International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 04 | April – 2025 DOI: 10.55041/ISJEM03827
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

# AgroYield: Country Wise Crop Yield Prediction Using Machine Learning

**Ranjana Ray[1], Saptarshi Mondal[1], Anindita Sarkar[1], Nisha Pandey[1]**

[1]*Electronics and Communication Engineering, JIS College of Engineering*

---

**Abstract -** Some of the current challenges in modern agriculture include unsustainable pesticide use along with the untimely rainfall, increasing temperatures, and climate change which directly impact the productivity of crops and food security. In light of these challenges, we have developed a solution that predicts a country's crop yields with unprecedented accuracy and reliability called AgroYield, this solution employs machine learning technology and leverages features such as crop type, country, year, average rainfall, temperature, and pesticide usage to make yield predictions in kg/hectare. The dataset was preprocessed with OneHotEncoding for categorical variables and StandardScaler for numerical inputs. After trialing multiple regression models, the Decision Tree Regressor was found to perform best. Alongside yield forecasting, AgroYield enables decision-making towards sustainable farming, fostering long-term ecological balance. By helping farmers, agronomists, and policymakers adapt to resource and environmental constraints, this solution aids in building a resilient agricultural ecosystem.

***Key Words*:** Crop Yield Prediction, Machine Learning, OneHotEncoding, StandardScaler, Decision Tree Regressor

## 1.INTRODUCTION

Global food security, economic growth, and social stability are all based on agriculture, especially in developing nations where it continues to be the main source of income for a sizable section of the populace. The demand for food is predicted to increase sharply as the world's population grows, putting agricultural systems under previously unheard-of pressure to boost output without sacrificing environmental sustainability.
However, there are many difficulties facing modern agriculture. Unpredictable weather patterns brought on by climate change, such as irregular rainfall and extremely high temperatures, have a direct effect on crop yield and health. The strain on agricultural productivity is further increased by soil degradation, excessive use of chemical pesticides, water scarcity, and a lack of arable land. These problems highlight the pressing need for data-driven, adaptable, and intelligent solutions that can support stakeholders in making prompt, well-informed decisions.

Crop yield, or the amount of crop produced per unit area, is one of the most important agricultural metrics. From individual farm-level strategies to national food security policies and international trade planning, precise yield forecasting is essential to many decision-making processes. Conventional statistical and empirical approaches frequently fail to capture
less accurate forecasts due to the dynamic interactions between various agro-environmental factors.

To overcome these issues, this research proposes AgroYield—a machine learning-based country-wise crop yield forecasting system. By combining past crop data with important environmental factors like rainfall, mean temperature, pesticide application, year, crop type, and country, AgroYield gives yield predictions in kg/hectare. The project uses preprocessing methods such as OneHotEncoding for categorical data and StandardScaler for numeric features to keep the data clean and improve model accuracy.

Different machine learning techniques were tested in terms of performance, such as Linear Regression, Lasso, Ridge, K-Nearest Neighbors, and Decision Tree Regressor. Of all these, the Decision Tree Regressor yielded the most precise and understandable outputs. The model and its pipeline for preprocessing are saved for the deployment in the real world such that users may provide environmental factors and get an instant yield prediction.

By offering a powerful, scalable, and interpretable framework for crop yield prediction, AgroYield enables farmers, agronomists, and policymakers to make anticipatory decisions, distribute resources effectively, and develop resilience to climate variability and agricultural threats. This project is an advance in leveraging artificial intelligence for sustainable farming and secure food supply in the long term.

## 2.SURVEY ON RECENT INVESTIGATIONS

### I. Background

The growing demand for food security in the face of climate variability and a rising global population has positioned crop yield prediction as a priority in agricultural research. Machine Learning (ML) has emerged as a transformative tool, offering data-driven insights to improve farming strategies and maximize productivity. Multiple recent studies have investigated ML's potential in addressing challenges related to yield forecasting, resource optimization, and sustainable agricultural practices.

### II. Methodology

Various methodologies have been employed across investigations, primarily involving supervised ML algorithms. Chlingaryan et al. (2018) integrated sensor data with ML models to estimate nitrogen status, while Elavarasan et al. (2018) examined climate-based features—such as rainfall and temperature—to model yield predictions. Techniques like regression analysis, decision trees, K-nearest neighbors, and ensemble learning models have been widely applied. Preprocessing steps typically include OneHotEncoding for categorical data and feature scaling for numerical values, similar to the methodology adopted in the *AgroYield* project.

### III. Results

These studies report promising outcomes in yield prediction accuracy and feature importance analysis. Liakos et al. (2018) showed consistent results across domains such as crop health, soil monitoring, and water usage. Li et al. (2018) demonstrated the efficiency of ML in estimating fruit ripeness, helping optimize harvest timing. The use of Decision Tree Regressors and ensemble models, as noted in Beulah (2019), yielded superior performance due to their ability to handle non-linear relationships and feature interactions.

### IV. Discussion and Implications

The research community has highlighted the critical role of multi-dimensional data—combining meteorological, geographical, and biological parameters—for improving model performance. The reviewed studies emphasize that yield prediction systems should not only focus on accuracy but also interpretability and adaptability to local conditions. The findings support the approach taken in *AgroYield*, where features like rainfall, pesticides, temperature, crop type, and country contribute to more granular predictions.

The real-world uses of machine learning-crop yield estimation models such as AgroYield are widespread and revolutionary, especially in responding to the complex issues of the agricultural industry. They have the capability to greatly improve planning and decision-making at several different levels, ranging from small-holder farmers to national governments and international agribusinesses. Some important civil uses are:

## 3. CIVIL APPLICATIONS

**Empowering Farmers with Data-Driven Crop Selection**

Through the analysis of historical and climatic data, ML models can suggest the best crops for a location and time of year. This enables farmers to make decisions based on data that optimizes yield potential, minimizes financial risk, and enhances profitability. It also facilitates more effective planning of sowing dates on the basis of forecasted weather and soil conditions.

**Supporting Policymakers in Resource Planning and Food Security Strategies**

Governments and agricultural ministries may employ yield forecasts to distribute subsidies, fertilizers and seeds in an efficient manner, and make contingency plans for anticipated low yields. The information also assists national food security policy by predicting shortages or surpluses and informing import/export policies accordingly.

**Optimizing Agribusiness Supply Chains**

Precise yield estimation enables agribusinesses to more accurately predict production quantities, organize logistics, operate storage facilities, and stabilize prices. This reduces post-harvest losses, optimizes inventory management, and improves coordination along the agricultural value chain.

**Improving Climate-Resilient Agriculture**

Machine learning techniques may assist farmers and stakeholders to foresee the effects of climate variability, for instance, heatwaves or droughts, and execute adaptation measures prior. These entail the realignment of planting times, selecting stress-resistant crop types, and taking up efficient irrigation practices.

**Supporting Financial Institutions in Agricultural Credit Risk Assessment**

Banks and microfinance institutions can employ yield forecasts to evaluate the risk profile of farm loans. Accurate forecasts assist in the formulation of insurance products and credit products that are suited to farmers' requirements, ensuring financial stability and inclusion.

**Enabling Smart Farming and Precision Agriculture**

Yield forecasts can be combined with IoT sensors, satellite imagery, and drone data to facilitate precision agriculture. This involves precision irrigation, fertilization, and pest control that enhance efficiency and reduce environmental footprint.

**Helping NGOs and Relief Agencies in Crisis Response**

In famine-prone, drought-stricken, or crop failure-prone areas, yield forecasting models can be used as early warning systems. Humanitarian organizations are thus able to mobilize food aid and resources ahead of time, minimizing human suffering and avoiding socio-economic dislocation.

**Impelling Research and Innovation in Agricultural Sciences**

Scientists can utilize the information and data from yield forecasting models to research crop behaviour, discover new

methods of cultivation, and determine the long-term effects of climate change on agriculture.

## 4. MATERIAL AND METHODS

### I. Flask

Flask is a micro web framework written in Python. It is designed to be lightweight, modular, and easy to use, making it ideal for both beginners and professionals who want to build web applications quickly and with flexibility.
Unlike larger frameworks (like Django), Flask does not come bundled with form validation, database abstraction layer, or other pre-built tools—though you can add them as needed. This makes Flask especially suitable for small to medium-sized projects and for integrating with external libraries like machine learning models.

