

# MACHINE LEARNING-BASED CROP YIELD PREDICTION SYSTEM USING ENVIRONMENTAL AND SOIL PARAMETERS

Adhwaitha.S.K<sup>1</sup>, Nandha Kumar.A.<sup>2</sup>, Dr.D.J.Anitha Merlin<sup>3</sup>

<sup>1,2</sup>UG Student Department of Computer Science and Data Science, Nehru Arts and Science College, Coimbatore, Tamil Nadu, India.

<sup>3</sup>Assistant Professor, Department of Computer Science and Data Science, Nehru Arts and Science College, Coimbatore, Tamil Nadu, India.

**ABSTRACT-** Predicting crop yields accurately really matters for boosting farm output and keeping food supplies steady. In this work, there's a machine learning setup that looks at environmental stuff and soil conditions to figure out crop production in a solid way. It digs into things like temperature, rainfall, humidity, how moist the soil is, the pH level there, and the nutrients available, all to see how they affect plants growing. I think using historical data from farms and weather helps train different algorithms. Like linear regression, decision trees, and random forest, those get put through the paces. Before that, data gets cleaned up and features are picked out carefully, which seems to make the models work better and predict more right. The experiments point to ensemble models doing superior over older approaches, at least in accuracy. That part stands out, though it's not always clear why in every case. Farmers and planners can lean on this system for choices about what to plant, when to irrigate, or how much fertilizer to use. It pushes for using resources smartly and farming in a way that's sustainable over time. Precision agriculture gets a boost from this, and it ties into bigger agricultural progress, I suppose. Some details on long term effects feel a bit open ended still.

## I. INTRODUCTION

To keep up with the growing need for food and make sure everyone has enough, we really have to push for better ways to farm, like using new tech and smarter methods. A lot depends on the local weather and climate, plus how often extreme events hit, and that all shapes how farmers handle their crops. When soil lacks nutrients, plants do not grow right, their processes slow down, and yields drop off. So, for keeping soil good over time, especially in places with low organic stuff, mixing organic and inorganic fertilizers seems key, and balancing them out matters a ton. The kind and amount of fertilizers really drive soil fertility, how crops develop, and if yields stay sustainable. Getting nutrients right and adding what is missing on time helps hit the best production levels, I suppose. Farmers sometimes pick crops without thinking much about the site, so they end up growing in tough spots and get lower outputs. That points to why we need solid systems that suggest the right crops for each place, accurate and quick ones. Bringing in advanced tech and automated stuff can

help with keeping the environment in check, managing soil better, and digging into all sorts of farm data. It lets people improve how they grow crops, based on what resources they have, how well things adapt, and boosting productivity overall. Lately, studies show big steps forward in predicting yields, classifying crops, spotting diseases, and checking health with machine learning and deep learning. Some folks built smart systems for recommending crops and forecasting yields using different algorithms, and they work pretty well for helping farmers decide things. But a bunch of these studies fall short, like not covering all the important factors or training models on mixed up crop types. Earlier work stuck to just a few soil or climate things, but crops actually respond to a whole mix of connected factors. To predict growth accurately, you need data from all over different areas, since nutrients and how things react to the environment change by crop and spot. How soil nutrients, pH, and climate together affect things like accuracy, precision, recall, and F1-score in models, that has not been looked at enough, it feels like. This study tries to fill some of those holes by using a big public dataset from India, pulled from Kaggle, that includes soil pH, NPK fertilizers, and stuff like rainfall, temperature, humidity. The idea is to test how well machine learning models do at giving reliable crop suggestions and pointing out fixes to get maximum yields. It could stretch to other places with similar setups and crop habits, maybe.

## II. METHODOLOGY

### A. Data

This dataset comes from the Kaggle site, citation 39, and the Indian Chamber of Food and Agriculture put it together over a period of time, that's 40. There are

2100 samples in total. It looks at 11 kinds of agricultural crops plus 10 horticultural ones. Along with that, there is data on NPK fertilizers, soil pH levels, and things like rainfall, temperature, humidity. The scientific names for the crops, they are all listed in the supplementary file S1. Now for the agricultural crops specifically, the average nitrogen use is 56 kg per hectare. Phosphorus comes in at 52.11, and potassium is 31.64. The environmental stuff averages temperature of 24.89 degrees C, with a standard deviation of 4.02. Humidity is 64.20 percent, sd 24.10. pH around 6.67, plus minus 0.85. Rainfall 90.90 mm, sd 61.64. Horticultural crops have different numbers, sort of. Mean nitrogen is 47.52 kg/ha, phosphorus 53.29, but potassium goes up to 69.09. For climate, temperature is 26.19 C with sd 5.83, humidity higher at 81.82 percent sd 14.80, pH 6.31 sd 0.57, and rain 112.67 mm sd 43.47. It seems like they need more water or something. We grouped the crops into three categories to test the models. Agricultural crops as AC, horticultural as HC, and then a combined agri-horticultural one called Co. That leads to the AC-model, HC-model, and the mixed crop model. What makes this dataset useful, I think, is how it includes crops from various regions geographically. And a broad range of types too. So probably it could apply to other places that have similar conditions, at least that's the idea. Not sure if it covers everything perfectly though.

