APPLICATION OF MACHINE LEARNING FOR CROP YIELD PREDICTION CHALLENGES AND FUTURE DIRECTIONS

Abstract

In recent years, the global agricultural landscape has witnessed a transformative shift, driven by technological advancements that aim to address the pressing challenges of food security and sustainability. It is critical aspect of precision agriculture and plays a pivotal role in optimizing agricultural practices, resource allocation, and decision-making for farmers, policymakers, and stakeholders alike. Accurate crop yield prediction empowers the agricultural community to anticipate potential fluctuations in productivity, optimize resource utilization, and mitigate risks associated with climate variability, pest outbreaks, and market fluctuations. Traditionally, crop yield prediction relied on empirical methods, historical data analysis, and localized expertise. However, these conventional approaches often suffer from limited accuracy and generalizability, making them less suitable for addressing the complexities of modern agriculture. In recent years, deep learning made revolution in by harnessing the power of neural networks to learn intricate patterns and relationships from vast and diverse datasets. The ability of deep learning models to process multimodal information, such as remote sensing data, climate records, and historical yield statistics, has opened new horizons in predicting crop yields at both regional and global scales. The objective of this chapter is to offer a thorough exploration of the utilization of deep learning methodologies. The emphasis is placed on addressing the obstacles encountered in this evolving domain and outlining potential paths for future advancements.

Keywords: crop yield prediction, machine learning, deep learning, regression models, time series forecasting.

Authors

Shalini Bhadola
Research Scholar
Deenbandhu Chhotu Ram University of Science And Technology
Murthal Sonipat (Haryana)

Dr. Kavita Rathi
Associate Professor
Deenbandhu Chhotu Ram University of Science And Technology
Murthal Sonipat (Haryana)

I. INTRODUCTION

Given the vital role that agriculture plays in providing a significant portion of the world's food supply, the issue of food shortage persists in many countries, particularly with the ever-growing population [1]. The challenges posed by population growth, unpredictable weather patterns, soil degradation, and the need to adopt climate-resilient agricultural practices demand effective crop growth and production strategies that can meet the increasing demand for food in a timely and dependable manner [2]. Ensuring sustainable agricultural food production is also critical in this context. The projected population growth in conjunction with dwindling resource availability has brought forth the concept of a 'peak society' [3], Figure 1, underscoring the pressing need to develop crops suitable for low-input systems and enhance resource management. As the global population and food demands are anticipated to rise while resources diminish, there is a critical urgency to address these challenges

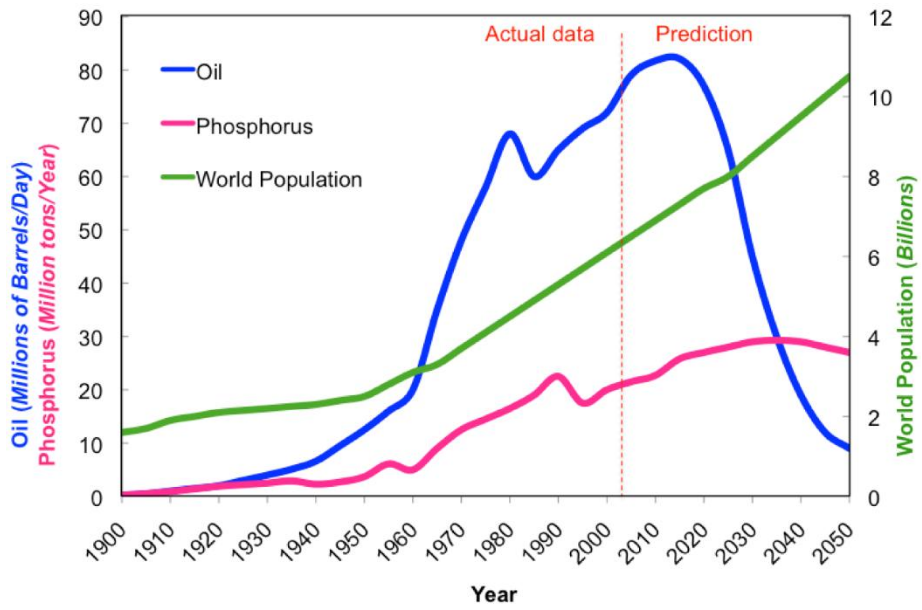

Figure 1: The projected population growth in conjunction with dwindling resource availability. [4]

The increasing population poses significant challenges to the field of agriculture, affecting both food production and sustainability. As the global population grows, the demand for food increases, putting pressure on agricultural systems to produce more food to meet these needs. This immediate challenge requires increased crop production and improved efficiency. The expanding population leads to urbanization and infrastructure development, which results in reduced agricultural land availability. This trend makes it difficult to find suitable areas for farming, leading to the degradation of arable land. More people require more water for drinking and agriculture. As a result, water resources come under strain, and overuse can lead to the depletion of groundwater and freshwater sources, affecting irrigation and crop growth.

Table 1: Key indicators related to population growth, food production, agricultural water usage, and fertilizer and pesticide usage over the years 1960, 2000, and projected for 2050 {4}.

Year	**1960**	**2000**	**2050**
Population (billions)	3	6	8.7–10
Food production (Mt)	$1.8e^9$	$3.5e^9$	$6.5\ e^9$
Agricultural water (km−**3)**	1500	7130	12-13,500
Nitrogen fertilizer	12	88	120
Phosphorus fertilizer	11	40	55–60
Pesticide	1	3.7	10.1

The data in table 1 {4} highlights the significant growth in the global population over the past decades and the projected further increase in the coming years. This puts tremendous pressure on food production to meet the rising demands. Food production has also increased over the years, but it needs to continue growing to keep up with the population growth and ensure food security. The substantial increase in agricultural water usage reflects the intensification of agriculture to support a growing population's food needs. Managing water resources sustainably will be crucial for future agricultural practices. The rise in fertilizer and pesticide usage indicates the increasing reliance on agrochemicals to enhance crop productivity. However, it also raises concerns about the environmental and health impacts of such intensive agricultural practices.