### II. Numpy

NumPy is a robust Python library employed for numerical computing, and within this project, it was mainly used to organize and manage the input data in numerical form prior to sending it to the machine learning model. NumPy arrays offer an optimal and organized way of dealing with input features, and they guarantee smooth compatibility with the preprocessing pipe and the model. Its optimized array operations provide substantial performance advantages over native Python lists, accelerating computations and reducing memory usage. Additionally, because Scikit-learn models need input data as NumPy arrays, NumPy was instrumental in facilitating seamless integration during preprocessing and prediction stages of the project.

### III. Pickle

Pickle is a Python package employed for serializing and deserializing Python objects, which makes it crucial for saving trained machine learning models and data preprocessing pipelines. In this project, Pickle was utilized to load the pre-trained machine learning model and the data preprocessing pipeline, such that the same transformations that were applied at training time are applied uniformly at prediction time. It facilitates model deployment by providing a way of saving the trained model to a .pkl file that can be reused later in production without needing to retrain it. Pickle also facilitates saving the entire transformation logic as part of the pipeline, which ensures consistency and reliability between various stages of the ML workflow. Its lightweight and efficient profile makes it a perfect choice for object persistence, particularly when incorporating machine learning functionality within web applications such as the one in this project.

### IV. Scikit-Learn

Scikit-learn (sklearn) is a powerful and popular Python machine learning and data preprocessing library, and in this project it played more than one key role. First, it offered the Decision Tree Regressor (DTR) model, which was used to train on past agriculture data to forecast crop yield. Secondly, Scikit-learn was employed to build the pipeline for preprocessing, which comprised operations like feature scaling, encoding, and data transformation to preprocess user inputs for precise prediction. Finally, it provided superior compatibility with NumPy arrays, helping seamlessly integrate during both the training phase and prediction phase of the machine learning process. In summary, Scikit-learn was a key component in facilitating effective model building, transformation, and deployment in this project.

### V. Matplotlib

Matplotlib is a robust Python library employed for the generation of static, animated, and interactive visualizations, and in this project, it was used to graphically depict the numerical relationships and trends in user inputs and model outputs. It provides fine-grained control over different plotting components like axes, labels, titles, colours, and grids, enabling precise and personalized data visualizations. Such visualizations enable people to better perceive the impact of variables such as rainfall, temperature, or use of pesticides on crop yield forecast, and make the output easier to interpret and understand.

### VI. Seaborn

Seaborn is a Python library for data visualization that is based on Matplotlib, and in this project, it was utilized to produce informative and aesthetically appealing statistical plots that made the data more interpretable. Seaborn makes it easier to generate complex visualizations like heatmaps, bar charts, and scatter plots with less code, which makes it simpler to investigate and communicate insights from the data. It integrates very well with structured data formats such as pandas DataFrames, enabling easy integration with the dataset employed in the project. In general, Seaborn assisted in better presenting relationships and patterns in the agricultural data, enabling better understanding and analysis.

### VII. Pandas

Pandas is a flexible and robust Python library for handling and analyzing data, and here, it has been utilized to process, manage, and analyze user-entered data prior to feeding the data to the machine learning model. Incorporating Pandas has provided the project with the functionality to easily handle data organization through DataFrames and Series to improve the structuring of features as well as providing consistency in formatting input. Pandas were also used to clean and validate user inputs, making sure that they were compatible with what the model required. It also allowed faster feature summarization and analysis, including the computation of averages, detecting minimum or maximum values, and analyzing correlations in the data. This enhanced interpretability and debugging while developing. Pandas also offers scalability, as the project can easily be extended to take CSV files for bulk predictions, where Pandas loads and transforms the data efficiently. Additionally, its easy

integration with NumPy and Scikit-learn facilitated easy transfer between data preparation and model training or prediction and thus became an integral part of the project's machine learning pipeline.

## VIII. Kaggle

Kaggle is a top data science and machine learning platform that provides high-quality datasets, collaboration tools, and community resources. In this project, Kaggle was the major source of the dataset for training and testing the crop yield prediction model. We particularly employed the yield_df.csv data from a publicly shared Kaggle dataset, which contains prominent agricultural attributes like rainfall, average temperature, pesticide use, year, and crop type. These attributes were crucial in developing and cross-validating the machine learning pipeline. Kaggle's simplicity of access, high-quality data, and friendly community made it a great platform for obtaining and testing real-world agricultural data.

The model for the prediction of crop yield was framed based on historic agricultural and weather data obtained from Kaggle (refer to Table 1 for input features). Predictors covered agro-climatic as well as biochemical characteristics like temperature, pesticide consumption, rainfall averages, crop kind, and nation-wise data. These were considered to train the Decision Tree Regressor (DTR) due to their interpretability coupled with the accuracy in dealing with nonlinear relationships amongst input features as well as target variable yield.

To maintain uniformity and decrease variance in the prediction pipeline, a preprocessing phase was carried out utilizing Scikit-learn's native utilities. OneHotEncoding was utilized for categorical variables like Country, Crop, and StandardScaler was utilized for numerical variables like Rainfall, Temperature, Pesticides. The trained model and the preprocessing pipeline were serialized with the help of Pickle to ensure effortless deployment using a web-based interface developed based on Flask.

For transparency and enabling reproducibility, the data set was divided equally into two groups: the training data set for one half and the other half as a test data set to evaluate models. This ensured all model comparison and evaluation would be based on previously unseen data points. Though the Decision Tree model was the focus of prediction, it was benchmarked by evaluating the performance of its test set against a baseline simple Multiple Linear Regression (MLR) model that was trained with the same training dataset. The performance of models was evaluated in terms of $R^2$ score and mean squared error (MSE) measures to reflect accuracy and variance in the predictions.

Visualization of input-output relationships and feature effects was done through Matplotlib and Seaborn, both supporting static plots and trend visualizations over variables. The entire pipeline from preprocessing to prediction was implemented through a Flask-based interface, supporting real-time user input and yield estimation on chosen features.

## 5. DATA COLLECTION

The data set utilized in this project was obtained from Kaggle, which is a commonly used data science and machine learning platform. In particular, we used the "Crop Yield Prediction Dataset," which was downloaded in CSV format with the file name **yield_df.csv**. The data set comprises around 5,000 entries with extensive information regarding the agricultural parameters influencing crop yield in different countries and years. The dataset contained records from **101 countries** out of a possible 195, offering diverse geographic and climatic representation for improved generalization of the model.

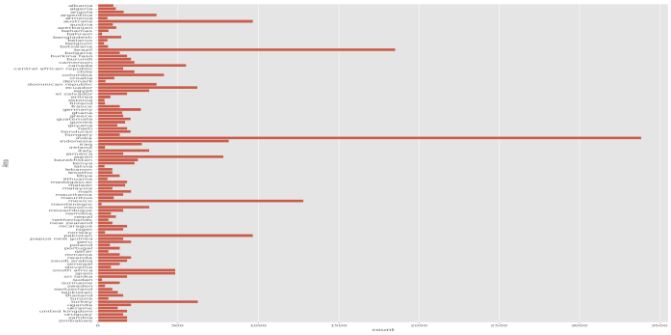

Fig1: Country-wise Distribution of Dataset Samples

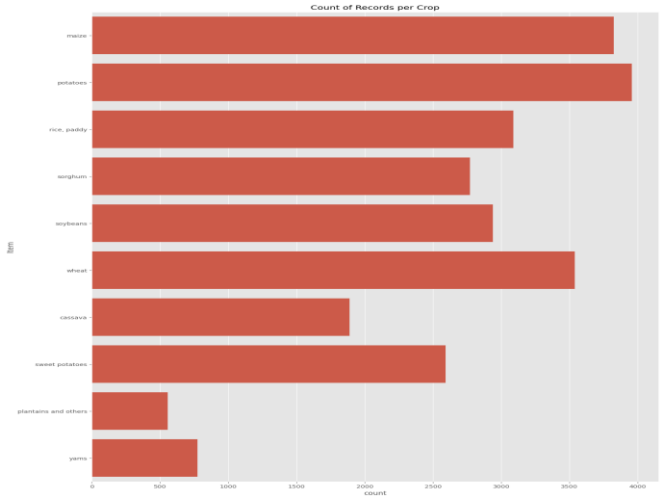


Fig2: Dataset Distribution by Crop Type

The data set comprises the following features (input variables):

**Country**: The country name where the crop was grown.

**Crop:** Crop type grown (e.g., Wheat, Maize, Rice, Paddy, Sorghum, Soybeans).

**Year:** The year that data was taken.

**Average_Rainfall:** The average annual rainfall total (in millimeters).

**Pesticides:** Amount of pesticide used in cultivation (in kilograms per hectare).

**Average_Temperature:** Average growing season temperature for the crop (in degrees Celsius).

The target variable in the dataset is

**Yield:** This is the real yield of the crops, in terms of tons per hectare, which is taken as the machine learning model prediction output.