#### *B. Prediction of crop using ML techniques*

By examining important variables like soil pH, NPK fertilizer levels, and climate, this study sought to use regression-based machine learning models to forecast the best crop choice (Fig. 1). Support Vector Machine (SVM), Random Forest (RF), eXtreme Gradient Boosting (XGBoost), K-Nearest Neighbors (KNN), and Decision Tree (DT) were the five algorithms used for this purpose.

##### *(i) Decision tree*

A decision tree is a type of classification model which classifies crops based on attribute-value tests like NPK values, pH value and rainfall. The tree is built upon these tests, using a training dataset, where each sample has input attributes and an associated crop category. At each step, the best attribute that correctly separates crop classes is chosen and recursively sorted until trees of varying complexity

are created. Pruning techniques are used to prevent overfitting and improve model generalization and portability. The most common hyperparameter optimization was performed using GridSearchCV.

##### *(ii) extreme gradient boosting*

The trained model uses XGBoost as its advanced machine learning algorithm because it implements gradient-boosted decision trees which deliver flexible performance, efficient computations, and accurate predictive results. The model builds its ensembles through the combination of multiple decision trees which use hierarchical data partitioning based on their chosen features. The training process develops new trees which correct previous model errors to enhance total prediction accuracy. The system assigns each data sample to a leaf node which links to a numerical weight that shows how much input features affect the sample. The XGBoost model performance required multiple hyperparameters to undergo precise tuning procedures. The learning rate was set to 0.1, the maximum tree depth to 17, the number of estimators to 200, the subsample rate to 0.5, the gamma parameter to 0, and the random seed to 50. These parameters were selected to balance model speed, accuracy, and flexibility, ensuring optimal predictive performance.

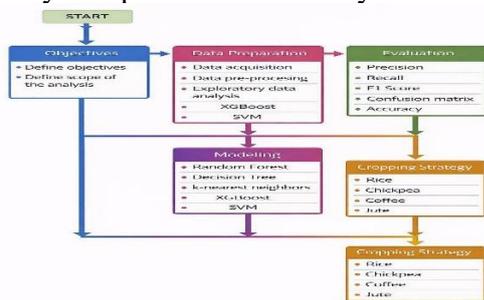
##### *(iii) Support vector machine*

Support Vector Machine (SVM) operates as an effective classification method which achieves high training precision through its straightforward model design. The system establishes optimal hyperplanes which create maximum separation between various classes to decrease generalization error risks. The algorithm decreases model complexity through its reduction of support vectors which represent the data points situated nearest to the decision boundary thus achieving better generalization results while preventing overfitting. SVM utilizes appropriate kernel functions to determine the best separating hyperplane in a high-dimensional feature space. The research selected a regularization parameter of  $C = 5$  to manage the relationship between acceptable misclassification levels and the complexity of the model. The system develops better generalization results through moderate  $C$  values which create better results, while higher  $C$  values result in

overfitting problems. The system used gamma value 0.1 together with the radial basis function (RBF) kernel to change input data into a higher-dimensional space. This setup allows SVM to effectively identify non-linear feature connections which leads to better prediction results and system performance.

*(iv) Random forest*

Random Forest (RF) functions as an ensemble learning method which combines multiple decision trees to enhance predictive accuracy while decreasing generalization errors. The model creates a tree collection through selecting different training samples and feature groups, and it uses majority voting to combine the results from all trees to make its final prediction. RF uses random feature selection and data partitioning techniques to create tree models which maintain their performance under unpredictable conditions and reduce their tendency to overfit. The system requires a splitting criterion which needs to examine each node division for quality through two standard methods of measurement which use either entropy measurement or Gini impurity measurement to assess node division quality. In this study, the RF model was optimized using the entropy criterion and tuned with a maximum depth of 6, maximum features of 5, minimum samples per split of 4, a random state of 0, and 15 estimators. The selected parameter settings aim to produce an optimal balance between model complexity and prediction accuracy.



**Fig. 1. Workflow for Conducting Analysis and Recommending Cropping Strategies Using Machine Learning Techniques**

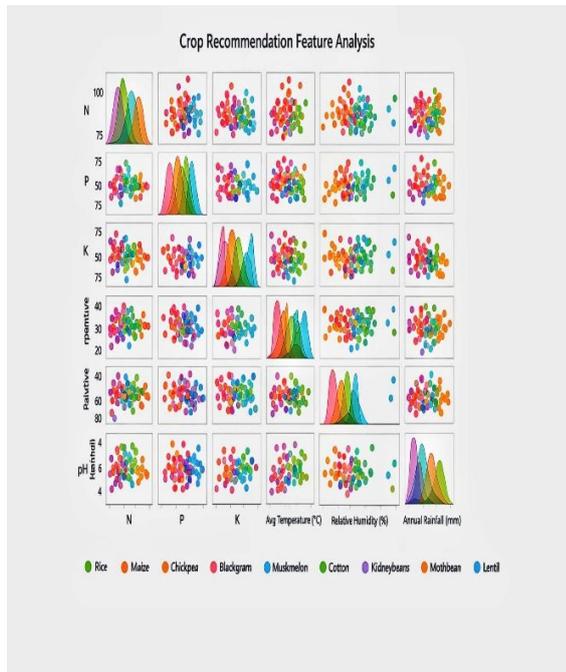
v. *K-nearest neighbours*

The K-Nearest Neighbours (KNN) algorithm serves as a popular classification technique which searches through the dataset to find the \*k\* training samples that have the closest resemblance to a target instance.

