



Crop yield prediction algorithm (CYPA) in precision agriculture based on IoT techniques and climate changes

Fatma M. Talaat¹

Received: 27 January 2023 / Accepted: 17 April 2023
© The Author(s) 2023

Abstract

Agriculture faces a significant challenge in predicting crop yields, a critical aspect of decision-making at international, regional, and local levels. Crop yield prediction utilizes soil, climatic, environmental, and crop traits extracted via decision support algorithms. This paper presents a novel approach, the Crop Yield Prediction Algorithm (CYPA), utilizing IoT techniques in precision agriculture. Crop yield simulations simplify the comprehension of cumulative impacts of field variables such as water and nutrient deficits, pests, and illnesses during the growing season. Big data databases accommodate multiple characteristics indefinitely in time and space and can aid in the analysis of meteorology, technology, soils, and plant species characterization. The proposed CYPA incorporates climate, weather, agricultural yield, and chemical data to facilitate the anticipation of annual crop yields by policymakers and farmers in their country. The study trains and verifies five models using optimal hyper-parameter settings for each machine learning technique. The DecisionTreeRegressor achieved a score of 0.9814, RandomForestRegressor scored 0.9903, and ExtraTreeRegressor scored 0.9933. Additionally, we introduce a new algorithm based on active learning, which can enhance CYPA's performance by reducing the number of labeled data needed for training. Incorporating active learning into CYPA can improve the efficiency and accuracy of crop yield prediction, thereby enhancing decision-making at international, regional, and local levels.

Keywords Precision agriculture · Crop yield prediction · Active learning · IoT

1 Introduction

Around two-thirds of all workers on the continent are employed in the agricultural sector, which has a significant impact on the economics of African nations. Changing rain patterns, droughts, flooding, and the geographical dispersion of pests and illnesses are just a few of the direct and indirect effects of climate change on agricultural output (Food and Agricultural Organization of United Nations, 2020). These findings indicate that one of the most difficult issues in precision agriculture is crop yield prediction. Agriculture's future is negatively impacted by the changing environmental circumstances, particularly global warming and climate variability [1, 2].

Since crop output varies on a wide range of variables, including climate, weather, soil, application of fertilizer, and seed variety [3], this challenge necessitates the use of many datasets [4]. Accurately forecasting crop production using statistical models is a time-consuming and challenging process [5]. It takes a lot of time and effort to anticipate the crop production using statistical models [6]. Big data's introduction in this decade has increased the adoption of more sophisticated analysis methods like machine learning. Depending on the study challenge and research questions, a machine learning model may be either descriptive or predictive [7, 8].

Predictive models are used to generate future forecasts, whereas descriptive models are used to learn from the data gathered and explain what has occurred [9]. It has been employed to find solutions in a variety of fields, including finance [10], biology [11], medicine [12], and, most recently, agriculture.

✉ Fatma M. Talaat
fatma.nada@ai.kfs.edu.eg

¹ Faculty of Artificial Intelligence, Kafrelsheikh University, Kafrelsheikh, Egypt

labeling that better represent the current conditions. Overall, incorporating active learning into the CYPA can improve the efficiency and accuracy of crop yield prediction, leading to better decision-making at the international, regional, and local levels.

Following is how the remaining work is structured. Section 2 presents some of the most recent research in the field of precision agriculture. The proposed framework is presented in Sect. 3. In Sect. 4, experimental evaluation is offered. Section 5 brings this effort to a close.

2 Related work

Machine learning (ML), a branch of artificial intelligence, allows computers to learn from data without having to be explicitly programmed. ML is now more useful because to big data technology. A considerable amount of data that is produced quickly from multiple sources is simply referred to as “big data”. ML, which aids in comprehending data, is fundamentally based on the construction of mathematical models [20].

The three main types of difficulties that machine learning is used to solve are supervised learning problems, unsupervised learning problems, and reinforcement problems [21–23]. A tuple of (X, Y) is the mathematical input for machine learning, where X represents the so-called independent characteristics and Y represents the dependent variable or target. We have a supervised learning problem when the variable Y is known before the training. In supervised learning, regression and classification problems are the two most frequent problems.

One of the regression difficulties is known to be the house price projection problem [24]. Unsupervised learning is used when the variable Y was unknown before to the training. Their methods are widely applied to clustering problems. Climate change is the long-term alteration of regional or global temperatures or weather patterns. It is challenging to combat global warming and reduce greenhouse gas emissions due to the legal and regulatory concerns posed by it [25].

Millions of people worldwide, especially in South Asia, Sub-Saharan Africa, and tiny islands, are projected to see an increase in food insecurity, hunger, and malnutrition issues [26]. Climate change is one of the main dangers to Africa’s agricultural development [27]. Crop yield forecasting continues to be a hot topic for researchers worldwide.

Using data from remote sensing, You et al. [28] developed a deep learning system for agricultural yield prediction. In order to forecast crop production in undeveloped countries annually, this method combined a Convolutional

Neural Network (CNN) with a Gaussian process component and dimensional reduction technique. The method was used on a dataset of soybeans that was created by fusing US soil, sensing, and climate data. The authors’ method demonstrates how the Gaussian approach was used to reduce the model’s Root Mean Square Error (RMSE), which was reduced with the Long Short-Term Memory (LSTM) model from 6,27 to 5, and 83 on average and with the CNN model from 5, 77 to 5, 57.

Paudel et al. [29] combined machine learning with agronomic crop modelling techniques to create a machine learning baseline for large-scale crop production prediction. They started with a workflow that prioritised accuracy, modularity, and reuse. The MARS Crop Yield Forecasting System (MCYFS) database was used to create the crop simulation outputs’ features along with weather, remote sensing, and soil data.

Sun et al. [30] presented a novel multilevel deep learning model coupling Recurrent Neural Network (RNN) and Convolutional Neural Network to extract both spatial and temporal data for agricultural productivity prediction (CNN). Their main goals were to evaluate the usefulness of the proposed strategy for US Corn Belt yield prediction and the influence of different data sets on the prediction task. Data on soil characteristics as well as time-series remote sensing data were used as inputs. For the purpose of predicting the county-level corn yield from 2013 to 2016, they carried out their studies in the US Corn Belt states.

