PREDICTIVE ANALYTICS IN AGRICULTURE: MACHINE LEARNING MODELS FOR CROP YIELD

Rameela. S [1], M Senthil Kumar [2]*

ABSTRACT

Machine learning applications are transforming the way data is processed and decisions are made, which is having a big effect on the global economy. Agriculture is one of the industries that is significantly impacted by the global food supply issue. This research aims to find out what modern farmers may gain by employing machine learning techniques. Optimising agricultural productivity while reducing waste is the primary goal of these algorithms, which use intelligent judgements for planting, watering, and harvesting. Study delves into the present situation of machine learning in farming, drawing attention to important obstacles and possibilities. It provides experimental evidence that changes to labels affect the precision of data processing algorithms. By analysing the massive amounts of data acquired from farms and utilising online, real-time data from IoT sensors, results suggest that farmers may make better decisions on variables influencing crop growth. Eventually, incorporating these technologies that increase agricultural yields while decreasing waste has the potential to change modern agriculture. Among the fifteen different algorithms, one newly developed algorithm that improves feature combining schemes is provided in order to help determine which method is most appropriate for usage in agriculture. According to the results, 99.46% classification accuracy can be achieved with the Hoeffding Tree and Naïve Bayes Classifier methods, and 99.59% classification accuracy can be achieved with the Bayes Net methodology. These results will demonstrate higher output rates.

Keywords: Random Forest, Logistic Regression, Naïve Bayes, ANN, Smart farming, Crop analysis and prediction.

Department of Computer Applications[1],
Karpagam Academy of Higher Education, Coimbatore, India[1]
rameelariya@gmail.com
Department of Computer Science[2],
Nehru Arts and Science College[2]
senthilmsc09@gmail.com

* Corresponding Author

I. INTRODUCTION

Agricultural sector is crucial in meeting nutritional needs of the world's expanding population. Farmers must make most effective use of them possible, maximising yields while minimising losses. Developed into a suitable means of accomplishing this goal [1,2]. One modern agricultural technique that uses state-of-the-art technology to maximise crop yields and minimise waste is precision agriculture, sometimes called smart farming. Enhances productivity. Optimal crop yields with minimum inputs of water, fertilizer, and energy are the hallmarks of "smart farming" [3]. Figure 1 depicts system for crop analysis and forecasting that makes use of machine learning and the Internet of Things.Smart farms have evolved throughout time to combine many components and technology that enhance agricultural methods and boost production [4-6]. One of key technologies that contribute to smart farming is the Internet of Things. Making use of information gathered via Internet of Things sensors. Use of IoT sensors by farmers to monitor crop watering and nutrient uptake may improve crop quality and yield. [8,9].

Machine learning applications have permeated many aspects of our life in recent years, including urbanization, education, health, and military, and they have proven useful when making decisions. Simultaneously, it began providing technical solutions and information as the foundation of recently developed search engine infrastructure, including ChatGPT, Google Bard [11] and other comparable AI. -based chatbots and applications.Numerous research firms indicate that emerging trends will continue to spread across a range of platforms. Implementation of machine learning models will transform many industries, including traffic predictions [13] and chip design [12].

Machine learning algorithms can sift through mountains of data collected by IoT devices and elsewhere. It's an area that's booming and has potential to change the way we forecast and assess agricultural yields. Computer systems may now learn and become better at what they do without human intervention thanks to machine learning methods. They do this by analysing data and making predictions utilising algorithms and models grounded in mathematics and statistics. [14].

II. LITERATURE SURVEY

Improved crop quality and maximum farmer profit may be achieved via the application of crop production prediction approaches that use ML processes and algorithms. An improvement in agricultural quality has a multiplier effect on the economy. A lot of research has gone into this [11– 15]. In order to forecast the production of palm oil, a survey of machine learning methods is presented in [11]. Authors weighed the pros, cons, and limitations of the suggested approaches by comparing the relevant studies. In addition, afterreviewing the literature, the authors proposed a novel architecture that makes use of ML methods for predicting palm oil production.