The class label of the target sample is then determined based on the majority category among these nearest neighbours. KNN classification relies on its closest data points because the algorithm uses only those points for its analysis. The KNN algorithm works best with datasets that have low dimensionality and small neighbourhood sizes because its effectiveness decreases when dealing with high-dimensional data. The researchers used the Minkowski distance metric to measure sample distances because this metric provides better accuracy results than all other distance measurement methods. The model was set to use \*k\* equal to 3 which meant that the system made classification choices by analyzing the three closest neighbours.

*C. Training, testing and performance evaluation*

Data scientists who work with imbalanced datasets face difficulties because their models will develop learning biases which lead to unreliable predictions for minority class instances. The researchers used an under-sampling method to create a balanced dataset which solved this problem. The researchers used machine learning algorithms to find the best cropping recommendations. The researchers split the dataset into two parts by allocating 70 percent of the data for model development and keeping 30 percent for testing purposes. The researchers used multiple evaluation methods to assess model performance because accuracy alone does not capture how well a model classifies data in situations with imbalanced classes. The research team employed multiple evaluation techniques which included recall, specificity, precision, F1-score, and AUC measurements to assess performance according to established standards [10,43]. AUC served as the primary evaluation metric because it delivered the most dependable predictions which enabled the researchers to evaluate various models.



**Fig. 2. The plot matrix illustrates the relationships among NPK levels, soil pH, temperature, humidity, and rainfall for agricultural crops grown in different regions of India, with diagonal cells showing the kernel density distribution of each variable.**

### III. RESULTS

#### A. Requirements for nutrients and environmental factors of agriculture and horticulture crops

The plot matrices show how NPK requirements depend on soil pH and temperature and humidity and rainfall for different agricultural and horticultural crops. Agricultural crops require specific temperature ranges for growth except black gram and moth bean and mung bean and muskmelon which need warmer weather. Muskmelon and mung bean and rice grow better in humid areas but chickpea and kidney bean thrive in dry environments. Mung bean shows high adaptability to a wide range of soil pH levels, while most agricultural crops prefer a pH range of 6–8. Kidney bean favors slightly acidic soil. Rice grows best when farmers apply NPK at moderate levels and when rain falls heavily and when soil pH approaches neutral. Leguminous crops need low nitrogen together with moderate to high phosphorus requirements while chickpea and black gram need more nitrogen. Black gram and chickpea and lentil

and kidney bean need more phosphate than muskmelon which needs only small amounts of phosphate in high-humidity conditions. Most legumes have low potassium requirements, except chickpea, which needs a comparatively higher amount. Cotton requires high nitrogen levels and grows best in moderate temperatures whereas coconut needs heavy rainfall and high humidity levels. Apple requires more nitrogen and phosphorus and potassium than any other horticultural crop. Grapes need more potassium and phosphorus than they need nitrogen while coconut requires less NPK than any other plant. Most crops prefer slightly acidic soils, whereas mango and orange grow best in neutral to slightly alkaline conditions.

#### B. Comparative analysis of ML classification performance

The research tested different machine learning algorithms which included Random Forest Decision Tree and Support Vector Machine to determine the best agricultural and horticultural crop recommendation system. The execution of the models followed the training and optimization process which used the specified parameters from the methodology section. The researchers examined the predicted results for agricultural horticultural and mixed crop models according to their actual test results. XGBoost achieved the highest test accuracy of 99.09% among agricultural crop models while KNN showed the lowest accuracy at 94.45%. The horticultural crop models showed identical results regarding their prediction accuracy and precision and recall metrics. The mixed crop modelling system which used agricultural and horticultural datasets achieved an accuracy rate of 98.5% which fell short of the success rates established through testing separate crop categories. The comparative analysis demonstrated that XGBoost-based agricultural and horticultural crop models performed better than all other methods whereas KNN-based systems showed reduced effectiveness. XGBoost performed better in mixed crop modelling which resulted in perfect AUC scores for most crops and high AUC scores for the other crops. Although most models achieved strong micro- and macro-average AUC scores XGBoost

demonstrated superior skills in identifying specific crop types.

### *C. Analyses of confusion matrix to visualize the recommendation response of the models*

The evaluation of crop recommendation models used confusion matrix analysis, which assesses actual class designations against forecasted designations. The diagonal elements represent correctly classified samples, while off-diagonal elements indicate misclassifications. The XGBoost-based method outperformed all other machine learning techniques when used to study agricultural crop models. The system made errors because it misidentified some black gram samples as lentil and moth bean. The Random Forest model misidentified lentil and moth bean samples as black gram, while it also mistook jute and lentil samples for rice and moth bean samples. The models, which include all systems except XGBoost, showed a pattern of misidentifying chickpea and cotton as kidney bean and maize. The models used for horticultural crop identification showed misclassification results, which included their confusion between watermelon and mango and between grapes and pomegranate. The KNN model displayed critical inaccuracies when it mistaken coconut for orange and grapes for pomegranate and pomegranate for orange. The XGBoost and Random Forest horticultural models achieved better results than all other modeling methods. The mixed crop model demonstrated a comparatively higher rate of misclassification. The Random Forest-based mixed model mistakenly identified apple and coconut samples as orange while it classified lentil as moth bean and mung bean and papaya as cotton. The results demonstrate that individual crop category models reached high accuracy, yet the combination of agricultural and horticultural crop models led to complex classification challenges that diminished prediction accuracy.