Using a 43-year benchmark dataset, Bali and Singla [6] projected wheat crop yield in northern India using a deep learning-based Recurrent Neural Network (RNN) model. Additionally, LSTM was used in their study to address the RNN model’s inherent vanishing gradient problem. Results from several models, including RNN-LSTM (RMSE = 147.12, MAE = 60.50), Artificial Neural Networks (RMSE = 732.14, MAE = 623.13), Random Forest (RMSE = 540.88, MAE = 449.36), and Multivariate Linear Regression (RMSE = 915.64, MAE = 796.07), demonstrated the effectiveness of the proposed model.

Authors in [31] proposes a novel edge-based spatiotemporal data preprocessing (ESDP) framework for enhancing the accuracy of air quality prediction. The proposed framework utilizes a combination of deep learning models and edge computing technologies to preprocess and analyze large volumes of spatiotemporal air quality data in real-time, thereby reducing the computational burden and improving the efficiency of air quality prediction. The authors demonstrate the effectiveness of the proposed framework by conducting experiments on a real-world dataset of air quality measurements from the city of Beijing, China. The results show that the ESDP framework outperforms other state-of-the-art approaches in terms of prediction accuracy and computational efficiency, thereby

demonstrating its potential to improve air quality monitoring and management. This article is relevant to our work as it proposes a novel framework that utilizes edge computing and deep learning techniques for enhancing the accuracy of spatiotemporal data prediction, which can be applied to various domains including precision agriculture.

3 Crop yield prediction algorithm (CYPA)

This section presented the implementation in two cases: (i) Using machine learning methods, (ii) Using Statistical methods (Correlation).

3.1 Using machine learning methods

In this study, the top 10 crop yields eaten globally are predicted using machine learning approaches. Among these crops are wheat, paddy, rice, sorghum, sweet potatoes, soy beans, plantains, cassava, and others.

The majority of agricultural operations are significantly impacted by our prediction models' most well-known and commonly used features which are: (i) Temperature: The nation's average annual temperature. Throughout the twenty-first century, yield declines among and across many soil and crop species may be caused by temperature increases [32]. (ii) Precipitation: the volume of precipitation on average each year. Variations in water availability may have an impact on agricultural productivity and profitability because water is a crucial input in agricultural production [33].

(iii) Pesticide: the quantity of pesticides used annually per square mile of land in a given nation. Crops are treated with pesticides, especially in low-income areas. The most typical justification for applying pesticides is to boost yield. (iv) Yield: the production rate per area, year, and nation; (v) Country: the area under discussion. (vi) The year the data were gathered. The approaches described in this work include Gaussian Nave Bayes (GNB), Gradient Boosting (GB), Decision Tree (DT), Random Forest (RF), and Multimodal Naive Bayes (MNB).

The overall scheme to describe the effect of the climate change on the crop yield is shown in Fig. 2. The applied machine learning models employed are I Gradient Boosting, (ii) Decision Tree, (iii) Random Forest, and (iv) SVR. Four models are trained and validated using the best hyper parameter settings for each of the machine learning techniques in the study. The pipeline for the crop yield

prediction system is shown in Fig. 3. Four steps make up this process: ETL, feature engineering, model training, evaluation, and model deployment, which is the last step.

i. *DecisionTreeRegressor*

DecisionTreeRegressor is a simple and interpretable algorithm that can handle both numerical and categorical data. It can also handle missing values and outliers effectively, and is resistant to over fitting. However, it can be prone to high variance and instability, and may not generalize well to new data if the tree is too deep or complex. To overcome this, techniques like pruning and ensemble methods, such as Random Forest, can be used. Overall, DecisionTreeRegressor is a powerful and widely used algorithm for regression tasks, especially when interpretability and simplicity are desired.

ii. *RandomForestRegressor*

RandomForestRegressor is a machine learning algorithm used for regression tasks, which predicts a continuous output variable based on input features. It is an ensemble learning method that combines multiple decision trees to create a more accurate and robust model. The algorithm works by constructing a large number of decision trees and training each tree on a random subset of the training data and a random subset of the input features. Each tree in the ensemble produces a prediction, and the final prediction is obtained by averaging the predictions of all the trees.

RandomForestRegressor can handle missing data and non-linear relationships between the input features and output variable. It is also robust to over fitting, which is a common problem in machine learning models, and can handle high-dimensional data with many input features. Additionally, it can provide a measure of feature importance, indicating which input features are most relevant for the prediction.

iii. *ExtraTreeRegressor*

The Extra Tree Model builds an ensemble algorithm using decision or regression trees that have not been pruned. It varies from other ensemble models based on trees. It begins by randomly splitting nodes based on the selected cut-points. Additionally, it grows the trees using the entire learnt sample set rather than a bootstrap replica. The Extra tree splitting process is described in Algorithm 1.

Algorithm 1: The Extra tree splitting algorithm**Part I: Splitfeature (X)**Input: LPI data $D = (X, Y)$, K Output: the labels of lncRNA-protein pairs or a split $[a \leq a_c]$

- 1: If (Stopsplit (X))==true) then
- 2: Return the labels of lncRNA-protein pairs
- 3: Else
- 4: Select K different LPI features a_1, \dots, a_K from X ;
- 5: Conduct K splits s_1, \dots, s_K where $s_i = \text{Picksplit}(X, a_i) (i = 1, 2, \dots, K)$
- 6: Obtain an optimal LPI feature split s^* by $\text{Score}(s^*, X) = \max_{i=1, \dots, K} \text{Score}(s_i, X)$
- 7: End if
- 8: Classify lncRNA-protein pairs based on the split s^*

Part II: Picksplit (X, a)Input: LPI data $D = (X, Y)$, a feature a

Output: a split

- 1: Suppose that a_{\max} and a_{\min} represent the maximum and minimum value of an LPI feature a in X , respectively
- 2: Draw a random cut-point a_c uniformly in $[a_{\min}, a_{\max}]$
- 3: Obtain the split $[a < a_c]$

