sustainability

MDPI

Article

Predicting Sustainable Crop Yields: Deep Learning and Explainable AI Tools

Ivan Malashin [1,*], Vadim Tynchenko [1,*], Andrei Gantimurov [1], Vladimir Nelyub [1,2], Aleksei Borodulin [1] and Yadviga Tynchenko [1,3]

[1] Artificial Intelligence Technology Scientific and Education Center, Bauman Moscow State Technical University, 105005 Moscow, Russia
[2] Scientific Department, Far Eastern Federal University, 690922 Vladivostok, Russia
[3] Laboratory of Biofuel Compositions, Siberian Federal University, 660041 Krasnoyarsk, Russia
* Correspondence: ivan.p.malashin@gmail.com (I.M.); vadimond@mail.ru (V.T.); Tel.: +7-926-875-7128 (I.M.)

Abstract: Optimizing agricultural productivity and promoting sustainability necessitates accurate predictions of crop yields to ensure food security. Various agricultural and climatic variables are included in the analysis, encompassing crop type, year, season, and the specific climatic conditions of the Indian state during the crop's growing season. Features such as crop and season were one-hot encoded. The primary objective was to predict yield using a deep neural network (DNN), with hyperparameters optimized through genetic algorithms (GAs) to maximize the R^2 score. The best-performing model, achieved by fine-tuning its hyperparameters, achieved an R^2 of 0.92, meaning it explains 92% of the variation in crop yields, indicating high predictive accuracy. The optimized DNN models were further analyzed using explainable AI (XAI) techniques, specifically local interpretable model-agnostic explanations (LIME), to elucidate feature importance and enhance model interpretability. The analysis underscored the significant role of features such as crops, leading to the incorporation of an additional dataset to classify the most optimal crops based on more detailed soil and climate data. This classification task was also executed using a GA-optimized DNN, aiming to maximize accuracy. The results demonstrate the effectiveness of this approach in predicting crop yields and classifying optimal crops.

Keywords: sustainable agriculture; yield optimization; machine learning; explainable AI

check for updates

Citation: Malashin, I.; Tynchenko, V.; Gantimurov, A.; Nelyub, V.; Borodulin, A.; Tynchenko, Y. Predicting Sustainable Crop Yields: Deep Learning and Explainable AI Tools. *Sustainability* **2024**, *16*, 9437. https://doi.org/10.3390/su16219437

Academic Editors: Dhiya Al-Jumeily OBE, Jamila Mustafina and Manoj Jayabalan

Received: 27 September 2024
Revised: 26 October 2024
Accepted: 29 October 2024
Published: 30 October 2024

1. Introduction

In recent years, agriculture has experienced a transformative evolution driven by technological advancements, heralding the era of precision agriculture [1]. This paradigm shift focuses on integrating technologies such as sensor systems [2], artificial intelligence (AI) [3], and machine learning (ML) [4] to optimize farming practices and boost agricultural productivity while minimizing environmental impact [5]. At the heart of this evolution is the capacity to leverage extensive datasets that encompass soil properties, climate variables, and agronomic factors, and their effects on crop growth, yield, and overall agricultural sustainability.

1.1. Seasonal Crop Cultivation in India

In India, a wide variety of crops are cultivated across different seasons, each with unique requirements and challenges. Rice is a staple crop grown during the Kharif (monsoon) season, typically from June to November [6]. It requires abundant water [7], making it highly vulnerable to droughts or erratic rainfall. Maize is adaptable and grown in both Kharif and Rabi (winter) seasons [8], yet faces threats from pests like fall armyworm and temperature fluctuations. The chickpea, a winter crop [9], is sown from October to November and harvested in March. It is drought-resistant but sensitive to diseases like root rot [10], especially in drier regions. Kidney beans, cultivated mainly in northern India [11], thrive during the Kharif season in cool, frost-free conditions, yet they face challenges (such as root diseases) and depend on well-distributed rainfall.

Pigeon peas are drought-tolerant legumes grown in the Kharif season but are vulnerable to pod borer infestations and uneven rainfall [12]. Similarly, moth beans and mung beans are short-duration legumes grown in dry areas during the Kharif season [13]; they are resilient but suffer from pod-shattering [14] and disease risks, particularly in low-water conditions. Black gram and lentils are primarily Rabi crops [15], requiring minimal water, but are sensitive to moisture stress and diseases like rust and wilt.

Pomegranates, cultivated year-round, are drought-tolerant yet require careful irrigation and disease management to combat fungal blight [16]. Bananas, grown in tropical regions throughout the year, face fungal threats, high water demand, and storm damage [17]. Mangoes, a summer fruit [18], are harvested from February to June but can be impacted by extreme temperatures and erratic rainfall, which affect flowering [19]. Grapes, grown from January to May, are highly sensitive to fungal infections, requiring precise irrigation and temperature control.

Watermelon [20] and muskmelon [21] are summer crops, often grown under high temperatures and sensitive to soil salinity and pests like aphids. Apples are temperate crops grown in the Himalayas [22] from June to September, but changing climate conditions, such as rising temperatures and irregular snowfall, challenge their yields. Oranges, grown mainly from October to March, require dry conditions but are sensitive to frost and water stress [23].

Papayas, cultivated year-round in tropical areas [24], require well-drained soil and consistent irrigation but are threatened by the papaya ringspot virus [25]. Coconuts are grown in coastal areas throughout the year and depend on regular rainfall [26], although drought and pests like the rhinoceros beetle pose risks. Cotton is a Kharif crop sown from June to September [27]; it requires warm, dry conditions but faces pest challenges from bollworms and is highly rainfall-dependent. Jute is a monsoon crop [28], grown from March to July, thriving in waterlogged soil but vulnerable to erratic rainfall affecting fiber quality [29]. Lastly, coffee [30], grown in high-altitude regions [31] and harvested from November to March, is sensitive to unpredictable rainfall, pests, and temperature variations driven by climate change.

Each of these crops represents a distinct set of agricultural challenges in India, including water dependence, vulnerability to pests and diseases, and sensitivity to fluctuating climate conditions, all of which are critical factors in achieving sustainable yields.

1.2. Literature Outlook

The integration of ML and remote sensing in soil and crop yield prediction is actively highlighted in the scientific literature. For instance, Khanal [32] uses aerial imagery [33] and ML to predict soil properties and corn yield at Molly Caren Farm in Ohio [34]. By analyzing multispectral images [35] and field data from seven plots in 2013, the research compares ML models like random forest and neural network models. Findings show these models outperform traditional methods, with neural networks excelling in predicting soil organic matter and cation exchange capacity, and Random Forest performing best for corn yield prediction. This approach demonstrates the potential of remote sensing and ML for accurate mapping, enhancing agricultural management tailored to local conditions.

Liu et al. [36] utilized Landsat and MODIS-NDVI data along with climatic, topographic data, and soil samples to map seven soil properties (texture, electrical conductivity, pH, nitrogen, phosphorus, potassium, and organic matter) in Punjab, Pakistan from 2000 to 2020. Comparing three statistical models—support vector machine (SVM), random forest regression [37] (RFR), and multiple linear regression (MLR)—RFR generally provided the highest accuracy. This highlights the effectiveness of ML over MLR in handling nonlinear relationships, revealing a decline in cultivated areas and high soil electrical conductivity due to salinity, with organic matter and nitrogen levels generally low.

Different types of soil support various crop growth based on their unique characteristics. Understanding these soil features is essential for optimal crop selection [38]. Rahman et al. [39] introduced a model that predicts soil series based on land type and recommends suitable crops accordingly, employing ML algorithms including weighted k-nearest neighbor (kNN) [40], bagged trees [41], and Gaussian kernel-based support vector machines (SVMs) [42] for soil classification. Experimental results demonstrate that the SVM approach outperforms existing methods, highlighting its effectiveness in agricultural decision-making.

Yadav et al. [43] developed a model for assessing soil fertility, recommending suitable crops, and predicting crop yield based on soil features. Various ML algorithms including SVM, random forest, Naive Bayes [44], Linear Regression, Multilayer Perceptron [45] (MLP), and ANN [46,47] were employed for soil classification and yield prediction. Results indicate that the ANN approach, utilizing deep learning (DL) architectures for enhanced accuracy, outperforms traditional methods in predicting crop yields based on soil characteristics.

Wickramasinghe et al. [48] explored the relationship between rice yield and climate variables in a key region of Sri Lanka [49] using various statistical and ML methods. They analyzed factors such as rainfall, temperature (min/max), evaporation, wind speed, and sunshine hours, leveraging three decades of rice yield and monthly climate data. Models developed included ANN, SVM regression [50] (SVMR), multiple linear regression [51] (MLR), Gaussian process regression [52] (GPR), power regression [53] (PR), and robust regression (RR). GPR outperformed others in yield prediction accuracy, validated with data from the 2019 Yala season.

Enhancing agricultural management through AI and ML was addressed by Diaz et al. [54], which highlighted the challenge of declining soil quality in intensive agriculture by exploring how AI and ML can estimate soil quality indicators [55] (SQIs) from agro-industrial data. The focus was on recent studies using remote sensing to predict crop yields at regional and local scales, evaluating different spectral bands, data preprocessing methods, and ML algorithms. The review proposed a model integrating SQI, environmental factors, and crop management data to enhance agricultural practices and optimize crop yield predictions through ML insights.

Burdett et al. [56] investigated the relationship between crop yield, soil properties, and topographic characteristics using high-resolution data in Southwestern Ontario, Canada. Analyzing a dataset of 145,500 observations on corn and soybean yields [57], soil nutrients, and topographic features, this study compared multiple analytical methods including multiple linear regression, ANN, decision trees, and random forests. Random forest emerged as the most effective method, achieving R^2 values of 0.85 for corn and 0.94 for soybeans, outperforming other techniques such as MLR. Cross-validation experiments demonstrated the ability of random forest models to predict yield variations in fields not included in the training dataset, indicating their potential in precision agriculture for identifying high-yield areas based on soil and topographic attributes. Another solution for precision agriculture is presented in [58], which introduces WH-DETR, a high-precision, end-to-end wheat spike detection network built on an enhanced RT-DETR architecture. This network achieves an impressive 95.7% Average Precision by utilizing multi-scale feature extraction, optimized convolution techniques, and the EIoU loss metric. WH-DETR outperforms existing methods. It is also interesting to consider the approach by Tang et al. [59], who propose a multi-scale inverse bottleneck residual network model based on ResNet-50 for accurate diagnosis of apple leaf diseases, achieving 98.73% accuracy in recognizing seven types of leaves, including six diseases. This model enhances computational efficiency and feature representation, outperforming classical methods by 1.82%.

Sidhu et al. [60] compared traditional linear regression (LR) with boosted regression trees (BRTs) for predicting crop yield responses to climate change. BRTs showed superior accuracy by identifying breakpoints in climate-yield relationships and handling complex interactions among variables affecting crop yields. In simulations and real data from India, BRTs predicted a less severe negative impact on rice, wheat, and pearl millet compared to LR, highlighting the importance of robust modeling techniques in diverse agricultural settings.

Li et al. [61] focused on enhancing crop yield projections under future climates using a combination of dynamic linear models [62] (DLM), RF, and nine global gridded crop models [63] (GGCM). By integrating RF with GGCM, significant improvements were achieved in predicting maize and soybean yields across China. Key factors influencing yields included chilling days, crop pests and diseases, and drought for maize, and crop pests and diseases, tropical days, and drought for soybean. The approach reduced uncertainties for maize and soybean, offering a robust framework to enhance future climate impact assessments on crop yields.

Understanding soil nutrients and properties helps in managing soil health effectively. Advances in sensing and computational technologies have made vast amounts of farmland data accessible, enabling the rapid adoption of ML techniques to analyze soil conditions. Motia et al. [64] discussed the use of ML in predicting and assessing soil properties to enhance agricultural soil health management [65] (SHM), highlighting key ML algorithms, tools, performance metrics, and identifying challenges and future research directions. ML shows promising potential for sustainable agricultural development through improved soil property prediction and management practices.

India, the world's second-largest exporter of agricultural products, relies heavily on accurate crop yield predictions for food security. Jhajharia et al. [66] used ML techniques to predict crop yield in the Rajasthan region of India, utilizing multi-source data including vegetation indices and weather data. Random forest proved to be the most accurate method, achieving a coefficient of determination (R^2) of 0.77, a root mean square error (RMSE) of 0.39 t/ha, and a mean absolute error (MAE) of 0.28 t/ha. The results provide a good estimation of crop yield prior to harvest, enabling farmers to prepare for environmental changes that may impact yield.

Agriculture is a key part of India's economy and a primary source of employment, exhibiting resilience during the COVID-19 pandemic with a 3.4% growth rate in 2020–2021. To sustain the growing population and ensure food security, researchers [67] leveraged technologies like remote sensing (RS) and ML to create smart, sustainable, and lucrative farming systems. This paper presents a comprehensive review of studies on the application of RS and ML in addressing agriculture-related challenges in India, covering crop management, soil management, and water management. The review, conducted from 2015 to 2022, highlights the potential of intelligent geospatial data analytics in Indian agriculture, with a focus on crop management, where RS sensors and ML techniques have yielded substantial improvements in agricultural monitoring, enabling effective management and valuable recommendations.

Table 1 summarizes various studies on modeling crop yield based on soil and climate properties using statistical and ML techniques. Each study focuses on different aspects such as predicting soil properties, assessing crop yield under climate variables, and exploring the application of ML in agriculture.

This study aimed to enhance the sustainability of agricultural practices by improving crop yield predictions through the integration of transformed data with state-specific climatic averages and seasonal adjustments. By incorporating aggregated climatic data—such as temperature, humidity, pressure, and precipitation—specific to each season and state, and omitting the year of cultivation from production data, this study sought to increase the accuracy of yield predictions. One-hot encoding (OHE) was applied to categorical features like crop type and season. A Deep Neural Network (DNN) was utilized for yield prediction, with hyperparameters optimized using genetic algorithms (GAs) to maximize the R^2 metric. Furthermore, this study aimed to interpret the DNN model through explainable AI (XAI) techniques like LIME, highlighting the importance of features such as crop type. Following this analysis, a classification model was developed to identify the most sustainable crops for cultivation based on soil and climatic conditions, again employing a GA-optimized DNN to enhance classification accuracy. Ultimately, this study aimed to demonstrate the effectiveness of this integrated approach in promoting sustainable agricultural practices by improving both crop yield predictions and crop classification.

Table 1. Summary of studies on modeling crop yield and soil properties using statistical and ML techniques.

