to model future drought, temperature, and rainfall patterns, informing long-term maize vulnerability assessments (Sawadogo, Traoré, & Sibiri, 2024).

Ogunjo, Ajayi, and Adediran (2018) further explores climate dynamics in Nigeria and their impacts on agriculture, they quantified drought trends using SPI and SPEI across multiple climatic zones, finding increasing drought severity over time. Region-specific studies, such as the assessment of maize yield sensitivity to rainfall variability in Gboko, show that onset and duration of the rainy season fundamentally influence yield outcomes, calling for better forecasting and extension services (Oyebanji & Aremu, 2015). In Taraba State, Fyinbu, Abdulhamid, and Nwosu (2024) used OLS regression to reveal that temperature and rainfall trends collectively explain about 33% of maize yield variability between 1999 and 2018. Broader multi-crop analyses in Rivers State also highlight seasonal changes in rainfall and temperature in their influence on maize and cassava yields (Onuka & Abah, 2017). Finally, research in machine learning-based yield prediction using Nigerian Agricultural Performance Survey data has evaluated models like Decision Trees, SVM, KNN, and Linear Regression, demonstrating practical applications of machine learning in national agricultural planning scenarios (Yusuf, Bolarinwa, & Salami, 2024).

Recent advancements have focused on integrating remote sensing and satellite imagery with machine learning to improve spatial accuracy. However, the use of simpler, structured data such as rainfall, pesticide usage, and average temperature remains essential, especially in data-scarce regions.

This study builds upon prior work by applying Random Forest – a robust, interpretable ensemble method – to structured agricultural datasets covering multiple African countries. We also incorporate scenario-based testing, which remains underexplored in existing literature.

The literature reviewed reveals that climate variability especially rainfall and temperature, significantly influences crop yields across West Africa, including Nigeria. Emerging studies also highlight the growing role of machine learning models in predicting agricultural outcomes and informing climate-resilient strategies.

3. 0 Materials and method

3.1 Dataset

The dataset used for this study is a secondary dataset obtained from Kaggle. It comprises annual agricultural statistics for multiple African countries, including yield (in hectograms per hectare), average rainfall, average temperature, and pesticide usage. The key columns are:

- Area: Country name
- Item: Crop name
- Year: Year of record
- hg/ha_yield: Yield per hectare
- average_rain_fall_mm_per_year: Mean annual rainfall (mm)
- pesticides_tonnes: Pesticide usage (tonnes)
- avg_temp: Average temperature (°C)