Part III: Stopsplit (X)Input: LPI data $D = (X, Y)$, n_{\min}

Output: a boolean value

- 1: If $|X| < n_{\min}$, then return True
- 2: If all features in X are the same, then return True
- 3: If the classes of all lncRNA-protein pairs in X is consistent, then return True
- 4: Else
- 5: return False
- 6: End if

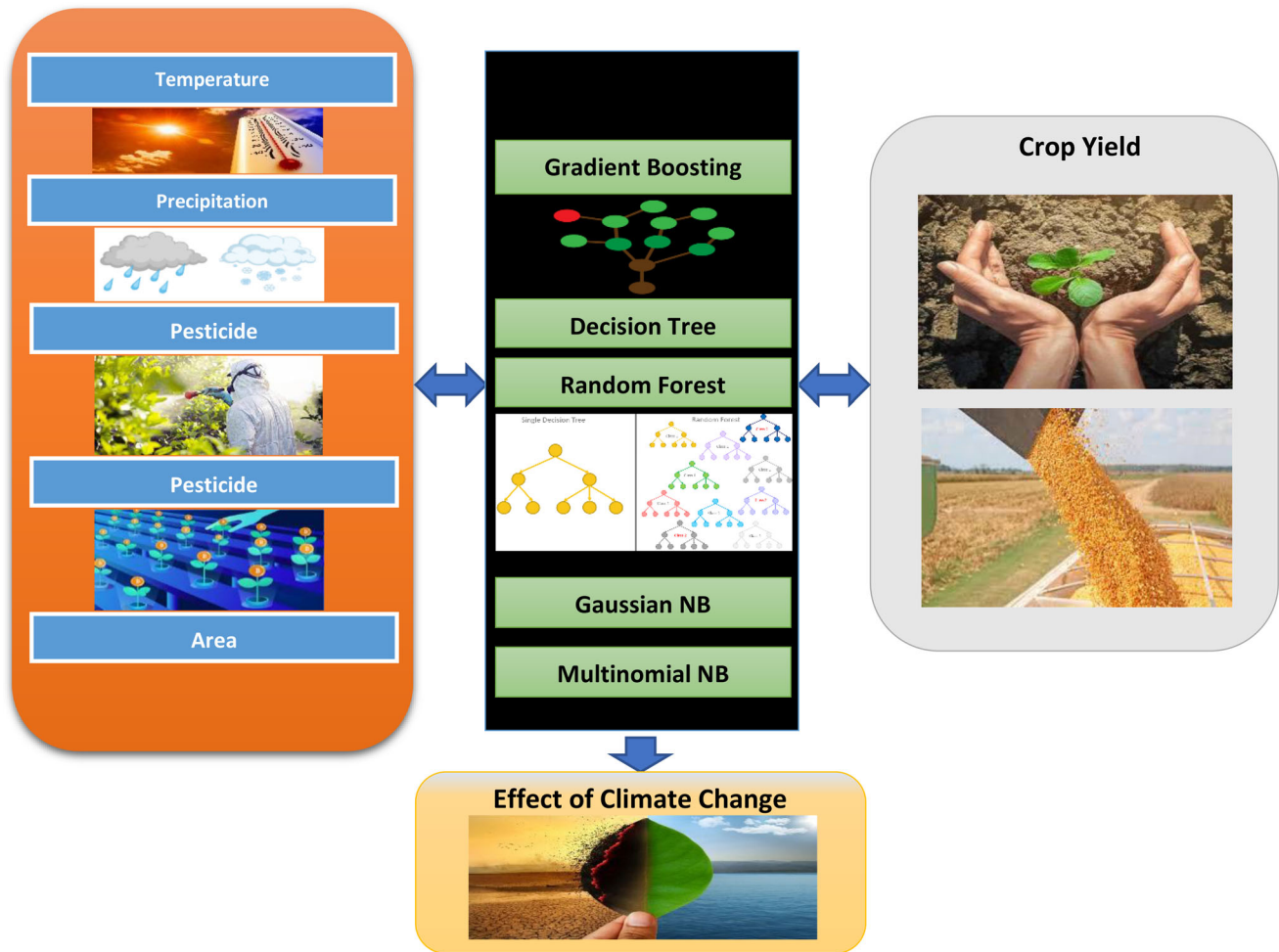


Fig. 2 The overall scheme to describe the effect of the climate change on the crop yield

3.1.1 Using statistical methods (correlation)

To test the impact of climate changes on crop yields, we follow two methods: (i) Pearson's Correlation Coefficient (PCC), and (ii) Multiple Regression (MR).

3.1.1.1 Pearson's correlation coefficient (PCC) To gauge how closely two variables are related to one another, correlation coefficients are used.

The Pearson's correlation coefficient is the most often used of the several correlation coefficient types. A correlation coefficient frequently used in linear regression is Pearson's correlation (sometimes known as Pearson's r) [26].

We determine the Pearson's correlation coefficient (r_a) between the most significant variables and crop yields (CY) as follows to examine the effect of climatic changes

on crop yields: (i) Using the formula in Eq. (1), we determine the relationship between the two variables Average Temperature (TA) and Crop Yields (CY).

$$r_t = \frac{n(\sum TACY) - (\sum TA)(\sum CY)}{\sqrt{[n \sum (TA)^2 - \sum (TA)^2][n \sum (CY)^2 - \sum (CY)^2]}} \quad (1)$$

(ii) we calculate r_p between the two variables Pesticides Tonnes (PT) and Crop Yields (CY) as shown in Eq. (2).

$$r_p = \frac{n(\sum PTCY) - (\sum PT)(\sum CY)}{\sqrt{[n \sum (PT)^2 - \sum (PT)^2][n \sum (CY)^2 - \sum (CY)^2]}} \quad (2)$$

(iii) we calculate r_r between the two variables Average Rain Fall (RF) and Crop Yields (CY) as shown in Eq. (3).