Reference	Focus	Applied Model	Limitations
Khanal [32]	Predicting soil properties and corn yield using aerial imagery and ML	Random forest, neural network	Limited to Molly Caren Farm, Ohio; empirical data-driven approach
Rahman et al. [39]	Predicting soil series and recommending crops based on land type	Weighted k-NN, Bagged Trees, SVM	Assumes homogeneous application of agro-inputs; potential variability in field conditions
Wickramasinghe et al. [48]	Modeling rice yield based on climate variables in Sri Lanka	ANN, SVMR, MLR, GPR, PR, RR	Relies on historical data up to early 2019; limited to specific region and time frame
Diaz et al. [54]	Estimating soil quality indicators and predicting crop yields using ML and remote sensing	Various ML algorithms	Challenges in data integration; variable performance across different agricultural contexts
Malik et al. [68]	Comparative analysis of soil properties and crop yield prediction using ML	KNN, Naïve Bayes, decision trees	Limited to tomato, potato, and chili crops; assumes uniform environmental conditions
Liu et al. [36]	Mapping soil properties in Punjab, Pakistan using remote sensing and ML	SVM, RFR, MLR	Relies on satellite and sensor data; potential inaccuracies in remote sensing data
Yadav et al. [43]	Assessing soil fertility and predicting crop yield based on soil features	SVM, random forest, Naive Bayes, linear regression, MLP, ANN	Limited to soil characteristics; assumes uniform crop management practices
Motia et al. [64]	Exploring ML techniques for predicting soil properties and enhancing soil health management	Various ML algorithms	Challenges in algorithm selection and integration with existing agricultural practices
Burdett et al. [56]	Analyzing crop yield responses to soil and topographic characteristics in Southwestern Ontario, Canada	MLR, ANN, decision trees, RFR	Limited to specific geographical area; potential biases in dataset sampling
Sidhu et al. [60]	Comparing LR and BRTs for predicting climate change impacts on crop yields	LR, BRTs	Challenges in capturing complex interactions and non-linear relationships
Li et al. [61]	Integrating DLM, RF, and GGCM for improving crop yield projections under future climates	DLM, RF, GGCM	Relies on global gridded crop models; uncertainties in climate projections
Jhajharia et al. [66]	Crop yield prediction in Rajasthan, India	Random forest	Limited to Rajasthan region; no mention of scalability
Pokhariyal et al. [67]	Review of RS and ML applications in Indian agriculture	Various ML models	Limited to studies from 2015 to 2022, no mention of future directions

2. Materials and Methods

2.1. Yield in India

The first aim of this study is to develop a predictive model for crop yield using a dataset that includes agricultural data from various states in India spanning from 1997 to 2020 [69]. The dataset comprises several key features, including crop types, cropping seasons, states, areas under cultivation, production quantities, annual rainfall, fertilizer usage, pesticide usage, and calculated yields.

During the process of feature engineering and data preprocessing, several techniques were applied to enhance the model's predictive performance. Initially, an encoding scheme was developed for the cropping seasons by associating each season with specific months to better capture the temporal effects on crop yield. The mappings are as follows: 'Whole Year' included all twelve months; 'Kharif' included June, July, August, and September; 'Rabi' included October, November, December, January, February, and March; 'Autumn' included September, October, November; and 'Summer' included March, April, May, June; and 'Winter' included December, January, and February. This encoding was intended to incorporate seasonal variations more effectively into the model.

Climate data from NASA, averaged by Indian state, were integrated to account for environmental factors affecting crop yield according to the season in which the crops were grown. This climatic data, including temperature and precipitation metrics, was merged with the primary dataset based on the state where each crop was cultivated. It should be noted that due to the absence of specific geographic coordinates for the locations where the crops were grown, climate data averaged by state had to be used, which represents a limitation of the data.

Categorical features, such as crop and season, were transformed using one-hot encoding to convert them into numerical values suitable for ML algorithms. This conversion enabled the effective processing of categorical information within the model.

The 'Year' and 'Production' features were excluded from the predictive modeling because 'Production' correlates with yield. Instead, the focus was placed on predicting 'Yield', as it represents a numerical variable (production per unit area) which was defined as the ratio of production to the area under cultivation. The primary objective was to predict this 'Yield' variable.

DNN architecture was employed to model yield predictions. The design and hyperparameters of the DNN were optimized using GA to maximize the R^2 score. This optimization process involved iterative adjustments of model parameters to enhance performance. The GA explored various hyperparameter configurations, resulting in a refined and effective predictive model.

The integration of feature engineering, climate data, and advanced ML techniques facilitated the development of a robust model for accurate crop yield predictions. This methodology highlights the significance of combining diverse data sources and optimizing model parameters to achieve high predictive accuracy in agricultural forecasting. The schematic pipeline of this process is shown in Figure 1.

Between 1997 and 2004 (Figure 2), the main trends in crop yields in India included several key points. Rice remained the primary cereal crop with the largest cultivation areas, ranging from 40 to 43 million hectares. The area under wheat cultivation also remained stable, averaging around 23–24 million hectares. There was an increase in the cultivation areas of pulse crops such as gram, arhar/tur, and moong (green gram). For instance, the area under gram cultivation grew from 2.4 million hectares in 1997 to 5.5 million hectares in 2004. Soybean and rapeseed/mustard showed significant growth in cultivation areas. Soybean cultivation increased from 4.5 million hectares in 1997 to over 7 million hectares in 2004. The areas under sorghum (jowar) and pearl millet (bajra) cultivation also showed significant fluctuations but remained important crops. The area under pearl millet ranged from 4.3 to 5.2 million hectares depending on the year. The area under sugarcane cultivation remained relatively stable, fluctuating around 3.3–4.4 million hectares. The area under maize cultivation increased from 4.8 million hectares in 1997 to 6.2 million hectares in 2004. Overall, the period from 1997 to 2004 saw significant changes in the structure of cultivated areas in India, indicating an adaptation of agriculture to changing market conditions and climatic factors.

From 2005 to 2012 (Figure 3), India saw a variety of trends in crop production and yield. During this period, the cultivation area for many crops fluctuated, with notable increases and decreases in specific years. For example, the area under rice cultivation showed a consistent increase from 43.4 million hectares in 2005 to 50.8 million hectares in 2011, before slightly decreasing to 43.4 million hectares in 2012. Wheat cultivation also saw a general upward trend, with the area increasing from 25.6 million hectares in 2005 to 27.2 million hectares in 2011, followed by a slight drop in 2012.

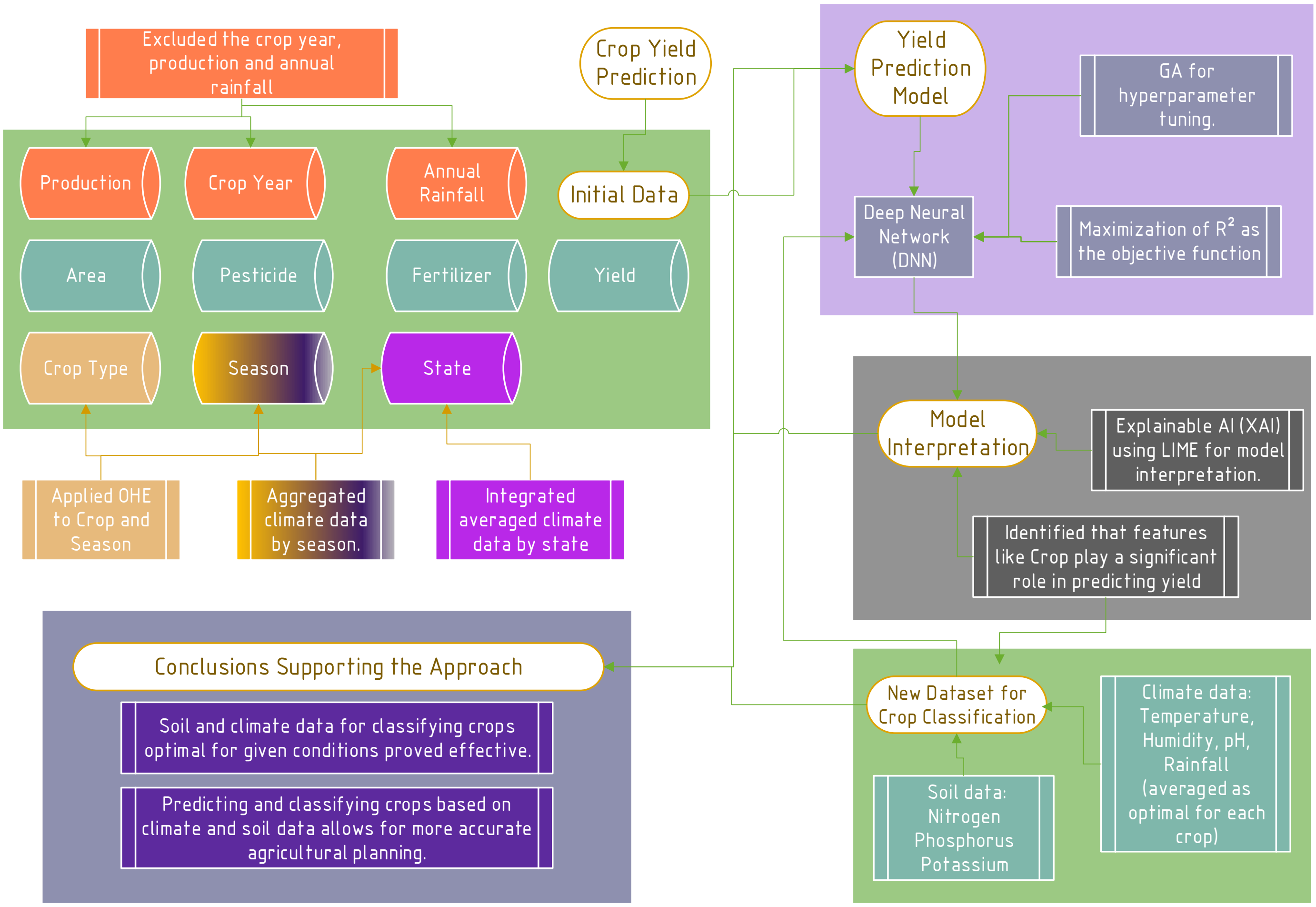

Figure 1. Summary of crop yield prediction and the classification process.

Significant changes were observed in the cultivation of pulses. The area under Arhar/Tur (pigeon pea) cultivation fluctuated, peaking at 4.3 million hectares in 2009 before declining to 2.9 million hectares in 2012. Moong (green gram) and Urad (black gram) also experienced similar fluctuations.

Oilseed crops such as soya beans, saw an increase in cultivation area, growing from 7.6 million hectares in 2005 to 9.6 million hectares in 2011, then dropping to 7.3 million hectares in 2012. Groundnut cultivation showed some volatility, with the area rising to over 6 million hectares in 2006 and 2008, but then dropping to around 4.4 million hectares in 2011 and 2012.

Cotton cultivation experienced significant growth, especially from 2009 onward, with the area increasing from 8.7 million hectares in 2005 to a peak of 11.6 million hectares in 2010, before settling at around 8.2 million hectares in 2012.

Sugarcane cultivation remained relatively stable, with a slight increase from 4.6 million hectares in 2005 to around 4.8 million hectares by 2012. Other crops such as potato, maize, rapeseed, and mustard also saw variations in their cultivation areas but remained key crops in Indian agriculture during this period.

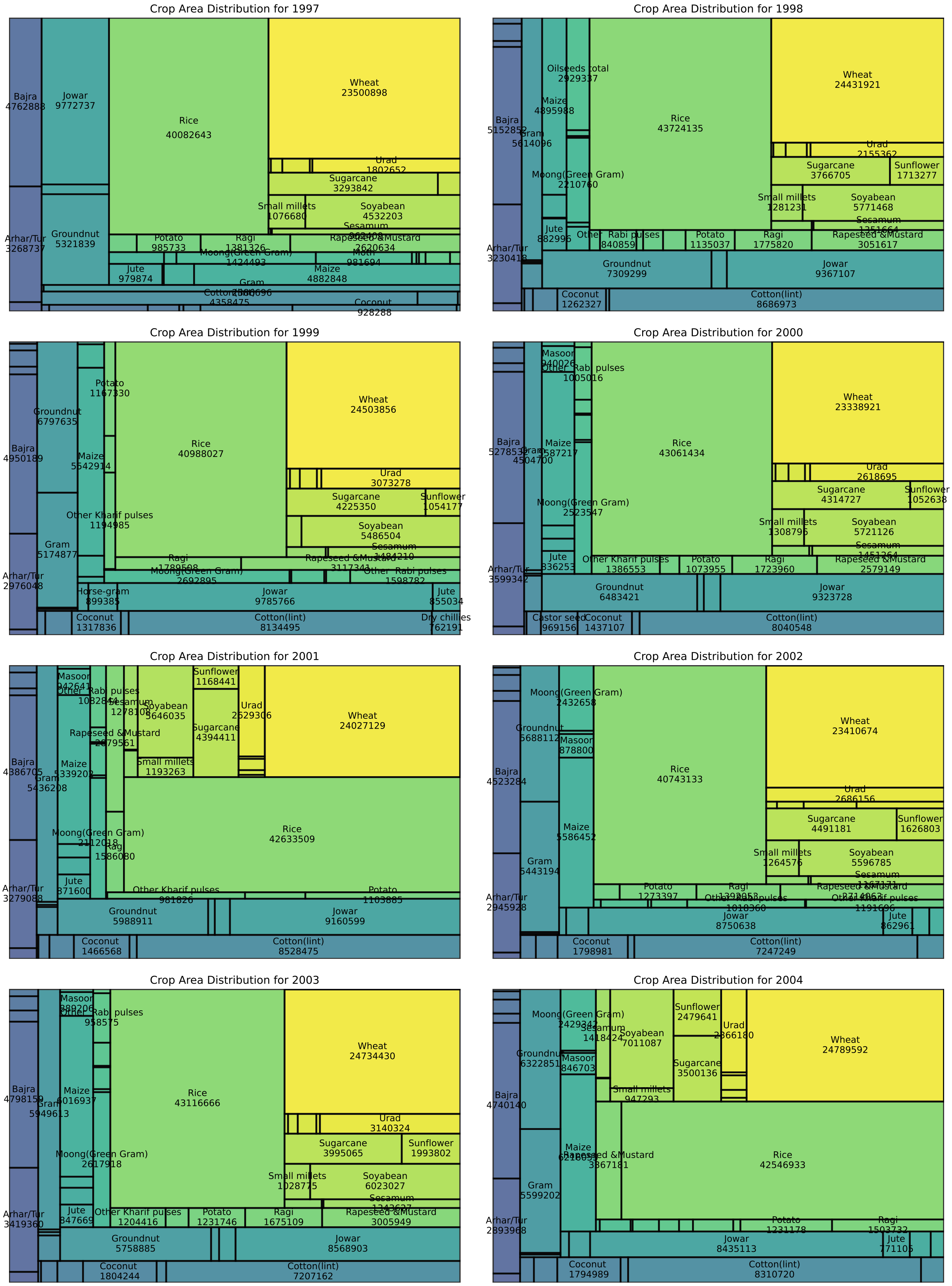

Figure 2. Trends in crop cultivation areas in India (1997–2004). The areas are indicated in hectares.

