International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Machine Learning Driven Precision Agriculture: Enhancing Farm Management through Predictive Insights

B. Priyanka [1], M. Kezia Joseph [2], B. Rajendra Naik[3]

Submitted: 12/05/2024 **Revised**: 25/06/2024 **Accepted**: 05/07/2024

Abstract: Machine learning (ML) is transforming all the fields including agriculture, by providing data driven insights, which has enhanced and transformed the decision making process. Recent advancements in sensor technology, Wireless Communications, GPS and Data Analytics has led to widespread use by farmers which has led to increased resource utilization and efficiency and means to practice sustainable farming. In the proposed method various ML models are used to forecast two crucial metrics for precision agriculture Growing Degree Days (GDD) and Evapotranspiration (ET), which can be used for effectively managing daily agricultural activities. They are useful in predicting growth stages of crops, pest warnings, fertilizer usage, and irrigation times. Hourly data collected though sensors and other sources such as temperature, humidity, wind speed and soil moisture is used in managing real time growth of crops. The study assessed machine learning models like Random Forest, Support Vector Regressor (SVR), Voting Regressor, Stacking Regressor, and Decision Trees for precision agriculture metrics. Random Forest performed best but struggled with ET ranges. Decision Tree showed potential overfitting and underperformed, Voting Regressor and Stacking Regressor showed high performance. Despite hyper parameter optimization, the artificial neural network (ANN) exhibited poor performance, suggesting issues with either model selection or data adherence. The study developed a Decision Support System (DSS) that uses GDD and ET forecasts to provide real-time recommendations for pest and disease risk, fertilizer usage, crop maturation stages and watering regimens. The aim of this system is to equip farmers with the necessary tools to efficiently and effectively oversee their farms, hence enhancing agricultural productivity and sustainability.

Keywords: *Crop Management, Evapotranspiration, Growing degree days, Machine Learning, Precision Agriculture, Random Forest, Smart Agriculture*

1. Introduction

Agriculture is the ancient art of growing crops and rearing cattle. It led to the formation of villages and the development of civilizations. It was one of the earliest means of employment and trade for humans. Over time, people invented various tools and machinery to increase productivity and simplify agriculture. Farmers still use traditional farming methods and rely on their experience. As the population across the globe rises, so does the need to produce more in order to keep up with the increasing demands. Although Agriculture 3.0 practices have led to increased yields, they also have adverse effects on the environment, such as soil exploitation, water and groundwater contamination, air pollution, and increased resistance in pests [1]. One of the primary reasons is a lack of knowledge and resource exploitation. There is a dire need to upgrade the farming method to include technology and sustainability. In today's technology- and AI-driven world, the use of data-driven AI tools and machine learning models to increase productivity, give better insights to include climate changes, determine pest cycles, and suggest fertilizer and pesticide amounts, using sensors for irrigation, and better crop monitoring is on the rise. With irrigation monitoring, farmers can now prevent under- and over-irrigation, prepare for rainfall and climate change, and also practice deficit irrigation in the right phase of the crop. The quality and quantity of the yield can be improved and predicted so that the farmers can prepare themselves beforehand and take proper measures to ensure maximum profits [2] [3]. Pesticide and fertilizer amounts and exact application times can be calculated, pests can be modelled, and preventive sprays can be suggested. Parameters such as temperature, humidity, soil temperature, soil moisture, wind speed, solar radiation, precipitation, soil nutrient values, historical datasets, satellite photos, RGB images, and spectral images are being utilized to construct machine learning models and extract valuable crop information [4]. The plant development and irrigation needs can be estimated by using growing degree days (GDD) and evapotranspiration (ET).

In this paper, we build various ML models to fit our data and analyze them using various metrics. The model gets insights on the crop stages, irrigation needs and fertilizer needs of the crops. Fig 1 shows the system overview of the proposed work to be carried out for the DSS model. The following section contains the previous works carried out, followed by the methodology and results.

1Assistant Professor, Department of Electronics and Communication Engineering, Sreenidhi Institute of Science and Technology
2Professor, Department of Electronics and Communication Engineering, Stanley College of Engineering and Technology for Women,
3Professor, Department of Electronics and Communication Engineering, University College of Engineering Osmania University
** Corresponding Author Email: priyanka.b@sreenidhi.edu.in*