In 1960, the global food production was around 1.8 billion metric tons. By the year 2000, it had increased to approximately 3.5 billion metric tons. Projections suggest that by 2050, food production is estimated to reach 6.5 billion metric tons. In 1960, the global agricultural water usage was 1500 cubic kilometers. By 2000, it had increased significantly to approximately 7130 cubic kilometers. Projections for 2050 show a wide range, indicating that agricultural water usage might be between 12,000 to 13,500 cubic kilometers. In 1960, the global consumption of nitrogen fertilizer was 12 million tons. By 2000, it had increased to around 88 million tons.In 1960, the global consumption of phosphorus fertilizer was 11 million tons. By 2000, it had increased to approximately 40 million tons. Projections for 2050 suggest it might be in the range of 55 to 60 million tons. In 1960, the global pesticide usage was around 1 million tons. By 2000, it had increased to approximately 3.7 million tons. Projections indicate that by 2050, pesticide usage might reach 10.1 million tons. The data underscores the challenges and importance of sustainable agricultural practices and innovative approaches to meet the food demands of a growing global population while safeguarding the environment and natural resources.

Global warming impacts weather patterns [5][6][7], leading to extreme events such as droughts, floods, and heatwaves. These unpredictable weather conditions make it challenging for farmers to plan and manage their crops effectively. Expanding agricultural practices often involve clearing natural habitats, leading to a loss of biodiversity. This reduction in biodiversity can harm ecosystems and disrupt essential pollinators and natural pest control mechanisms. Intensive agricultural practices can contribute to environmental issues like soil erosion, pollution from fertilizers and pesticides, and greenhouse gas emissions, exacerbating climate change. As population increases, the demand for agricultural inputs such as fertilizers, seeds, and machinery also rises. This can lead to increased input costs, making farming less profitable for small-scale farmers. To address these challenges, several potential solutions are being explored such as:

1. **Sustainable Farming Practices:** Encouraging and promoting sustainable agricultural practices, such as crop rotation, conservation tillage, and organic farming, can help maintain soil health, reduce environmental impact, and increase long-term productivity.

2. **Technology Adoption:** Embracing agricultural technologies, such as precision farming, IoT-based sensors, and drones, can optimize resource utilization, increase productivity, and reduce waste.

3. **Water Management:** Implementing efficient water management practices, like drip irrigation and rainwater harvesting, can conserve water resources and improve water use efficiency.

4. **Crop Breeding and GMOs:** Developing improved crop varieties through conventional breeding or genetically modified organisms (GMOs) [8] can enhance resistance to pests, diseases, and harsh environmental conditions.

5. **Agroforestry:** Integrating trees with crops and livestock can restore ecosystems, improve soil fertility, and provide additional sources of income.

6. **Crop Yield Prediction**: Crop yield prediction [9] is the process of estimating the potential harvest of crops before they are actually harvested. It involves the use of historical data, climate information, soil conditions, and various modelling techniques to forecast how much yield a particular crop may produce.

II. CROP YIELD PREDICTION

These requirements stated above underscore the significance of land assessment, crop protection, and accurate crop yield prediction on a global scale [10]. Accurate crop yield prediction is particularly crucial for policymakers to make informed decisions about export and import evaluations, thereby enhancing national food security. It serves as a vital tool for policymakers to ensure food availability and sustainability for their nations.

Crop yield prediction helps farmers make informed decisions regarding resource allocation, including water, fertilizers, and pesticides. This ensures that resources are used efficiently, reducing waste and environmental impact. With accurate yield predictions, farmers can anticipate potential shortfalls due to adverse weather conditions, pests, or diseases. They can take preventive measures or adopt alternative crops to reduce the impact of these risks. Predicting crop yields in advance enables better market planning. Farmers can estimate the supply and demand for their produce, making strategic decisions about when and where to sell their crops for the best prices. By accurately estimating crop yields, governments and organizations can proactively address potential food shortages and implement measures to ensure food security for the growing population.

Crop yield predictions provide valuable insights into which crops are likely to perform well in specific regions under prevailing conditions. This information helps farmers adapt their practices to optimize production. Data from crop yield predictions can aid researchers and policymakers in identifying areas with yield gaps and developing targeted interventions to improve agricultural productivity. The increasing population presents significant challenges to agriculture, ranging from meeting food demand to ensuring environmental sustainability. However, with the adoption of sustainable practices, technological advancements, and accurate crop yield prediction, it is possible to address these challenges and ensure food security and environmental preservation for future generations.

1. **Historical Background:** The historical background of crop yield prediction dates back to the early development of agriculture and human civilization [11]. However, the methods and techniques used for crop yield prediction have evolved significantly over time.

   - **Early History:** In ancient times, farmers relied on observations and experience to estimate crop yields. They learned to read natural signs such as changes in weather patterns, animal behavior, and plant growth stages to predict potential harvests. These traditional methods were often based on local knowledge and passed down through generations.

   - **Advancements in the 19th and 20th Centuries:** The Industrial Revolution brought significant advancements in agriculture. In the 19th century, scientific research and experimentation in agriculture began to play a more prominent role. Researchers and agronomists started collecting and analyzing data related to weather, soil, and crop yields to develop more systematic approaches to prediction. The advent of statistical analysis in the late 19th and early 20th centuries also contributed to the development of crop yield prediction models. Researchers began to apply statistical methods to historical data to establish relationships between crop yields and various influencing factors.

   - **Mid to Late 20th Century:** With the increasing availability of computing power, the field of crop yield prediction saw significant progress in the mid to late 20th century. Mathematical models, such as linear regression and time series analysis, were employed to make predictions based on historical data. These models considered factors like weather conditions, soil properties, and crop management practices. The introduction of remote sensing and satellite technology further revolutionized crop yield prediction. Satellite imagery provided valuable data on vegetation indices, land use, and weather patterns, enhancing the accuracy of predictions.

   - **Recent Developments:** In the late 20th and early 21st centuries, the rise of artificial intelligence (AI)[12] and machine learning [13] brought a new era of crop yield prediction. Machine learning algorithms, such as neural networks and support vector machines, allowed for more sophisticated analysis of complex datasets. Modern crop yield prediction models integrate a wide range of data sources, including historical and real-time weather data, soil information, satellite imagery, and crop management practices. AI-powered models can process vast amounts of data and identify intricate patterns to generate more accurate and detailed crop yield forecasts. Furthermore, the integration of Internet of things (IoT) devices [14], drones, and precision agriculture technologies has enabled farmers to collect real-time data on their fields, contributing to more precise on-farm crop yield predictions.

The historical background of crop yield prediction reflects the gradual progression from traditional observation-based methods to data-driven and technology-intensive approaches that we see today. Continued advancements in technology and data analytics will likely lead to even more sophisticated and accurate crop yield prediction models in the future.