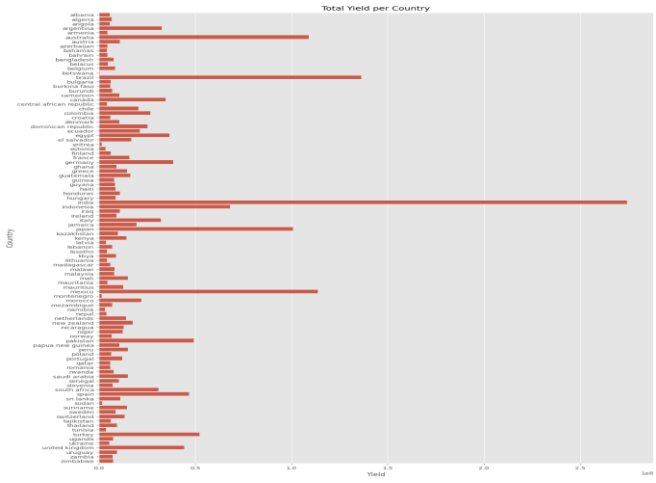


Fig3: Country-wise Aggregate Crop Yield

This dataset was selected because it covers a wide geography, has a variety of available agricultural factors, and is optimal for creating a strong regression-based prediction model. This data provided the basis for preprocessing, training, and model evaluation in the crop yield prediction pipeline.

### 6. DATA PREPROCESSING AND MODEL PIPELINE

### I. Handling missing values

In this research, we are developing a machine learning model that predicts crop yield on the basis of a number of features like area, average rainfall, use of pesticide, temperature, and crop type (e.g., Maize, Rice). But, as with most real datasets, the data that we have to work with may include missing or incomplete values. These missing values may be caused by a range of factors, including data collection errors, reporting gaps in the data for some periods of time, or inconsistency in the manner in which the data is reported.

## Why Missing Value Handling Is Significant in Crop Yield Prediction

**Incomplete Data:** Missing values in essential features like temperature, rainfall or pesticide usage can interfere with the learning process of the model. When any of these essential features are missing for a sample, the model would not be able to comprehend the relationships between these features and the target variable (crop yield). Consequently, the model's generalization capability to new, unseen data might be affected, making the predictions inaccurate.

**Model Performance:** Most machine learning models, such as decision trees, demand that all input data be present and absence-free. If the data set has missing data and the model can't process them directly, either the training will fail or the model will perform badly. That's why missing data handling before inserting data into the model is very important to obtain smooth model training and accurate predictions.

## Approach to Handling Missing Values

To address missing values in our project, we employed the dropna() function of the pandas library. This is a straightforward but efficient way of deleting rows that have any missing (NaN) data. In this manner, only rows containing complete data—i.e., no missing data in any feature—are fed into the model for training and testing.

In mathematical terms, this operation can be described as follows:

$$X = X \setminus \{X_i \mid \exists j, X_{ij} = \text{NaN}\}$$

This operation removes any rows $X_i$ where any feature value $X_{ij}$ is missing. The dataset is then reduced to only include complete rows, which ensures that the machine learning model is trained and tested on a full set of data points without any missing values.

While removing rows with missing values is a straightforward approach, it's not always the most suitable solution. If a large portion of the dataset contains missing values, dropping these rows could result in a significant reduction in the data size. This would lead to a smaller training set, which could negatively impact the model's ability to generalize effectively. In such cases, alternative strategies like **imputation**—where missing values are filled with values like the mean, median, or mode of the respective feature—could be considered. However, for our project, we opted for row deletion, assuming the missing values were few enough that removing them would not significantly affect the dataset.

#### Impact on the Project

In the context of our crop yield prediction model, the dataset contains crucial features such as area, rainfall, and temperature. Dropping rows with missing values ensures that the model only learns from complete and accurate data. This reduces the risk of the model making biased or unreliable predictions due to incomplete information. By performing this step, we help ensure that our machine learning model will make predictions

based on a consistent and complete understanding of the relationships between the features and the target variable (crop yield).

In summary, handling missing values by removing rows with NaN values is a critical step in ensuring that the machine learning model performs well. This preprocessing step prevents errors during model training and ensures that the model is trained on high-quality data. By handling missing values effectively, we improve the model's ability to generalize to new, unseen data, which is essential for making accurate crop yield predictions.

**II. Feature Selection**

In the process of building a robust crop yield prediction model, preprocessing of input features is essential to ensure the model performs optimally. One critical preprocessing step is feature scaling, particularly when dealing with real-world datasets that include features of different units and ranges, such as climatic and agricultural parameters.

**Importance of Feature Scaling in Crop Yield Prediction**

The dataset used in this study contains continuous variables such as:

- **Average Rainfall** (in mm per year),
- **Pesticides Used** (in tonnes),
- **Average Temperature** (in degrees Celsius), and
- **Year** (numerical, representing time).

These features vary significantly in scale. For example, the values for average_rain_fall_mm_per_year can range from a few hundred to several thousand, while avg_temp typically ranges between 10°C to 40°C. pesticides_tonnes may span a wide distribution, including zero values and very large numbers depending on the country and crop. This variation in scale can pose a challenge for many machine learning algorithms, where large-magnitude features might unduly influence the model's learning process.

Although Decision Tree Regressors, which were used in our final model, are generally robust to unscaled data (as they split nodes based on feature thresholds rather than distances), scaling still plays a crucial role during the training of alternative models (such as linear regression, SVMs, or neural networks). Moreover, it ensures consistency and interoperability when experimenting with ensemble techniques or pipelines that combine different algorithms.

**Scaling Method Used**

To standardize the numerical features, we applied Standardization using the **StandardScaler** class from the scikit-learn library. This transformation adjusts each feature so that it has:

- A **mean of 0**, and
- A **standard deviation of 1**.

This is mathematically defined as:

$$x' = \frac{xj - \mu j}{\sigma j}$$

Where:

- $x_j$ = original value of feature j
- $\mu_j$ = mean of feature j
- $\sigma_j$ = standard deviation of feature j
- $x'_j$ = standardized value

Categorical variables like Area and Item were excluded from scaling and were handled separately via **OneHotEncoding**, as they are not continuous numeric features.

Scaling the data helped provide an even and stabilized learning environment to the model. Although our finalized pipeline utilizes a Decision Tree Regressor that is inherently impervious to scaling, uniform format facilitated suitable experimentation with different regression models as well as stackable models down the line. More importantly, by promoting commonality across the distributions of the features, we mitigated overpowership that could occur amongst features such as rainfall in the rainy nations that have higher numbers.

Therefore, feature scaling played a critical role in boosting model generalizability, enhancing cross-model compatibility, and facilitating a more interpretable and reproducible training pipeline for country-wise crop yield prediction.

**III. Splitting the Dataset**

After the dataset is cleaned, encoded, and scaled, the second most critical operation in the machine learning pipeline is splitting the data into training and test sets. This operation is very important to ensure that the model learns from one part of the data and then gets tested on another part, unknown data, to measure its generalization performance.

**Rationale for Splitting the Dataset**

In any supervised learning problem—like our objective to predict crop yield (hg/ha_yield)—it's important to measure how accurately a model performs on fresh, unseen data. Training and testing on the same data would cause overfitting, with the model remembering the data instead of learning the patterns behind the data.