Because soil properties greatly affect agricultural output, the writers of [12] investigated soil quality through the lens of crop production and forecast. The researchers measured several soil factors, such as temperature, precipitation, moisture, pH, humidity, and NPK (nitrogen, phosphorus, and potassium) concentrations.

Authors think that supervised machine learning techniques like K-Nearest Neighbour, SVM, Random Forest, and ANN might help farmers choose crops and harvest them more efficiently. In light of the country's diminishing natural resources, this strategy has the potential to improve economic circumstances. Using machine learning and smartphone photos, the research aims to track the development of cotton and chilli plants. Using a dataset spanning 32 districts, they tested several algorithms, like Decision Tree, Random Forest, and Gaussian Naïve Bayes, from TamilNadu Agricultural University. With an accuracy of 99.45%, the authors determined that Naïve Bayes was the top classifier. They imply that further categorisation and performance assessment in practical contexts might result in even more favourable results.

The scientists asserted that deep learning-based agricultural monitoring measures outperform certain older methods used in developing nations based on comparison and results [11]. For further information on practical uses of agricultural industry, the writers presented a unique approach using support vector machines in. [12].

Several viewpoints on this subject have been discussed in the literature study. Research on the use of ML algorithms to improve crop prediction, yield, and productivity has been extensive. In particular, the authors of [11] boasted a remarkable 99.45% accuracy rate in their analysis and performance rating; nevertheless, their results were confined to 22 distinct crops and were not grounded on actual data. Our study's overarching goal is to enhance smart farming's machine learning applications by:

- ❖ Providing results from experiments that show how changing labels impacts the precision of data analysis algorithms.
- ❖ Results of this study demonstrate that farmers may improve their decision- making on the elements that impact crop development by analysing a wide range of field data, including real-time data from IoT sensors.
- ❖ Results of this study demonstrate which machine learning algorithms are very accurate classifiers. It is worth mentioning that Bayes Net approach attained an unbelievable 99.59% accuracy, & Naïve
- ❖ Bayes Classifier and Hoffding Tree methods generated remarkable 99.46% accuracy. Algorithms' efficacy and reliability in accurately identifying input data are shown by these results.

III. SMART FARMING

Using cutting-edge technologies to improve sustainability, reduce environmental harm, and raise productivity, innovative farming techniques have completely changed the agricultural industry. Machine learning (ML) has enabled many applications that simplify agricultural processes and enhance decision-making, and it is a crucial part of this change. ML improves sustainability by increasing productivity and making livestock more comfortable via predictive modelling and real-time health monitoring. [3].

Numerous machine learning (ML) approaches, like Random Forests, SVM, and ANNs, can be used to forecast crop yield [11-13].

Already, machine learning is making a significant contribution to farmers' profitability by giving them knowledge that will increase agricultural productivity and efficiency. Farmers are therefore being urged to effectively use machine learning algorithms and techniques, particularly when gathering, processing, and evaluating data. Most problems that farmers have can be solved by ML technology [13]. Better weather forecasting, less waste, more production, and higher profit margins are all possible outcomes. This is why modern technology such as GPS-based soil sampling, self- driving automobiles, variable rate systems, automated gear, software, sensors, cameras, robotics, drones, and GPS guidance are highly recommended for use by farmers. [14].More than half of planters are unaware of the alternatives available, and two thirds find using technology to be extremely challenging [15]. Modernizing agriculture and boosting productivity can be achieved by equipping farmers with machine learning skills.

IV. BENEFITS AND DIFFICULTIES OF CROP ANALYSIS AND PREDICTION

Although machine learning is finding uses in many other fields, studies in the agricultural industry are still in their early stages and face their own set of obstacles. This section highlights the main advantages and difficulties associated with using machine learning for crop analysis and forecasting, drawing from recent studies [14-15]. Figure 2 illustrates the application of AI in crop analysis and the different stages of prediction.