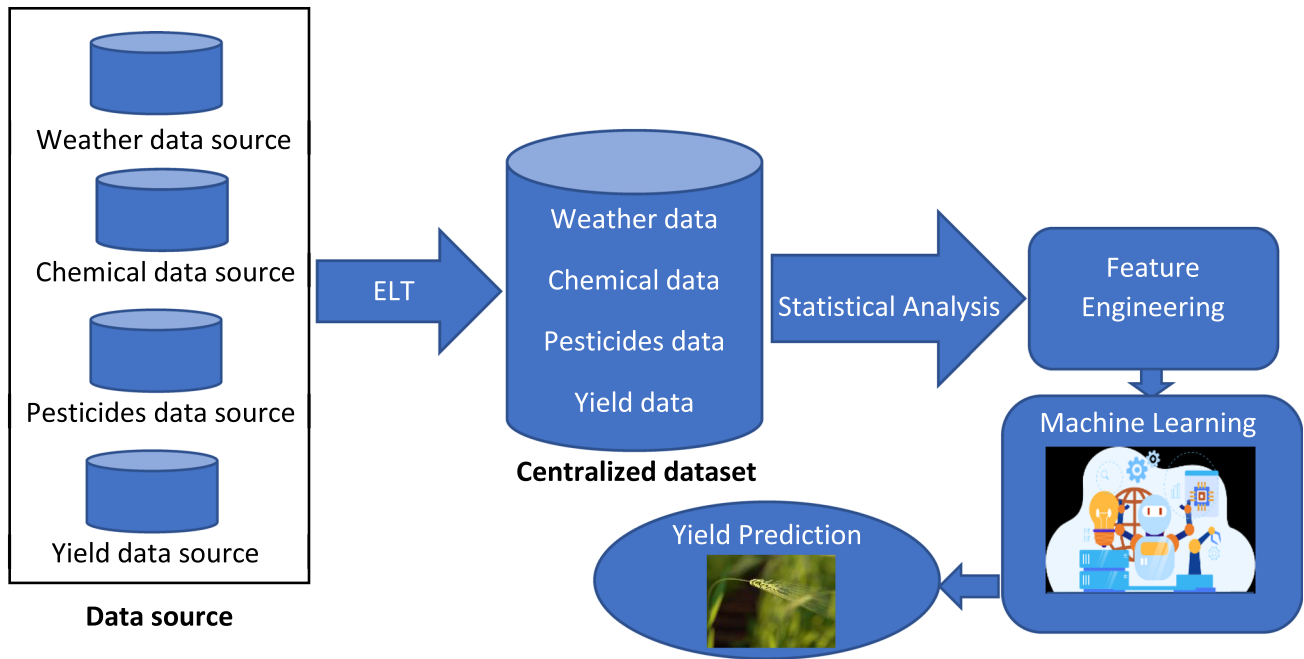


Fig. 3 An overview of the crops yield prediction pipeline

$$rr = \frac{n(\sum RFCY) - (\sum RF)(\sum CY)}{\sqrt{[n \sum (RF)^2 - \sum (RF)^2][n \sum (CY)^2 - \sum (CY)^2]}} \quad (3)$$

where, r_t is coefficient between TA and CY, r_p is coefficient between PT and CY, r_r is coefficient between RF and CY. TA is the average temperature, PT is the pesticides tonnes, RF is the average rain fall, n is number of points.

3.1.1.2 Multiple regression (MR) The second phase is using multiple regression analysis to determine how much the chosen features have an impact on drop yield. Following are the very important explanatory variables, listed in decreasing order of partial correlation coefficient, from a multiple regression analysis utilizing three factors as the explanatory variables: Average rainfall (RA), average pesticide use (PT), and average temperature (TA) (RF). MR formula and computation are shown in Eq. (4).

$$NCY = \beta_0 + \beta_1(TA) + \beta_2(PT) + \beta_3(RF) + \epsilon \quad (4)$$

where NCY is the new crop yield (the predicted value of crop yield), β_0 is the NCY-intercept (constant value), β_1 is the change in NCY each 1 increment change in TA, β_2 is the change in NCY each 1 increment change in PT, β_3 is the change in NCY each 1 increment change in RF, ϵ is the model error term.

4 Implementation and Experiments

The utilized dataset and the suggested algorithm's outcomes are introduced in this section.

4.1 Dataset

Crop Yield Prediction Dataset [35] is used in this study. Pesticides & Yield are collected from FAO. Rainfall & Avg. Temperature are collected from World Data Bank. Yield_df.csv is final dataset processed by cleaning & merging of pesticides, yield, rainfall, and avg. temperature. Table 1 provides a sample of utilized dataset.

Figure 4 displays the HeatMap for the dataset that was used.