III. CURRENT CHALLENGES OF CROP YIELD PREDICTION

Efficient Crop Yield Prediction requires collaborative efforts from researchers, policymakers, and technology developers. Improving data infrastructure, encouraging technology adoption among farmers, and refining prediction models through ongoing research and validation are essential steps to overcome these obstacles. Additionally, enhancing communication channels between stakeholders can help bridge the gap between research and practical implementation in the field. Currently following research challenges were found with Efficient Crop Yield Prediction:

1. **Data Availability and Quality:** Crop yield prediction relies heavily on diverse and high-quality data, including historical weather records, soil information, satellite imagery, and farm management practices. However, accessing reliable and comprehensive data can be a challenge, especially in developing regions or for small-scale farmers who might not have access to advanced technology.

2. **Data Integration and Standardization:** Integrating data from various sources and formats can be complex. Different data formats, scales, and units can lead to discrepancies and errors in the prediction models. Standardizing data and creating interoperable systems remain challenging.

3. **Limited Spatial Resolution:** Some satellite-based data sources might have limitations in spatial resolution, making it difficult to obtain detailed information for small or fragmented land holdings. This can impact the accuracy of predictions for specific locations.

4. **Weather Variability and Extreme Events:** Climate change has increased the frequency and intensity of extreme weather events, such as droughts, floods, and heatwaves. These unpredictable weather patterns can disrupt crop growth and make yield prediction more challenging.

5. **Pest and Disease Outbreaks:** Crop yield prediction models often do not account for sudden pest or disease outbreaks that can severely impact production. Integrating real-time data on pest and disease occurrences remains a challenge.

6. **Crop Management Practices:** Variability in crop management practices among farmers can significantly affect crop yields. Incorporating these diverse practices into prediction models requires more comprehensive and localized data.

7. **Complex Interactions and Non-Linearities:** Crop growth and yield are influenced by complex interactions between various factors, such as temperature, humidity, soil nutrients, and pests. Additionally, non-linear responses to changing conditions add complexity to prediction models.

8. **Limited Adoption of Technology:** While advanced technologies like IoT devices and drones hold the potential to improve data collection, their adoption is not widespread among all farmers. This limits the availability of real-time data for prediction models.

9. **Model Calibration and Validation:** Ensuring the accuracy and reliability of prediction models is an ongoing challenge. Proper calibration and validation of models with ground-truth data are essential to minimize errors and biases.

IV. MACHINE LEARNING & CROP YIELD PREDICTION

Machine learning constitutes a segment within the broader field of artificial intelligence (AI), concentrating on the development of algorithms and models that empower computers to learn from data, thereby making forecasts or choices devoid of explicit programming. In lieu of direct programming for particular tasks, machine learning algorithms undergo training using extensive datasets, allowing them to discern patterns and arrive at informed decisions [15][16].

ML models can analyse historical data to make predictions about future events. This is particularly valuable in areas such as weather forecasting, commodity market predictions, plant disease diagnosis and crop yield estimation. learning to handle large datasets and complex relationships makes it an efficient tool in addressing a wide range of real-world problems & driving innovation across various industries including agriculture.

Crop yield prediction relies heavily on diverse and high-quality data, including historical weather records, soil information, satellite imagery, and farm management practices. However, accessing reliable and comprehensive data can be a challenge, especially in developing regions or for small-scale farmers who might not have access to advanced technology. Integrating data from various sources and formats can be complex. Different data formats, scales, and units can lead to discrepancies and errors in the prediction models. Standardizing data and creating interoperable systems remain challenging. Some satellite-based data sources might have limitations in spatial resolution, making it difficult to obtain detailed information for small or fragmented land holdings. This can impact the accuracy of predictions for specific locations.

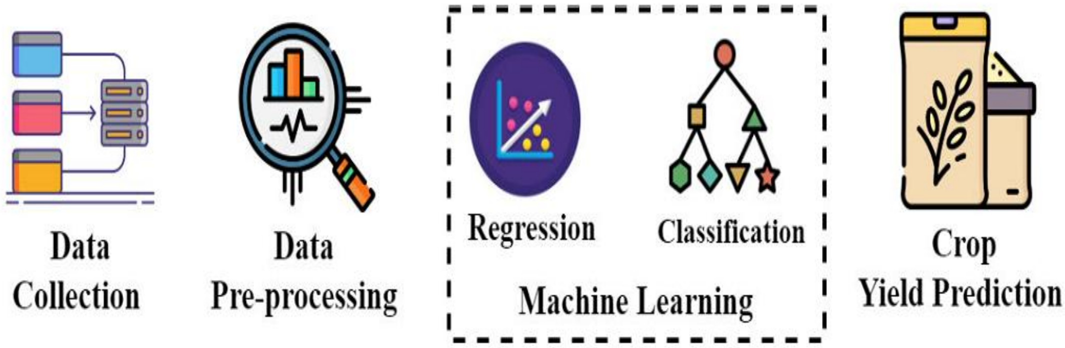

Figure 2: The generalized process of the ML crop yield estimation [17]

Climate change has increased the frequency and intensity of extreme weather events, such as droughts, floods, and heatwaves. These unpredictable weather patterns can disrupt crop growth and make yield prediction more challenging. Crop yield prediction models often do not account for sudden pest or disease outbreaks that can severely impact production. Integrating real-time data on pest and disease occurrences remains a challenge. Variability in crop management practices among farmers can significantly affect crop yields. Incorporating these diverse practices into prediction models requires more comprehensive and localized data.

Crop growth and yield are influenced by complex interactions between various factors, such as temperature, humidity, soil nutrients, and pests. Additionally, non-linear responses to changing conditions add complexity to prediction models. While advanced technologies like IoT devices and drones hold the potential to improve data collection, their adoption is not widespread among all farmers. This limits the availability of real-time data for prediction models. Ensuring the accuracy and reliability of prediction models is an ongoing challenge. Proper calibration and validation of models with ground-truth data are essential to minimize errors and biases.

The use of data from farmers and agricultural enterprises for prediction models raises ethical and privacy concerns. Balancing data accessibility and privacy protection is an ongoing challenge in the agricultural technology domain. Several types of machine learning algorithms can be applied to crop yield prediction, each with its strengths and suitability for different scenarios. Here are some of the common types of ML algorithms used for crop yield prediction [18]:

1. Regression models
2. Time series models
3. Traditional ML algorithms
4. Deep Learning Models

1. **Regression Models:** Regression models are used to predict continuous numerical values based on input features. In the context of crop yield estimation, regression models are used to predict the yield of crops (the continuous numerical output). Here are some examples of regression models commonly used for crop yield estimation:

   - **Linear Regression:** Linear regression is one of the simplest and widely used regression models [19][20]. It develops linear relation for input and the target (crop yield). For example, a linear regression model can predict yield of the crops with factors like average temp, precipitation, and nutrients.