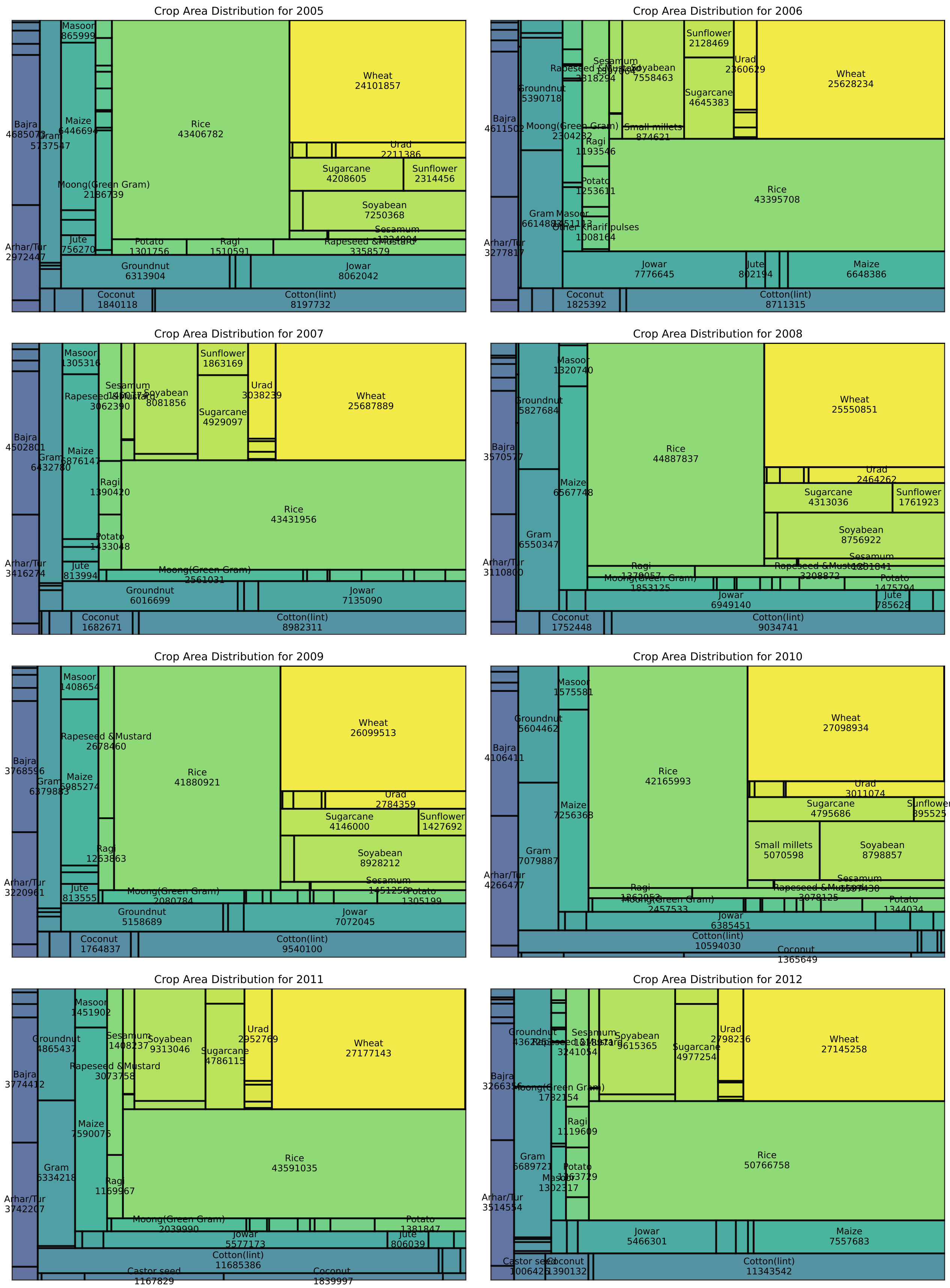

Figure 3. Trends in crop cultivation areas in India (2005–2012). The areas are indicated in hectares.

Observations reveal that rice and wheat consistently occupy the largest areas among all crops across the years. Cotton, soybeans, sugarcane, and chickpeas also cover significant areas, indicating their importance in agriculture. Trends over the years show that the area under rice fluctuated, with the lowest point in 2015 at 42,993,617 hectares and peaking in 2019 at 47,114,882 hectares. Similarly, the area under wheat varied, reaching its highest point in 2021 at 29,718,900 hectares. The cotton area displayed notable fluctuations, with a peak in 2019 at 12,723,411 hectares. Soybean cultivation increased until 2019, reaching 11,090,458 hectares, after which it began to decline. These trends suggest that the area dedicated to rice and wheat has remained substantial despite variations, highlighting their enduring significance. The fluctuations in cotton and soybean cultivation might be influenced by factors such as weather conditions, government policies, and market dynamics. The analysis of data from 2013 to 2020 (Figure 4) shows that, despite the changes, rice and wheat continue to dominate in terms of cultivated area, reflecting their role in agriculture. The decrease in the area under rice and wheat in 2020, with rice falling to 257,251 hectares and wheat to 297,189 hectares, likely reflects incomplete data, as the area under cotton was not available and soybean cultivation dropped to 7246 hectares. Examining these trends over a longer period could offer deeper insights into the long-term shifts in agricultural practices and the factors driving them.

2.2. Crop Dataset

Optimal crop cultivation requires selecting the most suitable environmental and soil conditions for each specific crop, ensuring that each crop thrives under the best possible circumstances for its growth and yield. The dataset [70] enables users to develop predictive models that recommend the most suitable crops for cultivation based on a variety of soil and environmental parameters. Figures 5 and 6 present histograms illustrating the distribution of cultivated crops in relation to key environmental characteristics.

One important climatic characteristic is ambient temperature (Figures 5a and 6a), measured in degrees Celsius. Temperature is an environmental factor influencing plant growth, development, and yield [71,72]. Each crop species has an optimal temperature range in which it flourishes, and deviations from this range can adversely affect growth and productivity. By closely monitoring temperature, farmers can select crops that are best adapted to the climatic conditions of a particular region, thereby enhancing agricultural efficiency and output.

Humidity (Figures 5b and 6b) refers to the relative humidity of the environment, expressed as a percentage [73]. It influences the rate of transpiration—the process by which plants lose water vapor through their leaves. High humidity can reduce transpiration, potentially affecting water uptake and nutrient transport. Conversely, low humidity increases transpiration, which may lead to water stress. Understanding the impact of humidity levels is essential for effective irrigation management and maintaining optimal plant health.

pH (Figures 5c and 6c) measures the potential of hydrogen in the soil [74,75] and indicates its acidity or alkalinity. Soil pH significantly affects nutrient availability and microbial activity. Each crop has specific pH requirements for optimal growth. Maintaining the correct soil pH ensures that nutrients are accessible to plants and supports a healthy microbial community.

Rainfall (Figures 5d and 6d) quantifies the amount of precipitation in millimeters and serves as a primary water source for crops [76,77]. It impacts soil moisture levels and irrigation needs. Adequate rainfall is a key factor for crop growth, while excessive or insufficient rainfall can lead to issues such as waterlogging or drought stress. Understanding rainfall patterns aids in planning irrigation strategies and choosing crops suited to specific rainfall conditions.

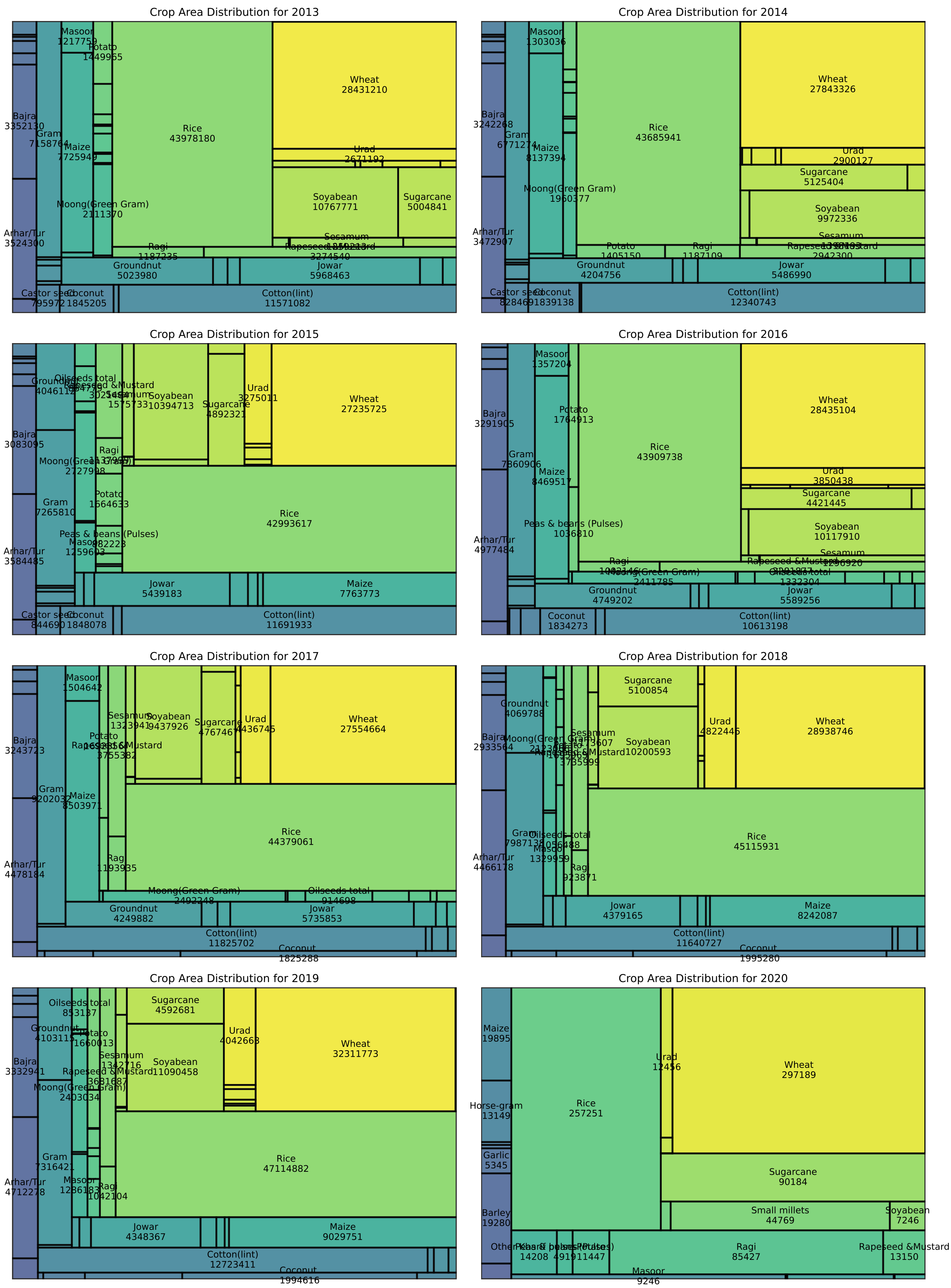

Figure 4. Trends in crop cultivation areas in India (2013–2020); 2020 has incomplete data. The areas are indicated in hectares.

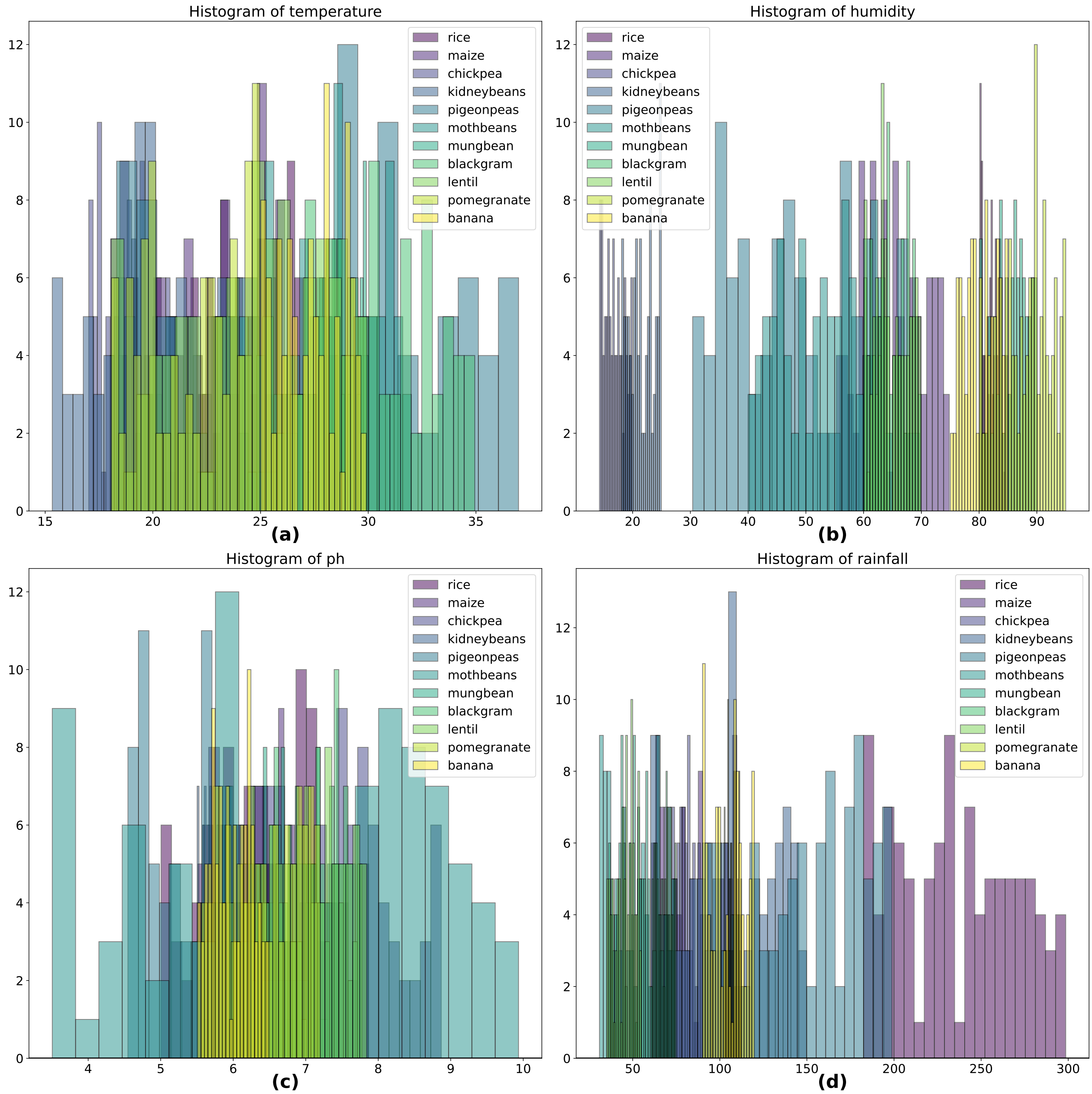

Figure 5. Histograms depicting the distribution of cultivated crops across various environmental factors: temperature (**a**), humidity (**b**), pH (**c**), and rainfall (**d**). The data represents crops including rice, maize, chickpea, kidney beans, pigeon peas, moth beans, mung bean, black gram, lentil, pomegranate, and banana.

Histograms of crop distribution reveal how climatic factors such as temperature and rainfall vary among different crops. For example, crops like bananas and coconuts require higher rainfall, between 250 and 300 mm, while pigeon peas and moth beans thrive with as little as 50 mm. Papayas and grapes exhibit high heat tolerance, with grape temperatures ranging widely from 8 to 43 degrees Celsius. Humidity preferences also differ, with mangoes requiring humidity levels above 70%, whereas coconuts and musk melons thrive in conditions exceeding 90%.