## IV. DISCUSSION

The selection of appropriate crops for cultivation depends largely on soil nutrient availability and crop requirements for essential elements such as nitrogen, phosphorus, and potassium which are affected by

environmental factors that include soil pH and rainfall and temperature and humidity. Farmers select crops for each particular location based on their personal experience. The decisions which farmers need to make regarding their crops have become more difficult because of climate change and soil fertility decline, which frequently results in unproductive results. Machine learning techniques provide efficient solutions for creating crop recommendations which use data from soil nutrient levels and current weather conditions. The research tested how different machine learning algorithms perform in agricultural settings by using actual field data which included multiple measurement variables from various Indian regions that researchers had collected from a public database. The results demonstrate that an ML-based decision support system which used a non-mechanistic method can accurately recommend appropriate cropping patterns for different environmental conditions. The use of multi-environmental datasets played a crucial role in achieving these outcomes.

### *A. Performances of the models*

The results indicate that XGBoost serves as an extremely efficient model which predicts both agricultural and horticultural crop categories. The KNN model showed lower accuracy when it tried to identify particular agricultural crops because its AUC scores failed to match the actual performance. XGBoost achieved better performance than other methods when it recommended horticultural crops because it produced high macro and micro average AUC results together with exceptional classification accuracy across various crop categories. The AUC scores reached their highest values for coconut and pomegranate and orange, while most other horticultural crops achieved near-perfect classification performance. The findings match earlier research results which used advanced machine learning and optimization methods to achieve high predictive accuracy for crop recommendation and yield prediction. Research studies have shown that XGBoost Random Forest KNN and Decision Tree models achieve high AUC values when used for crop classification based on soil nutrient data. Ensemble-based machine learning

methods have proved their capacity to make accurate agricultural yield predictions. The results demonstrate that the models used in this research study successfully predict crop suitability across various NPK levels and climatic conditions throughout India, which shows their value for precision agriculture and decision support system applications.

### *B. Possible causes of miss-recommendation*

The machine learning models demonstrated variations in their performance when applied to different crop categories. The agricultural crop prediction models produced identical patterns of incorrect recommendations which all models showed. Rice was recommended instead of jute because both crops shared similar NPK fertilizer and temperature requirements. The misclassification between cotton and maize took place because both crops had identical rainfall and phosphorus–potassium fertilizer and temperature requirements. The horticultural crop prediction exhibited the same pattern which led to orange being misidentified as coconut because both fruits shared similar NPK and humidity requirements. The mixed crop model showed diminished success rates because it failed to identify effective cropping methods. The combination of agricultural and horticultural data resulted in higher misclassification rates which led to greater confusion about the data. The research results show that separate models for each crop category produce better results than a unified model for all crops. The similarity of agronomic and environmental requirements across various crop types creates challenges for accurate classification when all crops are modelled together.

### *C. Implication of the study*

The research study investigates the selective environmental requirements and nutrient needs of multiple agricultural and horticultural crops through its complete examination of NPK nutrient needs and climatic requirements for their cultivation. The findings show that rice and jute need high rainfall and high humidity levels for their growth, while other crops thrive under conditions of moderate water availability. The study confirmed that optimal

relative humidity levels match previous research findings, which showed that low humidity can disrupt flowering. Leguminous crops needed low nitrogen and medium to high phosphorus levels, while their potassium requirements varied. Horticultural crops grow best within particular temperature and soil pH ranges, while different climate and nutrient and soil condition requirements exist between different horticultural crops. The potassium and phosphorus requirements of grapes and apples exceed those needed for coconut growth, which needs less fertilizer. Understanding how nutrients interact with environmental factors is essential for making correct decisions about fertilizer use. The results show that machine learning models possess strong capabilities to determine which crops will thrive in particular environmental settings. Previous studies investigated machine learning applications for crop recommendation and yield prediction, but the current study achieved superior results. Different algorithms and feature selection methods produced moderate to high accuracy results in earlier research, while deep learning and ensemble methods demonstrated effective performance. The study results show similar patterns found in previous research because they confirm existing trends.

## V. CONCLUSION

The current farming system has been revolutionized through precision agriculture which provides this study to agricultural research by using machine learning models to evaluate crop suitability based on available nutrients and climate conditions and soil pH. The need for precise nutrient control exists because insufficient nutrient supply will decrease crop yield together with long-term soil fertility. Agricultural and horticultural crops exhibit different nutritional needs because they respond differently to environmental changes which include NPK levels and climate conditions. The study requires separate models because different crop categories need separate modeling according to its findings. The results indicate that specifically trained models for individual crop classes enable crop recommendations that become more efficient and faster and more precise. The findings provide

essential guidance to farmers who work in rural zones because they help farmers to stop using trial-and-error methods. The proposed approach supports the creation of decision support tools which help users determine optimal crop yield according to local environmental conditions. The systems deliver accurate crop recommendations which enable farms to achieve higher productivity while using resources more efficiently, which leads to greater profits and enhanced food security. The research produces strong findings but the study depended on publicly accessible data, which limits its findings. The study findings will gain more reliability and broader applicability through the inclusion of datasets from multiple geographic areas. The research should conduct extensive field testing throughout different environmental situations to g.