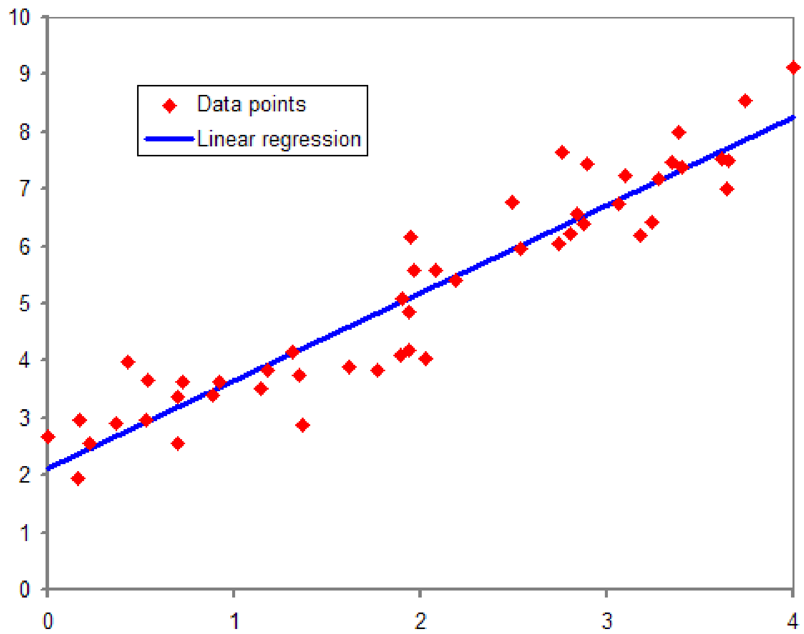

Figure 3: Example of Linear Regression model

   - **Multiple Linear Regression:** This is an extension of linear regression that considers multiple input features simultaneously[21][22]. It can handle more complex relationships between the input variables and crop yield, allowing for a more comprehensive analysis.

   - **Polynomial Regression:** Polynomial regression is used when the relationship between the input features and crop yield is best described by a polynomial equation[23]. It allows for curved fits to the data, accommodating non-linear relationships.

   - **Support Vector Regression (SVR):** SVR is a regression variant of the Support Vector Machine (SVM) algorithm [24]. It is particularly useful for cases where the data has non-linear patterns. SVR uses support vectors to determine a hyperplane that best fits the data.

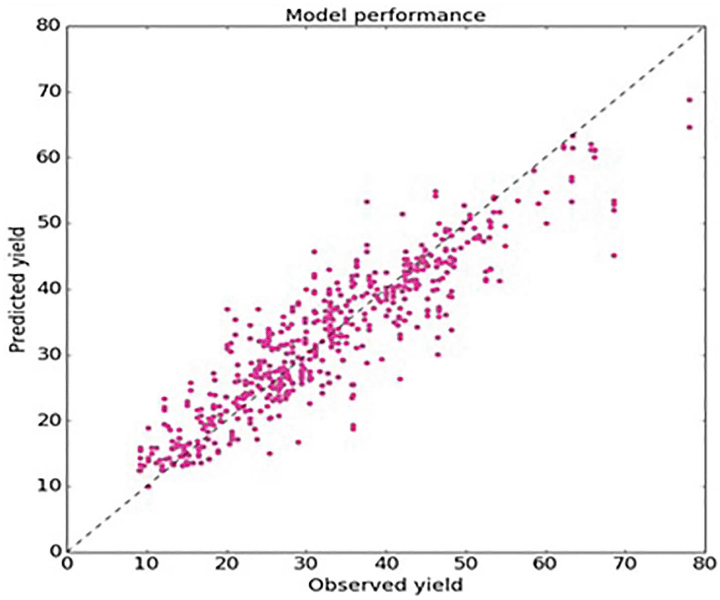

Figure 4: Application of SVR algorithm for crop yield estimation

- **Decision Tree Regression:** Decision tree regression [25] creates a tree-like structure to predict crop yield based on a series of rules and conditions. It is useful for capturing non-linear relationships and interactions between features.

  Other Approaches include ensemble learning technique builds multiple decision trees and combines their predictions to achieve higher accuracy[26]. It is robust and can handle a large number of input features. Gradient boosting regression is another ensemble learning[27] method that combines weak learners (typically decision trees) sequentially to improve model performance. It is known for its ability to capture complex patterns and achieve high accuracy.

  Neural networks can be used for regression tasks by training a network with multiple layers of interconnected neurons. Deep learning models like Multi-Layer Perceptrons (MLPs) [28] and Long Short-Term Memory (LSTM) [29] networks are examples of neural network regression models. Gaussian process regression is a probabilistic regression method that estimates the uncertainty of predictions. It is particularly useful when dealing with limited data and provides confidence intervals for yield estimates.

2. **Time Series Models:** Time series models [30] are a class of statistical models that are specifically designed to handle data collected over time, where observations are recorded at regular intervals. Time series analysis involves studying the patterns, trends, and seasonality in the data to make predictions or forecasts about future values. In the context of crop yield estimation, time series models are used to predict crop yields based on historical yield data collected over several time periods[31]. These models can take into account the temporal dependencies and patterns in crop yield variations, which are influenced by factors like weather conditions, crop management practices, and other

seasonal variations. Here are some examples of time series models commonly used for crop yield estimation:

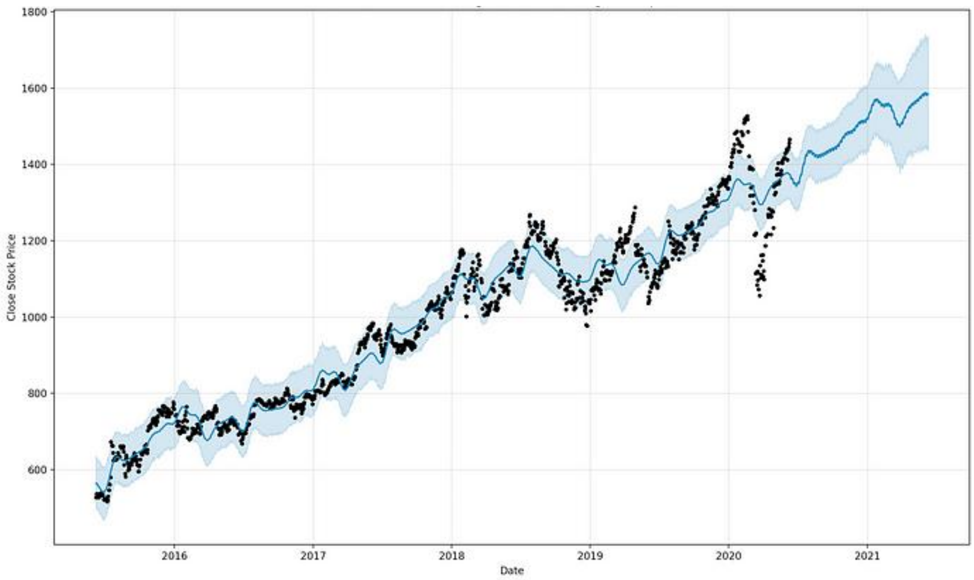

Figure 5: Time series-based forecasting using prophet method [32]

- **Autoregressive Integrated Moving Average (ARIMA):** ARIMA [33] is a popular time series model that combines autoregression, differencing, and moving average components. It is useful for capturing both short-term and long-term trends and seasonality in the data.