4.2 Performance assessment

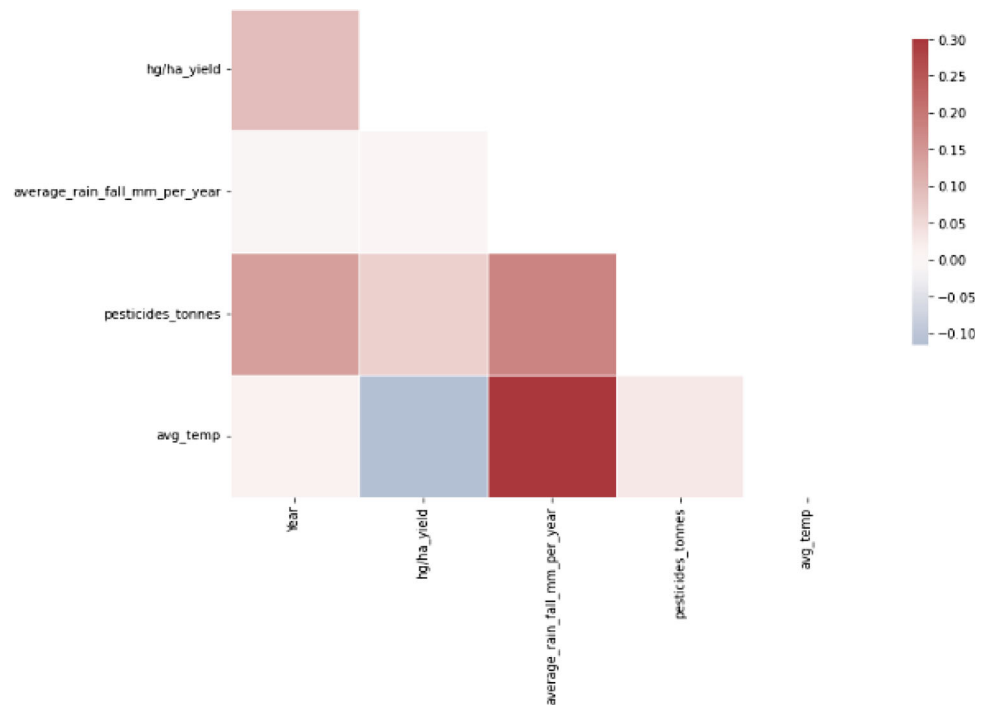
This part of paper presented the results of the implementation in two cases: (i) Results of using machine learning methods, (ii) Results of using Statistical methods (Correlation).

4.2.1 Results of using machine learning methods

Table 2 displays the anticipated values for each crop using Gradient Boosting Regressor and Decision Tree Regressor. Figure 5 compares real values to expected ones when using a decision tree regression model, and Fig. 6 compares

Table 1 A sample of utilized dataset

	Area	Item	Year	hg/ ha_yield	Average_rain_fall_mm_per_year	Pesticides_tonnes	Avg_temp
0	Albania	Maize	1990	36,613	1485	121	16.37
1	Albania	Potatoes	1990	66,667	1485	121	16.37
2	Albania	Rice, paddy	1990	23,333	1485	121	16.37
3	Albania	Sorghum	1990	12,500	1485	121	16.37
4	Albania	Soybeans	1990	7000	1485	121	16.37
5	Albania	Wheat	1990	30,197	1485	121	16.37

Fig. 4 HeatMap**Table 2** Predicted data for each crop

Crop	Decision tree regressor	Gradient boosting regressor
Cassava	92.8892	73.5649
Maize	88.4270	34.3709
Plantains and others	77.2187	53.8415
Potatoes	90.9645	69.3540
Rice, paddy	89.5408	50.4945
Sorghum	80.2281	46.6916
Soybeans	85.3711	-36.0522
Sweet potatoes	83.9677	56.5816
Wheat	92.3558	52.3807
Yams	92.9960	71.8454

actual values to projected ones when using a gradient boosting regression model.

Using DecisionTreeRegressor, the score = 0.9814, using RandomForestRegressor, the score = 0.9903, and using ExtraTreeRegressor, the score = 0.9933.

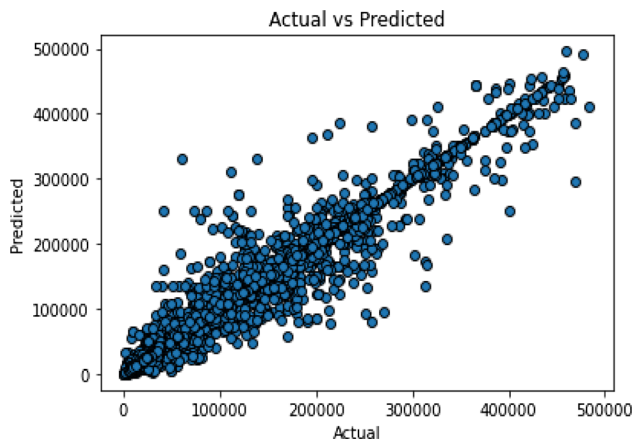


Fig. 5 Actual Data versus Predicted Data using Decision Tree Regressor

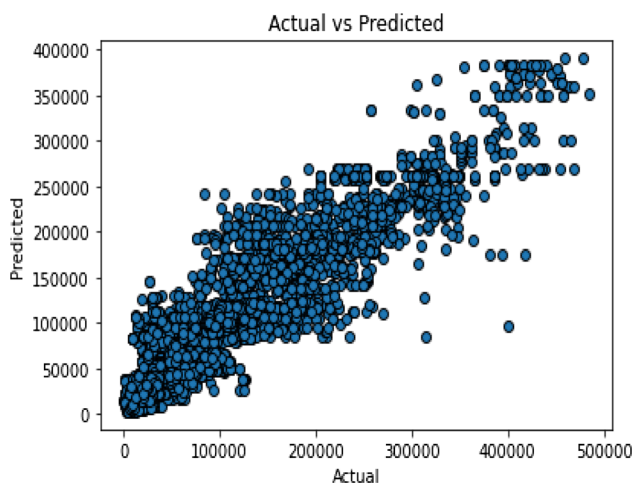


Fig. 6 Actual Data versus Predicted Data using Gradient Boosting Regressor

4.2.2 Results of using statistical methods (Correlation)

To test the impact of climate changes on crop yields, we follow two methods: (i) Pearson's Correlation Coefficient (PCC), and (ii) Multiple Regression (MR).

4.2.2.1 Calculating PCC In order to investigate the impact of average temperature on crop yield, we calculate PCC (r_t) between TA and CY as shown in Fig. 7.

From Fig. 7, it is shown that there is a negative relationship (inverse correlation) between the average temperature (TA) and crop yield (CY). The $r_t = -0.11477$ (negative correlation). We also calculate PCC between the pesticides tonnes (PT) and crop yield (CY) as shown in Fig. 8.

From Fig. 8, it is shown that there is a weak relationship between the pesticides tonnes (PT) and crop yield (CY). The $r_p = +0.06408$ (positive correlation) means that both variables move in tandem—that is, in the same direction. We also PCC between the average rain fall (RF) and crop yield (CY) as shown in Fig. 9.