To counter this, the dataset is split into two broad subsets:

**Training set:** Employed by the model to learn the associations between input features and the target variable.

**Testing set:** Employed to independently test the model's predictive performance on unseen data.

X = df.drop('hg/ha_yield', axis=1):

This line extracts all features (independent variables) from the dataset except the target variable hg/ha_yield.

y = df['hg/ha_yield']:

This isolates the target variable, which is the **crop yield measured in hectograms per hectare**.

train_test_split(...):

This function from the scikit-learn library is used to randomly divide the dataset:

- test_size=0.2: 20% of the data is set aside for testing, while 80% is used for training. This ratio balances model training and evaluation.
- random_state=0: Ensures **reproducibility**—using the same random state will always result in the same split.
- shuffle=True: Ensures the data is shuffled before splitting. This is crucial to avoid **bias** caused by sequential data patterns (e.g., years or countries appearing in sorted order).

$X \in \mathbb{R}^{(nd)}$ : Feature matrix with n samples and d features.

$y \in \mathbb{R}^n$ : Target vector representing crop yield.

The dataset is split into:

Training set: $( X_{train}, y_{train} )$

Testing set: $( X_{test}, y_{test} )$

$X_{train} \cup X_{test} = X$ and $y_{train} \cup y_{test} = y$

$|X_{test}| = 0.2 \times n$ , $|X_{train}| = 0.8 \times n$

Splitting the dataset allowed for the unbiased testing of the machine learning model. In keeping the testing data apart throughout training, we were able to quantify how good the model generalized to new situations, which in real-world crop yield prediction application is crucial when the model should perform under diverse years, nations, and types of crops. This step helps ensure that the performance measures like $R^2$ score, MAE (Mean Absolute Error), and RMSE (Root Mean Squared Error) are a true indicator of the model's predictive ability and not its capacity to memorize past data.

## IV. Model Selection and Evaluation

In this crop yield prediction project, we tried to develop a good machine learning model that can be used to forecast the yield of different crops with respect to environmental and agricultural factors like country, year, crop type, mean rainfall, use of pesticides, and temperature. To do that, we contrasted the performances of five diverse regression models. Each model was tested on how well it generalizes on unseen data based on standard evaluation metrics: Mean Absolute Error (MAE), Mean Squared Error (MSE), and the Coefficient of Determination ($R^2$ score).

### Linear Regression

A basic algorithm that presumes a linear relationship between the independent and dependent variables. It was employed as a base model to measure the performance improvement offered by more sophisticated methods.

### Lasso Regression

A regularized linear regression that uses L1 regularization. It is useful in model simplification and variable selection by limiting less important feature coefficients to zero.

### Ridge Regression

Like Lasso, Ridge Regression uses L2 regularization, which imposes a penalty on the size of coefficients. However, unlike Lasso, it does not drop variables completely and is therefore applicable when all the features are suspected to have a contribution to prediction.

### Decision Tree Regressor

A non-parametric model which splits the feature space into regions using feature values and hence extracts non-linear relationships well. It supports categorical and numerical data and is not sensitive to scaling of features.

### K-Nearest Neighbors (KNN) Regressor

A k-nearest neighbors algorithm that uses the average target value of the k-nearest training instances for prediction. It is intuitive and works well in the presence of localized patterns but extremely sensitive to data sparsity and scaling.

1. **Mean Absolute Error (MAE)**

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 04 | April – 2025 DOI: 10.55041/ISJEM03827
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

$$MAE = \frac{1}{n}\Sigma|y_i - \hat{y}_i|$$

- Measures the average magnitude of errors without considering their direction.
- Lower MAE = better accuracy.

**2. Mean Squared Error (MSE)**

$$MSE = \frac{1}{n}\Sigma(y_i - \bar{y}_i)^2$$

- Penalizes larger errors more than MAE due to squaring.
- Lower MSE = better model performance.

**3. Coefficient of Determination ($R^2$ Score)**

$$R^2 = 1 - \frac{\Sigma_{i=1}^{n}(y_i - \hat{y}_i)^2}{\Sigma_{i=1}^{n}(y_i - \bar{y})^2}$$

- Compares the model's performance to a baseline mean prediction.
- $R^2$ = 1 implies perfect predictions.
- $R^2$= 0 means predictions are no better than the mean.

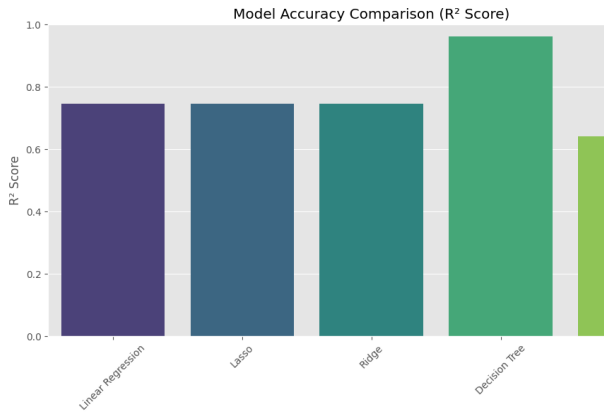


Fig4: $R^2$ Score Comparison Across Models

**Decision Tree Regressor Evaluation**

In the context of predicting crop yields using various agricultural and climatic features (e.g., country, crop type, rainfall, temperature, pesticide use, year), the **Decision Tree Regressor** demonstrated exceptional performance based on standard evaluation metrics:

**1. Mean Absolute Error (MAE) = 5,332.39 hg/ha**

This metric represents the **average absolute difference** between the actual crop yields and the predicted values produced by the model. In simpler terms:

- On average, the model's prediction differs from the true yield by only **~5,300 hectograms per hectare (hg/ha)**.
- Given that typical crop yields in the dataset span a wide range (from thousands to hundreds of thousands of hg/ha), this low MAE highlights that the **model's predictions are highly accurate and close to real values**.

**2. Mean Squared Error (MSE) = 266,386,401.04 (hg/ha)²**

MSE calculates the **average of the squares of prediction errors**. Squaring the errors penalizes larger mistakes more heavily than smaller ones.

- Despite the large numeric value (due to squaring and the scale of yield units), this **MSE is quite low in the context of yield prediction**.
- It suggests that **the model rarely makes large prediction mistakes**, and overall variance in errors is well-controlled.

**3. $R^2$ Score (Coefficient of Determination) = 0.96**

The $R^2$ score measures the proportion of the variance in the target variable (crop yield) that is **explained by the model**.

- An $R^2$ score of **0.96** means that **96% of the total variability in crop yield can be explained by the input features used in the model**.
- Only 4% of the variation is left unexplained, which could be due to random noise, unmeasured variables, or data anomalies.

| Model | MAE (hg/ha) | MSE (hg/ha)² | $R^2$ |
|---|---|---|---|
| **Linear Regression** | 22,000+ | 484,000,000 | 0.50 |
| **Lasso** | 29,909 | 1,819,003,478 | 0.75 |
| **Ridge** | 29,888 | 1,819,010,174 | 0.75 |
| **Decision Tree** | 5,332 | 266,386,401 | **0.96** |
| **KNN** | 31,682 | 2,576,076,749 | 0.64 |

Table1: Performance Metrices of Different Models

Among all tested models (Linear Regression, Lasso, Ridge, KNN, Decision Tree), the Decision Tree Regressor significantly outperformed others across **all major evaluation metrics**:

- **Lowest MAE** → Smallest average error per prediction
- **Lowest MSE** → Most consistent predictions with minimal large errors
- **Highest $R^2$ Score** → Most explanatory power and predictive accuracy

Thus, the Decision Tree Regressor is the model of choice within this research to predict crop yield based on its high accuracy, minimal error levels, and strong ability to map intricate, non-linear relationships characteristic of agricultural data.

## 6. DISCUSSION

Our results clearly show that the Decision Tree Regressor (DTR) model performs extremely well for predicting crop yield from different environmental and agricultural factors. The Decision Tree Regressor performed better than the other models every time, including Linear Regression (LR), Lasso, Ridge , and K-Nearest Neighbours (KNN) across all of the main metrics, including Mean Absolute Error (MAE), Mean Squared Error (MSE), and $R^2$. With a remarkably high $R^2$ value of 0.96, the Decision Tree model accounts for 96% of the variation in crop yield, which means that it does an excellent job of capturing the pattern among the features and the target variable.