- **Seasonal Autoregressive Integrated Moving-Average (SARIMA):** SARIMA [34] is an extension of ARIMA that accounts for seasonal variations in the time series data. It can be effective in modeling crop yield variations that exhibit seasonal patterns.

- **Prophet:** is a time-series [35] model designed to handle strong seasonal patterns and provides a flexible and intuitive framework for forecasting crop yields.

- **Exponential Smoothing (ETS):** ETS is a family of time-series models that assigns exponentially decreasing weights to past observations. It is useful for capturing trends and seasonality in the data [36].

- **Seasonal Decomposition of Time Series (STL):** is a decomposition which separates the time series into components, including trend, seasonal, and residual components [37]. It can help identify underlying patterns in crop yield data. [38]. GRU **is s**imilar to LSTM, GRU is a type of RNN that can capture long-term dependencies in time series data. It is computationally more efficient than LSTM and can be used for crop yield estimation. **Vector Autoregression (VAR) [40]** is a time series model used when multiple time series variables influence each other. It is suitable for crop yield estimation when there are interactions between different crops or factors. **Dynamic**

Linear Models (DLM) is a flexible framework for modeling time series data with time-varying parameters. It can handle changing relationships in the data over time, making it suitable for dynamic crop yield variations [41].

3. **Traditional ML Algorithms:** Machine learning, a subfield of Artificial Intelligence (AI) dedicated to learning from data, offers a practical solution for improving crop yield prediction through the analysis of multiple factors. By harnessing machine learning (ML), we can identify patterns, uncover correlations, and extract valuable insights from datasets. To create predictive models, we train them using datasets that contain outcomes based on past experiences. During the training phase, historical data is used to determine the model's parameters, incorporating several features. In the testing phase, a portion of the historical data not used during training is employed to evaluate the model's performance. Several types of machine learning algorithms can be applied to crop yield prediction, each with its strengths and suitability for different scenarios. Here are some of the common types of ML algorithms used for crop yield prediction:

- **Decision Trees (DT):** DTs are algorithms which can work on regression and the classification tasks. In crop yield prediction, decision trees can be used to model the complex interactions between multiple factors that influence crop growth and yield [42].

- **Random Forest (RF):** RFs builds multiple trees and for predictions to achieve higher accuracy and reduce overfitting. It is well-suited for handling large and diverse datasets, making it useful for crop yield prediction [43].

- **Support Vector Machines (SVM):** SVM is a powerful algorithm for both classification and regression tasks. In crop yield prediction, SVM can be applied to estimate yields based on historical data and other relevant features [44].

- **K-Nearest Neighbors (KNN):** it is intuitive algorithm; it predicts the target variable based on the average or majority of the k-nearest data points. KNN can be applied to predict crop yields based on the yields of neighboring farms with similar characteristics [45].

- **Neural Networks:** Neural networks, particularly deep learning models, have gained popularity for complex tasks, including crop yield prediction. They can handle large amounts of data and learn intricate patterns and relationships between input features and crop yields [46].

- **Gradient Boosting Machines (GBM):** GBM is ensemble model combines weak learners for a better model. It is effective in capturing non-linear relationships and can be used for accurate crop yield prediction [47].

- **Long Short-Term Memory (LSTM) Networks:** LSTM is a type of recurrent neural network (RNN) specifically designed for sequential data. It can be useful for crop yield prediction when dealing with time-series data, such as weather patterns or crop growth stages [48].

- **Gaussian Processes:** Gaussian processes are probabilistic models that can be used for regression tasks. They provide uncertainty estimates, which can be valuable for understanding the confidence of yield predictions [49].

- **XGBoost:** XGBoost is an optimized gradient boosting library known for its high performance and efficiency. It is commonly used for both regression and classification tasks, including crop yield prediction [50].

4. **Deep Learning Models:** Deep learning falls within the realm of machine learning and revolves around employing artificial neural networks to simulate and tackle intricate issues (51). Drawing inspiration from the arrangement and operation of the human brain—where interconnected neurons carry out information processing and transmission—deep learning constructs, termed deep neural networks, comprise numerous tiers of interconnected neurons. This architecture empowers them to grasp layered portrayals of data, grasping intricate structures and associations. These models encompass multiple strata, including an initial input layer, one or more concealed strata, and a concluding output layer. Every layer incorporates neurons that handle and morph the data.

   Deep learning algorithms possess the ability to autonomously extract pertinent attributes from raw data, eliminating the necessity for manual crafting of characteristics (52). This proficiency becomes especially advantageous when contending with data sets that are intricate and high-dimensional. Profound learning models exhibit the capability to manage substantial volumes of data and adapt efficiently as data and computational resources increase. This capacity to scale renders them apt for scenarios involving extensive datasets, such as prognosticating crop yields. Profound learning models stand out in their aptitude for grasping layered portrayals of data, enabling them to apprehend and decipher intricate motifs within the input information. The realm of deep learning spans a broad spectrum of tasks, encompassing image recognition, natural language processing, and prediction of chronological sequences. This versatility positions it as a flexible strategy suited for a diverse array of applications.

   Predicting crop yields entails navigating intricate relationships among diverse factors like weather patterns, soil conditions, and agricultural practices. Deep learning models excel at comprehending and modeling these intricate interactions, thus yielding more precise prognostications. A significant advantage lies in the automatic extraction of pertinent attributes from raw data, bypassing the need for manual feature engineering [51]. This proves especially beneficial in the realm of crop yield prediction, where identifying relevant characteristics might not be immediately evident. Notably, recurrent neural networks (RNNs) and Long Short-Term Memory (LSTM) networks, which are types of deep learning models, suitably analyze time-series data. They account for the temporal aspects of weather patterns and crop growth, thereby enhancing yield forecasts.