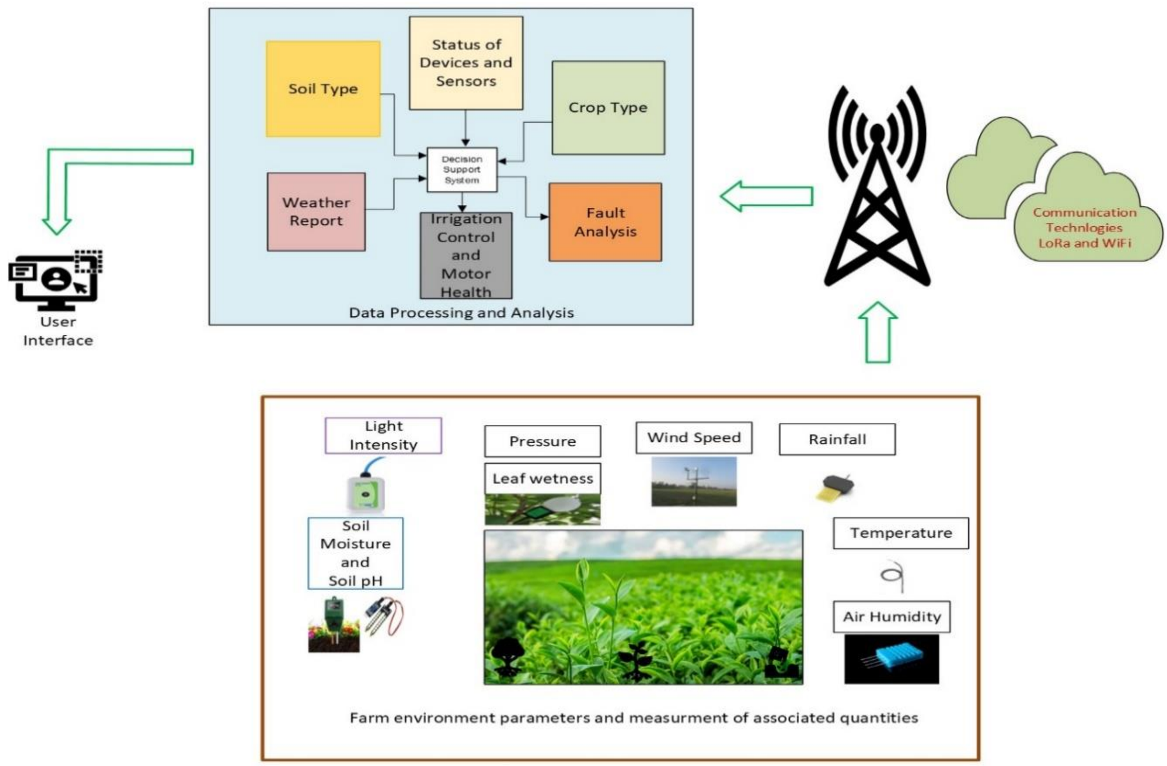

Fig 1. System Overview

2. Literature Review

Significant efforts are being made in the field of smart agriculture to enhance crop monitoring, water management, soil conditioning, disease prediction and identification, insect detection and identification, weed detection, animal management, yield prediction, and harvest management, among others [9].

Data acquisition using sensors is important for taking real time decisions and monitoring. Abba, S *et al.* developed a low-cost autonomous sensor interface for an Internet of Things (IoT)-based smart irrigation monitoring and control system. The system makes use of a water pump to supply water, a moisture sensor to determine the water content of the soil, and a WiFi module to enable internet-based data access. Data is sent to thing speak and is analysed and decision to switch on and off the motor are taken. The model can be used in large scale farms for easy monitoring and tracking of crops [5]. System developed by Farzad Kiani and Amir Seyyedabbasi collects temperature, humidity and soil moisture data via sensor nodes and sends it via gateway and is displayed on a GUI. The farm is divided into four parts and hourly data is collected from nodes in each part of the farm [6]. The goal of the model by Saha, G. C. et al. is to create an Internet of Things (IoT)-based system that monitors temperature and moisture content in agricultural fields and gives farmers useful data for crop management. To collect real-time data, the system makes use of a GSM module, an ESP8266 Wi-Fi module, a moisture sensor, and an LM 35 temperature sensor [7].

With the widespread use and rapid progress of machine learning and artificial intelligence techniques, they are being used to simplify various aspects of the world. One of the author modelled a web interface that facilitates easy access for farmers. Ten different algorithms were tested for crop recommendation in which the Random Forest classifier that was hyper tuned with Randomized cross validation proved to be the most effective model. Based on variables like the pH of the soil and typical rainfall, the system forecasted five crops. Longitude and latitude values are entered by the user and sent to the Weather API. The system can also have a pesticides and weeds detector. Weeds are predicted by the Weeds RESNET152V2 pre-trained algorithm with an accuracy of 0.89, and insects are predicted by the same method with an accuracy of 0.98. Agricultural costs were predicted using data sets from 2010 to 2018, which had data for 11 crops [8]. Regression models such as Gradient Boosting Regressor, Decision Tree Regressor, Bagging Regressor and XGBoost regressors were used and their R2 scores were compared.

The data-driven approach for developing Precision Agriculture solutions for data modelling and collection systems is presented in [9]. In this research soil moisture, a crucial component of the crop growth cycle, is selected as an example. Utilizing the MicaZ mote and VH400 soil moisture sensor, a reactive wireless sensor node is created for the collection side of soil moisture and the prototype gadget is tested in field soil. For data analysis, machine learning methods SVM (support vector machine) and RVM (relevance vector machine) are used for predicting soil moisture. Methodology is evaluated using Illinois historical data as the machine learning algorithms require enormous data sizes. When predicting soil moisture it produces low error rates (15%) and strong correlations (95%) between anticipated values and actual values across in nine distinct sites for 2 weeks period.