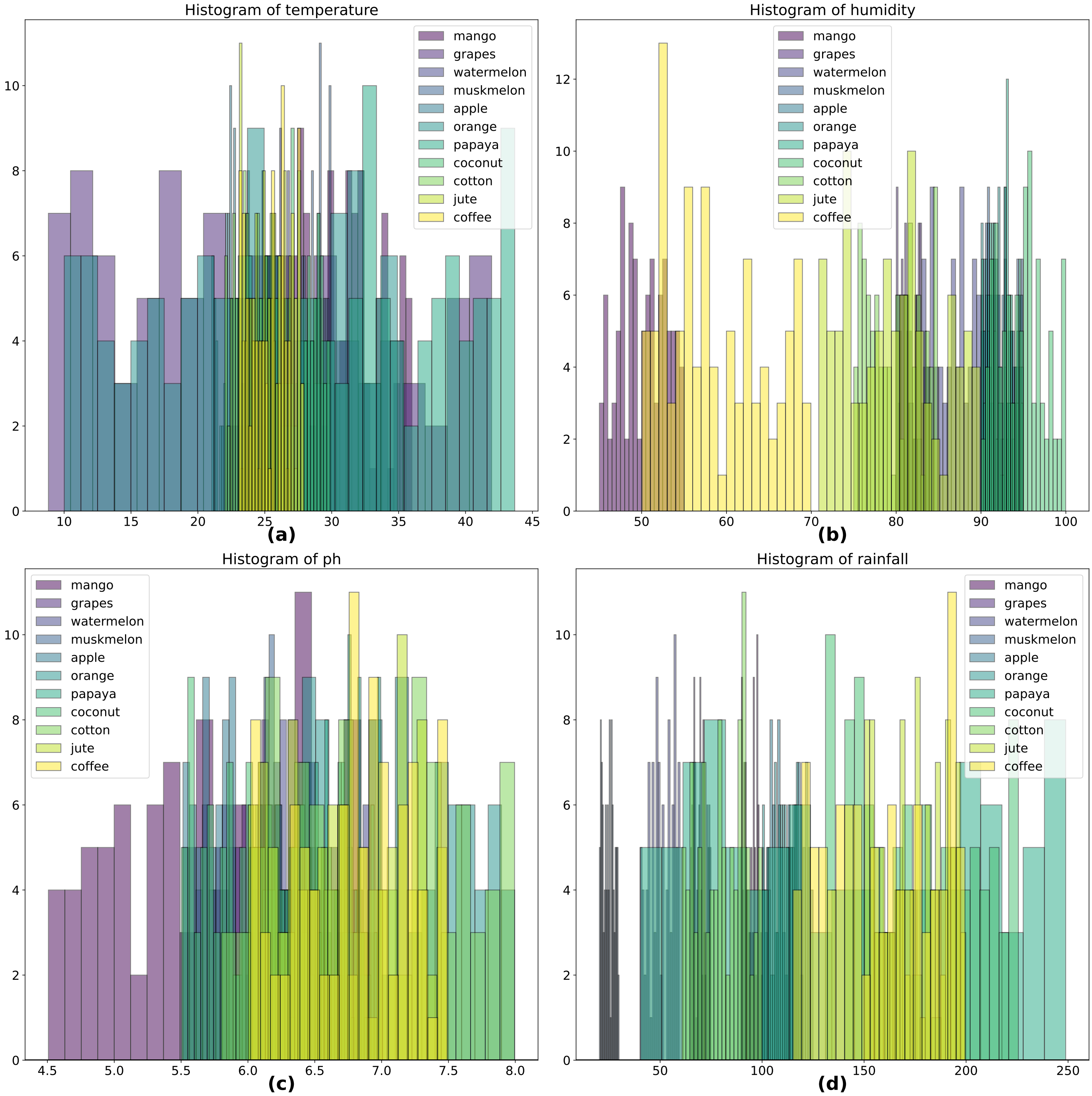

Figure 6. Histograms depicting the distribution of cultivated crops across various environmental factors: temperature (**a**), humidity (**b**), pH (**c**), and rainfall (**d**). The data represent crops including mango, grapes, watermelon, muskmelon, apple, orange, papaya, coconut, cotton, jute, and coffee.

Figures 7 and 8 show histograms of nitrogen (N), phosphorus (P), and potassium (K) levels across various crops, highlighting distinct nutrient requirements. For instance, nitrogen levels vary widely, with maize and bananas showing higher concentrations compared to pulses like chickpeas and lentils. Phosphorus levels also differ, with rice requiring more phosphorus than pomegranate. Potassium distribution reveals varying crop needs, with chickpeas having higher potassium levels compared to moth beans.

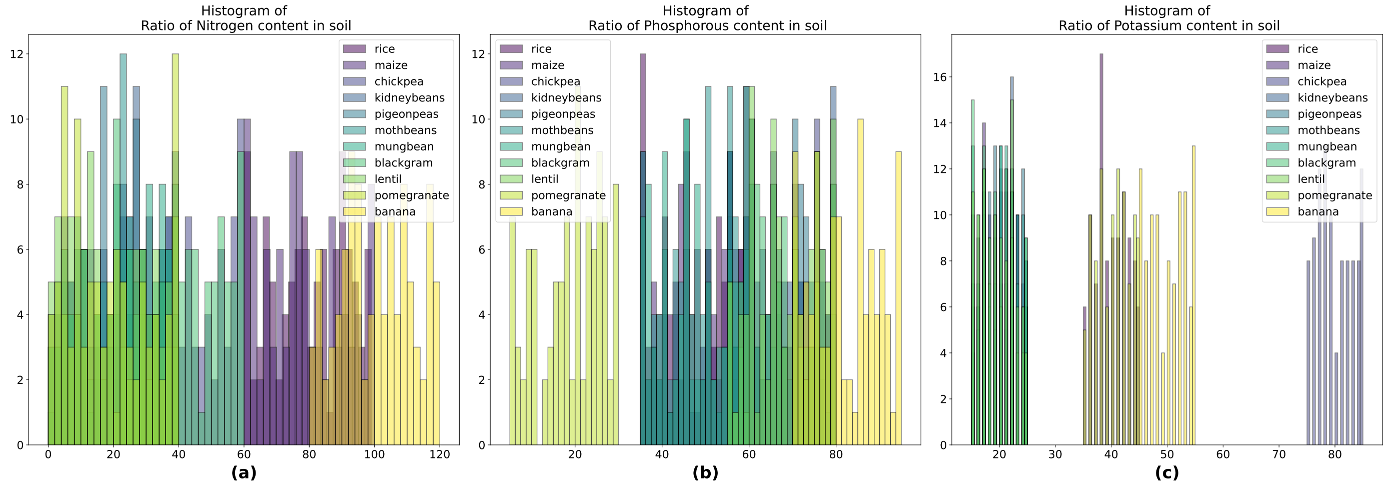

Figure 7. Histograms illustrating the distribution of cultivated crops based on key macronutrient features: nitrogen (**a**), phosphorus (**b**), and potassium (**c**). The crops analyzed include rice, maize, chickpea, kidney beans, pigeon peas, moth beans, mung bean, black gram, lentil, pomegranate, and banana.

Nitrogen (Figures 7a and 8a) is a component of chlorophyll [78], the pigment responsible for photosynthesis, and amino acids, which are the building blocks of proteins. Adequate nitrogen levels are vital for healthy crop development, significantly impacting overall growth and yield.

Phosphorus (Figures 7b and 8b) plays a role in energy transfer [79], photosynthesis, and the movement of nutrients within plants. It is also essential for root and flower development. Maintaining optimal phosphorus levels is crucial for ensuring robust plant health and maximizing productivity.

Potassium (Figures 7c and 8c) regulates various physiological processes, including water uptake [80], enzyme activation, and photosynthesis. Sufficient potassium levels enhance a plant's disease resistance, improve water use efficiency, and boost both crop quality and yield.

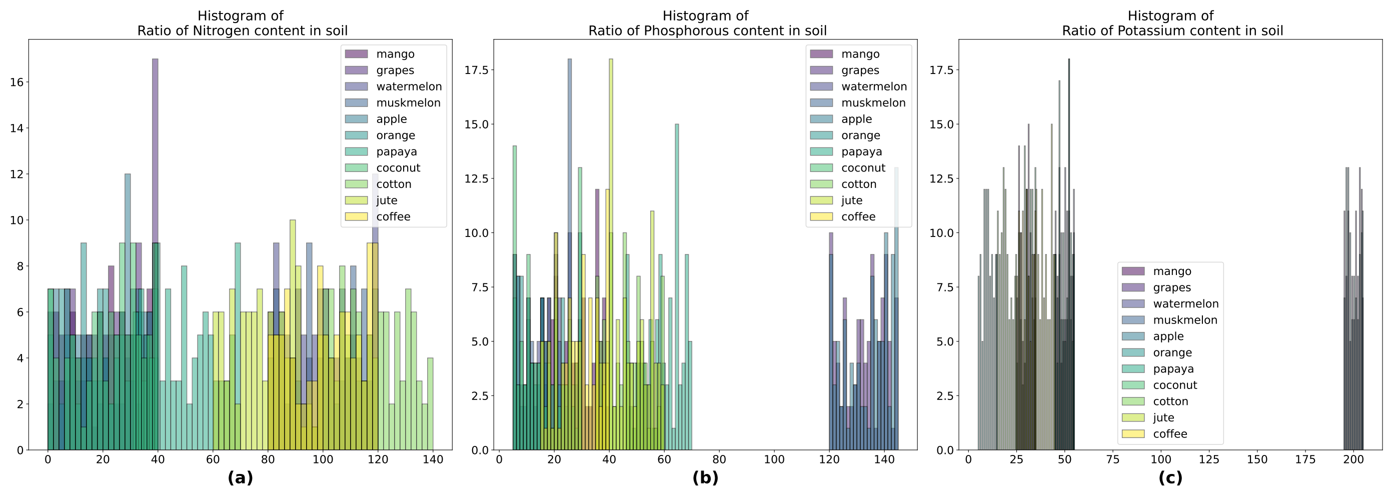

Figure 8. Histograms illustrating the distribution of cultivated crops based on key macronutrient features: nitrogen (**a**), phosphorus (**b**), and potassium (**c**). The crops analyzed include mango, grapes, watermelon, muskmelon, apple, orange, papaya, coconut, cotton, jute, and coffee.

Soil nutrient levels, specifically nitrogen (N), phosphorus (P), and potassium (K), showcase variability, spanning deficient to nutrient-rich soils. Nitrogen, which ranges from

0 to 140 mg/kg with a mean of 50.55 mg/kg, is vital for vegetative growth and chlorophyll synthesis, making it relevant for leafy crops like rice and maize. Phosphorus levels, with a range from 5 to 145 mg/kg and a mean of 53.36 mg/kg, influence root development, especially in the early growth phases. Phosphorus is strongly correlated with potassium (correlation coefficient of 0.74), indicating that these nutrients often co-exist in balanced soils. This synergy could favor crops with combined nutrient demands, such as root vegetables or legumes, which require robust root systems for optimal growth. Potassium, varying from 5 to 205 mg/kg with a mean of 48.15 mg/kg, is essential for water regulation and disease resistance. Crops like bananas and other tropical fruits, which are highly potassium-dependent, could benefit from this range of potassium levels.

The temperature range captured in the dataset, from 8.8 °C to 43.7 °C, with an average of 25.6 °C, reflects diverse climatic conditions that accommodate both cool and warm-weather crops. Warmer temperatures, for instance, could support heat-tolerant crops like maize and cotton, while cooler temperatures below 20 °C would be ideal for fruits like apples and grapes. Humidity, ranging from 14.3% to nearly 100% and averaging at 71.5%, further enhances this diversity. High humidity levels may support tropical crops, while lower levels favor drought-resistant varieties like chickpeas and pulses. The positive correlation between potassium and humidity (0.19) suggests that soils in humid environments often have higher potassium levels, potentially supporting potassium-dependent crops that thrive in moist conditions.

Soil pH, ranging from 3.5 to 9.9 with a mean of 6.47, represents an extensive spectrum from acidic to moderately alkaline soils. This variation is significant, as pH influences nutrient availability and soil microbial activity. Neutral to slightly acidic soils (pH 6–7) are generally ideal for a broad range of crops, while strongly acidic soils (below pH 5.5) are favorable for crops like coffee and citrus, which are well-suited to such conditions. Conversely, moderately alkaline soils can support crops that thrive at a higher pH, like barley and some legumes.

Rainfall spans from 20.2 to 298.6 mm, with an average of 103.46 mm, covering arid to monsoon-like conditions. This variability allows the dataset to account for crops with high water demands, like rice and sugarcane, as well as those adapted to low rainfall, such as chickpeas and lentils. Additionally, the positive correlation between rainfall and humidity (0.09) underscores the interdependence of these conditions, which can impact crop resilience to water availability.

2.3. DNN Methodology

An approach to optimizing hyperparameters for a DL model applied to a regression and classification task using GA [81] was employed.

GA offers a robust advantage in hyperparameter optimization by adaptively exploring complex, high-dimensional search spaces, reducing the risk of local optima, and handling mixed parameter types more efficiently than, for example, grid search or Bayesian optimization. Grid search [82], though effective in exploring a pre-defined parameter space, becomes computationally expensive when the parameter space is large. It exhaustively evaluates each possible combination of parameters, which can be infeasible for models with high-dimensional hyperparameter spaces, resulting in a substantial time and resource cost. Bayesian optimization [83], while more efficient, relies on probabilistic modeling (e.g., Gaussian processes) to select promising parameters and can struggle with highly non-linear, irregular search spaces due to its reliance on the assumptions embedded in its surrogate model. Bayesian methods also have limited flexibility in adjusting to dynamic fitness landscapes, which can lead to suboptimal performance in scenarios with complex, multi-modal parameter relationships.

Initially, data were loaded and preprocessed. The dataset, imported from a CSV file, was divided into feature and target variables. The target variable was encoded using the LabelEncoder method and transformed into a one-hot encoded format suitable for the classification model.

The data were then split into training and testing sets, and the StandardScaler method was applied to standardize the features, ensuring a mean of zero and a standard deviation of one.

The model architecture employed was a TensorFlow Keras DNN. The hyperparameter space included a wide range of configurations, with the number of layers varying from 1 to 20 and neurons per layer ranging from 1 to 128. This flexibility enabled fine-tuning of the network's depth and density to determine the optimal setup. Activation functions [84] used included ReLU, Sigmoid, Tanh, Softmax, Softplus, Softsign, ELU, SELU, GELU, hard Sigmoid, and linear, enabling the selection of the most suitable functions for the data. Various optimization algorithms [85] were employed to adjust the learning rate during training, including Adam, SGD, RMSprop, Adagrad, Adadelta, Adamax, and Nadam. Learning rates were parameterized at 0.0001, 0.001, 0.01, and 0.1 to fine-tune gradient descent for optimal convergence. Categorical cross-entropy was used as the loss function for multiclass classification, determining how the network penalizes errors.

The overall hyperparameter space $\mathcal{H}$ for a TensorFlow Keras DNN can be described mathematically as follows:

- Number of layers:
$$L \in \{1, 2, \ldots, 20\}$$
where L denotes the total number of layers in the network.
- Number of neurons per layer:
$$N_l \in \{1, 2, \ldots, 128\} \quad \text{for} \quad l \in \{1, 2, \ldots, L\}$$
where N_l represents the number of neurons in the l-th layer.
- Activation functions:
$$A \in \{\text{Linear}, \text{Hard Sigmoid}, \text{GELU}, \text{SELU}, \text{ELU}, \text{Softsign},$$
$$\text{Softplus}, \text{Softmax}, \text{Tanh}, \text{Sigmoid}, \text{ReLU}\}$$
where A denotes the activation function applied in each neuron.
- Optimizers:
$$O \in \{\text{Nadam}, \text{Adamax}, \text{Adadelta}, \text{Adagrad}, \text{RMSprop}, \text{SGD}, \text{Adam}\}$$
where O represents the optimization algorithm used for training the network.
- Learning rates:
$$\eta \in \{0.0001, 0.001, 0.01, 0.1\}$$
where η is the learning rate parameter used in the optimization process.
- Loss function for classification task:
$$L_{f_c} = \text{Categorical cross-entropy}$$
where L_{f_c} is the loss function employed for evaluating the performance of the classification model.
- Loss function for the regression task:
$$L_{f_r} = \text{MSE, MAE, logcosh, huberloss}$$
where L_{f_r} is the loss function employed for evaluating the performance of the regression model.

The overall hyperparameter space $\mathcal{H}$ for a TensorFlow Keras DNN can be described as follows:

$$\mathcal{H} = \{(L, N, A, O, \eta, L_f)\}$$

The fitness function [86] evaluated the quality of models trained with different hyperparameter combinations. It involved building and compiling the model with the selected hyperparameters, training it on the training data with fixed epochs and batch size, and testing it on a held-out test set. Model predictions were compared with true class labels to calculate accuracy, which served as the primary criterion for the fitness function. Additionally, loss history and accuracy at each training step were recorded, providing a detailed evaluation of the training process and stability.

Results from each generation of the algorithm were stored, allowing analysis of the hyperparameter evolution [87] and the effectiveness of the GA. This comprehensive assessment guided the evolutionary process toward identifying the best hyperparameters for the classification task.

Model quality was assessed using accuracy. The evaluation involved compiling the model with the specified optimizer and loss function, training it on the training set, and making predictions on the test set. Prediction accuracy was computed using the accuracy score function from the scikit-learn library.