   Given the multifaceted nature of crop yield prediction, involving data from sources such as satellite imagery, weather data, and soil information, deep learning models adeptly manage diverse data types and integrate insights from various origins [52]. These models exhibit a capacity for continuous enhancement in performance as they encounter more data. Frequent updates to the model contribute to heightened accuracy and the ability to adapt to shifting environmental conditions.

Despite the advantages of deep learning, it's essential to consider that its success depends on the availability of large and diverse datasets. Deep learning models can be computationally intensive and require substantial computational resources for training and inference. Therefore, in scenarios with limited data or computational constraints, traditional machine learning approaches may still be effective for crop yield prediction. A hybrid approach, combining both deep learning and conventional methods, might be a promising direction for achieving the best results in crop yield prediction.

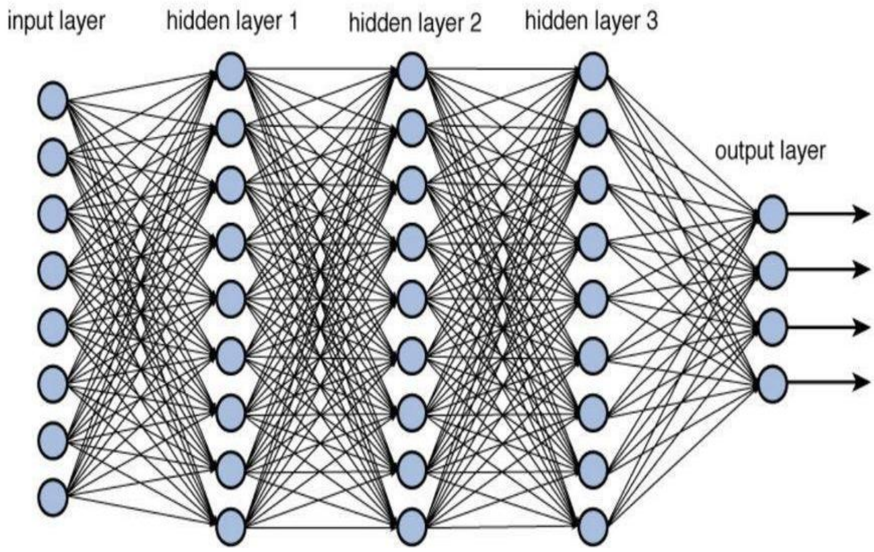

Figure 6: Architecture of deep neural network for machine learning [53]

Deep learning-based crop yield estimation methods leverage the power of neural networks to predict crop yields using historical data and relevant features. Here are a few deep learning methods commonly used for crop yield estimation:

- **Convolutional Neural Networks (CNNs) for Crop Image Analysis:** CNNs are widely used in computer vision tasks, including crop image analysis. Remote sensing techniques, such as satellite imagery, aerial photography, and drone-based imagery, provide valuable information about crop health and growth. CNNs can analyze these images to detect patterns related to crop yield and assess factors such as plant health, crop density, and pest infestations.

- **Recurrent Neural Networks (RNNs) for Time-Series Crop Yield Prediction:** RNNs are designed to handle sequential data, making them suitable for time-series crop yield prediction. They can model temporal dependencies in historical yield data and other time-varying factors like weather conditions, soil moisture, and crop management practices. Long Short-Term Memory (LSTM) networks and Gated Recurrent Units (GRUs) are popular RNN variants used in this context [54].

- **Transformer Models for Crop Yield Forecasting:** Transformer models, originally designed for natural language processing, have been adapted for time-series forecasting tasks. These models use self-attention mechanisms to capture long-range dependencies in time-series data, making them effective for crop yield forecasting. They can handle large sequences of historical yield data and incorporate other features like weather variables for accurate predictions [55].

- **Graph Neural Networks (GNNs) for Crop Yield Prediction in Agroecological Systems:** GNNs are specialized neural networks that operate on graph-structured data. In the context of crop yield estimation, GNNs can model complex interactions between different components of an agroecological system, such as soil properties, crop types, and climate variables. This approach enables a more holistic and data-driven understanding of crop yield patterns [56].

- **Autoencoders for Feature Learning in Crop Yield Prediction:** Autoencoders are unsupervised deep learning models that learn to compress and reconstruct data. They can be used for feature learning, where they encode input features into a lower-dimensional representation. This can help in reducing the dimensionality of high-dimensional agricultural data, making it more manageable for crop yield prediction models [57].

- **Ensemble Deep Learning Models for Robust Crop Yield Estimation:** Ensemble methods combine multiple deep learning models to improve predictive performance and generalization. Bagging, boosting, and stacking techniques can be applied to deep learning models, creating an ensemble that leverages the diversity of individual models to achieve more accurate crop yield predictions [58].

Table 2: Comparison of various Crop Yield estimation Methods

Aspect	**Traditional ML**	**Regression Models**	**Time Series Models**	**Deep Learning Models**
Data Type	Tabular or structured data	Tabular or structured data	Time series data	Tabular, sequential, image, etc.
Target Variable Type	Numerical or categorical	Numerical	Numerical	Numerical or categorical
Feature Engineering	Manual feature engineering	Manual feature engineering	Not applicable	May automatically learn features
Handling Temporal Dependencies	Not specialized	Not specialized	Specialized	Specialized for sequential data
Seasonality and Trends	Not specialized	Not specialized	Specialized	Can handle trends and seasonality
Data Size	Handles small to large data	Handles small to large data	Handles time series data	Handles large datasets
Interpretability	Generally interpretable	Generally interpretable	May lack interpretability	May lack interpretability

Computational Complexity	Lower complexity	Lower complexity	Higher complexity	Higher complexity
Model Flexibility and Complexity	Moderate to high flexibility	Moderate to high flexibility	Specialized	High flexibility and complexity
Applicability	Various domains	Various domains	Time series forecasting	Various domains
Performance on Time Series Data	Might not capture temporal patterns effectively	Might not capture temporal patterns effectively	Specially designed for time series data	Can capture temporal dependencies
Example Models	Decision Trees, Random Forests, SVM, etc.	Linear Regression, Polynomial Regression, SVR, etc.	ARIMA, SARIMA, Prophet, LSTM, etc.	

Table 3 performs comparison of various Crop Yield estimation Methods, It's important to note that the choice of the appropriate model depends on the specific characteristics of the data, the nature of the problem, and the desired level of interpretability and performance. In practice, a combination of different models from various categories might be employed to address the complexities and challenges in crop yield estimation or any other specific domain.