Whereas models such as Linear Regression and Lasso had difficulty making accurate predictions with their own weakness in dealing with intricate relationships among features, the Decision Tree Regressor could effectively capture non-linear relationships, which is normally the situation in agricultural systems. This is particularly important in modelling crop yield, where the correspondence between environmental drivers (e.g., temperature, rainfall) and crop yield can be very non-linear and subject to sophisticated interactions that straightforward linear models are unlikely to handle well.

The fact that Decision Trees can process both numerical and categorical data without requiring heavy preprocessing, coupled with the interpretability of decision trees, makes them an exceptionally good candidate for this problem. Visualizing the splits in decision trees allows stakeholders in agribusiness to understand what factors are most relevant to predicting crop yield, thus making the model more explainable and actionable.

Our discussion also emphasizes the significance of data splitting during training and testing. Although Decision Trees do not necessarily need to have different training and testing datasets owing to their data bootstrapping and bagging behaviour, we followed the traditional approach of data splitting so that the model was tested thoroughly on unseen samples. This division of the training and test sets helps our results to be valid and applicable in real-world situations where unseen data are encountered.

In addition, we saw that the accuracy of the Decision Tree Regressor increases with the size of the dataset, just like in the results of other research on machine learning. Adding more data for training may help improve the generalization capability of the model, particularly for the prediction of crop yields in different regions or under different climatic conditions.

In summary, the Decision Tree Regressor was an outstanding selection for crop yield forecasting, offering very accurate forecasts and great interpretability. This model's capability to process big data, intricate relationships, and high predictive accuracy indicates its robust potential for agricultural prediction. Additional research can also focus on hyperparameter tuning of the model, cross-validation, and using finer data (like satellite images or composition of the soil) to optimize the model performance and usefulness in actual field scenarios.

## 7. CONCLUSION

Here, we created a machine learning pipeline to forecast crop yield (in hectograms per hectare) from important agricultural and environmental characteristics like mean rainfall, use of pesticides, temperature, crop variety, region (country), and year. We used several regression models—Linear Regression, Lasso, Ridge, K-Nearest Neighbours, and Decision Tree Regressor—to test the efficacy of these models in forecasting crop yield. Of all the models that were tried, the Decision Tree Regressor performed the best, with an $R^2$ of 0.96, MAE of 5,332.39, and MSE of 266,386,401.04. This shows that the model could explain 96% of crop yield variance and make very accurate predictions, performing much better than the other methods. The capacity of the model to process numerical and categorical inputs, extract intricate and non-linear patterns, and yield interpretable results is indicative of its potential application in real-world agricultural prediction problem-solving. These outcomes reinforce the promise of machine learning models—particularly tree-based methods—to underpin data-driven decision-making in agriculture.

Through the use of a well-designed pipeline involving data cleaning, preprocessing, feature encoding, scaling, model comparison, and performance evaluation, this project lays the groundwork for future applications and research in predictive agriculture. Future enhancements may involve hyperparameter tuning, ensemble techniques (such as Random Forest or Gradient Boosting), and the incorporation of other data sources to increase robustness and generalizability. In summary, our method effectively showcases the ability of machine learning

to accurately and effectively predict crop yield, thus promoting smarter, more sustainable agricultural planning.

## 8. REFERENCES

[1]A.T.M.S. Ahamed, N.T. Mahmood, N. Hossain, M.T. Kabir, K. Das, F. Rahman, R.M. Rahman
Applying data mining techniques to predict annual yield of major crops and recommend planting different crops in different districts in Bangladesh
2015 IEEE/ACIS 16th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, SNPD 2015 - Proceedings (2015)
»https://doi.org/10.1109/SNPD.2015.7176185

[2] I. Ahmad, U. Saeed, M. Fahad, A. Ullah, M. Habib-ur-Rahman, A. Ahmad, J. Judge
Yield forecasting of spring maize using remote sensing and crop modeling in Faisalabad-Punjab Pakistan
J. Indian Soc. Remote Sens., 46 (10) (2018), pp. 1701-1711
»https://doi.org/10.1007/s12524-018-0825-8

[3] I. Ali, F. Cawkwell, E. Dwyer, S. Green
Modeling managed grassland biomass estimation by using multitemporal remote sensing data—a machine learning approach
IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., 10 (7) (2017), pp. 3254-3264
» https://doi.org/10.1109/JSTARS.2016.2561618

[4] S. Bargoti, J.P. Underwood

Image segmentation for fruit detection and yield estimation in apple orchards
J. Field Rob., 34 (6) (2017), pp. 1039-1060
» https://doi.org/10.1002/rob.21699

[5] R. Beulah
A survey on different data mining techniques for crop yield prediction
Int. J. Comput. Sci. Eng., 7 (1) (2019), pp. 738-744
» https://doi.org/10.26438/ijcse/v7i1.738744

[6] Y. Chen, W.S. Lee, H. Gan, N. Peres, C. Fraisse, Y. Zhang, Y. He
Strawberry yield prediction based on a deep neural network using high-resolution aerial orthoimages
Remote Sens., 11 (13) (2019), p. 1584
» https://doi.org/10.3390/rs11131584

[4] S. Bargoti, J.P. Underwood
Image segmentation for fruit detection and yield estimation in apple orchards
J. Field Rob., 34 (6) (2017), pp. 1039-1060
» https://doi.org/10.1002/rob.21699

[5] R. Beulah
A survey on different data mining techniques for crop yield prediction
Int. J. Comput. Sci. Eng., 7 (1) (2019), pp. 738-744
» https://doi.org/10.26438/ijcse/v7i1.738744

[6] Y. Chen, W.S. Lee, H. Gan, N. Peres, C. Fraisse, Y. Zhang, Y. He
Strawberry yield prediction based on a deep neural network using high-resolution aerial orthoimages
Remote Sens., 11 (13) (2019), p. 1584
» https://doi.org/10.3390/rs11131584

[7] A. Chlingaryan, S. Sukkarieh, B. Whelan
Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: a review
Comput. Electron. Agric., 151 (2018), pp. 61-69
» https://doi.org/10.1016/j.compag.2018.05.012

[8] A. Crane-Droesch
Machine learning methods for crop yield prediction and climate change impact assessment in agriculture
Environ. Res. Lett., 13 (11) (2018), Article 114003
» https://doi.org/10.1088/1748-9326/aae159

[9] S. De Alwis, Y. Zhang, M. Na, G. Li
Duo attention with deep learning on tomato yield prediction and factor interpretation
Pacific Rim International Conference on Artificial Intelligence, Springer, Cham (2019), pp. 704-715
» https://doi.org/10.1007/978-3-030-29894-4_56

[10] D. Elavarasan, P.D. Vincent
Crop yield prediction using deep reinforcement learning model for sustainable agrarian applications
IEEE Access, 8 (2020), pp. 86886-86901
» https://www.scopus.com/inward/record.url?eid=2-s2.0-85085182944&partnerID=10&rel=R3.0.0

[12] Y.L. Everingham, C.W. Smyth, N.G. Inman-Bamber
Ensemble data mining approaches to forecast regional sugarcane crop production
Agric. For. Meteorol., 149 (3–4) (2009), pp. 689-696
» https://doi.org/10.1016/J.AGRFORMET.2008.10.018

[13] J.L. Fernandes, N.F.F. Ebecken, J.C.D.M. Esquerdo
Sugarcane yield prediction in Brazil using NDVI time series and neural networks
Int. J. Remote Sens., 38 (16) (2017), pp. 4631-4644
» https://doi.org/10.1080/01431161.2017.1325531

[14] A. Goldstein, L. Fink, A. Meitin, S. Bohadana, O. Lutenberg, G. Ravid
Applying machine learning on sensor data for irrigation recommendations: revealing the agronomist's tacit knowledge

Precis. Agric., 19 (3) (2018), pp. 421-444
» https://doi.org/10.1007/s11119-017-9527-4