From Fig. 9, it is shown that there is a weak relationship between the average rain fall (RF) and crop yield (CY). The $r_p = +0.000962$ (positive correlation) means that both variables move in tandem—that is, in the same direction.

4.2.2.2 Calculating MR We employ Multiple Linear Regression (MLR) [36] to forecast the quantity of citations from the aforementioned components (TA, PT, and RF), as shown in Fig. 10.

From Fig. 10, it is shown that the residuals are normally distributed.

Comparing the proposed algorithm (CYPA) with the state of the art algorithm proposed in [37] is illustrated in Table 3.

One limitation of our study is that it is based on data collected from a specific geographical region, and the performance of the CYPA algorithm may vary for different regions. Additionally, the CYPA algorithm relies on accurate and timely data collection, and any errors or delays in data collection can affect the accuracy of crop yield prediction.

Moreover, the presented algorithm utilizes machine learning techniques, which are dependent on the quality and quantity of data used for training. Therefore, the accuracy of the CYPA algorithm may be affected by incomplete or biased data, as well as any changes in environmental conditions that are not captured in the training dataset.

5 Conclusion

This paper presented a novel approach to predicting crop yields using IoT techniques in precision agriculture. The proposed Crop Yield Prediction Algorithm (CYPA) incorporates multiple data sources such as climate, weather, agricultural yield, and chemical data to anticipate annual crop yields by policymakers and farmers in their country. The study demonstrated the efficacy of CYPA by training and verifying five models using optimal hyperparameter settings for each machine learning technique. The results indicate that CYPA can achieve high accuracy

Fig. 7 Crop Yield (CY) versus average temperature (TA)

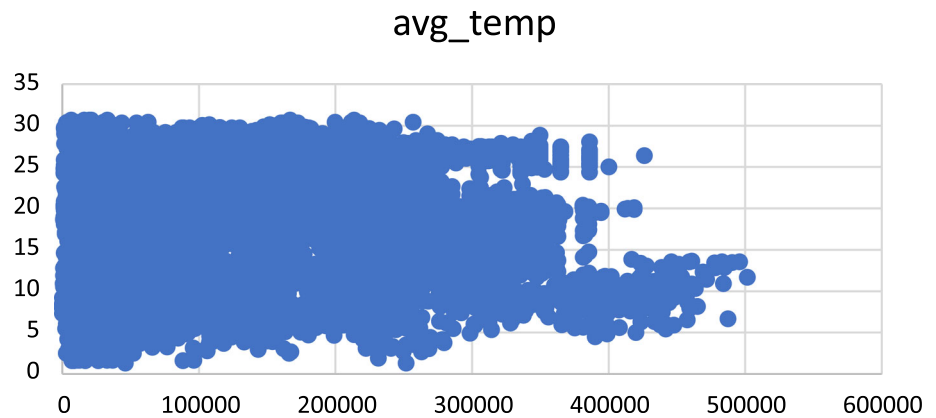


Fig. 8 Crop Yield (CY) versus pesticides tonnes (PT)

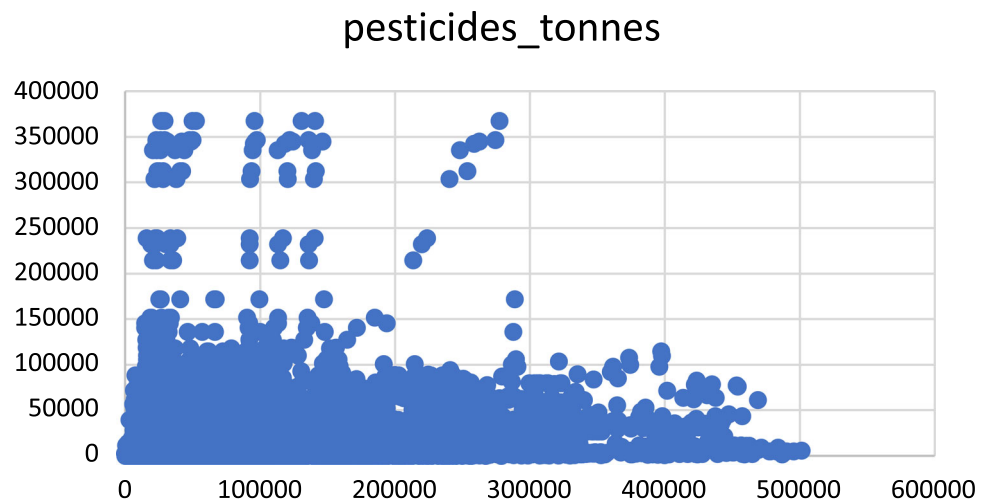
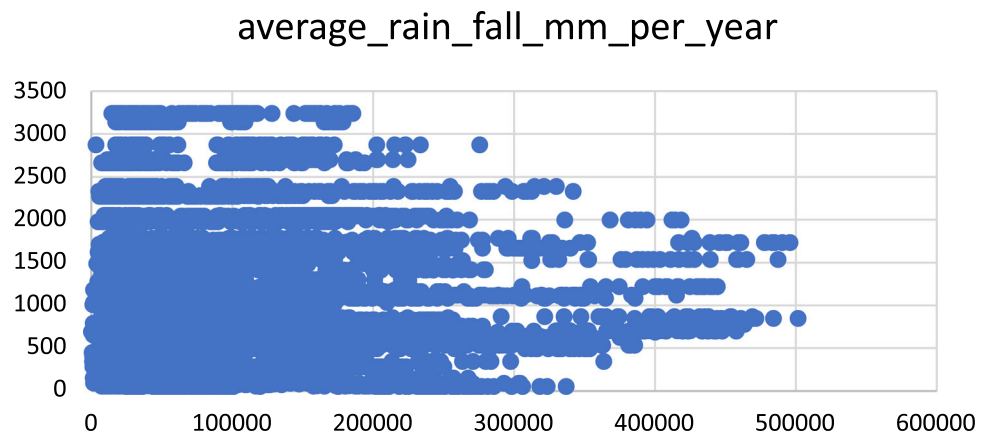
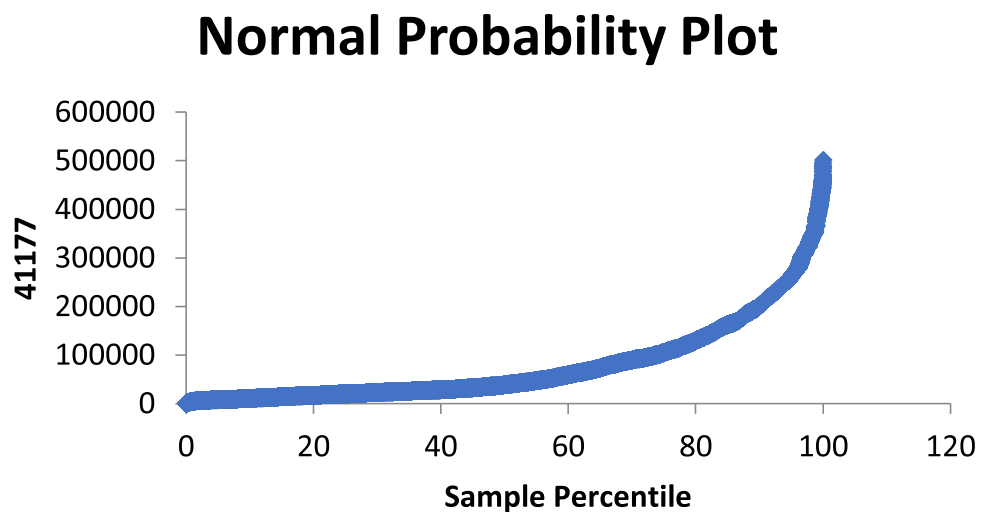