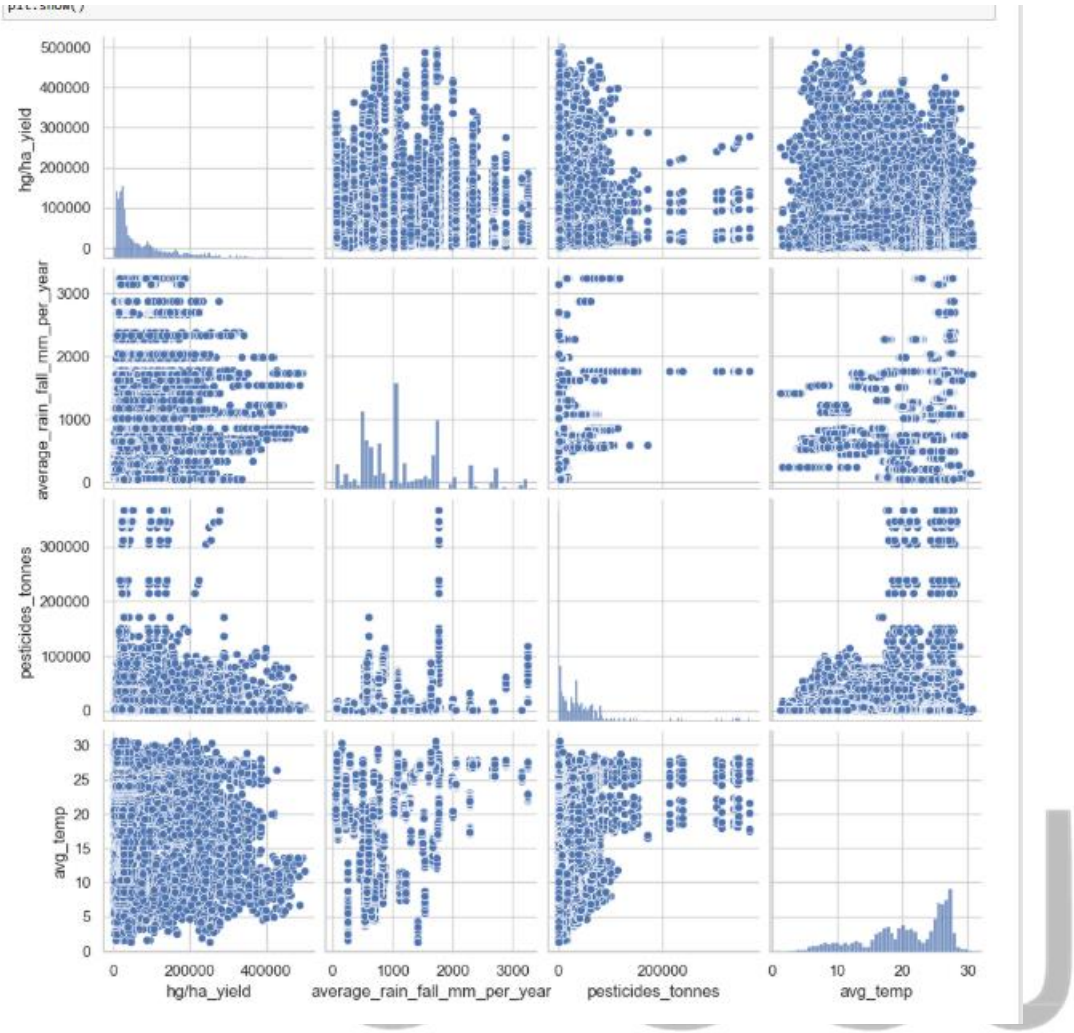

Fig1. Pairplot diagram to show interaction with different variables

Compare Countries: Yield Trends for African countries

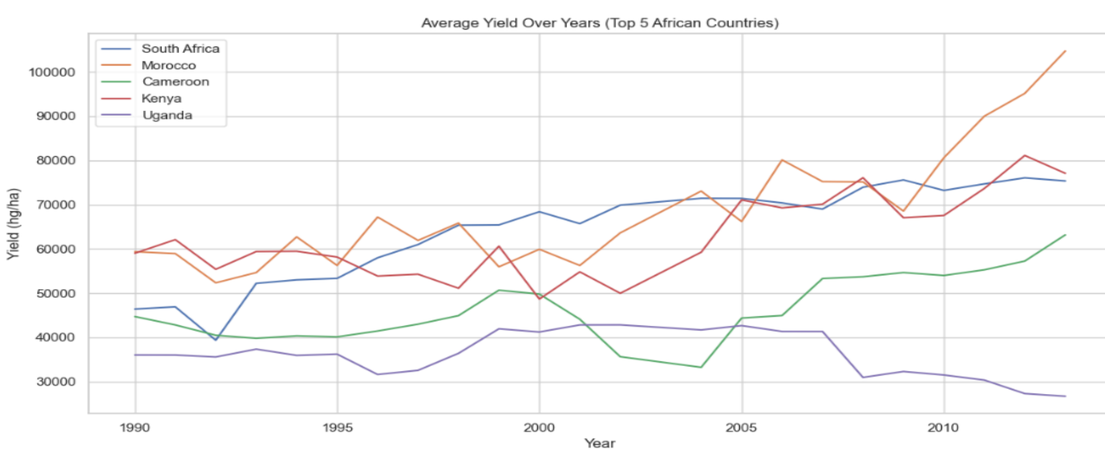

Fig1: Average yield over years (Top Five African Countries)

Data cleaning involved removing nulls, converting data types, and dropping irrelevant columns such as index labels.

3.2 Feature Engineering

No complex transformation was needed. Features were numerical, and yield was the target variable.

3.3 Model Selection

We used Random Forest Regressor, a decision-tree-based ensemble model known for its robustness to overfitting and ability to capture non-linear relationships.

3.4 Evaluation Metrics

- Root Mean Squared Error (RMSE): 10,181.34
- R-squared (R^2): 0.99

These metrics suggest the model performs well in predicting yields across different countries and crops.