The described GA process could be defined mathematically as follows:

- Generate an initial population $\mathcal{P}_0$ of candidate hyperparameter sets $\{\mathbf{h}_i\}_{i=1}^N$, where each $\mathbf{h}_i$ includes the following:

$$\mathbf{h}_i = (L_i, \mathbf{N}_i, \mathbf{A}_i, O_i, \eta_i, L_f)$$

- For each hyperparameter set $\mathbf{h}_i$, train the model on the training dataset for E epochs and batch size B.

$$\text{Train}(\mathbf{h}_i, E, B)$$

- Test the model on a held-out test set to compute the DOR for the classification task:

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}}$$

  and, for the regression task:

$$R^2 = 1 - \frac{\sum_{j=1}^{M}(y_j - \hat{y}_j)^2}{\sum_{j=1}^{M}(y_j - \bar{y})^2}$$

  where y_j are true values, $\hat{y}_j$ are predicted values, and $\bar{y}$ denotes the mean of the true values.
- Define the fitness function $F(\mathbf{h}_i)$ as accuracy or R^2 for the required task:

$$F(\mathbf{h}_i) = \text{Accuracy} \quad F(\mathbf{h}_i) = R^2$$

- Select individuals with higher fitness scores using a selection strategy (e.g., roulette wheel or tournament selection).
- Perform crossover between pairs of selected individuals to create offspring. Let $\mathbf{h}_i$ and $\mathbf{h}_j$ be the parents; the offspring $\mathbf{h}'_{ij}$ is generated by the following:

$$\mathbf{h}'_{ij} = \text{Crossover}(\mathbf{h}_i, \mathbf{h}_j)$$

- Apply mutation to the offspring to introduce genetic diversity:

$$\mathbf{h}''_{ij} = \text{Mutation}(\mathbf{h}'_{ij})$$

- Form a new population $\mathcal{P}_{t+1}$ by combining the offspring with the best individuals from the current generation:

$$\mathcal{P}_{t+1} = \text{SelectBest}(\mathcal{P}_t \cup \text{Offspring})$$

- Continue iterations until a stopping criterion is met (e.g., the maximum number of generations G or convergence):

$$t = \min(G, \text{Convergence})$$

- Identify the best hyperparameter set $\mathbf{h}_{\text{best}}$ based on the highest fitness score:

$$\mathbf{h}_{\text{best}} = \arg\max_{\mathbf{h}_i \in \mathcal{P}_G} F(\mathbf{h}_i)$$

- Analyze the evolution of hyperparameters across generations to evaluate the GA's effectiveness:

$$\text{Analyze}(\mathcal{P}_0, \mathcal{P}_1, \ldots, \mathcal{P}_G)$$

3. Results

Figure 9 illustrates the model accuracy evolution across various neural network configurations, sorted in ascending order. Initially, over 250 individuals exhibited low accuracy, below 0.2. Subsequently, a rapid increase in accuracy was observed up to 0.95 by the 350th individual, stabilizing around 0.97 ± 0.01 up to the 450th individual. The maximum accuracy of 0.998 was achieved at the final point.

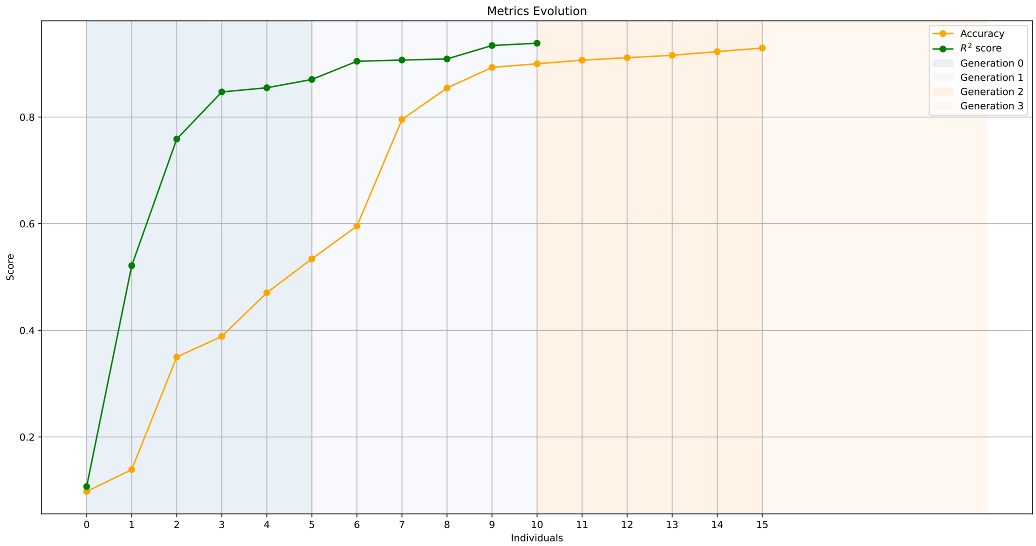

Figure 9. Evolution of metrics sorted in ascending order.

Table 2 presents the best hyperparameter configurations obtained after the optimization process. It includes the number of layers, neurons per layer, activation functions, optimizer, learning rate (alpha), and the corresponding metrics achieved by each configuration. Metric scores are derived from a rigorous evaluation process involving 5-fold cross-validation, which ensures that the reported results are robust and generalizable across different subsets of the data. This approach helps to mitigate overfitting and provides a reliable estimate of model performance.

Figure 10 presents a comparison between actual and predicted yield values for the architecture 'Yield prediction' from Table 2. This analysis aims to illustrate the model's accuracy in forecasting agricultural outcomes. With an. With an R^2 value of 0.92, there is a strong correlation between the predicted and actual yield predictions.

The integration of explainable AI (XAI) techniques, such as local interpretable model-agnostic explanations (LIME), into DNNs aimed at yield prediction is important for applications in agriculture. Understanding the contribution of various factors to crop yield can enhance decision-making and policy formulation. The LIME diagram (Figure 11) highlights the top 70 features impacting the model's predictions. These features are categorized into two main groups based on their impact direction and magnitude. Negative impact features negatively influence yield predictions, with significant contributors including various crops (e.g., Soybean, Cotton, Wheat, Rice), seasonal factors (e.g., summer, winter, Kharif, Rabi), and environmental parameters (e.g., temperature, humidity, precipitation). Positive impact features positively influence yield predictions, although the specific positive impact features are not explicitly listed in the provided diagram snippet.

Table 2. DNN hyperparameter configurations and scores on the test set.

Task	Layers	Neurons Per Layer	Activation Functions	Optimizer	Alpha	Loss Function	Metrics Type	Score
Yield prediction	2	[24, 126]	[Softmax, Softsign]	Adamax	0.1	MSE	R^2 score	0.92 ± 0.04
							RMSE	2.45 ± 0.12
							MPE	1.75 ± 0.09
Crop selection	2	[79, 111]	[linear, GELU]	Adam	0.001	categorical cross-entropy	Accuracy	0.87 ± 0.11
Crop selection	6	[36, 53, 62, 9, 39, 9]	[linear, Tanh, ReLU, Tanh, Sigmoid, ...]	Adagrad	0.100	categorical cross-entropy	Accuracy	0.91 ± 0.07

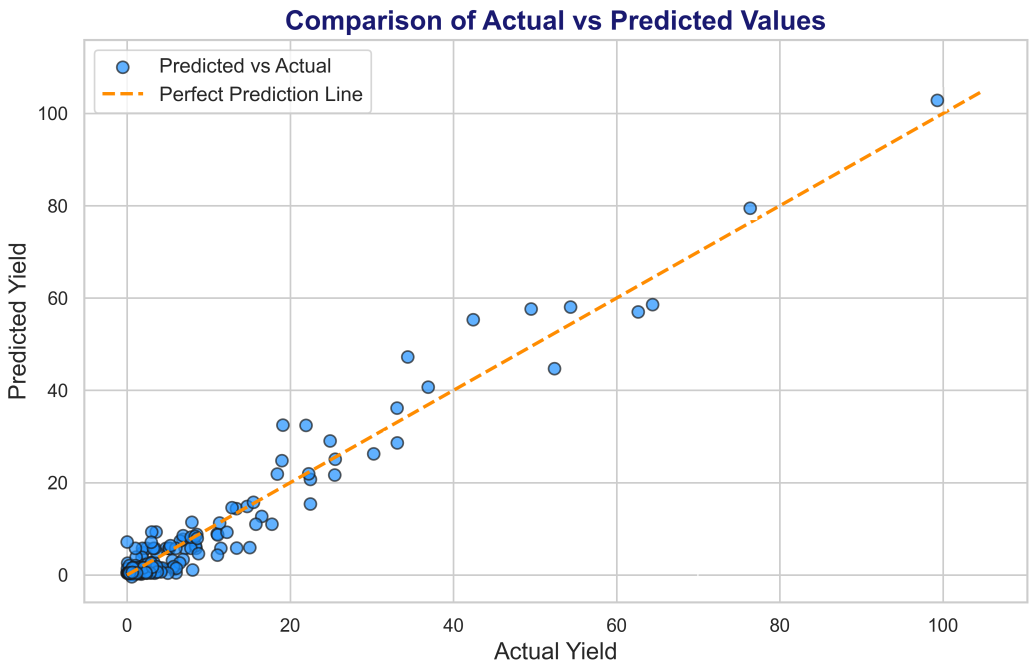

Figure 10. Comparison of actual and predicted yield values, demonstrating the model's high accuracy with an R^2 of 0.92, which confirms its effectiveness in forecasting.

Analyzing the key features, crop-specific features like soybean, cotton, wheat, and rice exhibit significant negative impacts on yield prediction. This could indicate the model's sensitivity to the areas under these crops, suggesting their critical role in determining overall yield. The impact of different seasons (e.g., summer, winter, Kharif, Rabi) on yield prediction highlights the importance of temporal factors in agricultural productivity. Variables such as temperature, humidity, and precipitation reflect the model's reliance on climatic conditions to predict crop yields accurately.

To improve the model, incorporating real-time data from remote sensing technologies, such as satellites, can enhance the model's responsiveness to dynamic environmental changes. While GA has proven effective for HPO, it is computationally intensive. Future research could explore hybrid approaches combining GA with other optimization algorithms to enhance efficiency and scalability. Developing region-specific and crop-specific models can improve prediction accuracy by tailoring models to local environmental conditions and agricultural practices, thereby enhancing their relevance and applicability. Expanding the dataset to include diverse geographical regions and more comprehensive environmental variables can improve the model's generalizability, ensuring that it performs well across different contexts and is not overly reliant on region-specific data. Ensuring robust and scalable predictions across various regions and contexts necessitates the integration of multiple environmental variables. This includes climatic factors, soil conditions, and agricultural practices, which can be challenging but essential for accurate yield predictions.

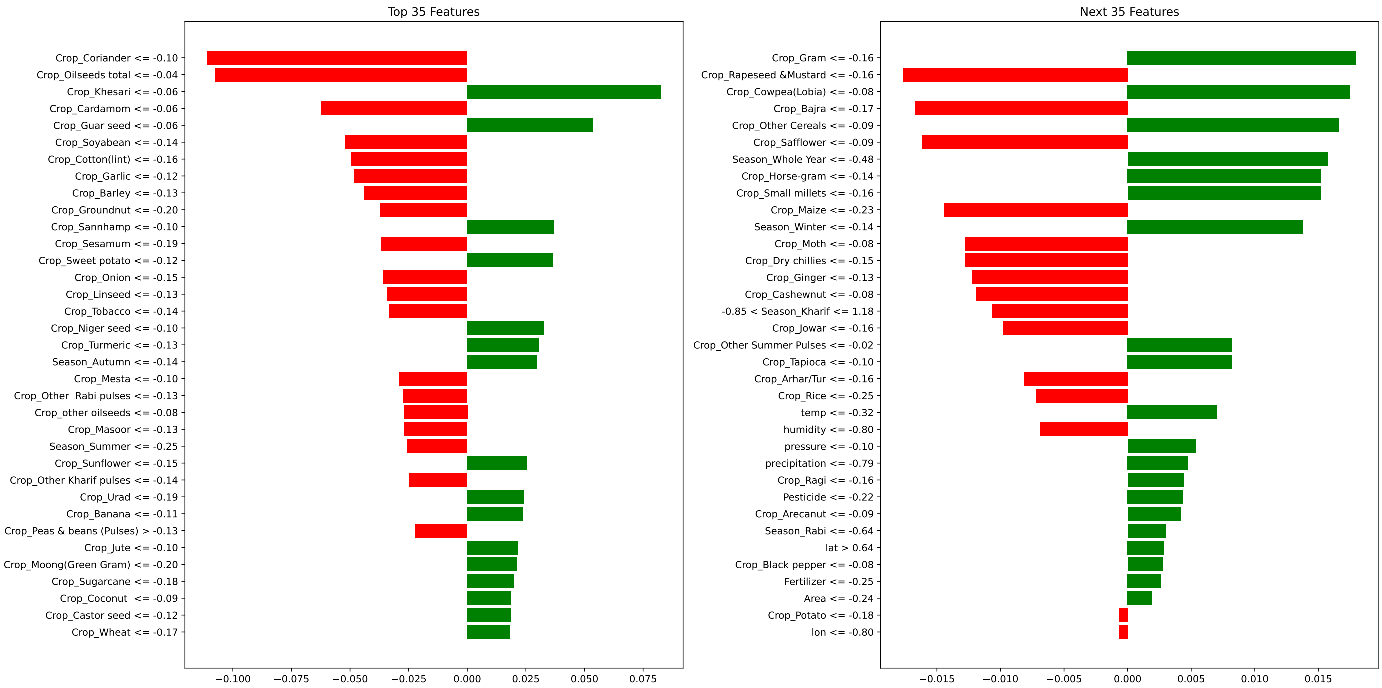

Figure 11. LIME explanations of the 'Yield prediction' DNN architecture from Table 2.

The confusion matrices for crop selection models (from Table 2) are depicted in Figure 12, illustrating minimal inter-class misclassification, as evidenced by the near absence of off-diagonal elements. The labels used in the tables are as follows: apple (0), banana (1), black gram (2), chickpea (3), coconut (4), coffee (5), cotton (6), grapes (7), jute (8), kidney beans (9), lentil (10), maize (11), mango (12), moth beans (13), mung bean (14), muskmelon (15), orange (16), papaya (17), pigeon peas (18), pomegranate (19), rice (20), and watermelon (21).

The difference between the architectures is particularly notable for moth beans (13), where the two-layer model correctly identifies 21 entries, while the six-layer model achieves 24 correct identifications. There is also some minor confusion observed with the lentil class (10).