V. CONCLUSION AND FUTURE DIRECTIONS

Crop yield estimation is a critical aspect of agricultural planning and management. Accurate predictions of crop yields can aid policymakers, farmers, and other stakeholders in making informed decisions regarding food security, resource allocation, and sustainable agricultural practices. In recent years, regression techniques have emerged as powerful tools for crop yield estimation, leveraging historical yield data and relevant features to make predictions. This review chapter aims to provide an overview of the existing research on crop yield estimation using machine learning techniques and identify the strengths, limitations, and potential areas for future development in this domain.

Crop yield estimation using machine learning is a dynamic and evolving field with several potential future research directions. There are some promising areas for further investigation Such as, exploring methods to effectively integrate diverse data sources, such as satellite imagery, weather data, soil information, and crop management practices, for more accurate and robust crop yield predictions. Developing techniques to handle missing or incomplete data from various sources will be essential. Enhance crop yield prediction models to provide uncertainty estimates along with predictions. Uncertainty quantification will help policymakers and farmers make informed decisions, considering the reliability and confidence of the yield estimates. Investigate techniques for transferring knowledge from one region or crop to another to address data scarcity challenges. Domain adaptation methods can help generalize models trained on data from one region or season to be applied effectively in different settings. Develop interpretable deep learning models that can provide insights into the factors influencing crop yield predictions. Explainable AI techniques will be crucial in gaining trust and understanding the decision-making process.

Also exploring federated learning approaches to enable collaborative crop yield prediction without sharing raw data. This is particularly relevant when dealing with sensitive data or data distributed across multiple farms or organizations. Investigate the impact of climate change on crop yields and develop models that can adapt to changing environmental conditions. Understanding the long-term effects of climate change on crop production will be crucial for sustainable agriculture. Development of data augmentation techniques to generate synthetic data to supplement limited real-world datasets. This can help improve the performance of models, especially in regions or years with limited historical data. Explore advanced ensemble techniques that combine the strengths of various machine learning models and deep learning architectures to achieve more accurate and reliable crop yield predictions.

Development of online and incremental learning approaches that can continuously update crop yield prediction models with new data can also be explored. This is important for adapting to dynamic agricultural conditions and ever-changing datasets. Integrate crop yield prediction models with decision support systems that provide actionable insights and recommendations for farmers and policymakers, aiding in optimal crop management and resource allocation. Consider the influence of social and economic factors on crop yield estimation, such as market demand, trade policies, and socio-economic indicators. These future research directions hold the potential to further advance the field of crop yield estimation using machine learning and contribute to the sustainable growth and security of global agriculture.

The application of deep learning in crop yield prediction has garnered significant interest within the scientific community and agricultural stakeholders. This motivates us exploring deep learning models to improve prediction accuracy, and uncover hidden patterns in agricultural data. Moreover, as deep learning models offer the advantage of adaptability, accommodating new data streams and refine predictions over time can lead to dynamic and responsive yield forecasts. As the world faces the challenges of a growing population, changing climate, and limited arable land, harnessing the power of deep learning for crop yield prediction becomes paramount in ensuring food security and sustainable agricultural practices.

REFERENCES

[1]. Paul Paredes et al. "Soil water balance models for determining crop water and irrigation requirements and irrigation scheduling focusing on the FAO56 method and the dual Kc approach." Agricultural water management 241 (2020): 106357.
[2]. Gurdeep Singh et al. "Impact of climate change on agriculture and its mitigation strategies: A review." Sustainability 13.3 (2021): 1318.
[3]. Heinberg Richard "Peak everything: waking up to the century of declines". New Society Publishers, 2010.
[4]. Gomez-Galera et al. "Feeding future populations with nutritionally complete crops." ISB News Report (2010).
[5]. Maik Eichelbaum et al "Preparation and Real World Applications of Titania Composite Materials for Photocatalytic Surface, Air, and Water Purification: State of the Art." Inorganics 10.9 (2022): 139.
[6]. Zaval Lisa, et al. "How warm days increase belief in global warming." Nature Climate Change 4.2 (2014): 143-147.
[7]. Michael Dillon et al. "Global metabolic impacts of recent climate warming." Nature 467.7316 (2010): 704-706.

[8]. Farid Ahmed et al. "Detection of genetically modified organisms in foods." TRENDS in Biotechnology 20.5 (2002): 215-223.
[9]. Van Klompenburg et al. "Crop yield prediction using machine learning: A systematic literature review." Computers and Electronics in Agriculture 177 (2020): 105709.
[10]. Gopal Bhargavi et al. "A novel approach for efficient crop yield prediction." Computers and Electronics in Agriculture 165 (2019): 104968.
[11]. Federico Giovanni "Feeding the world: an economic history of agriculture", 1800-2000. Princeton University Press, 2010.
[12]. Tauger Mark et al. Agriculture in world history. Routledge, 2010.
[13]. Rashid Mamunur et al. "A comprehensive review of crop yield prediction using machine learning approaches with special emphasis on palm oil yield prediction." IEEE access 9 (2021): 63406-63439.
[14]. Reshma et al. "IoT based classification techniques for soil content analysis and crop yield prediction." 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC). IEEE, 2020.
[15]. Droesch Andrew et al. "Machine learning methods for crop yield prediction and climate change impact assessment in agriculture." Environmental Research Letters 13.11 (2018): 114003.
[16]. Chlingaryan Anna et al. "Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review." Computers and electronics in agriculture 151 (2018): 61-69.
[17]. Rashid Mamunur et al. "A comprehensive review of crop yield prediction using machine learning approaches with special emphasis on palm oil yield prediction." IEEE access 9 (2021): 63406-63439.
[18]. Verma Ruchi. "Crop Analysis and Prediction." 2022 5th International Conference on Multimedia, Signal Processing and Communication Technologies (IMPACT). IEEE, 2022.
[19]. Sellam Poovammal et al. "Prediction of crop yield using regression analysis." Indian Journal of Science and Technology 9.38 (2016): 1-5.
[20]. Shastry Aditya et al. "Prediction of crop yield using regression techniques." International Journal of Soft Computing 12.2 (2017): 96-102.
[21]. Gopal Bhargavi et al. "A novel approach for efficient crop yield prediction." Computers and Electronics in Agriculture 165 (2019): 104968.
[22]. Rahimov Nodir et al."The application of multiple linear regression algorithm and python for crop yield prediction in agriculture." Harvard educational and scientific review 2.1 (2022).
[23]. Shah Ayush, et al. "Smart farming system: Crop yield prediction using regression techniques." Proceedings of International Conference on Wireless Communication: ICWiCom 2017. Springer Singapore, 2018.
[24]. Sonai Muthu et al. "An intelligent decision support system for crop yield prediction using hybrid machine learning algorithms." F1000Research 10 (2021).
[25]. Nigam Aruvansh, et al. "Crop yield prediction using machine learning algorithms." 2019 Fifth International Conference on Image Information Processing (ICIIP). IEEE, 2019.
[26]. Kamath Pallavi, et al. "Crop yield forecasting using data mining." Global Transitions Proceedings 2.2 (2021): 402-407.
[27]. Balakrishnan Narayanan et al."Crop production-ensemble machine learning model for prediction." International Journal of Computer Science and Software Engineering 5.7 (2016): 148.
[28]. Ahmed Shakeel. "A Software Framework for Predicting the Maize Yield Using Modified Multi-Layer Perceptron." Sustainability 15.4 (2023): 3017.
[29]. Gavahi Keyhan et al."DeepYield: A combined convolutional neural network with long short-term memory for crop yield forecasting." Expert Systems with Applications 184 (2021): 115511.
[30]. Tierney Graham, et al. "Multivariate Bayesian dynamic modeling for causal prediction." arXiv preprint arXiv:2302.03200 (2023).
[31]. Reddy Chandra et al. "An applied time series forecasting model for yield prediction of agricultural crop." Soft Computing and Signal Processing: Proceedings of 2nd ICSCSP 2019 2. Springer Singapore, 2020.
[32]. Lim Bryan et al. "Time-series forecasting with deep learning: a survey." Philosophical Transactions of the Royal Society A 379.2194 (2021): 20200209.
[33]. Bang Shivam, et al. "Fuzzy Logic based Crop Yield Prediction using Temperature and Rainfall parameters predicted through ARMA, SARIMA, and ARMAX models." 2019 Twelfth international conference on contemporary computing (IC3). IEEE, 2019.
[34]. Desai Mittal, et al. "Time Series Prediction of Wheat Crop based on FB Prophet Forecast Framework." ITM Web of Conferences. Vol. 53. EDP Sciences, 2023.