[15] A. Gonzalez-Sanchez, J. Frausto-Solis, W. Ojeda-Bustamante
Predictive ability of machine learning methods for massive crop yield prediction
Spanish J. Agric. Res., 12 (2) (2014), pp. 313-328
» https://doi.org/10.5424/sjar/2014122-4439

[16] S. Ji, W. Xu, M. Yang, K. Yu
3D convolutional neural networks for human action recognition
IEEE Trans. Pattern Anal. Mach. Intell., 35 (1) (2012), pp. 221-231
» https://doi.org/10.1090/amsip/051.1/16

[17] H. Jiang, H. Hu, R. Zhong, J. Xu, J. Xu, J. Huang, T. Lin
A deep learning approach to conflating heterogeneous geospatial data for corn yield estimation: a case study of the US Corn Belt at the county level
Glob. Change Biol., 26 (3) (2020), pp. 1754-1766
» https://doi.org/10.1111/gcb.14885

[18] S. Khaki, L. Wang
Crop yield prediction using deep neural networks
Front. Plant Sci., 10 (2019), p. 621
» https://www.scopus.com/inward/record.url?eid=2-s2.0-85067349640&partnerID=10&rel=R3.0.0

[19] S. Khaki, L. Wang, S.V. Archontoulis
A cnn-rnn framework for crop yield prediction
Front. Plant Sci., 10 (2020), p. 1750

» https://www.scopus.com/inward/record.url?eid=2-s2.0-85079196653&partnerID=10&rel=R3.0.0

[20] S. Khanal, J. Fulton, A. Klopfenstein, N. Douridas, S. Shearer
Integration of high resolution remotely sensed data and machine learning techniques for spatial prediction of soil properties and corn yield
Comput. Electron. Agric., 153 (2018), pp. 213-225
» https://doi.org/10.1016/J.COMPAG.2018.07.016

[21]L. Kouadio, R.C. Deo, V. Byrareddy, J.F. Adamowski, S. Mushtaq, V. Phuong Nguyen
Artificial intelligence approach for the prediction of Robusta coffee yield using soil fertility properties
Comput. Electron. Agric., 155 (2018), pp. 324-338
» https://doi.org/10.1016/J.COMPAG.2018.10.014

[22] S. Lee, Y. Jeong, S. Son, B. Lee
A self-predictable crop yield platform (SCYP) based on crop diseases using deep learning
Sustainability, 11 (13) (2019), p. 3637
» https://www.scopus.com/inward/record.url?eid=2-s2.0-85068648790&partnerID=10&rel=R3.0.0

[23] K. Matsumura, C.F. Gaitan, K. Sugimoto, A.J. Cannon, W.W. Hsieh
Maize yield forecasting by linear regression and artificial neural networks in Jilin, China
J. Agric. Sci., 153 (3) (2015), pp. 399-410
» https://doi.org/10.1017/S0021859614000392

[24] R.J. McQueen, S.R. Garner, C.G. Nevill-Manning, I.H. Witten
Applying machine learning to agricultural data
Comput. Electron. Agric., 12 (4) (1995), pp. 275-293
» https://doi.org/10.1016/0168-1699(95)98601-9

[25]V. Mnih, K. Kavukcuoglu, D. Silver, A.A. Rusu, J. Veness, M.G. Bellemare, S. Petersen
Human-level control through deep reinforcement learning
Nature, 518 (7540) (2015), pp. 529-533
» https://doi.org/10.1038/nature14236

[26] B. Mola-Yudego, J. Rahlf, R. Astrup, I. Dimitriou
Spatial yield estimates of fast-growing willow plantations for energy based on climatic variables in northern Europe
GCB Bioenergy, 8 (6) (2016), pp. 1093-1105
» https://doi.org/10.1111/gcbb.12332 70

[27] L.H. Nguyen, J. Zhu, Z. Lin, H. Du, Z. Yang, W. Guo, F. Jin
Spatial-temporal multi-task learning for within-field cotton yield prediction
Pacific-Asia Conference on Knowledge Discovery and Data Mining, Springer, Cham (2019), pp. 343-354

» https://doi.org/10.1007/978-3-030-16148-4_27

[28] X.E. Pantazi, D. Moshou, T. Alexandridis, R.L. Whetton, A.M. Mouazen
Wheat yield prediction using machine learning and advanced sensing techniques
Comput. Electron. Agric., 121 (2016), pp. 57-65
» https://doi.org/10.1016/j.compag.2015.11.018

[29] M. Rahnemoonfar, C. Sheppard
Real-time yield estimation based on deep learning
Autonomous Air and Ground Sensing Systems for Agricultural Optimization and Phenotyping II, Vol. 10218, International Society for Optics and Photonics (2017), p. 1021809
» https://www.scopus.com/inward/record.url?eid=2-s2.0-85021780491&partnerID=10&rel=R3.0.0

[30]J.R. Romero, P.F. Roncallo, P.C. Akkiraju, I. Ponzoni, V.C. Echenique, J.A. Carballido
Using classification algorithms for predicting durum wheat yield in the province of Buenos Aires

Comput. Electron. Agric., 96 (2013), pp. 173-179
» https://doi.org/10.1016/j.compag.2013.05.006

[31] A. Shah, A. Dubey, V. Hemnani, D. Gala, D.R. Kalbande
Smart Farming System: Crop Yield Prediction Using Regression Techniques
Springer, Singapore (2018), pp. 49-56
»https://doi.org/10.1007/978-981-10-8339-6_6

[32] A. Shekoofa, Y. Emam, N. Shekoufa, M. Ebrahimi, E. Ebrahimie
Determining the most important physiological and agronomic traits contributing to maize grain yield through machine learning algorithms: a new avenue in intelligent agriculture
PLoS ONE, 9 (5) (2014), Article e97288
» https://doi.org/10.1371/journal.pone.0097288

[33] D. Šmite, C. Wohlin, T. Gorschek, R. Feldt 71
Empirical evidence in global software engineering: a systematic review
Empirical Softw. Eng., 15 (1) (2010), pp. 91-118
» https://doi.org/10.1007/s10664-009-9123-y

[34] Y.X. Su, H. Xu, L.J. Yan
Support vector machine-based open crop model (SBOCM): case of rice production in China
Saudi J. Biol. Sci., 24 (3) (2017), pp. 537-547
»https://www.sciencedirect.com/science/article/pii/S1319562X17300335

[35] J. Sun, L. Di, Z. Sun, Y. Shen, Z. Lai
County-level soybean yield prediction using deep CNN-LSTM model
Sensors, 19 (20) (2019), p. 4363
» https://www.scopus.com/inward/record.url?eid=2-s2.0-85073105002&partnerID=10&rel=R3.0.0

[36] P. Taherei-Ghazvinei, H. Hassanpour-Darvishi, A. Mosavi, K.W. Yusof, M. Alizamir, S. Shamshirband, K. Chau
Sugarcane growth prediction based on meteorological parameters using extreme learning machine and artificial neural network
Eng. Appl. Comput. Fluid Mech., 12 (1) (2018), pp. 738-749
» https://doi.org/10.1080/19942060.2018.1526119

[37] A.S. Terliksiz, D.T. Altýlar
Use Of deep neural networks for crop yield prediction: a case study Of Soybean Yield in Lauderdale County, Alabama, USA
2019 8th International Conference on Agro-Geoinformatics (Agro-Geoinformatics), IEEE (2019), pp. 1-4
» https://doi.org/10.1109/agro-geoinformatics.2019.8820257

[38] P. Vincent, H. Larochelle, Y. Bengio, P.A. Manzagol
Extracting and composing robust features with denoising autoencoders
Proceedings of the 25th international conference on Machine learning (2008), pp. 1096-1103
» https://doi.org/10.1145/1390156.1390294

[39] X. Wang, J. Huang, Q. Feng, D. Yin
Winter wheat yield prediction at county level and uncertainty analysis in main wheat-producing regions of china with deep learning approaches
Remote Sens., 12 (11) (2020), p. 1744 72
» https://doi.org/10.3390/rs12111744

[40] X. Xu, P. Gao, X. Zhu, W. Guo, J. Ding, C. Li, X. Wu
Design of an integrated climatic assessment indicator (ICAI) for wheat production: a case study in Jiangsu Province, China
Ecol. Ind., 101 (2019), pp. 943-953
» https://doi.org/10.1016/j.ecolind.2019.01.059