Fig. 9 Crop Yield (CY) versus average rain fall (RF)



in predicting crop yields, as demonstrated by the scores of DecisionTreeRegressor, RandomForestRegressor, and ExtraTreeRegressor. Additionally, the paper introduces a new algorithm based on active learning that can enhance CYPA's performance by reducing the number of labeled

data needed for training. Incorporating active learning into CYPA can improve the efficiency and accuracy of crop yield prediction, thereby enhancing decision-making at international, regional, and local levels. Overall, this study highlights the potential of IoT and machine learning

Fig. 10 Normal probability plot (TA,PT,RF)**Table 3** Predicted data for each crop

Algorithm	RMSE	MAE
RNN model [37]	147.12	60.50
CYPA	132.01	34.79

techniques in addressing the critical challenge of predicting crop yields, thereby facilitating informed decision-making for policymakers and farmers alike. When utilizing DecisionTreeRegressor, RandomForestRegressor, and ExtraTreeRegressor, the score is equal to 0.9814, 0.9903, and 0.9933, respectively. In the future, the proposed algorithm can be used with OCNN [37–41] and make use of Resnet [42].

Authors' contributions Single Author.

Funding Open access funding provided by The Science, Technology & Innovation Funding Authority (STDF) in cooperation with The Egyptian Knowledge Bank (EKB). The authors received no specific funding for this study.

Data availability <https://www.kaggle.com/code/nghianguyen39/crop-yield-prediction/data>.

Declarations

Conflict of interest The authors declare that they have no conflict of interest to report regarding the present study.

Ethical approval There is no any ethical conflict.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the

source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Bhadouria R, Singh R, Singh VK, Borthakur A, Ahamad A, Kumar G, Singh P (2019) Chapter 1 - agriculture in the era of climate change: consequences and effects. In: Choudhary KK, Kumar A, Singh AK (eds) Climate change and agricultural ecosystems. Woodhead Publishing, Sawston, pp 1–23
- Lal R (2018) Climate-resilient agriculture for ensuring food security. *Sustainability* 10(2):1–15
- Prasad R, Singh V (2019) Prediction models for crop yield and factors affecting crop yield: a review. *J Agric Sci Technol* 21(5):1005–1021
- Xu X, Gao P, Zhu X, Guo W, Ding J, Li C, Zhu M, Wu X (2019) Design of an integrated climatic assessment indicator (ICAI) for wheat production: a case study in Jiangsu Province, China. *Ecol Indic* 101:943–953. <https://doi.org/10.1016/j.ecolind.2019.01.059>
- Akter T, Islam MS (2021) Crop yield prediction using machine learning techniques: a review. *J King Saud Univ-Comput Inform Sci* 33(4):480–488
- Bali N, Singla A (2021) Deep learning based wheat crop yield prediction model in Punjab region of North India. *Appl Artif Intell* 35(15):1304–1328. <https://doi.org/10.1080/08839514.2021.1976091>
- van Klompenburg T, Kassahun A, Catal C (2020) Crop yield prediction using machine learning: a systematic literature review. *Comput Electron Agric* 177:105709. <https://doi.org/10.1016/j.compag.2020.105709>
- Mohan S, Kumar P (2018) Machine learning techniques for agriculture crop yield prediction: a review. *Int J Comput Sci Eng* 6(10):405–410
- Alpaydin E (2010) Introduction to machine learning, 2nd edn. MIT Press, Cambridge