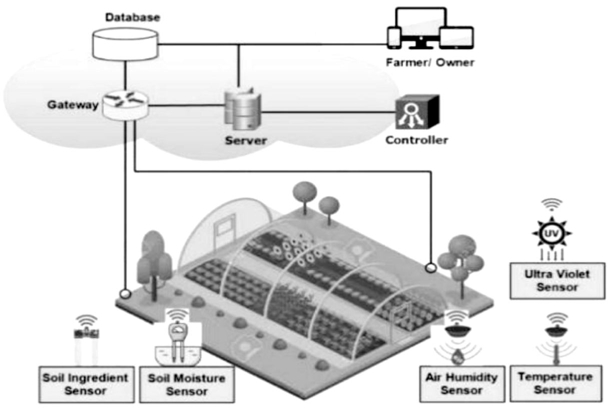

Figure 1: AI in agriculture

Benefits:

The principal benefits that farmers stand to gain by implementing machine learning on their fields are as follows:

- Greater effectiveness: Because a greater amount of data can be extracted, this method is more precise and successful at spotting trends while sparing farmers' time and money.
- Enhanced crop yield: Weather patterns, soil conditions, and historical data from ML algorithms are just a few of sources that farmers can leverage for making better decisions, which might increase crop yields.

B. Obstacles

Machine learning may greatly improve crop analysis & predicting, but there are a number of other issues to take into account. The primary challenges include the following:

- Data quality: Because of environmental elements such as soil, climate, and terrain, it may be difficult to acquire high-quality data in agriculture. Consequently, data cleansing and collecting may be complex operations. Some of the most important challenges in using deep learning for fruit detection and identification are discussed in reference [15]. We have discovered most of the characteristics which lead to slowness, poor robustness, and low accuracy in fruit identification and detection.
- Data volume: For proper training, ML models frequently require massive amounts of data. It may be particularly challenging for smaller farms to manage and gather data on a large scale in the agricultural industry. Key issues with big data, as stated in Ref. [12], are its bulk, velocity, diversity, and validity.

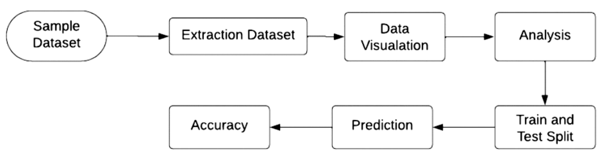

Figure 2: Proposed system

Our research suggested using several machine learning techniques, namely Multilayer Neural Networks, Random Forest, and Naïve Bayes Classifier, all of which were selected based on their distinct features and capabilities. By focussing on the most important aspects that improve accuracy, we performed a thorough examination of the data.The principal benefits that farmers stand to gain by implementing machine learning on their fields are as follows:

- Greater effectiveness: Because a greater amount of data can be extracted, this method is more precise and successful at spotting trends while sparing farmers' time and money.
- Enhanced crop yield: Weather patterns, soil conditions, and historical data from ML algorithms are just a few of sources that farmers can leverage for making better decisions, which might increase crop yields.

B. Obstacles

Machine learning may greatly improve crop analysis & predicting, but there are a number of other issues to take into account. The primary challenges include the following:

- Data quality: Because of environmental elements such as soil, climate, and terrain, it may be difficult to acquire high-quality data in agriculture. Consequently, data cleansing and collecting may be complex operations. Some of the most important challenges in using deep learning for fruit detection and identification are discussed in reference [15]. We have discovered most of the characteristics which lead to slowness, poor robustness, and low accuracy in fruit identification and detection.
- Data volume: For proper training, ML models frequently require massive amounts of data. It may be particularly challenging for smaller farms to manage and gather data on a large scale in the agricultural industry. Key issues with big data, as stated in Ref. [12], are its bulk, velocity, diversity, and validity.

We also stressed the significance of data cleansing and transformation as pretreatment steps to guarantee high-quality data for our study. Fig-3 offers a higher-level overview of our crop analysis and method prediction system that uses algorithms from the Internet of Things (IoT) and machine learning.

We used these algorithms in our experiment: - An approach to supervised learning known as the Naïve Bayes Classifier uses the Bayes theorem to categorise objects. It finds use in data mining and ML programs for tasks like text analysis, medical diagnosis, spam filtering, and similar ones. Underlying premise of Naïve Bayes is that there is no relationship between attributes that make up a class. In practical situations, the Naïve Bayes method is effective. This is pseudocode for the Naïve Bayes Classifier:

{X: Training set; m: Observation count; n: Feature count; Labels y for training set} are the input parameters.