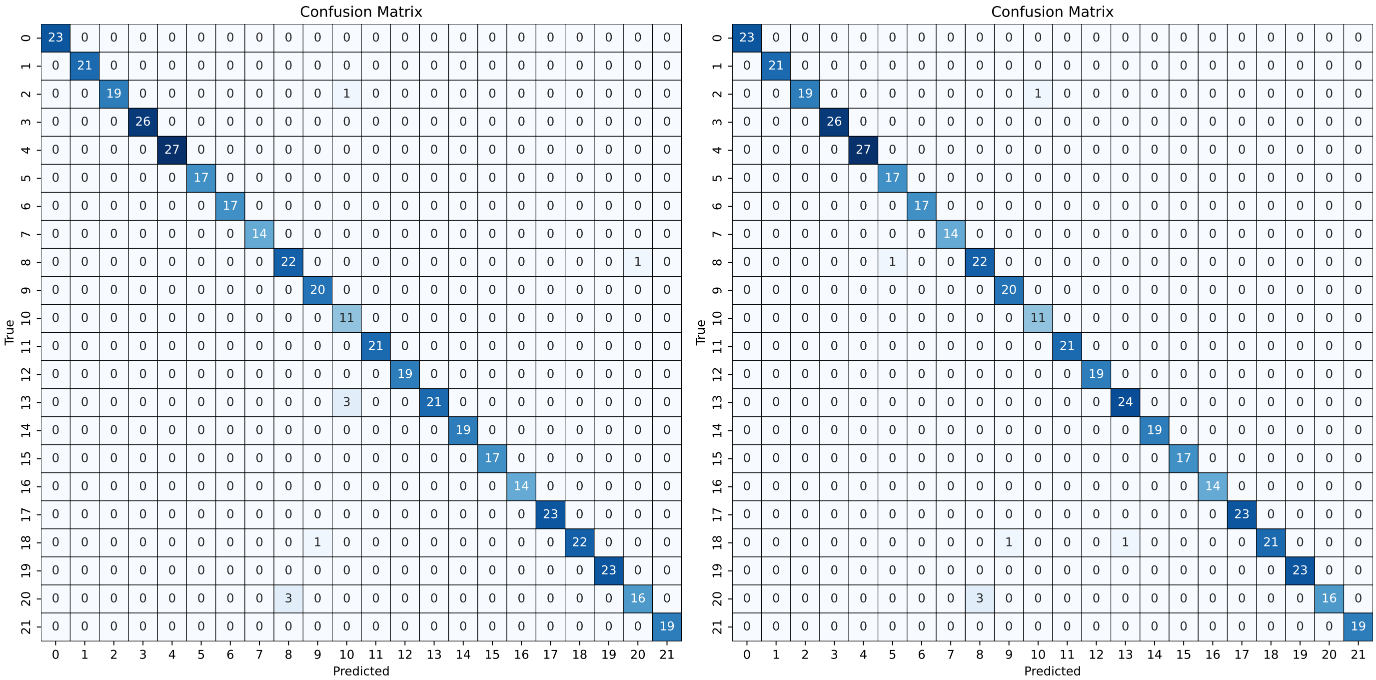

Figure 12. Confusion matrices for the model's crop selection (from Table 2) for the test set.

4. Discussion

Yield prediction and the selection of optimal conditions for crop growth are vital components of modern agriculture. The application of ML techniques to these tasks has the potential to transform agricultural practices by providing accurate, data-driven insights that can improve productivity and sustainability. The workflow diagram in Figure 13 outlines the structured approach for crop yield prediction and optimization using ML. Divided into key stages, this framework highlights the steps from data acquisition to interpretation and future directions.

ML algorithms are increasingly being used to predict crop yields based on various environmental, climatic, and agricultural factors. These algorithms analyze historical data to identify patterns and make predictions about future yields. Commonly used ML techniques in this domain include DNNs, support vector machines [88] (SVMs), and random forests [89] (RFs). DNNs are capable of modeling complex, non-linear relationships between input features and yield outcomes. They are particularly effective when large datasets are available, as they can learn intricate patterns from the data. SVMs are used for regression and classification tasks and have been applied to yield prediction due to their ability to handle high-dimensional spaces and define clear decision boundaries. RFs are ensemble methods that combine multiple decision trees to improve predictive accuracy and control overfitting [90]. They are useful in yield prediction due to their robustness and interpretability.

ML models utilize a wide array of features to predict crop yields. These features can be broadly categorized into environmental, climatic, and agricultural factors. Environmental factors such as soil type, pH levels, nutrient content, and topography influence crop growth and yield. Climatic factors, including temperature, humidity, precipitation, and solar radiation, are critical determinants of crop performance. Agricultural practices like crop variety, planting density, irrigation practices, and fertilizer usage directly impact yield outcomes.

This study highlighted the integration of GA optimization with DNN for yield prediction. The model was optimized with specific parameters, including the number of layers, activation functions, optimizer, and learning rate, achieving a high R^2 score of 0.92 ± 0.04. Local interpretable model-agnostic explanations (LIME) were employed to provide insights into the model's decision-making process, identifying key features such as crop type, seasonal factors, and environmental conditions. While LIME plots offered insights into the model's decision-making, future research should explore more advanced methods like Shapley additive explanations [91] (SHAP). Utilizing SHAP can offer deeper insights into feature contributions and enhance model transparency. Comparative experiments between LIME and SHAP could validate the robustness of explanations and improve the credibility of the model. This focus on interpretability is essential for building trust in machine learning applications, especially in critical domains where understanding model behavior is vital.

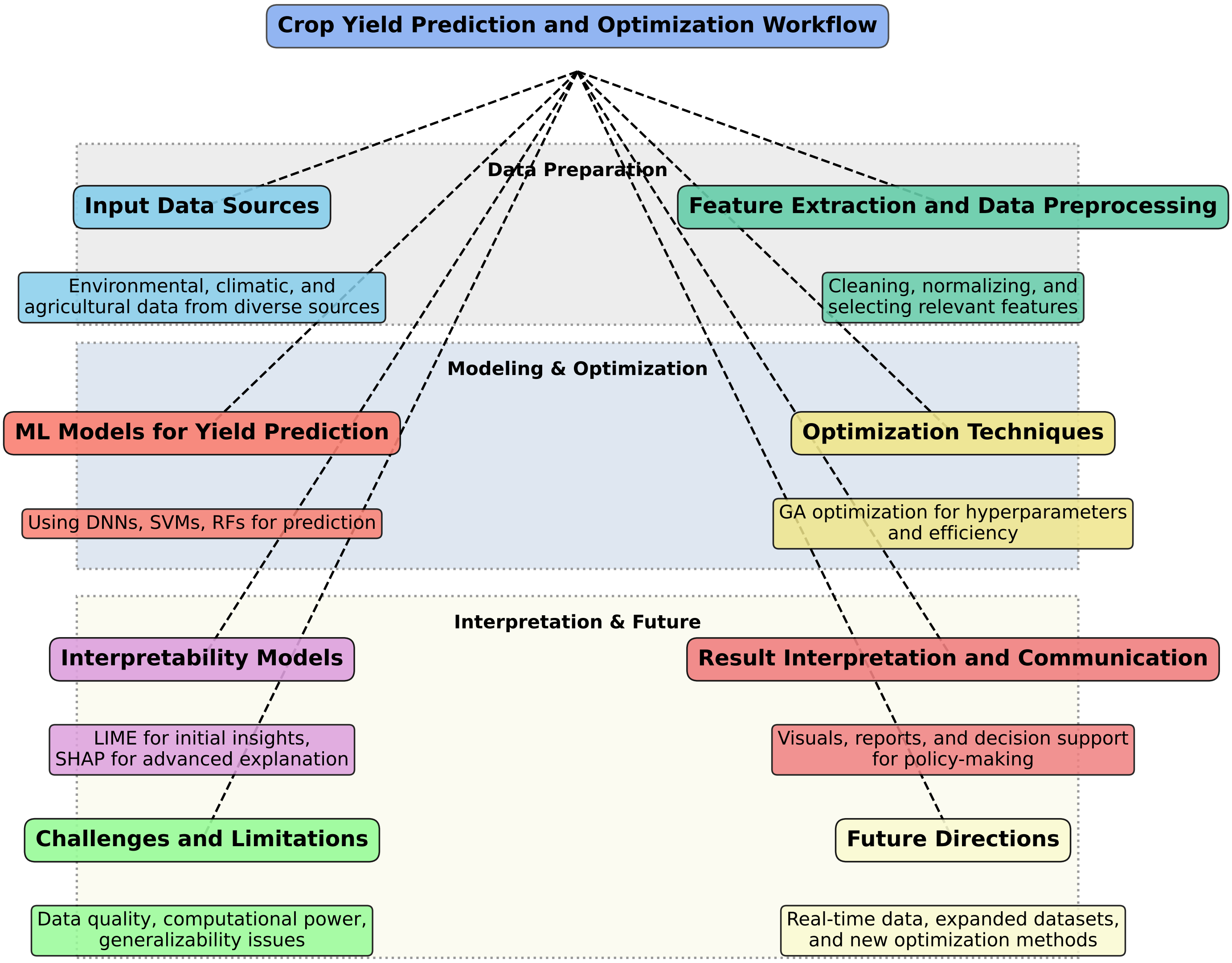

Figure 13. Workflow diagram for crop yield prediction and optimization using ML techniques.

Translating complex model explanations into accessible graphics or concise reports is essential for enhancing the usability of research findings among users and policymakers. Such visual representations can demystify intricate analytical results, making them more comprehensible to non-expert audiences. For example, Berget [92] introduces a spatial multi-agent programming model designed to evaluate policy options for the diffusion of innovations and changes in resource use. Utilizing a multi-agent/cellular automata approach, the model explicitly captures the social and spatial interactions of heterogeneous farm households, incorporating economic and hydrologic processes to analyze the impacts of water use for irrigation and the adoption of agricultural innovations. Agricultural systems science offers insights for tackling complex issues and making informed decisions. As researchers strive to create next-generation models, it is notable to reflect on the rich history of agricultural modeling, which encompasses diverse approaches and collaborations [93]. By learning from past developments, the community can effectively advance future models and decision support systems while avoiding common pitfalls.

Despite the successes, several challenges remain in the application of ML to yield prediction and optimal crop growth conditions. High-quality, comprehensive datasets are essential for training accurate ML models. Data gaps [94] can particularly limit model effectiveness. While complex models like DNNs offer high predictive accuracy, they often operate as "black boxes" [95]. Techniques like LIME are essential for interpreting these models, but further advancements in explainable AI are needed. Models trained on data from specific regions or conditions may not generalize well to other areas. Incorporating diverse datasets from various geographical regions can improve model robustness. Training complex ML models, especially those involving GA optimization, requires significant computational power. Developing more efficient algorithms and leveraging cloud-based solutions can mitigate this challenge.

Additionally, the model's limitations include reduced accuracy in predicting yields for less common crops [96], such as pigeon peas, moth beans, and other niche varieties, which may not have as extensive historical or environmental data available. Moreover, crops grown in challenging or extreme conditions—such as drought-prone or highly saline areas—are harder to model accurately [97]. This is due to limited data on crop resilience and the complex interactions between environmental stressors and crop performance that the model may not fully capture.

Expanding the model's applicability to developing countries could provide more agricultural benefits. By integrating insights from these findings into decision-support tools, smallholder farmers in resource-limited settings could optimize crop selection and timing. Such models, when combined with mobile technology and IoT-based soil and weather monitoring, could empower farmers with predictive insights, helping them adapt to climate variability, improve yield forecasts, and make informed decisions to increase food security and reduce losses [98].

To improve the model, incorporating real-time data from remote sensing technologies, such as satellites and IoT devices [99], can enhance model accuracy and responsiveness to changing conditions. Combining GA with other optimization algorithms can improve efficiency and scalability, reducing computational demands. Additionally, integrating GA optimization significantly enhances the performance of DNNs by fine-tuning their hyperparameters, thereby improving predictive accuracy. While utilizing random choice for HPO can still wield significant influence over DNN performance, their optimal values often remain elusive due to the exponential space and complex interactions involved. Traditional methods are often impractical for this purpose, but evolutionary algorithms like GA offer scalable solutions.

It should also be noted that the dataset used in this study was primarily limited to specific regions and crop types, which presents a potential limitation in the model's generalizability. This regional and crop-specific focus means that while the model performs well within the confines of the dataset, its applicability to broader geographical areas and a more diverse range of crops may be constrained. Future research could address this limitation by expanding the dataset to include additional regions with varying climates, soil conditions, and a wider variety of crop types. Such an expansion would not only improve the model's robustness and predictive accuracy across diverse agricultural settings but also enhance its adaptability to different environmental conditions, making it more broadly applicable to real-world agricultural challenges. This broader dataset could enable more comprehensive yield prediction and crop recommendations that are relevant across various regions, ultimately contributing to more sustainable and adaptable agricultural practices.

Big data applications in agriculture are advancing but face challenges in scalability and real-world readiness due to the inherent complexity of agricultural data (volume, velocity, variety, veracity). Although the adoption of big data in agriculture is growing, practical deployment remains limited, as most solutions require domain-specific customization and remain at a low technological maturity level [100]. Addressing these gaps requires both practical engineering and systems-thinking approaches to deliver scalable, affordable, and user-friendly solutions tailored to agriculture's unique needs. However, the current study did not delve into the specific challenges associated with existing data processing technologies. Issues such as processing speed, data integration from disparate sources, and the security of data storage and transmission are factors that can impact the feasibility and effectiveness of big data applications in agriculture. Without addressing these challenges, the potential of big data to transform agricultural practices may be constrained, as processing bottlenecks and security vulnerabilities could limit its accessibility and reliability for widespread agricultural use. Future studies should explore these technical challenges in greater depth to provide solutions that enhance data processing efficiency and security, thereby facilitating the broader adoption of big data technologies in the agricultural sector.

For example, Shrestha et al. [101] integrated nature-inspired enhancements and Monte Carlo-based methods to enhance DNN architecture for superior accuracy on datasets compared to traditional genetic algorithms. Their study leveraged augmented datasets comprising comprehensive rainfall, climate, and fertilizer data sourced from India, serving as robust training and validation sets for the model. By employing GA, the DNN achieves superior classification results, effectively identifying optimal crop choices tailored to diverse environmental scenarios. Another approach, suggested by Liu et al. [36], involves a hybrid intelligent genetic algorithm (HIGA) that integrates DNNs and GAs. A two-step training approach refines the DNN with data from the optimization process, enhancing predictive accuracy for efficient truss optimization. Evaluating three classic truss problems validates the method's effectiveness while exploring various settings demonstrates its robustness and applicability in improving optimization performance and efficiency.

DNN layers are complex loops that can be organized, tiled, and scheduled in various ways on DNN accelerators. Optimal per-layer mappings are beneficial, yet selecting the right mapping remains challenging due to the vast map space. Kao et al. [102] introduced GAMMA, a genetic algorithm-based approach tailored for hardware mapping. Unlike previous methods limited to specific accelerators or mappings, GAMMA explores a flexible map space efficiently. Comparative evaluations demonstrate GAMMA consistently outperforms other methods in finding optimized mappings.

Despite its strengths, this approach exhibits several limitations that warrant consideration for future research directions. Firstly, while the DNN model proves effective in predicting crop suitability based on available environmental data, its reliance on historical datasets may limit its adaptability to real-time or dynamically changing conditions [103]. Incorporating real-time data, potentially obtained from remote sensing, particularly satellites [104], could enhance global and regional analysis of crop conditions and production. For example, Wu et al. [105] employ a hierarchical approach that integrates climatic and remote sensing indicators at multiple scales—from global environmental conditions to detailed assessments at the national and sub-national levels. This methodology provides accurate and timely information, supporting food producers with insights into crop health, farming intensity, and production trends.

The complexity of integrating multiple environmental variables into the model poses challenges in ensuring the robustness and scalability of the predictions across different geographical regions and varying agricultural contexts. Furthermore, the use of GA for HPO, while effective, requires significant computational resources [106] and time-intensive experimentation. Future studies could explore more efficient optimization techniques or hybrid approaches that combine GA with other classification algorithms to further enhance model performance and scalability. Additionally, incorporating more diverse and comprehensive datasets from global regions could improve the generalizability and applicability of the model beyond specific geographical boundaries.

By advancing these methodologies, it is possible to enhance the precision and applicability of crop yield predictions across different regions and crops. Exploring the integration of advanced optimization techniques and leveraging diverse data sources will be pivotal in developing robust, scalable models capable of addressing the dynamic challenges in agriculture. This multifaceted approach holds the potential to significantly improve agricultural productivity and sustainability worldwide.

5. Conclusions

The application of GA optimization enhances the predictive accuracy and performance of DNNs by fine-tuning their hyperparameters. By utilizing augmented datasets from India that include rainfall, climate, and fertilizer data, the model is effectively trained and validated, achieving a resulting R^2 score of 0.92 ± 0.04. GA significantly improves hyperparameter optimization, allowing the DNN to deliver superior classification results and identify optimal crops suited to diverse environmental conditions.