#### FUNDING

This research did not receive any specific grant from public, commercial, or nonprofit funding agencies. The Center for Research in Environment, iGen, and Livelihood (CREGL) provided significant operational assistance which enabled researchers to finish their work.

#### DATA AVAILABILITY STATEMENT

The data supporting the findings of this study will be made available upon request.

#### CREDIT AUTHORSHIP CONTRIBUTION STATEMENT

Biplob Dey managed all aspects of the original draft creation while he performed tasks of visualization validation and methodology development and formal analysis and data curation and conceptualization. Jannatul Ferdous handled all tasks of review and editing work together with visualization and formal analysis. Romel Ahmed performed review and editing work while he supervised projects and managed them and conducted research and developed project ideas.

#### DECLARATION OF COMPETING INTEREST

The authors declare that there are no financial or personal associations or interests that may have influenced the presentation of the work in this study.

#### ACKNOWLEDGEMENT

Our organization expresses genuine appreciation to the Indian Chamber of Food and Agriculture, the Indian Agricultural Extension Department, and the anonymous farmers who assisted in creating this important dataset.

#### REFERENCES

- [1] S.W. Wang, W.K. Lee, Y. Son, An assessment of climate change impacts and adaptation in South Asian agriculture, *Int. J. Clim. Chang. Strateg. Manag.* 9 (2017) 517–534, <https://doi.org/10.1108/IJCCSM-05-2016-0069>.
- [2] K.K. Verma, X.P. Song, A. Joshi, D.D. Tian, V.D. Rajput, M. Singh, J. Arora, T. Minkina, Y.R. Li, Recent trends in nano-fertilizers for sustainable agriculture under climate change for global food security, *Nanomaterials* 12 (2022) 1–25, <https://doi.org/10.3390/nano12010173>.
- [3] X. Liu, Y. Xu, S. Sun, X. Zhao, P. Wu, Y. Wang, what is the potential to improve food security by restructuring crops in Northwest China? *J. Clean. Prod.* 378 (2022) 134620 <https://doi.org/10.1016/J.JCLEPRO.2022.134620>.
- [4] A. Bouguettaya, H. Zarzour, A. Kechida, A.M. Taberkit, Deep learning techniques to classify agricultural crops through UAV imagery: a review, *Neural Comput. Appl.* 34 (2022) 9511–9536, <https://doi.org/10.1007/S00521-022-07104-9>, 2022 3412.
- [5] United Nations, *World Population Prospects 2019*, 2019. New York, <https://population.un.org/wpp/Download/Standard/Population/>.
- [6] W.H. Meyers, N. Kalaitzandonakes, W.H. Meyers, N. Kalaitzandonakes, *World population, Food Growth, and Food Security Challenges* 15 (2015) 161–177, <https://doi.org/10.1108/S1574-871520>.
- [7] Shoukat fiza, M. Awais, M. Khalid, Asad Sohail, the role of genetically-modified (GM) crops in food security, *Life Sci. J.* 19 (2022) 26–30, <https://doi.org/10.7537/marslsj190222.04>.
- [8] G. Clarkson, P. Dorward, S. Poskitt, R.D. Stern, D. Nyirongo, K. Fara, J.M. Gathenya, C.G. Staub, A. Trotman, G. Nsengiyumva, F. Torgbor, D. Giraldo, Stimulating small-scale farmer innovation and adaptation with participatory integrated climate services for agriculture (PICSA): lessons from successful implementation in africa, Latin America,