[35]. Akram Muhammad, et al. "Forecast accuracy of exponential smoothing methods applied to four major crops of Pakistan." Journal of Stat Theory and Application (2015).
[36]. Theodosiou Marina et al. "Forecasting monthly and quarterly time series using STL decomposition." International Journal of Forecasting 27.4 (2011): 1178-1195.
[37]. Park Soo-Hwan et al. "Development of a soil moisture prediction model based on recurrent neural network long short-term memory (RNN-LSTM) in soybean cultivation." Sensors 23.4 (2023): 1976.
[38]. Jin Xue-Bo, et al. "Hybrid deep learning predictor for smart agriculture sensing based on empirical mode decomposition and gated recurrent unit group model." Sensors 20.5 (2020): 1334.
[39]. Anetor Friday, et al. "Credit supply and agricultural production in Nigeria: a vector autoregressive (VAR) approach." Journal of Economics and Sustainable Development 7.2 (2016).
[40]. Akram Akram, et al. "New approach to forecasting agro-based statistical models." Journal of Statistical Theory and Applications (2016): 387-399.
[41]. Keerthana Mummaleti, et al. "An ensemble algorithm for crop yield prediction." 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV). IEEE, 2021.
[42]. Jeong Jig Han, et al. "Random forests for global and regional crop yield predictions." PloS one 11.6 (2016): e0156571.
[43]. Gandhi Niketa, et al. "Rice crop yield prediction in India using support vector machines." 2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE). IEEE, 2016.
[44]. Paul Monali et al."Analysis of soil behaviour and prediction of crop yield using data mining approach." 2015 International Conference on Computational Intelligence and Communication Networks (CICN). IEEE, 2015.
[45]. Panda Sudhanshu et al. "Application of vegetation indices for agricultural crop yield prediction using neural network techniques." Remote sensing 2.3 (2010): 673-696.
[46]. Konstantinov Andrei et al. "Interpretable machine learning with an ensemble of gradient boosting machines." Knowledge-Based Systems 222 (2021): 106993.
[47]. Xiao Qingxin, et al. "Prediction of crop pests and diseases in cotton by long short term memory network." Intelligent Computing Theories and Application: 14th International Conference, ICIC 2018, Wuhan, China, August 15-18, 2018, Proceedings, Part II 14. Springer International Publishing, 2018.
[48]. Gustau Camps-Valls et al. "Crop yield estimation and interpretability with Gaussian processes." IEEE Geoscience and Remote Sensing Letters 18.12 (2020): 2043-2047.
[49]. Mariadas, Dennis AL, et al. "Extreme Gradient Boosting (XGBoost) Regressor and Shapley Additive Explanation for Crop Yield Prediction in Agriculture." 2022 12th International Conference on Computer and Knowledge Engineering (ICCKE). IEEE, 2022.
[50]. LeCun Hinton et al. "others,“Deep learning. nature 521 (7553), 436-444,” Google Sch." Google Sch. Cross Ref Cross Ref (2015).
[51]. Mathew Amitha et al."Deep learning techniques: an overview." Advanced Machine Learning Technologies and Applications: Proceedings of AMLTA 2020 (2021): 599-608.
[52]. Agarwal Sonal et al. "A hybrid approach for crop yield prediction using machine learning and deep learning algorithms." Journal of Physics: Conference Series. Vol. 1714. No. 1. IOP Publishing, 2021.
[53]. Nishant Potnuru et al. "Crop yield prediction based on Indian agriculture using machine learning." 2020 International Conference for Emerging Technology (INCET). IEEE, 2020.
[54]. Agarwal Aakash. "Study of machine learning algorithms for potential stock trading strategy frameworks." International Journal of Financial, Accounting, and Management 3.3 (2021): 275-287.
[55]. Lee Sue Han et al. "Deep-plant: Plant identification with convolutional neural networks." 2015 IEEE international conference on image processing (ICIP). IEEE, 2015.
[56]. Kurumatani Koichi. "Time series forecasting of agricultural product prices based on recurrent neural networks and its evaluation method." SN Applied Sciences 2.8 (2020): 1434.
[57]. Bi Luning, et al. "A transformer-based approach for early prediction of soybean yield using time-series images." Frontiers in Plant Science 14 (2023): 1173036.
[58]. Seireg Hayam et al. "Ensemble machine learning techniques using computer simulation data for wild blueberry yield prediction." IEEE Access 10 (2022): 64671-64687.