Future research could expand the DNN-GA yield prediction model's scope by adapting it to diverse and less common crops, allowing for better regional customization and climate resilience modeling. Integrating real-time IoT data (e.g., soil moisture, temperature) would support precision agriculture by providing timely insights, while improvements in feature engineering could highlight key variables that most affect yield, enhancing both model simplicity and interpretability. Developing computationally efficient versions of the model would also make it accessible to farmers in low-resource settings, where data and processing power are limited. Additionally, the model could have cross-applications in agricultural insurance and economic risk assessment, offering reliable forecasts that inform crop-specific insurance products and sustainable farming practices. Finally, embedding this model in agronomic advisory systems would help farmers make data-driven decisions, from planting schedules to resource management, bolstering productivity and sustainability in the agricultural sector.

In conclusion, this study highlights the potential of DNNs and GA in promoting sustainability within precision agriculture. However, ongoing research should address challenges related to data availability, model interpretability, generalizability, and computational resource requirements. Incorporating real-time data, developing hybrid optimization techniques, creating region-specific models, and integrating diverse datasets will further enhance the adaptability, accuracy, and applicability of predictive models. By tackling these challenges, the agricultural sector can advance toward more sustainable and efficient practices on a global scale.

Author Contributions: Conceptualization, I.M.; formal analysis, A.G. and Y.T.; funding acquisition, A.B., A.G., V.N. and Y.T.; methodology, V.T.; project administration, A.B., A.G., V.N. and Y.T.; resources, I.M.; software, V.T., A.B. and V.N.; supervision, A.B., A.G., V.N. and Y.T.; validation, I.M. and V.T.; visualization, I.M.; writing—original draft, I.M.; writing—review and editing, V.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data is contained within the article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Charania, I.; Li, X. Smart farming: Agriculture's shift from a labor intensive to technology native industry. *Internet Things* **2020**, *9*, 100142. [CrossRef]
2. Catini, A.; Papale, L.; Capuano, R.; Pasqualetti, V.; Di Giuseppe, D.; Brizzolara, S.; Tonutti, P.; Di Natale, C. Development of a sensor node for remote monitoring of plants. *Sensors* **2019**, *19*, 4865. [CrossRef] [PubMed]
3. Taneja, A.; Nair, G.; Joshi, M.; Sharma, S.; Sharma, S.; Jambrak, A.R.; Roselló-Soto, E.; Barba, F.J.; Castagnini, J.M.; Leksawasdi, N.; et al. Artificial intelligence: Implications for the agri-food sector. *Agronomy* **2023**, *13*, 1397. [CrossRef]

4. Liakos, K.G.; Busato, P.; Moshou, D.; Pearson, S.; Bochtis, D. Machine learning in agriculture: A review. *Sensors* **2018**, *18*, 2674. [CrossRef]
5. Koul, B.; Yakoob, M.; Shah, M.P. Agricultural waste management strategies for environmental sustainability. *Environ. Res.* **2022**, *206*, 112285. [CrossRef]
6. Mandal, S.; Choudhury, B.; Satpati, L. Monsoon variability, crop water requirement, and crop planning for kharif rice in Sagar Island, India. *Int. J. Biometeorol.* **2015**, *59*, 1891–1903. [CrossRef]
7. Xu, D.; Duan, X.; Wang, B.; Hong, B.; Ho, T.H.D.; Wu, R. Expression of a late embryogenesis abundant protein gene, HVA1, from barley confers tolerance to water deficit and salt stress in transgenic rice. *Plant Physiol.* **1996**, *110*, 249–257. [CrossRef]
8. Ramadhan, A.J.; Tiwari, A.K.; Kumar, B.; Supriya, S.; Mishra, H.; Gautam, S.; Gautam, R.; Abotaleb, M.; Alkattan, H. Comparative Economics of Maize Crop in Kharif and Rabi Season. In *BIO Web of Conferences*; EDP Sciences: Hulis, France, 2024; Volume 97, p. 00134.
9. Mir, A.H.; Bhat, M.A.; Dar, S.A.; Sofi, P.A.; Bhat, N.A.; Mir, R.R. Assessment of cold tolerance in chickpea (*Cicer* spp.) grown under cold/freezing weather conditions of North-Western Himalayas of Jammu and Kashmir, India. *Physiol. Mol. Biol. Plants* **2021**, *27*, 1105–1118. [CrossRef]
10. Chilakala, A.R.; Pandey, P.; Durgadevi, A.; Kandpal, M.; Patil, B.S.; Rangappa, K.; Reddy, P.C.O.; Ramegowda, V.; Senthil-Kumar, M. Drought attenuates plant responses to multiple rhizospheric pathogens: A study on a dry root rot-associated disease complex in chickpea fields. *Field Crop. Res.* **2023**, *298*, 108965. [CrossRef]
11. Khan, A.; Singh, A.V.; Pareek, N.; Arya, P.; Upadhayay, V.K.; Kumar Jugran, A.; Kumar Mishra, P.; Goel, R. Credibility assessment of cold adaptive Pseudomonas jesenni MP1 and P. palleroniana N26 on growth, rhizosphere dynamics, nutrient status, and yield of the kidney bean cultivated in Indian Central Himalaya. *Front. Plant Sci.* **2023**, *14*, 1042053. [CrossRef]
12. Jeevarathinam, G.; Chelladurai, V. Pigeon pea. In *Pulses: Processing and Product Development*; Springer: Cham, Switzerland, 2020; pp. 275–296.
13. Nasir, M.; Sidhu, J.S.; Sogi, D.S. Processing and nutritional profile of mung bean, black gram, pigeon pea, lupin, moth bean, and Indian vetch. In *Dry Beans and Pulses: Production, Processing, and Nutrition*; Wiley: Hoboken, NJ, USA, 2022; pp. 431–452.
14. Parker, T.A.; Lo, S.; Gepts, P. Pod shattering in grain legumes: Emerging genetic and environment-related patterns. *Plant Cell* **2021**, *33*, 179–199. [CrossRef] [PubMed]
15. Rahman, M.M.; Bakr, M.; Mia, M.; Idris, K.; Gowda, C.; Kumar, J.; Deb, U.; Malek, M.; Sobhan, A. *Legumes in Bangladesh*; International Crops Research Institute for the Semi-Arid Tropics: Hyderabad, India, 2000.
16. Saroj, P.; Kumar, R. Recent advances in pomegranate production in India—A review. *Ann. Hortic.* **2019**, *12*, 1–10. [CrossRef]
17. Sarma, U.; Govila, V.K.; Yadav, A. The traditional and therapeutic use of banana and its plant based delicacies in ethnic Assamese cuisine and religious rituals from Northeast India. *J. Ethn. Foods* **2020**, *7*, 21. [CrossRef]
18. Jena, R.C.; Agarwal, K.; Chand, P.K. Fruit and leaf diversity of selected Indian mangoes (*Mangifera indica* L.). *Sci. Hortic.* **2021**, *282*, 109941. [CrossRef]
19. Kumar, M.; Ponnuswami, V.; Kumar, P.J.; Saraswathy, S. Influence of season affecting flowering and physiological parameters in mango. *Sci. Res. Essays* **2014**, *9*, 1–6. [CrossRef]
20. Priyanka, R.; Nagendran, K.; Aravintharaj, R.; Balaji, C.; Mohankumar, S.; Renukadevi, P.; Karthikeyan, G. Characterization and management of watermelon bud necrosis virus infecting watermelon in India. *Eur. J. Plant Pathol.* **2019**, *153*, 759–770. [CrossRef]
21. Kaur, A.; Sharma, M.; Manan, J. Comparative performance of muskmelon (*Cucumis melo*) hybrids at farmers' field in district Kapurthala. *J. Krishi Vigyan* **2017**, *6*, 24–31. [CrossRef]
22. Zahoor, S.; Dutt, V.; Mughal, A.; Pala, N.A.; Qaisar, K.; Khan, P. Apple-based agroforestry systems for biomass production and carbon sequestration: Implication for food security and climate change contemplates in temperate region of Northern Himalaya, India. *Agrofor. Syst.* **2021**, *95*, 367–382. [CrossRef]
23. Gupta, G.; George, P. Profitability of Nagpur santra (oranges) cultivation. *Indian J. Agric. Econ.* **1974**, *29*, 134–142.
24. Koul, B.; Pudhuvai, B.; Sharma, C.; Kumar, A.; Sharma, V.; Yadav, D.; Jin, J.O. *Carica papaya* L.: A tropical fruit with benefits beyond the tropics. *Diversity* **2022**, *14*, 683. [CrossRef]
25. Premchand, U.; Mesta, R.K.; Devappa, V.; Basavarajappa, M.P.; Venkataravanappa, V.; Narasimha Reddy, L.R.C.; Shankarappa, K.S. Survey, detection, characterization of papaya ringspot virus from southern India and management of papaya ringspot disease. *Pathogens* **2023**, *12*, 824. [CrossRef] [PubMed]
26. Devi, M.; Ghatani, K. The use of coconut in rituals and food preparations in India: A review. *J. Ethn. Foods* **2022**, *9*, 37. [CrossRef]
27. Gothandan, G.; Selvi, B.S.; Velmurugan, S.; Meenakshi, P. Performance of coriander genotypes (*Coriandrum sativum* L.) for herbage yield and quality parameters under coconut during Kharif season. *Pharma Innov. J.* **2021**, *10*, 1324–1328.
28. Barman, D.; Kar, G.; Mitra, S.; Alam, N.; Manna, T.; Kheroar, S.; Zaman, A.; Jena, S. Dry and wet spell probability analysis using Markov chain model for planning jute-based cropping systems in eastern India. *Indian J. Soil Conserv.* **2024**, *52*, 1–8.
29. Kar, G.; Blaise, D.; Srivastava, T.; Kar, C.S.; Verma, P.; Reddy, A.; Singh, P. Commercial Crops (Jute, Cotton and Sugarcane). In *Trajectory of 75 Years of Indian Agriculture After Independence*; Springer: Berlin/Heidelberg, Germany, 2023; pp. 331–362.
30. VA, N.A.; Panakaje, N. Growth of Coffee Cultivation, Consumption & Production in India—An Overview. *Int. J. Case Stud. Bus. IT Educ. (IJCSBE)* **2022**, *6*, 755–770.
31. Kamal, B.; Acharya, B.; Srivastava, A.; Pandey, M. Effect of different altitudes in qualitative and quantitative attributes of green coffee beans (*Coffea arabica*) in Nepal. *Int. J. Hortic. Agric. Food Sci.* **2021**, *5*, 1–7.

32. Khanal, S.; Fulton, J.; Klopfenstein, A.; Douridas, N.; Shearer, S. Integration of high resolution remotely sensed data and machine learning techniques for spatial prediction of soil properties and corn yield. *Comput. Electron. Agric.* **2018**, *153*, 213–225. [CrossRef]
33. Olson, D.; Anderson, J. Review on unmanned aerial vehicles, remote sensors, imagery processing, and their applications in agriculture. *Agron. J.* **2021**, *113*, 971–992. [CrossRef]
34. Weatherington-Rice, J.; Aller, L.; Bennett, T.; Hunter, D.; Fausey, N.; Allred, B.; Tornes, L.; Brockman, C.; Angle, M.; Hall, G.; et al. The use of long term, large scale pumping data to determine the regional recharge rates through fractured till and lacustrine materials. In *Groundwater 2000*; CRC Press: Boca Raton, FL, USA, 2020; pp. 515–516.
35. Plaza, J.; Criado, M.; Sánchez, N.; Pérez-Sánchez, R.; Palacios, C.; Charfolé, F. UAV multispectral imaging potential to monitor and predict agronomic characteristics of different forage associations. *Agronomy* **2021**, *11*, 1697. [CrossRef]
36. Liu, J.; Yang, K.; Tariq, A.; Lu, L.; Soufan, W.; El Sabagh, A. Interaction of climate, topography and soil properties with cropland and cropping pattern using remote sensing data and machine learning methods. *Egypt. J. Remote Sens. Space Sci.* **2023**, *26*, 415–426. [CrossRef]
37. Zhang, L.; Wu, Z.; Sun, X.; Yan, J.; Sun, Y.; Liu, P.; Chen, J. Mapping topsoil total nitrogen using random forest and modified regression Kriging in Agricultural Areas of Central China. *Plants* **2023**, *12*, 1464. [CrossRef] [PubMed]
38. Raja, S.; Sawicka, B.; Stamenkovic, Z.; Mariammal, G. Crop prediction based on characteristics of the agricultural environment using various feature selection techniques and classifiers. *IEEE Access* **2022**, *10*, 23625–23641. [CrossRef]
39. Rahman, S.A.Z.; Mitra, K.C.; Islam, S.M. Soil classification using machine learning methods and crop suggestion based on soil series. In Proceedings of the 2018 21st International Conference of Computer and Information Technology (ICCIT), Dhaka, Bangladesh, 21–23 December 2018; pp. 1–4.
40. Maesa, S.; Basri, H.; Sugianto, S. Classification of Rice Growth Phases Using the K-Nearest Neighbor Algorithm in the Irrigation area of Seulimeum Sub District, Aceh, Indonesia. In *IOP Conference Series: Earth and Environmental Science*; IOP Publishing: Bristol, UK, 2024; Volume 1297, p. 012006.
41. Wu, C.; Wang, M.; Wang, C.; Zhao, X.; Liu, Y.; Masoudi, A.; Yu, Z.; Liu, J. Reed biochar improved the soil functioning and bacterial interactions: A bagging experiment using the plantation forest soil (*Fraxinus chinensis*) in the Xiong'an new area, China. *J. Clean. Prod.* **2023**, *410*, 137316. [CrossRef]
42. Devyatkin, D.A. Estimation of vegetation indices with Random Kernel Forests. *IEEE Access* **2023**, *11*, 29500–29509. [CrossRef]
43. Yadav, J.; Chopra, S.; Vijayalakshmi, M. Soil analysis and crop fertility prediction using machine learning. *Mach. Learn.* **2021**, *8*, 41–49.
44. Gunawan, G.; Sya'bani, A.Z.; Anandianshka, S. Expert system for diagnosing diseases in corn plants using the navies bayes method. *J. Mantik* **2024**, *8*, 849–859. [CrossRef]
45. Biazar, S.M.; Shehadeh, H.A.; Ghorbani, M.A.; Golmohammadi, G.; Saha, A. Soil temperature forecasting using a hybrid artificial neural network in florida subtropical Grazinglands agro-ecosystems. *Sci. Rep.* **2024**, *14*, 1535. [CrossRef]
46. Qiu, J.; Tabasi, E.; Hammoud, A.; Benjeddou, O.; Zarei, M.; Khordehbinan, M.W. Determining the fracture stiffness of modified Hot and Warm Mix Asphalt using semi-circular bending (SCB) geometry. *Theor. Appl. Fract. Mech.* **2024**, *129*, 104237. [CrossRef]
47. Ebrahimzadeh, E.; Fayaz, F.; Rajabion, L.; Seraji, M.; Aflaki, F.; Hammoud, A.; Taghizadeh, Z.; Asgarinejad, M.; Soltanian-Zadeh, H. Machine learning approaches and non-linear processing of extracted components in frontal region to predict rTMS treatment response in major depressive disorder. *Front. Syst. Neurosci.* **2023**, *17*, 919977. [CrossRef]
48. Wickramasinghe, L.; Weliwatta, R.; Ekanayake, P.; Jayasinghe, J. Modeling the relationship between rice yield and climate variables using statistical and machine learning techniques. *J. Math.* **2021**, *2021*, 6646126. [CrossRef]
49. Chandrasiri, C.K.; Tsusaka, T.W.; Ho, T.D.; Zulfiqar, F.; Datta, A. Impacts of climate change on paddy yields in different climatic zones of Sri Lanka: A panel data approach. *Asia-Pac. J. Reg. Sci.* **2023**, *7*, 455–489. [CrossRef]
50. Gatera, A.; Kuradusenge, M.; Bajpai, G.; Mikeka, C.; Shrivastava, S. Comparison of random forest and support vector machine regression models for forecasting road accidents. *Sci. Afr.* **2023**, *21*, e01739. [CrossRef]
51. Sahin, G.; Isik, G.; van Sark, W.G. Predictive modeling of PV solar power plant efficiency considering weather conditions: A comparative analysis of artificial neural networks and multiple linear regression. *Energy Rep.* **2023**, *10*, 2837–2849. [CrossRef]
52. De Caro, D.; Ippolito, M.; Cannarozzo, M.; Provenzano, G.; Ciraolo, G. Assessing the performance of the Gaussian Process Regression algorithm to fill gaps in the time-series of daily actual evapotranspiration of different crops in temperate and continental zones using ground and remotely sensed data. *Agric. Water Manag.* **2023**, *290*, 108596. [CrossRef]
53. Yang, H.; Gebhardt, W.; Ororbia, A.G.; Desell, T. Predicted Embedding Power Regression for Large-Scale Out-of-Distribution Detection. *arXiv* **2023**, arXiv:2303.04115.
54. Diaz-Gonzalez, F.A.; Vuelvas, J.; Correa, C.A.; Vallejo, V.E.; Patino, D. Machine learning and remote sensing techniques applied to estimate soil indicators–review. *Ecol. Indic.* **2022**, *135*, 108517. [CrossRef]
55. García-López, A.M.; Delgado, A.; Anjos, O.; Horta, C. Digestate not only affects nutrient availability but also soil quality indicators. *Agronomy* **2023**, *13*, 1308. [CrossRef]
56. Burdett, H.; Wellen, C. Statistical and machine learning methods for crop yield prediction in the context of precision agriculture. *Precis. Agric.* **2022**, *23*, 1553–1574. [CrossRef]
57. Osei, E.; Jafri, S.H.; Saleh, A.; Gassman, P.W.; Gallego, O. Simulated climate change impacts on corn and soybean yields in Buchanan County, Iowa. *Agriculture* **2023**, *13*, 268. [CrossRef]