- the caribbean and South Asia, *Clim. Serv.* 26 (2022) 100298, <https://doi.org/10.1016/J.CLISER.2022.100298>.
- [9] R. Islam, R. Ahmed, B. Dey, S. Haque, S. Aktar, M.S. Bhuiyan, M.S. Arif, M.A. Habib Ador, M.M. Ul Haque, N. Saha, Salinity hazard drives the alteration of occupation, land use and ecosystem service in the coastal areas: evidence from the south-western coastal region of Bangladesh, *Heliyon* 9 (2023) e18512, <https://doi.org/10.1016/j.heliyon.2023.e18512>.
- [10] B. Dey, M. Masum Ul Haque, R. Khatun, R. Ahmed, Comparative performance of four CNN-based deep learning variants in detecting Hispa pest, two fungal diseases, and NPK deficiency symptoms of rice (*Oryza sativa*), *Comput. Electron. Agric.* 202 (2022) 107340, <https://doi.org/10.1016/j.compag.2022.107340>.
- [11] J.-H. Chen, J.-T. Wu, C. Young, The Combined Use of Chemical, Organic Fertilizers And/or Biofertilizer for Crop Growth and Soil Fertility, 2007, <https://doi.org/10.30058/SE.200706.0001>.
- [12] M.A. Saleque, M.J. Abedin, N.I. Bhuiyan, S.K. Zaman, G.M. Panaullah, Long-term effects of inorganic and organic fertilizer sources on yield and nutrient accumulation of lowland rice, *Field Crops Res.* 86 (2004) 53–65, [https://doi.org/10.1016/S0378-4290\(03\)00119-9](https://doi.org/10.1016/S0378-4290(03)00119-9).
- [13] D. Batool, M. Shahbaz, H. Shahzad Asif, K. Shaukat, T.M. Alam, I.A. Hameed, Z. Ramzan, A. Waheed, H. Aljuaid, S. Luo, A hybrid approach to tea crop yield prediction using simulation models and machine learning, *Plants* 11 (2022) 1925, <https://doi.org/10.3390/PLANTS11151925>, 2022, Vol. 11, Page 1925.
- [14] H. Burdett, C. Wellen, Statistical and machine learning methods for crop yield prediction in the context of precision agriculture, *Precis. Agric.* 23 (2022) 1553–1574, <https://doi.org/10.1007/s11119-022-09897-0>.
- [15] S. Fei, M.A. Hassan, Y. Xiao, X. Su, Z. Chen, Q. Cheng, F. Duan, R. Chen, Y. Ma, UAV-based multi-sensor data fusion and machine learning algorithm for yield prediction in wheat, *Precis. Agric.* (2022) 1–26, <https://doi.org/10.1007/s11119-022-09938-8>.
- [16] D. Paudel, H. Boogaard, A. de Wit, M. van der Velde, M. Claverie, L. Nisini, S. Janssen, S. Osinga, I.N. Athanasiadis, Machine learning for regional crop yield forecasting in Europe, *Field Crops Res.* 276 (2022) 108377, <https://doi.org/10.1016/J.FCR.2021.108377>.
- [17] A. Nihar, N.R. Patel, A. Danodia, Machine-learning-based regional yield forecasting for sugarcane crop in Uttar Pradesh, India, *J. Indian Soc. Remote Sens.* 50 (2022) 1519–1530, <https://doi.org/10.1007/S12524-022-01549-0>.
- [18] R. Tufail, A. Ahmad, M.A. Javed, S.R. Ahmad, A machine learning approach for accurate crop type mapping using combined SAR and optical time series data, *Adv. Sp. Res.* 69 (2022) 331–346, <https://doi.org/10.1016/J.ASR.2021.09.019>.
- [19] O.P. Duke, T. Alabi, N. Neeti, J. Adewopo, Comparison of UAV and SAR performance for Crop type classification using machine learning algorithms: a case study of humid forest ecology experimental research site of West Africa, *Https://Doi.Org/10.1080/01431161.2022.2109444*. 43 (2022) 4259–4286. <https://doi.org/10.1080/01431161.2022.2109444>.
- [20] S. Nandhini, K. Ashokkumar, Machine learning technique for crop disease prediction through crop leaf image, *Appl. Math. Inf. Sci.* 16 (2022) 149–158, <https://doi.org/10.18576/amis/160202>.
- [21] M. Sharif, M.A. Khan, Z. Iqbal, M.F. Azam, M.I.U. Lali, M.Y. Javed, Detection and classification of citrus diseases in agriculture based on optimized weighted segmentation and feature selection, *Comput. Electron. Agric.* 150 (2018) 220–234, <https://doi.org/10.1016/J.COMPAG.2018.04.023>.
- [22] E.C. Too, L. Yujian, S. Njuki, L. Yingchun, A comparative study of fine-tuning deep learning models for plant disease identification, *Comput. Electron. Agric.* 161 (2019) 272–279, <https://doi.org/10.1016/j.compag.2018.03.032>.
- [23] J. Chen, D. Zhang, A. Zeb, Y.A. Nanekaran, Identification of rice plant diseases using lightweight attention networks, *Expert Syst. Appl.* 169 (2021), <https://doi.org/10.1016/J.ESWA.2020.114514>.
- [24] H. Chu, C. Zhang, M. Wang, M. Gouda, X. Wei, Y. He, Y. Liu, Hyperspectral imaging with shallow convolutional neural networks (SCNN) predicts the early herbicide stress in wheat cultivars, *J. Hazard Mater.* 421 (2022) 126706, <https://doi.org/10.1016/j.jhazmat.2021.126706>.
- [25] P. Bhatt, S. Sarangi, S. Pappula, Comparison of CNN models for application in crop health assessment with participatory sensing, *GHTC 2017 - IEEE Glob. Humanit. Technol. Conf. Proc.* (2017) 1–7, <https://doi.org/10.1109/GHTC.2017.8239295>, 2017-January.
- [26] P.R. Sai Sankar, S.D.p.s Ramakrishna, M.M. Venkata Rakesh, P. Raja, V.T. Hoang, C. Szczepanski, Intelligent health assessment system for paddy crop using CNN, 2021 3rd, *Int. Conf. Signal Process. Commun. ICPSC 2021* (2021) 382–387, <https://doi.org/10.1109/ICSPC51351.2021.9451644>.
- [27] N. Kaur, Devendran, S. Verma, Kavita, N.Z. Jhanjhi, De-noising diseased plant leaf image, in: 2022 2nd *Int. Conf. Comput. Inf. Technol.*, IEEE, Tabuk, Saudi Arabia, 2022, pp. 130–137, <https://doi.org/10.1109/ICCIT52419.2022.9711604>.
- [28] Z. Doshi, S. Nadkarni, R. Agrawal, N. Shah, AgroConsultant: intelligent crop recommendation system using machine learning algorithms, *Proc. - 2018 4th Int. Conf. Comput. Commun. Control*