10. Aziz S, Dowling MM, Hammami H, Piepenbrink A (2019) Machine learning in finance: a topic modeling approach. SSRN. <https://doi.org/10.2139/ssrn.3327277>
11. Camacho DM, Collins KM, Powers RK, Costello JC, Collins JJ (2018) Next- generation machine learning for biological networks. *Cell* 173(7):1581–1592. <https://doi.org/10.1016/j.cell.2018.05.015>
12. Doupe P, Faghmous J, Basu S (2019) Machine learning for health services researchers. *Value Health* 22(7):808–815. <https://doi.org/10.1016/j.jval.2019.02.012>
13. Atzori L, Iera A, Morabito G (2010) The internet of things: a survey. *Comput Netw* 54(15):2787–2805
14. Gubbi J, Buyya R, Marusic S, Palaniswami M (2013) Internet of Things (IoT): a vision, architectural elements, and future directions. *Futur Gener Comput Syst* 29(7):1645–1660
15. Lee I, Lee K (2015) The Internet of Things (IoT): applications, investments, and challenges for enterprises. *Bus Horiz* 58(4):431–440
16. Srirama SN (2017) Mobile web and cloud services enabling Internet of Things. *CSI Trans ICT* 5:109–117
17. Talaat FM (2022) Effective deep Q-networks (EDQN) strategy for resource allocation based on optimized reinforcement learning algorithm. *Multimed Tools Appl* 81(28):39945–39961. <https://doi.org/10.1007/s11042-022-13000-0>
18. Perera C, Qin Y, Estrella JC, Reiff-Marganiec S, Vasilakos AV (2017) Fog computing for sustainable smart cities. *ACM Comput Surv* 50:1–43
19. Atlam HF, Walters RJ, Wills GB (2018) Fog computing and the internet of things: a review. *Big Data Cogn Comput* 2(2):10
20. VanderPlas J (2016) Python data science handbook: essential tools for working with data. O'Reilly Media Inc, Newton
21. Al-Haija QA, Krichen M, Elhaija WA (2022) Machine-learning-based Darknet traffic detection system for IoT applications. *Electronics* 11(4):556. <https://doi.org/10.3390/electronics11040556>
22. Mihoub A, Snoun H, Krichen M, Salah RBH, Kahia M (2020) Predicting COVID-19 spread level using socio-economic indicators and machine learning techniques. In: 2020 First international conference of smart systems and emerging technologies (SMARTTECH), IEEE, pp 128–133
23. Srinivasan S, Ravi V, Sowmya V, Krichen M, Noureddine DB, Anivilla S, So-man KP (2020) Deep convolutional neural network based image spam classification. In: 2020 6th conference on data science and machine learning applications (CDMA), IEEE, pp 112–117
24. Truong Q, Nguyen M, Dang, H, Mei B (2020) Housing price prediction via improved machine learning techniques, procedia computer science. In 2019 International conference on identification, information and knowledge in the Internet of Things, vol 174, pp 433–442 <https://doi.org/10.1016/j.procs.2020.06.111>
25. McElDowney JF (2021) Climate change and the law. The impacts of climate change. Elsevier, pp 503–519. <https://doi.org/10.1016/B978-0-12-822373-4.00018-5>
26. Costa de Oliveira A, Marini N, Farias DR (2014) Climate change: new breeding pressures and goals. *Encyclopedia of agriculture and food systems*. Elsevier, pp 284–293. <https://doi.org/10.1016/B978-0-444-52512-3.00005-X>
27. Williams TO, Mul ML, Cofie OO, Kinyangi J, Zougmore RB, Wamukoya G, Nyasimi M, Mapfumo P, Speranza CI, Amwata D et al. Climate smart agriculture in the African context (2015).
28. You J, Li X, Low M, Lobell D, Ermon S (2017) Deep Gaussian process for crop yield prediction based on remote sensing data. *Proc AAAI Conf Artif Intell*. <https://doi.org/10.1609/aaai.v31i1.11172>
29. Paudel D et al (2021) Machine learning for large-scale crop yield forecasting. *Agric Syst* 187:103016
30. Sun J, Lai Z, Di L, Sun Z, Tao J, Shen Y (2020) Multilevel deep learning network for county-level corn yield estimation in the U.S. corn belt. *IEEE J Select Top Appl Earth Observ Remote Sens* 13:5048–5060. <https://doi.org/10.1109/JSTARS.2020.3019046>
31. Ojagh S, Cauteruccio F, Terracina G, Liang SHL (2021) Enhanced air quality prediction by edge-based spatiotemporal data preprocessing. *Comput Electr Eng* 96:107572. <https://doi.org/10.1016/j.compeleceng.2021.107572>
32. Hatfield JL et al (2011) Climate impacts on agriculture: implications for crop production. *Agron J* 103(2):351–370. <https://doi.org/10.2134/agronj2010.0303>
33. Torres M, Howitt R, Rodrigues L (2019) Analyzing rainfall effects on agricultural income: Why timing matters. *Economia* 20(1):1–14. <https://doi.org/10.1016/j.econ.2019.03.006>
34. Kumari K, Yadav S (2018) Linear regression analysis study. *J Pract Cardiovasc Sci* 4(1):33. https://doi.org/10.4103/jpcs.jpcs_8_18
35. <https://www.kaggle.com/code/nghianguyen39/crop-yield-prediction/data>
36. Talaat FM, Gamel SA (2022) Predicting the impact of no. Of authors on no. Of citations of research publications based on neural networks. *J Amb Intell Human Comput*. <https://doi.org/10.1007/s12652-022-03882-1>
37. Talaat FM, Ali HA, Saraya MS et al (2022) Effective scheduling algorithm for load balancing in fog environment using CNN and MPSO. *Knowl Inf Syst* 64:773–797. <https://doi.org/10.1007/s10115-021-01649-2>
38. Talaat FM (2022) Effective prediction and resource allocation method (EPRAM) in fog computing environment for smart healthcare system. *Multimed Tools Appl* 81(6):8235–8258. <https://doi.org/10.1007/s11042-022-12223-5>
39. Talaat FM, Ali SH, Saleh AI, Ali HA (2020) Effective cache replacement strategy (ECRS) for real-time fog computing environment. *Cluster Comput* 23(4):3309–3333. <https://doi.org/10.1007/s10586-020-03089-z>
40. Talaat FM (2022) Effective deep Q-networks (EDQN) strategy for resource allocation based on optimized reinforcement learning algorithm. *Multimed Tools Appl* 81(17). <https://doi.org/10.1007/s11042-022-13000-0>
41. Talaat FM, Alshathri S, Nasr AA (2022) A new reliable system for managing virtual cloud network. *Comput Mater Continua* 73(3):5863–5885
42. Hassan E, El-Rashidy N (2022) Review: mask R-CNN models. *Nile J Commun Comput Sci* 3(1):17–27. <https://doi.org/10.21608/njccs.2022.280047>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.