[41] Q. Yang, L. Shi, J. Han, Y. Zha, P. Zhu
Deep convolutional neural networks for rice grain yield estimation at the ripening stage using UAV-based remotely sensed images
Field Crops Res., 235 (2019), pp. 142-153
»https://www.sciencedirect.com/science/article/pii/S037842901831390X

[42] L. Zhang, Z. Zhang, Y. Luo, J. Cao, F. Tao
Combining optical, fluorescence, thermal satellite, and environmental data to predict county-level maize yield in china using machine learning approaches
Remote Sens., 12 (1) (2020), p. 21
»https://www.sciencedirect.com/science/article/pii/S0920586118308654

[43] H. Zhong, Xiaocheng Li, D. Lobell, S. Ermon, M.L. Brandeau
Hierarchical modeling of seed variety yields and decision making for future planting plans
Environ. Syst. Decis., 38 (2018), pp. 458-470
» https://doi.org/10.1007/s10669-018-9695-4

[44] A. Modi, P. Sharma, D. Saraswat, R. Mehta
Review of crop yield estimation using machine learning and deep learning techniques
Scalable Computing, 23 (2) (2022), pp. 59-79
» https://doi.org/10.12694/scpe.v23i2.2025

[45] N. Goel, S. Kaur, Y. Kumar
Chapter 23 - machine learning-based remote monitoring and predictive analytics system for crop and livestock
AI, Edge and IoT-Based Smart Agriculture, Elsevier Inc (2022)
» https://doi.org/10.1016/B978-0-12-823694-9.00016-5

[46] T. van Klompenburg, A. Kassahun, C. Catal
Crop yield prediction using machine learning: a systematic

literature review 73
Comput. Electron. Agric., 177 (July) (2020), Article 105709
» https://doi.org/10.1016/j.compag.2020.105709

[47]X. Xu, P. Gao, X. Zhu, W. Guo, J. Ding, C. Li, M. Zhu
Design of an integrated climatic assessment indicator (ICAI) for wheat production : a case study in Jiangsu Province , China
Ecol. Indicat., 101 (July 2018) (2019), pp. 943-953
» https://doi.org/10.1016/j.ecolind.2019.01.059

[48] P. Filippi
An Approach to Forecast Grain Crop Yield Using Multi - Layered , Multi - Farm Data Sets and Machine Learning 0123456789 (2019)
» https://doi.org/10.1007/s11119-018-09628-4

[49] Y. Jin, H. Wang, C. Sun
Data-driven evolutionary optimization
Stud. Comput. Intell. (SCI), 975 (2021), pp. 103-143
» https://doi.org/10.1007/978-3-030-74640-7_4

[50] D. Paudel, H. Boogaard, A. de Wit, S. Janssen, S. Osinga, C. Pylianidis, I.N. Athanasiadis
Machine learning for large-scale crop yield forecasting
Agric. Syst., 187 (December 2020) (2021), Article 103016
» https://doi.org/10.1016/j.agsy.2020.103016

[51] S. Iniyan, V. Akhil Varma, C. Teja Naidu
Crop yield prediction using machine learning techniques
Adv. Eng. Software, 175 (October 2022) (2023), Article 103326
» https://doi.org/10.1016/j.advengsoft.2022.103326

[52] A. Chlingaryan, S. Sukkarieh, B. Whelan
Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: a review
Comput. Electron. Agric., 151 (November 2017) (2018), pp. 61-69
» https://doi.org/10.1016/j.compag.2018.05.012

[53] K.G. Liakos, P. Busato, D. Moshou, S. Pearson, D. Bochtis
Machine learning in agriculture: a review
Sensors, 18 (8) (2018), pp. 1-29 74
» https://doi.org/10.3390/s18082674

[54] L.J. Young
Agricultural crop forecasting for large geographical areas
Annual Review of Statistics and Its Application, 6 (August 2018) (2019), pp. 173-196
» https://doi.org/10.1146/annurev-statistics-030718-105002

[55] D. Elavarasan, D.R. Vincent, V. Sharma, A.Y. Zomaya, K. Srinivasan
Forecasting yield by integrating agrarian factors and machine learning models: a survey
Comput. Electron. Agric., 155 (October) (2018), pp. 257-282
» https://doi.org/10.1016/j.compag.2018.10.024

[56] A. Kamilaris, F.X. Prenafeta-Boldú
Deep learning in agriculture: a survey
Comput. Electron. Agric., 147 (July 2017) (2018), pp. 70-90
» https://doi.org/10.1016/j.compag.2018.02.016

[57] A. Koirala, K.B. Walsh, Z. Wang, C. McCarthy
Deep learning – method overview and review of use for fruit detection and yield estimation
Comput. Electron. Agric., 162 (January) (2019), pp. 219-234
» https://doi.org/10.1016/j.compag.2019.04.017

[58] Q. Zhang, Y. Liu, C. Gong, Y. Chen, H. Yu
Applications of deep learning for dense scenes analysis in agriculture: a review
Sensors, 20 (5) (2020), pp. 1-33
» https://doi.org/10.3390/s20051520

[59] B. Darwin, P. Dharmaraj, S. Prince, D.E. Popescu, D.J. Hemanth
Recognition of bloom/yield in crop images using deep learning models for smart agriculture: a review
Agronomy, 11 (4) (2021), pp. 1-22
» https://doi.org/10.3390/agronomy11040646

[60] P. Maheswari, P. Raja, O.E. Apolo-Apolo, M. Pérez-Ruiz
Intelligent fruit yield estimation for orchards using deep learning based semantic segmentation techniques—a review
Front. Plant Sci., 12 (June) (2021), pp. 1-18 75
» https://doi.org/10.3389/fpls.2021.684328

[61] A. Monteiro, S. Santos, P. Gonçalves
Precision agriculture for crop and livestock farming—brief review
Animals, 11 (8) (2021), pp. 1-18
» https://doi.org/10.3390/ani11082345

[62]M. Rashid, B.S. Bari, Y. Yusup, M.A. Kamaruddin, N. Khan
A comprehensive review of crop yield prediction using machine learning approaches with special emphasis on palm oil yield prediction
IEEE Access, 9 (2021), pp. 63406-63439
» https://doi.org/10.1109/ACCESS.2021.3075159

[63]A.S.M.M. Hasan, F. Sohel, D. Diepeveen, H. Laga, M.G.K. Jones
A survey of deep learning techniques for weed detection from images
Comput. Electron. Agric., 184 (December 2020) (2021),

Article 106067
» https://doi.org/10.1016/j.compag.2021.106067

[64]P. Muruganantham, S. Wibowo, S. Grandhi, N.H. Samrat, N. Islam
A systematic literature review on crop yield prediction with deep learning and remote sensing
Rem. Sens., 14 (9) (2022)
» https://doi.org/10.3390/rs14091990

[65] A. Oikonomidis, C. Catal, A. Kassahun
Deep learning for crop yield prediction: a systematic literature review
N. Z. J. Crop Hortic. Sci., 51 (1) (2023), pp. 1-26
» https://doi.org/10.1080/01140671.2022.2032213

[66]A. Bouguettaya, H. Zarzour, A. Kechida, A.M. Taberkit
Deep learning techniques to classify agricultural crops through UAV imagery: a review
Neural Comput. Appl., 34 (12) (2022), pp. 9511-9536
» https://doi.org/10.1007/s00521-022-07104-9

[67] O.D. Arigbe, M.B. Oyeneyin, I. Arana, M.D. Ghazi
Real-time relative permeability prediction using deep learning
J. Petrol. Explor. Prod. Technol., 9 (2) (2019), pp. 1271-1284
» https://doi.org/10.1007/s13202-018-0578-5

[68] D.K. Bolton, M.A. Friedl
Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics
Agric. Forest Meteorol., 173 (2013), pp. 74-84
»https://www.sciencedirect.com/science/article/pii/S0168192313000129

[69] P. Bose, N.K. Kasabov, L. Bruzzone, R.N. Hartono
Spiking neural networks for crop yield estimation based on spatiotemporal analysis of image time series