58. Yang, Z.; Yang, W.; Yi, J.; Liu, R. WH-DETR: An Efficient Network Architecture for Wheat Spike Detection in Complex Backgrounds. *Agriculture* **2024**, *14*, 961. [CrossRef]
59. Tang, L.; Yi, J.; Li, X. Improved multi-scale inverse bottleneck residual network based on triplet parallel attention for apple leaf disease identification. *J. Integr. Agric.* **2024**, *23*, 901–922. [CrossRef]
60. Sidhu, B.S.; Mehrabi, Z.; Ramankutty, N.; Kandlikar, M. How can machine learning help in understanding the impact of climate change on crop yields? *Environ. Res. Lett.* **2023**, *18*, 024008. [CrossRef]
61. Li, L.; Zhang, Y.; Wang, B.; Feng, P.; He, Q.; Shi, Y.; Liu, K.; Harrison, M.T.; Li Liu, D.; Yao, N.; et al. Integrating machine learning and environmental variables to constrain uncertainty in crop yield change projections under climate change. *Eur. J. Agron.* **2023**, *149*, 126917. [CrossRef]
62. Holden, T.D. Existence and uniqueness of solutions to dynamic models with occasionally binding constraints. *Rev. Econ. Stat.* **2023**, *105*, 1481–1499. [CrossRef]
63. Li, L.; Wang, B.; Feng, P.; Jägermeyr, J.; Asseng, S.; Müller, C.; Macadam, I.; Liu, D.L.; Waters, C.; Zhang, Y.; et al. The optimization of model ensemble composition and size can enhance the robustness of crop yield projections. *Commun. Earth Environ.* **2023**, *4*, 362. [CrossRef]
64. Motia, S.; Reddy, S. Exploration of machine learning methods for prediction and assessment of soil properties for agricultural soil management: A quantitative evaluation. In *Journal of Physics: Conference Series*; IOP Publishing: Bristol, UK, 2021; Volume 1950, p. 012037.
65. Ghimire, R.; Thapa, V.R.; Acosta-Martinez, V.; Schipanski, M.; Slaughter, L.C.; Fonte, S.J.; Shukla, M.K.; Bista, P.; Angadi, S.V.; Mikha, M.M.; et al. Soil health assessment and management framework for water-limited environments: Examples from the Great Plains of the USA. *Soil Syst.* **2023**, *7*, 22. [CrossRef]
66. Jhajharia, K.; Mathur, P. Prediction of crop yield using satellite vegetation indices combined with machine learning approaches. *Adv. Space Res.* **2023**, *72*, 3998–4007. [CrossRef]
67. Pokhariyal, S.; Patel, N.; Govind, A. Machine Learning-Driven Remote Sensing Applications for Agriculture in India—A Systematic Review. *Agronomy* **2023**, *13*, 2302. [CrossRef]
68. Malik, P.; Sengupta, S.; Jadon, J.S. Comparative analysis of soil properties to predict fertility and crop yield using machine learning algorithms. In Proceedings of the 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 28–29 January 2021; pp. 1004–1007.
69. United States Department of Agriculture. *India Production Summary*; United States Department of Agriculture: Washington, DC, USA, 2024.
70. Mahale, Y.; Khan, N.; Kulkarni, K.; Wagle, S.A.; Pareek, P.; Kotecha, K.; Choudhury, T.; Sharma, A. Crop recommendation and forecasting system for Maharashtra using machine learning with LSTM: A novel expectation-maximization technique. *Discov. Sustain.* **2024**, *5*, 134. [CrossRef]
71. Staniak, M.; Szpunar-Krok, E.; Kocira, A. Responses of soybean to selected abiotic stresses—Photoperiod, temperature and water. *Agriculture* **2023**, *13*, 146. [CrossRef]
72. Rehan, M.; Al-Turki, A.; Abdelmageed, A.H.; Abdelhameid, N.M.; Omar, A.F. Performance of plant-growth-promoting rhizobacteria (PGPR) isolated from sandy soil on growth of tomato (*Solanum lycopersicum* L.). *Plants* **2023**, *12*, 1588. [CrossRef] [PubMed]
73. Ratke, R.F.; Zuffo, A.M.; Steiner, F.; Aguilera, J.G.; de Godoy, M.L.; Gava, R.; de Oliveira, J.T.; Filho, T.A.d.S.; Viana, P.R.N.; Ratke, L.P.T.; et al. Can soil moisture and crop production be influenced by different cropping systems? *AgriEngineering* **2023**, *5*, 112–126. [CrossRef]
74. Wang, Y.J.; Yang, T.; Kim, H.J. pH dynamics in aquaponic systems: Implications for plant and fish crop productivity and yield. *Sustainability* **2023**, *15*, 7137. [CrossRef]
75. Zhang, H.; Wang, L.; Fu, W.; Xu, C.; Zhang, H.; Xu, X.; Ma, H.; Wang, J.; Zhang, Y. Soil Acidification Can Be Improved under Different Long-Term Fertilization Regimes in a Sweetpotato–Wheat Rotation System. *Plants* **2024**, *13*, 1740. [CrossRef]
76. Chen, M.; Shen, Y.; Wang, H.; Cheng, X.; Luo, Y. Analysis of the Rainfall Pattern and Rainfall Utilization Efficiency during the Growth Period of Paddy Rice. *Agronomy* **2024**, *14*, 1332. [CrossRef]
77. Hlisnikovsk̀y, L.; Menšík, L.; Roman, M.; Kunzová, E. The Evaluation of a Long-Term Experiment on the Relationships between Weather, Nitrogen Fertilization, Preceding Crop, and Winter Wheat Grain Yield on Cambisol. *Plants* **2024**, *13*, 802. [CrossRef]
78. Zeng, J.; Liu, F.; Zhu, Y.; Li, J.; Ruan, Y.; Quan, X.; Wang, C.; Wang, X. Degree of shade tolerance shapes seasonality of chlorophyll, nitrogen and phosphorus levels of trees and herbs in a temperate deciduous forest. *J. For. Res.* **2024**, *35*, 72. [CrossRef]
79. Islam, M.; Siddique, K.H.; Padhye, L.P.; Pang, J.; Solaiman, Z.M.; Hou, D.; Srinivasarao, C.; Zhang, T.; Chandana, P.; Venu, N.; et al. A critical review of soil phosphorus dynamics and biogeochemical processes for unlocking soil phosphorus reserves. *Adv. Agron.* **2024**, *185*, 153–249.
80. Gonzalez-Porras, C.V.; Teixeira, G.C.M.; Prado, R.d.M.; Ferreira, P.M.; Palaretti, L.F.; Oliveira, K.S. Silicon via fertigation with and without potassium application, improve physiological aspects of common beans cultivated under three water regimes in field. *Sci. Rep.* **2024**, *14*, 2051. [CrossRef]
81. Dabba, A.; Tari, A.; Meftali, S. A novel grey wolf optimization algorithm based on geometric transformations for gene selection and cancer classification. *J. Supercomput.* **2024**, *80*, 4808–4840. [CrossRef]

82. Tripathi, A.; Rani, P. An improved mser using grid search based pca and ensemble voting technique. *Multimed. Tools Appl.* **2024**, *83*, 80497–80522. [CrossRef]
83. Wang, X.; Jin, Y.; Schmitt, S.; Olhofer, M. Recent advances in Bayesian optimization. *ACM Comput. Surv.* **2023**, *55*, 1–36. [CrossRef]
84. Ramadevi, B.; Kasi, V.R.; Bingi, K. Fractional ordering of activation functions for neural networks: A case study on Texas wind turbine. *Eng. Appl. Artif. Intell.* **2024**, *127*, 107308. [CrossRef]
85. Blekos, K.; Brand, D.; Ceschini, A.; Chou, C.H.; Li, R.H.; Pandya, K.; Summer, A. A review on quantum approximate optimization algorithm and its variants. *Phys. Rep.* **2024**, *1068*, 1–66. [CrossRef]
86. Alhijawi, B.; Awajan, A. Genetic algorithms: Theory, genetic operators, solutions, and applications. *Evol. Intell.* **2024**, *17*, 1245–1256. [CrossRef]
87. Ayan, E. Genetic algorithm-based hyperparameter optimization for convolutional neural networks in the classification of crop pests. *Arab. J. Sci. Eng.* **2024**, *49*, 3079–3093. [CrossRef]
88. Ruiz-Gonzalez, R.; Gomez-Gil, J.; Gomez-Gil, F.J.; Martínez-Martínez, V. An SVM-based classifier for estimating the state of various rotating components in agro-industrial machinery with a vibration signal acquired from a single point on the machine chassis. *Sensors* **2014**, *14*, 20713–20735. [CrossRef]
89. Silva, J.V.; van Heerwaarden, J.; Reidsma, P.; Laborte, A.G.; Tesfaye, K.; van Ittersum, M.K. Big data, small explanatory and predictive power: Lessons from random forest modeling of on-farm yield variability and implications for data-driven agronomy. *Field Crop. Res.* **2023**, *302*, 109063. [CrossRef]
90. Beltran-Royo, C.; Llopis-Ibor, L.; Pantrigo, J.J.; Ramírez, I. DC Neural Networks avoid overfitting in one-dimensional nonlinear regression. *Knowl.-Based Syst.* **2024**, *283*, 111154. [CrossRef]
91. Makumbura, R.K.; Mampitiya, L.; Rathnayake, N.; Meddage, D.; Henna, S.; Dang, T.L.; Hoshino, Y.; Rathnayake, U. Advancing water quality assessment and prediction using machine learning models, coupled with explainable artificial intelligence (XAI) techniques like shapley additive explanations (SHAP) for interpreting the black-box nature. *Results Eng.* **2024**, *23*, 102831. [CrossRef]
92. Berger, T. Agent-based spatial models applied to agriculture: A simulation tool for technology diffusion, resource use changes and policy analysis. *Agric. Econ.* **2001**, *25*, 245–260.
93. Jones, J.W.; Antle, J.M.; Basso, B.; Boote, K.J.; Conant, R.T.; Foster, I.; Godfray, H.C.J.; Herrero, M.; Howitt, R.E.; Janssen, S.; et al. Brief history of agricultural systems modeling. *Agric. Syst.* **2017**, *155*, 240–254. [CrossRef] [PubMed]
94. Tenorio, F.A.; Edreira, J.I.R.; Monzon, J.P.; Aramburu-Merlos, F.; Dobermann, A.; Gruere, A.; Brihet, J.M.; Gayo, S.; Conley, S.; Mourtzinis, S.; et al. Filling the agronomic data gap through a minimum data collection approach. *Field Crop. Res.* **2024**, *308*, 109278. [CrossRef] [PubMed]
95. Brożek, B.; Furman, M.; Jakubiec, M.; Kucharzyk, B. The black box problem revisited. Real and imaginary challenges for automated legal decision making. *Artif. Intell. Law* **2024**, *32*, 427–440. [CrossRef]
96. Khaki, S.; Wang, L. Crop yield prediction using deep neural networks. *Front. Plant Sci.* **2019**, *10*, 621. [CrossRef]
97. Vadez, V.; Grondin, A.; Chenu, K.; Henry, A.; Laplaze, L.; Millet, E.J.; Carminati, A. Crop traits and production under drought. *Nat. Rev. Earth Environ.* **2024**, *5*, 211–225. [CrossRef]
98. Naik, A.; Jogi, M.; Shreenivas, B. Assessing the impact of climate change on global crop yields and farming practices. *Arch. Curr. Res. Int.* **2024**, *24*, 696–712. [CrossRef]
99. Amertet Finecomess, S.; Gebresenbet, G.; Alwan, H.M. Utilizing an Internet of Things (IoT) Device, Intelligent Control Design, and Simulation for an Agricultural System. *IoT* **2024**, *5*, 58–78. [CrossRef]
100. Osinga, S.A.; Paudel, D.; Mouzakitis, S.A.; Athanasiadis, I.N. Big data in agriculture: Between opportunity and solution. *Agric. Syst.* **2022**, *195*, 103298. [CrossRef]
101. Shrestha, A.D. *Enhanced Deep Network Designs Using Mitochondrial DNA Based Genetic Algorithm and Importance Sampling*; University of Bridgeport: Bridgeport, CT, USA, 2019.
102. Kao, S.C.; Krishna, T. Gamma: Automating the hw mapping of dnn models on accelerators via genetic algorithm. In Proceedings of the 39th International Conference on Computer-Aided Design, Virtual, 2–5 November 2020; pp. 1–9.
103. Bracho-Mujica, G.; Rötter, R.P.; Haakana, M.; Palosuo, T.; Fronzek, S.; Asseng, S.; Yi, C.; Ewert, F.; Gaiser, T.; Kassie, B.; et al. Effects of changes in climatic means, variability, and agro-technologies on future wheat and maize yields at 10 sites across the globe. *Agric. For. Meteorol.* **2024**, *346*, 109887. [CrossRef]
104. Segarra, J. Satellite Imagery in Precision Agriculture. In *Digital Agriculture: A Solution for Sustainable Food and Nutritional Security*; Springer: Berlin/Heidelberg, Germany, 2024; pp. 325–340.
105. Wu, B.; Gommes, R.; Zhang, M.; Zeng, H.; Yan, N.; Zou, W.; Zheng, Y.; Zhang, N.; Chang, S.; Xing, Q.; et al. Global crop monitoring: A satellite-based hierarchical approach. *Remote Sens.* **2015**, *7*, 3907–3933. [CrossRef]
106. Keshri, R.; Vidyarthi, D.P. Energy-efficient communication-aware VM placement in cloud datacenter using hybrid ACO–GWO. *Clust. Comput.* **2024**, *27*, 13047–13074. [CrossRef]