- Autom. ICCUBEA (2018),  
<https://doi.org/10.1109/ICCUBEA.2018.8697349>,  
2018
- [29] V.C. Waikar, S.Y. Thorat, A.A. Ghute, P.P. Rajput, M.S. Shinde, Crop prediction based on soil classification using machine learning with classifier ensembling, *Int. Res. J. Eng. Technol.* 7 (2020) 4857–4861. [www.irjet.net](http://www.irjet.net). (Accessed 23 December 2022).
- [30] A. Dubois, F. Teytaud, S. Verel, Short term soil moisture forecasts for potato crop farming: a machine learning approach, *Comput. Electron. Agric.* 180 (2021) 105902, <https://doi.org/10.1016/J.COMPAG.2020.105902>.
- [31] M.U. Ahmed, I. Hussain, Prediction of wheat production using machine learning algorithms in northern areas of Pakistan, *Telecomm. Policy.* 46 (2022) 102370, <https://doi.org/10.1016/J.TELPOL.2022.102370>.
- [32] K. Bakthavatchalam, B. Karthik, V. Thiruvengadam, S. Muthal, D. Jose, K. Kotecha, V. Varadarajan, IoT framework for measurement and precision agriculture: predicting the crop using machine learning algorithms, *Technol.* 10 (2022) 13, <https://doi.org/10.3390/TECHNOLOGIES10010013>, 2022, Vol. 10, Page 13.
- [33] R.K. Rajak, A. Pawar, M. Pendke, P. Shinde, S. Rathod, A. Devare, Crop Recommendation System to Maximize Crop Yield Using Machine Learning Technique, *Academia.Edu.*, 2017. <https://www.academia.edu/download/55495855/IRJE-T-V4I12179.pdf>. (Accessed 23 December 2022).
- [34] N.H. Kulkarni, G.N. Srinivasan, B.M. Sagar, N.K. Cauvery, improving crop productivity through A crop recommendation system using ensembling technique, *Proc. 2018 3rd Int. Conf. Comput. Syst. Inf. Technol. Sustain. Solut. CSITSS* (2018) 114–119, <https://doi.org/10.1109/CSITSS.2018.8768790>, 2018.
- [35] D. Modi, A.V. Sutagundar, V. Yalavigi, A. Aravatagimath, Crop recommendation using machine learning algorithm, in: 2021 5th Int. Conf. Inf. Syst. Comput. Networks, 2021, <https://doi.org/10.1109/ISCON52037.2021.9702392>. ISCON 2021.
- [36] P. Parameswari, N. Rajathi, K.J. Harshanaa, Machine learning approaches for crop recommendation, in: 2021 Int. Conf. Adv. Electr. Electron. Commun. Comput. Autom. ICAECA 2021, 2021, <https://doi.org/10.1109/ICAECA52838.2021.9675480>.
- [37] R.K. Rajak, A. Pawar, M. Pendke, P. Shinde, S. Rathod, A. Devare, Crop recommendation system to maximize crop yield using machine learning technique, *Int. Res. J. Eng. Technol.* 4 (2017) 950–953. [www.irjet.net](http://www.irjet.net). (Accessed 23 December 2022).
- [38] G. Suresh, D.A.S. Kumar, D.S. Lekashri, D.R. Manikandan, C.-O. Head, Efficient crop yield recommendation system using machine learning for digital farming, *Int. J. Mod. Agric.* 10 (2021) 906–914. <http://modern-journals.com/index.php/ijma/article/view/688>. (Accessed 23 December 2022).
- [39] Kaggle. <https://www.kaggle.com/>, 2022. (Accessed 25 December 2022).
- [40] ICFA, Indian Chamber of Food and Agriculture, 2022. <https://www.icfa.org.in/>. (Accessed 25 December 2022).
- [41] Y. Hua, F. Li, S. Yang, Application of support vector machine model based on machine learning in art teaching, *Wireless Commun. Mobile Comput.* 2022 (2022), <https://doi.org/10.1155/2022/7954589>.
- [42] B. Dey, K.A.M. Abir, R. Ahmed, M.A. Salam, M. Redowan, M.D. Miah, M.A. Iqbal, Monitoring groundwater potential dynamics of north-eastern Bengal Basin in Bangladesh using AHP-Machine learning approaches, *Ecol. Indicat.* 154 (2023) 110886, <https://doi.org/10.1016/j.ecolind.2023.110886>.
- [43] B. Dey, R. Ahmed, J. Ferdous, M.M.U. Haque, R. Khatun, F.E. Hasan, S.N. Uddin, Automated plant species identification from the stomata images using deep neural network: a study of selected mangrove and freshwater swamp forest tree species of Bangladesh, *Ecol. Inform.* 75 (2023) 102128, <https://doi.org/10.1016/J.ECOINF.2023.102128>.
- [44] P.S.S. Gopi, M. Karthikeyan, Red fox optimization with ensemble recurrent neural network for crop recommendation and yield prediction model, *Multimed. Tool. Appl.* (2023) 1–21, <https://doi.org/10.1007/S11042-023-16113-2>.
- [45] T. Swathi, S. Sudha, Crop classification and prediction based on soil nutrition using machine learning methods, *Int. J. Inf. Technol.* 15 (2023) 2951–2960, <https://doi.org/10.1007/S41870-023-01345-0>.
- [46] S.G. Kundu, A. Ghosh, A. Kundu, G. G P, a ml-ai enabled ensemble model for predicting agricultural yield, *Cogent Food Agric.* 8 (2022), <https://doi.org/10.1080/23311932.2022.2085717>.
- [47] S.H. Ewald, S.A. Abed, A. Chabuk, N. Al-Ansari, Water footprint of rice in Iraq, *IOP Conf. Ser. Earth Environ. Sci.* 722 (2021) 012008, <https://doi.org/10.1088/1755-1315/722/1/012008>.
- [48] C.S. Murthy, M.K. Poddar, K.K. Choudhary, P. Srikanth, V. Pandey, S. Ramasubramanian, G.S. Kumar, Remote sensing based crop insurance for jute (*Corchorus olitorius*) crop in India, *Remote Sens. Appl. Soc. Environ.* 26 (2022) 100717, <https://doi.org/10.1016/j.rsase.2022.100717>.
- [49] V. Sridevi, V. Chellamuthu, Impact of weather on rice – a review, *Int. J. Appl. Res.* 1 (2015) 825–831.
- [50] H. Singh Jatav Sri Karan, J. Jena, S. Maitra, A. Hossain, B. Pramanick, H.I. Gitari, S. Praharaj, T. Shankar, J. Bharati Palai, A. Rathore, T. Kumar Mandal, H. Singh Jatav, Role of Legumes in Cropping System for Soil Ecosystem Improvement