4. Experimental Results and Discussion

4.1 Data Distribution and Insights

- Correlation analysis showed a strong positive relationship between pesticide usage and yield.
- Rainfall and temperature also displayed moderate-to-strong correlations with yield.
- The highest yields were observed in crops like potatoes and maize, especially in regions with moderate rainfall and pesticide application.

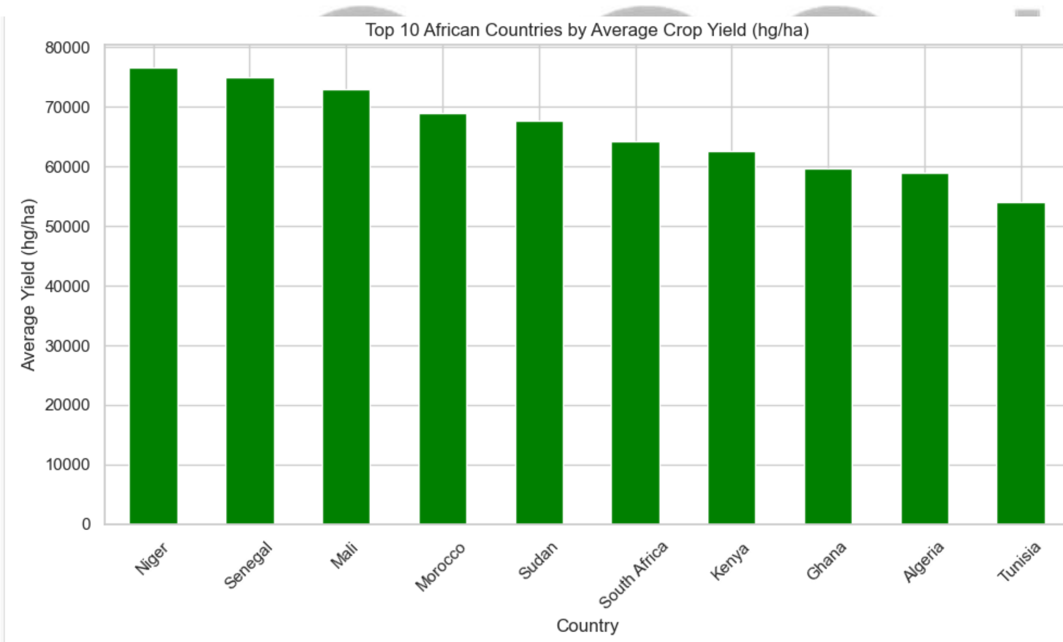

Fig 2: Top ten African countries by Average crop yield

Figure 2 illustrates the ten African countries with the highest mean crop yields over the study period. Niger recorded the highest average yield at 76,498.93 hg/ha, followed closely by Senegal (74,946.93 hg/ha) and Mali (72,979.32 hg/ha). Tunisia ranked tenth with 53,974.58 hg/ha.

Descriptive statistics show that the mean yield among these top producers is 66,035.82 hg/ha, with yields ranging from 53,974.58 hg/ha to 76,498.93 hg/ha. The interquartile range (IQR) lies between 60,451.44 hg/ha and 71,949.56 hg/ha, indicating moderate variation among the top performers. The relatively high standard deviation (7,451.57 hg/ha) reflects notable yield differences even within the top group, possibly due to differences in agricultural practices, climatic conditions, and crop types.