IEEE Trans. Geosci. Remote Sensing, 54 (11) (2016), pp. 6563-6573
» https://www.scopus.com/inward/record.url?eid=2-s2.0-84979916681&partnerID=10&rel=R3.0.0

[70]Y. Cai, K. Guan, D. Lobell, A.B. Potgieter, S. Wang, J. Peng, T. Xu, S. Asseng, Y. Zhang, L. You, B. Peng
Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches
Agric. Forest Meteorol., 274 (2019), pp. 144-159
»https://www.sciencedirect.com/science/article/pii/S0168192319301224

[71] Cao, Z. Zhang, F. Tao, L. Zhang, Y. Luo, J. Zhang, J. Han, J. Xie
Integrating multi-source data for rice yield prediction across China using machine learning and deep learning approaches
Agric. Forest Meteorol., 297 (2021), Article 108275
»https://www.sciencedirect.com/science/article/pii/S0168192320303774

[72] K. Chandrasekar, M.V.R. Sesha Sai, P.S. Roy, R.S. Dwevedi, L. Surface Water
Land surface water index (LSWI) response to rainfall and NDVI using the MODIS vegetation index product
Int. J. Remote Sens., 31 (15) (2010), pp. 3987-4005
» https://doi.org/10.1080/01431160802575653

[73]P.C. Doraiswamy, J.L. Hatfield, T.J. Jackson, B. Akhmedov, J. Prueger, A. Stern
Crop condition and yield simulations using Landsat and MODIS
Remote Sens. Environ., 92 (4) (2004), pp. 548-559
»https://www.sciencedirect.com/science/article/pii/S0034425704001853

[74]P.C. Doraiswamy, T.R. Sinclair, S. Hollinger, B. Akhmedov, A. Stern, J. Prueger
Application of MODIS derived parameters for regional crop yield assessment
Remote Sens. Environ., 97 (2) (2005), pp. 192-202
»https://www.sciencedirect.com/science/article/pii/S0034425705001355

[75] H. Elbern, H. Schmidt, O. Talagrand, A. Ebel
4D-variational data assimilation with an adjoint air quality model for emission analysis
Environ. Model. Softw., 15 (6–7) (2000), pp. 539-548
»https://www.sciencedirect.com/science/article/pii/S1364815200000499

[76] N.K. Fageria
Green manuring in crop production
J. Plant Nutr., 30 (5) (2007), pp. 691-719
» https://doi.org/10.1080/01904160701289529

[77]C. Ferencz, P. Bognár, J. Lichtenberger, D. Hamar, G. Timár, G. Molnár, S.Z. Pásztor, P. Steinbach, B. Székely, O.E. Ferencz, I. Ferencz-Árkos, G.Y. Tarcsai
Crop yield estimation by satellite remote sensing
Int. J. Remote Sens., 25 (20) (2004), pp. 4113-4149
» https://www.scopus.com/inward/record.url?eid=2-s2.0-6344285791&partnerID=10&rel=R3.0.0

[78]P. Filippi, E.J. Jones, N.S. Wimalathunge, P.D.S.N. Somarathna, L.E. Pozza, S.U. Ugbaje, T.G. Jephcott, S.E. Paterson, B.M. Whelan, T.F.A. Bishop
An approach to forecast grain crop yield using multi-layered, multi-farm data sets and machine learning
Precis. Agric., 20 (5) (2019), pp. 1015-1029
» https://doi.org/10.1007/s11119-018-09628-4

[79] X. He, K. Zhao, X. Chu
AutoML: a survey of the state-of-the-art

Knowl. Based Syst., 212 (2021), Article 106622
»https://www.sciencedirect.com/science/article/pii/S0950705120307516

[80] S. Khaki, L. Wang
Crop yield prediction using deep neural networks
Front. Plant Sci., 10 (2019), p. 621
» https://www.scopus.com/inward/record.url?eid=2-s2.0-78850673496 40&partnerID=10&rel=R3.0.0

[81] S. Khaki, L. Wang, S.V. Archontoulis
A cnn-rnn framework for crop yield prediction
Front. Plant Sci., 10 (2020), p. 1750
» https://doi.org/10.3389/fpls.2019.01750

[82] M. Kim, J. Ko, S. Jeong, J. Yeom, H. Kim
Monitoring canopy growth and grain yield of paddy rice in South Korea by using the GRAMI model and high spatial resolution imagery
GISci. Remote Sens., 54 (4) (2017), pp. 534-551
» https://doi.org/10.1080/15481603.2017.1291783

[83] N. Kim, K. Ha, N. Park, J. Cho, S. Hong, Y. Lee
A comparison between major artificial intelligence models for crop yield prediction: case study of the midwestern United States, 2006–2015
ISPRS Int. J. Geo Inf., 8 (5) (2019), p. 240
» https://doi.org/10.3390/ijgi8050240

[84] N. Kim, S. Na, C. Park, M. Huh, J. Oh, K. Ha, J. Cho, Y. Lee
An artificial intelligence approach to prediction of corn yields under extreme weather conditions using satellite and meteorological data
Appl. Sci., 10 (11) (2020), p. 3785
» https://doi.org/10.3390/app10113785

[85] F. Kogan, L. Salazar, L. Roytman
Forecasting crop production using satellite-based vegetation health indices in Kansas, USA
Int. J. Remote Sens., 33 (9) (2012), pp. 2798-2814
» https://doi.org/10.1080/01431161.2011.621464

[86]P. Kumar, R. Prasad, A. Choudhary, D.K. Gupta, V.N. Mishra, A.K. Vishwakarma, A.K. Singh, P.K. Srivastava
Comprehensive evaluation of soil moisture retrieval models under different crop cover types using C-band synthetic aperture radar data
Geocarto Int., 34 (9) (2019), pp. 1022-1041
» https://doi.org/10.1080/10106049.2018.1464601

[87] D.B. Lobell, G.P. Asner
Climate and management contributions to recent trends in U.S. agricultural yields 79
Science, 299 (5609) (2003), p. 1032
» https://doi.org/10.1126/science.1078475

[88]D.B. Lobell, M.J. Roberts, W. Schlenker, N. Braun, B.B. Little, R.M. Rejesus, G.L. Hammer
Greater sensitivity to drought accompanies maize yield increase in the U.S Midwest
Science, 344 (6183) (2014), pp. 516-519
» https://doi.org/10.1126/science.1251423

[89] S.J. Maas
Within-season calibration of modeled wheat growth using remote sensing and field sampling
Agron. J., 85 (3) (1993), pp. 669-672
»https://doi.org/10.2134/agronj1993.00021962008500030028x

[90] X. Zhang, Q. Zhang
Monitoring interannual variation in global crop yield using long-term AVHRR and MODIS observations
ISPRS J. Photogramm., 114 (2016), pp. 191-205
» https://doi.org/10.3390/f7090191

[91] V.C. Nguyen, S. Jeong, J. Ko, C.T. Ng, J. Yeom
Mathematical integration of remotely sensed information into a crop modelling process for mapping crop productivity
Remote Sens., 11 (18) (2019), p. 2131
» https://doi.org/10.3390/rs11182131

[92] M.C. Peel, B.L. Finlayson, T.A. McMahon
Updated world map of the Köppen-Geiger climate classification
Hydrol. Earth Syst. Sci., 11 (5) (2007), pp. 1633-1644
» https://doi.org/10.5194/hess-11-1633-2007

[93] E. Quintero, A.E. Thessen, P. Arias-Caballero, B. Ayala-Orozco
A statistical assessment of population trends for data deficient mexican amphibians
PeerJ, 2 (2014), Article e703
» https://doi.org/10.7717/peerj.703

[94] J.H. Ryu, K.S. Han, Y.W. Lee, N.W. Park, S. Hong, C.Y. Chung, J. Cho
Different agricultural responses to extreme drought events in neighboring counties of south and North Korea
Remote Sens., 11 (15) (2019), p. 1773
» https://doi.org/10.3390/rs11151773

[95] X. Song, G. Zhang, F. Liu, D. Li, Y. Zhao, J. Yang
Modeling spatio-temporal distribution of soil moisture by deep learning-based cellular automata model
J. Arid Land, 8 (5) (2016), pp. 734-748

» https://doi.org/10.1007/s40333-016-0049-0