- Improvement of Cropping System View Project Precision Agriculture View Project Role of Legumes in Cropping System for Soil Ecosystem Improvement, Nova Science Publishers, Inc., 2022.
- [51] T. Thorat, B.K. Patle, S.K. Kashyap, Intelligent insecticide and fertilizer recommendation system based on TPF-CNN for smart farming, *Smart Agric. Technol.* 3 (2023) 100114, <https://doi.org/10.1016/J.ATECH.2022.100114>.
- [52] S.P. Raja, B. Sawicka, Z. Stamenkovic, G. Mariammal, Crop prediction based on characteristics of the agricultural environment using various feature selection techniques and classifiers, *IEEE Access* 10 (2022) 23625–23641, <https://doi.org/10.1109/ACCESS.2022.3154350>.
- [53] P. Charoen-Ung, P. Mittrapiyanuruk, Sugarcane yield grade prediction using random forest with forward feature selection and hyper-parameter tuning, *Adv. Intell. Syst. Comput.* 769 (2019) 33–42, [https://doi.org/10.1007/978-3-319-93692-5\\_4](https://doi.org/10.1007/978-3-319-93692-5_4).
- [54] A.T.M.S. Ahamed, N.T. Mahmood, N. Hossain, M.T. Kabir, K. Das, F. Rahman, R.M. Rahman, applying data mining techniques to predict annual yield of major crops and recommend planting different crops in different districts in Bangladesh, in: 2015 IEEE/ACIS 16th Int. Conf. Softw. Eng. Artif. Intell. Netw. Parallel/ Distributed Comput. SNPDP 2015 - Proc, 2015, <https://doi.org/10.1109/SNPD.2015.7176185>.
- [55] N.N. Thilakarathne, M.S.A. Bakar, P.E. Abas, H. Yassin, A cloud enabled crop recommendation platform for machine learning-driven precision farming, *Sensors* 22 (2022), <https://doi.org/10.3390/s22166299>.
- [56] M. Stevanović, A. Popp, H. Lotze-Campen, J.P. Dietrich, C. Müller, M. Bonsch, C. Schmitz, B.L. Bodirsky, F. Humpenöder, I. Weindl, The impact of high-end climate change on agricultural welfare, *Sci. Adv.* 2 (2016) 1–10, <https://doi.org/10.1126/sciadv.1501452>.
- [57] L.K. Paine, P. Todd L, D.J. Undersander, K.C. Rineer, G.A. Bartelt, S.A. Temple, D.W. Sample, R.M. Klemme, Some ecological and socio-economic considerations for biomass energy crop production, *Biomass Bioenergy* 10 (1996) 231–242, [https://doi.org/10.1016/0961-9534\(95\)00072-0](https://doi.org/10.1016/0961-9534(95)00072-0).
- [58] E.B. Barbier, Cash crops, food crops, and sustainability: the case of Indonesia, *World Dev.* 17 (1989) 879–895, [https://doi.org/10.1016/0305-750X\(89\)90009-0](https://doi.org/10.1016/0305-750X(89)90009-0).