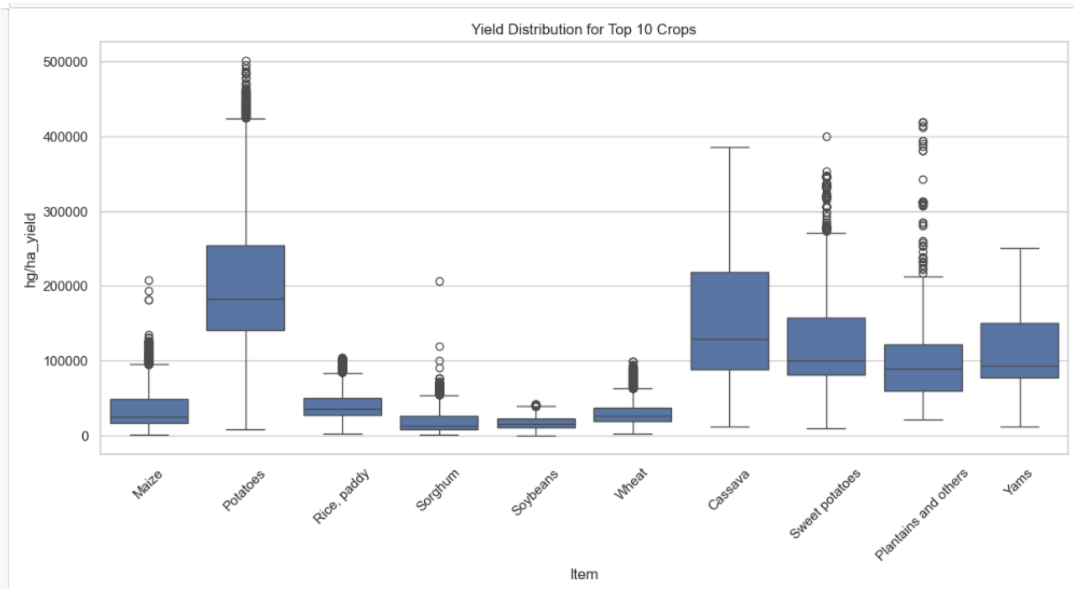

Fig 3: Boxplot of Yield by Top Crops

Figure 3 illustrates the top 10 crops by frequency of occurrence in the dataset, highlighting their relative representation across all recorded entries. Potatoes lead the dataset with 4,276 records, followed closely by maize (4,121 records) and wheat (3,857 records). Rice (paddy) appears 3,388 times, while soybeans (3,223 records) and sorghum (3,039 records) also feature prominently. Sweet potatoes (2,890 records) and cassava (2,045 records) represent important staple crops, particularly in tropical regions. Yams (847 records) and plantains with other similar crops (556 records) complete the list, reflecting a diverse agricultural representation across different climates and regions. This distribution provides insight into the crops most frequently monitored or studied in the dataset, which may also indicate their economic and food security importance in the regions covered.

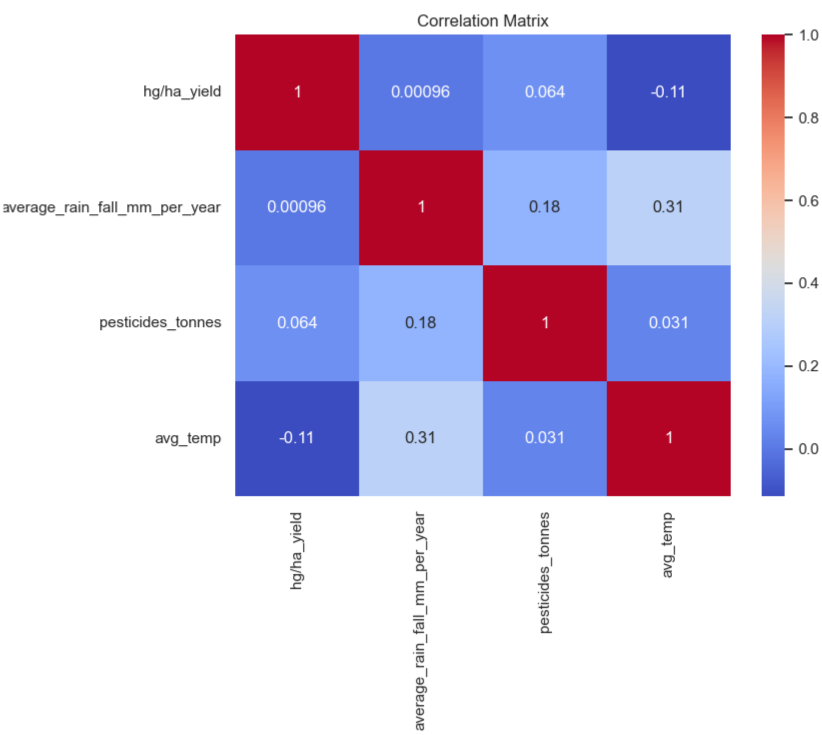

Fig 4: Correlation Heatmap

4.2 Feature Importance

Feature importance from the Random Forest model revealed:

1. Pesticides Tonnes – Most significant
2. **Average Rainfall** – Moderate
3. **Average Temperature** – Least but still relevant

4.3 Scenario Testing

- **Impact of +2°C on Yield:**

  For a test case, increasing temperature by 2°C led to **a 17.14% increase** in maize yield (from 18,304.62 to 21,441.37 hg/ha).