The output will say, "Classification of new data point x; x's anticipated classification }."

For every unique class value c in y, get prior probability of class c.

Count(c)* m -1=P(c) (1)

wherein count(c) is the number of observations that belong to class c.

Use the formula below to determine the conditional probability of feature i given class c for each feature i in X & every unique class value c in Y:

Count(c,i)1* Count(c)-1=P(c|i) (2) wherein count (i, c) is the number of observations for class c and feature i.

Determine the posterior probability of every class in light of the newly discovered data item x.

- Initializing posterior prospectP(c|x) to P(c)
- With every characteristic i in x: =, multiply P(c|x) by P(i|c) if x[i] is seen into training data.

As anticipated class for x, select class having the highest posterior probability.

The supervised machine learning method known as Random Forest is useful for regression and classification jobs alike. Predictions are made by merging outputs of numerous Decision Trees, since its ensemble learning process.

By generating a large number of Decision Trees and averaging their predictions, Random Forest achieves reliable results with high accuracy. The creation of more trees contributes to the robustness of the outcomes.

V. EXPERIMENTAL RESULTS

To help farmers choose the most suitable crops for their area, we model agricultural data using fifteen different ML methods.

Temperature, soil pH, precipitation, humidity, and the soil's potassium (P) to phosphorus (K) ratio are some of the soil factors contained in the dataset. Apple, banana, cotton, black gramme, watermelon, chickpeas, coconut, grapes, jute, kidney beans, grape, lentil, orange, and a slew of other crops are included in the 2200 records that make up the crop prediction dataset. Temperature, soil pH, precipitation, humidity, and the soil's potassium (P) to phosphorus (K) ratio are some of the soil factors contained in the dataset. Training set makes use of 67% of the data, while the testing set makes utilisation of another 35%. The rates of accuracy and error are shown in Table 1. All of the other classification algorithms achieve accuracy levels over 90%, with the Regression algorithm achieving a result of 97.33%. A more formalised version of error metrics is as follows:

$$(Po - Pe)*(1 - Pe)\text{-}1 = K \text{ (Kappa)} \quad (3)$$

WherePe is the theoretical possibility of a planned pact and Po is the impartial observed agreement.

MAE\ value:

$$I = N \sum i{=}1 \; |yi - xi \,|\, n \quad (4)$$

If yi is the forecast, xi is the true value, and n is the total quantity of data.

$$RMSE = n \sum i{=}1 \; (xi - x\ 0\ i) \quad (5)$$

where xi 0 is the predicted value and xi is the observed value.

$$RAE = n \sum i{=}1 \; |yi - y\}\ 0\}\ i\}|\ n \sum i{=}1 \; |yi - y| \quad (6)$$

where y is the data's average value.

Root Relative Squared Error (RRSE) is equal to

$$A = n \sum i = 1.Pi - Tj \sim 2\ n \sum i{=}1 \; (Ti - T) \quad (7)$$

If the goal value is T and the projected value is P.

Table 1: Error Values

Method	Accuracy	RMSE	RAE	RRSE
Bayes	98.49	0.017	1.12	7.64
Logistic	97.67	0.028	1.88	11.02
Regression	97.33	0.032	10.42	19.33

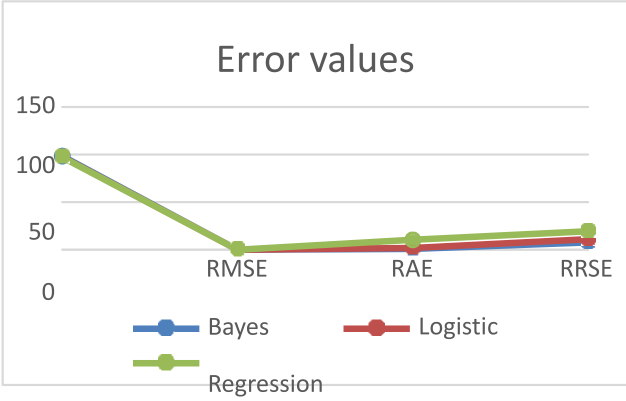

Figure 3: Error Values Variation

Table 2 Displays build and test times for each algorithm. Because the multilayer perception algorithm uses an artificial neural network and has multiple hidden layers along with input and output layers, it can be seen that its construction time is shorter than other Longer than the algorithm. Construction time is increased by multiple iterations of the MLP algorithm that searches for the optimal model given the input and output sets. The testing time of KSTAR and LWL algorithms is longer than others. When it comes to accuracy and processing speed, NBC algorithm is the most effective method for categorizing agricultural data.

Table 2: Build and Test times

Method	Build Time	Test Time
Bayes	0.44	0.23
Logistic	3.76	0.01
Regression	2.3	0.04

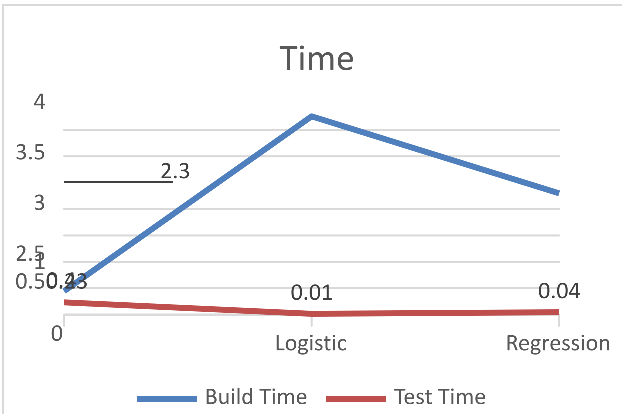

Figure 4: Test Duration in Seconds

Table 3 The precision and error measurement of the Multilayer perception (MLP) algorithm vary with different training data sample sizes. With 20% of the data used for training, the accuracy is 92.45%, while it increases to 98.45% when 90% of the data is utilized. Achieving high accuracy and efficient time complexity is essential for the MLP algorithm. Table 4 shows the Multilayer perception (MLP) algorithm's build and test timings for several scenarios where the percentages of training and testing data are changed. It is noteworthy that compared to other methods, the MLP approach requires longer build times [59, 60]. The findings indicate that using 60% of the training set results in the shortest build time, while a 20% training set leads to the longest build time. Despite this variation, test times remain consistently low across all scenarios, with the highest test time recorded at 0.02 when using a 20% training set.

Table 3: Multilayer perception

Training set	Accuracy	RMSE	RAE	RRSE
20%	92.45	0.061	18.05	33.59
60%	96.85	0.042	5.41	19.66
90%	98.45	0.031	3.31	14.78