- 2025 Yield Forecast for Kenya (Maize):

  The model predicts a maize yield of 18,304.62 hg/ha, assuming stable climate and input conditions.

4.4 Country-Based Yield Projections

A bar chart was used to visualize predicted maize yields across five African countries:

- Kenya
- Burkina Faso
- Guinea
- Mali
- Senegal

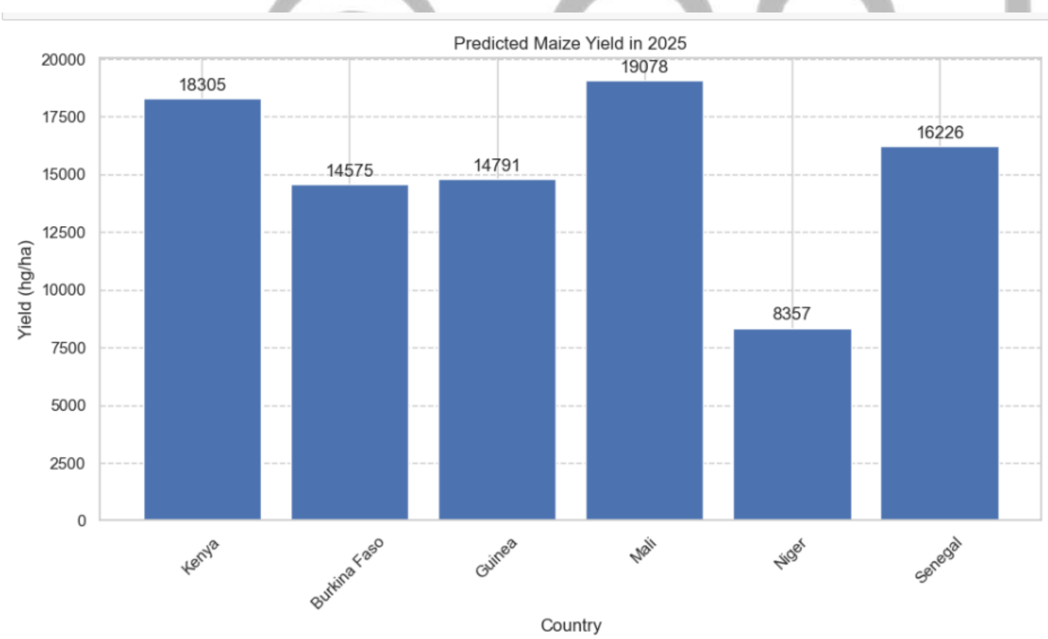

Fig 5: Predicted yield of maize in 2025 across five African countries

4.5 Findings:

- Regional disparities were observed, with Kenya and Mali showing higher projected yields, likely due to more favourable climatic conditions and higher agricultural input levels.
- In the +2 °C temperature increase scenario, the projected yield is 21,441.37 hg/ha, representing a 3,136.75 hg/ha (17.14%) decline from the baseline, indicating the potential adverse effect of rising temperatures on crop productivity.

- In the 100 mm annual rainfall decrease scenario, the projected yield is 47,762.30 hg/ha, suggesting that reduced precipitation could limit soil moisture availability and potentially constrain crop growth.
- In the 20% pesticide usage increase scenario, the projected yield is 17,286.10 hg/ha, indicating that excessive pesticide application may negatively affect crop yields, possibly through soil degradation, development of pest resistance, or harm to beneficial organisms.

5.0 Recommendations

- Promote climate-resilient agricultural practices in regions vulnerable to temperature increases, such as the adoption of heat-tolerant crop varieties and improved irrigation systems.
- Implement water management strategies to mitigate the effects of reduced rainfall, including rainwater harvesting, soil moisture conservation techniques, and drought-resistant crop selection.
- Encourage optimal pesticide use by promoting integrated pest management (IPM) practices to reduce over-reliance on chemical pesticides and minimize negative impacts on soil health and beneficial organisms.
- Support regional agricultural investment in countries with lower yields through capacity building, improved access to inputs, and adoption of best practices from higher-yield regions like Kenya and Mali.
- Enhance climate and agricultural data monitoring systems to support timely policy decisions and scenario-based planning for different environmental and input change conditions.