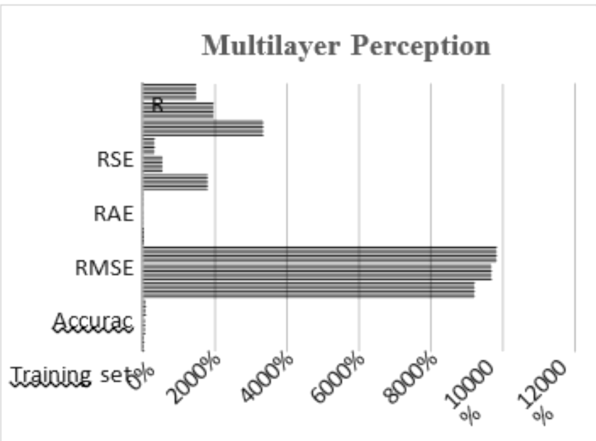

Figure 5: Multilayer perception accuracy and error rates

Table 4: Build and Test times for Multilayer perception

Training set	Build Time	Test Time
20%	16.02	0.02
60%	13.11	0.03
90%	12.22	0.01

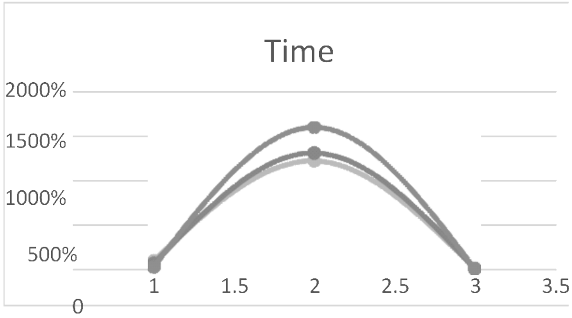

Figure 6: Multilayer perception duration

After implementing machine learning classification upon a dataset with 7 features and a single label representing crop type, we recorded the accuracy of various algorithms in Table 1. To further investigate the minimum number of features required for optimal accuracy in algorithm learning and prediction, additional simulations were conducted.

The results of four scenarios, where we chose three to four features and assessed crop detection accuracy for each set, are presented in Table 5. The highest accuracy, reaching 96.04%, was achieved with the second set of features (temperature, humidity, pH, and rainfall) using the Bayes Net algorithm. In contrast, utilising features N, P, and K resulted in poorest prediction outcomes.

Table 5: Accuracy of Various Set

Method	N(T,H,R)	P(T,H,R)	K(T,H,R)
Bayes	88.60	96.04	84.67
Logistic	73.72	84.15	71.85
Regression	85.47	94.97	83.21

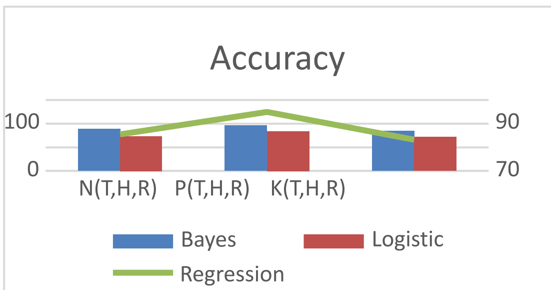

Figure 7: Accuracy at different set

Table 5's findings highlight the significance of feature selection for accurate crop identification utilising machine learning techniques. The traits we discovered will be useful in future analyses of agricultural data. Data set may be used as a roadmap to identify important traits for crop resolution; it includes weather conditions such as temperature, humidity, pH, and precipitation.

VI. CONCLUSION

In order to maximise harvest yields and minimise waste via data-driven decision-making, our research highlighted the crucial role of IoT devices and ML algorithms in contemporary agriculture. Experimental results showed how changes in labels impact the efficiency of data processing algorithms, and our findings included parameters like accuracy, mistake rates, build time, and test length for every classification technique.

The findings show that farmers may greatly improve their decision-making when it comes to variables impacting crop development by analysing a variety of agricultural data, including real-time information from IoT devices. Despite the difficulties, our findings point to the rising importance of machine learning approaches in agricultural production prediction.

We used a number of machine learning algorithms to examine crops in this experiment; by doing so, we were able to make predictions regardless of how challenging it was to identify individual crop kinds. The significance of feature selection in machine learning for proper interpretation of agricultural data is emphasised by our study. Using the BayesNet method and features such as temperature, humidity, pH, and precipitation, the research reached an accuracy high of 96.04%. The study's findings highlight the promising future of these technologies in contemporary farming. More research and development in this field has the potential to increase food security throughout the world by increasing agricultural yields while decreasing food waste.

To improve our evaluation of agricultural data, we will add sensor data from other geographies and Internet of Things (IoT) devices that use GPS in future studies. These results will be analysed using a machine learning approach, which will enable us to construct a thorough data assessment pool. Furthermore, we will separately investigate other plant species belonging to the same variety. This will allow us to use various machine learning algorithms to determine which species produces the best output.

REFERENCES

[1] Li, L.; Wang, B.; Feng, P.; Liu, D.L.; He, Q.; Zhang, Y.; Wang, Y.; Li, S.; Lu, X.; Yue, C.; et al. Developing a machine learning model with multi- source environmental data to predict wheat yield in China. Comput. Electron. Agric. 2022, 194, 106790. [CrossRef]

[2] Van Klompenburg, T.; Kassahun, A.; Catal, C. Crop yield prediction using machine learning: A systematicliterature review. Comput. Electron. Agric. 2020, 177, 105709. [CrossRef]

[3] Kuradusenge, M.; Hitimana, E.; Hanyurwimfura, D.; Rukundo, P.; Mtonga, K.; Mukasine, A.; Uwitonze, C.; Ngabonziza, J.;Uwamahoro, A. Crop Yield Prediction Using Machine Learning Models: Case of Irish Potato and Maize. Agriculture 2023, 13, 225.[CrossRef]

[4] Xu, W.; Kaili, Z.; Tianlei, W. Smart Farm Based on Six-Domain Model. Proceedings of the IEEE 4th International Conference on Electronics Technology (ICET), Chengdu, China, 7–10 May 2021; pp. 417–421.

[5] Moysiadis, V.; Tsakos, K.; Sarigiannidis, P.; Petrakis, E.G.M.; Boursianis, A.D.; Goudos, S.K. A Cloud Computing web-based application for Smart Farming based on microservices architecture. In Proceedings of the 11th International Conference onModern Circuits and Systems Technologies (MOCAST), Bremen, Germany, 8–10 June 2022; pp. 1–5.

[6] Ranjan, P.; Garg, R.; Rai, J.K. Artificial Intelligence Applications in Soil & Crop Management. In Proceedings of the IEEEConference on Interdisciplinary Approaches in Technology and

Management for Social Innovation (IATMSI), Gwalior, India,21–23 December 2022; pp. 1–5.

[7] Oré, G.; Alcântara, M.S.; Góes, J.A.; Oliveira, L.P.; Yepes, J.; Teruel, B.; Castro, V. Crop Growth Monitoring with Drone-BorneDInSAR. Remote Sens. 2020, 12, 615. [CrossRef]

[8] Gehlot, A.; Sidana, N.; Jawale, D.; Jain, N.; Singh, B.P.; Singh, B. Technical analysis of crop production prediction using MachineLearning and Deep Learning Algorithms. Proceedings of the International Conference on Innovative Computing, IntelligentCommunication and Smart Electrical Systems (ICSES), Chennai, India, 24–25 September 2022; pp. 1–5.

[9] Vashisht, S.; Kumar, P.; Trivedi, M.C. Improvised Extreme Learning Machine for Crop Yield Prediction. Proceedings of the 3rdInternational Conference on Intelligent Engineering and Management (ICIEM), London, UK, 27–29 April 2022; pp. 754–757.

[10] OpenAI. New and Improved Content Moderation Tooling.OpenAI. 2022. Available online: https://openai.com/blog/new-andimproved- content-moderation-tooling/ (accessed on 1 April 2023).

[11] Google. Bard Chatbox.Google. Available online: https://bard.google.com (accessed on 2 April 2023).

[12] Dean, J. The deep learning revolution and its implications for computer architecture and chip design. Proceedings of the IEEEInternational Solid-State Circuits Conference-(ISSCC), San Francisco, CA, USA, 16–20 February 2020.

[13] Cui, Y.W.; Henrickson, K.; Ke, R.; Pu, Z.; Wang, Y. Traffic graph convolutional recurrent neural network: A deep learning framework for network-scale traffic learning and forecasting. IEEE Trans. Intell. Transp. Syst. 2019, 21, 4883–4894. [CrossRef]

[14] Shahrin, F.; Zahin, L.; Rahman, R.; Hossain, A.J.; Kaf, A.H.; Abdul Malek Azad, A.K.M. Agricultural Analysis and Crop YieldPrediction of Habiganj using Multispectral Bands of Satellite Imagery with Machine Learning. Proceedings of the 11thInternational Conference on Electrical and Computer Engineering (ICECE), Dhaka, Bangladesh, 17–19 December 2020; pp. 21–24.

[15] Tawseef, A.S.; Tabasum, R.; Faisal, R.L. Towards leveraging the role of machine learning and artificial intelligence in precision agriculture and smart farming. Comput.Electron.Agric. 2022, 198, 107119.