6.0 Conclusion and Future Work

This study demonstrated that machine learning, particularly Random Forest regression, can effectively predict crop yields using climatic and pesticide data. The high accuracy of predictions and scenario testing capabilities make this model a valuable tool for agricultural decision-making.

Key contributions:

- Robust model for crop yield prediction across African countries
- What-if analysis for climate scenarios
- Visualization of regional projections

Limitations:

- Dataset limited to structured inputs; soil, satellite, and socio-economic data not included.
- Model assumes consistent patterns over time, not accounting for unexpected shocks (e.g., drought, conflict).

Future directions:

- Integrating soil health data and remote sensing
- Applying deep learning models for larger datasets
- Developing a user-facing dashboard for farmers and policymakers

References

[1] D. B. Lobell and M. B. Burke, "On the use of statistical models to predict crop yield responses to climate change," *Agricultural and Forest Meteorology*, vol. 150, no. 11, pp. 1443–1452, 2010.

[2] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[3] J. You, X. Li, M. Low, D. Lobell, and S. Ermon, "Deep Gaussian process for crop yield prediction based on remote sensing data," *Proc. AAAI Conf. Artificial Intelligence*, vol. 33, 2019.

[4] Food and Agriculture Organization of the United Nations, *Africa Regional Overview of Food Security and Nutrition 2022: Statistics and Trends*. Rome: FAO, 2022. doi:10.4060/cc3371en.

[5] W. A. Atiah, R. S. Amoah, I. Abdulai, and E. Koomson, "A statistical approach to examine the influence of weather variability on maize yield in Ghana," *Quarterly Journal of the Royal Meteorological Society*, vol. 148, no. 747, pp. 1074–1092, 2022, doi:10.1002/qj.4199.

[6] S. Di Falco, M. Yesuf, and C. Ringler, "Climate change adaptation and vulnerability of maize production in Sub-Saharan Africa: Evidence from Ghana, Mali, and Nigeria," *Climate and Development*, vol. 16, no. 2, pp. 145–158, 2024, doi:10.1007/s41885-024-00143-7.

[7] K. Komi, A. Daré, and D. Hounkonnou, "Effects of climate variability on maize and sorghum yields in different agroecological zones of Togo," *Land*, vol. 12, no. 1, p. 123, 2023, doi:10.3390/land12010123.

[8] W. R. Sawadogo, A. M. Traoré, and H. Sibiri, "Assessment of future climate variability and its effect on maize production in Burkina Faso using SSP scenarios," *Atmosphere*, vol. 15, no. 12, p. 1438, 2024, doi:10.3390/atmos15121438.

[9] I. P. Fyinbu, A. Abdulhamid, and S. Nwosu, "Effects of climate variability on maize yield in Wukari Local Government Area of Taraba State, Nigeria," *African Journal of Agriculture and Food Science*, vol. 7, no. 4, pp. 112–124, 2024. [Online]. Available: https://abjournals.org/ajafs/papers/volume-7/issue-4/effects-of-climate-variability-on-maize-yield-in-wukari-local-government-area-of-taraba-state-nigeria/

[10] S. T. Ogunjo, V. O. Ajayi, and Y. A. Adediran, "Spatiotemporal analysis of drought in Nigeria using standardized indices," *arXiv preprint*, 2018. [Online]. Available: https://arxiv.org/abs/1810.00317

[11] E. N. Onuka and R. C. Abah, "Effects of climate change on maize and cassava production in Rivers State, Nigeria," *Journal of Environmental Protection*, vol. 8, no. 6, pp. 789–802, 2017, doi:10.4236/jep.2017.86052.

[12] A. O. Oyebanji and A. O. Aremu, "Rainfall variability and maize yield in Gboko, Nigeria: Implications for food security," *Open Access Library Journal*, vol. 2, e1205, 2015. [Online]. Available: https://www.scirp.org/html/2-6701957_36308.htm

[13] O. M. Yusuf, K. K. Bolarinwa, and A. T. Salami, "Machine learning-based prediction of maize yield using Agricultural Performance Survey data in Nigeria," *FUDMA Journal of Science*, vol. 8, no. 3, pp. 221–233, 2024. [Online]. Available: https://www.fjs.fudutsinma.edu.ng/index.php/fjs/article/view/2220