Crop Yield Prediction needs are assessed in a study used by Tamil Nadu paddy data which has 745 data points. Artificial Neural Network, Support Vector Regression, K-Nearest Neighbour, and Random Forest (RF) ml algorithms are used on the same dataset. The RF algorithm has the highest accuracy based on error analysis values. A survey talks about agriculture 4.0 and its major requirements. It gives a systematic literature review of thirteen chosen decision-support agricultural systems. These systems were graded, and advantages and disadvantages were highlighted [10]. Future challenges and improvements were suggested and a hybrid MLR-ANN algorithm is proposed for yield prediction of paddy crops. MLR intercept and coefficients are used to initialize ANN input bias and weights. The model is compared to Support Vector Regression, k-Nearest Neighbor, MLR, ANN and Random Forest models and is found to be more accurate [11] [12].

Growing Degree-days offer a method to approximate the pace of growth of plants, insects, fungus, and other creatures by utilising the highest and lowest temperatures recorded on an hourly or daily basis. Degree-days quantify the cumulative heat accumulated by plants during a specific period. Evapotranspiration (ET) refers to the collective processes of water evaporation from both soil and plant surfaces, as well as the water transpiration via plant tissues. Food and Agriculture Organizations of United Nations (FAO) Penman-Monteith equation is used to estimate ET based on Weather conditions and the crop conditions. ET can be used to estimate the water requirements so that maximum yield can be achieved with the available water [13].

In the proposed system we use ML algorithms to analyse and calculate GDD and evapotranspiration to give daily crop updates to the farmers

3. Methodology

Smart crop management deals with geographical, environmental, periodical and technological feasible quantities, that increase the internal visibility of farming. Every aspect of the surroundings, including soil qualities, climates, water availability, nutrients, humidity levels, and temperatures. Because of their losses farmers find it extremely difficult to farm. With fewer inputs and greater efficiency, smart farming generates more outputs, using digital technologies to enhance farming methods. Various technological advances are altering the limitations and forms of agricultural intelligence. The two main themes are machine learning methods and Internet of Things sensors.

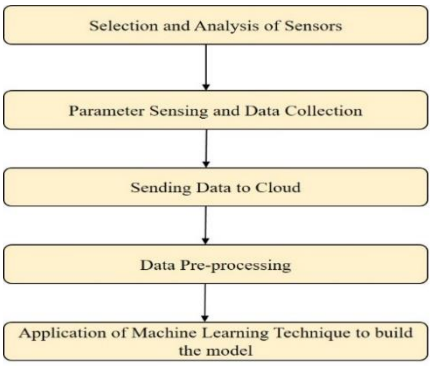

Fig 2. Work flow chart

3.1 Data Collection

Open-source datasets were employed for the proposed system to validate the results. The publicly accessible data from NREL was utilised for the decision support system, with location specific parameters to suit the environment were used. Sensing quantities that require more cost were extracted from open source data and farm sensitive data was captured for approximately 3 and half months. The parameters such as atmospheric temperature, soil temperature, Humidity and soil moisture (primary and secondary root) content was captured from the sensor module prototype.

3.2 Data Pre-processing

Obtained datasets had missing and erroneous values that were properly identified to clear possible inefficiencies in the data. The dataset errors are the common causes of the poor performance of the model, improper analysis of data could lead to inaccurate decision making system for farmers and troubling them in handling technology. GDD and ET are calculated based on FAO drainage paper 56 [13].

3.3 Proposed Method

The process of choosing a model starts with the analysis of sensors from the survey, the literature suggests that the ML model choice of Random Forest, Support Vector Regression (SVR), Voting Regressor, Stacking Regressor, Artificial Neural Networks (ANN) and Decision Tress would be efficient for the datasets collected. The data availability was for each hour for parameters like temperature, humidity, wind speed, soil temperature, soil moisture at the primary root and soil moisture at the secondary root. The prime output that every researcher is targeting to predict is crop yield, but the difficulty in accurate prediction and real-time crop yield data availability is the primary concern.

To primarily address the day-to-day issues of farmers rather than crop yield, Growing degree days (GDD) and Evapotranspiration (ET) can be considered as the response variables for prediction. In the proposed work primarily GDD and ET are considered for the analysis and prediction. The GDD though is a formulated quantity, but in addition, it is being examined with respect to other parameters such as humidity, soil moisture, Soil temperature and wind speed. It was observed that the GDD and ET have the additional capability to predict the stage of maturity of the crop, Pest alert and fertilizer usage for the crop with irrigation schedule, which could be more important for a farmer to protect the crop apart from the yield prediction.

So the proposed methodology is to use GDD and ET values prediction from real-time data, the method considers the implementation of above said models for training to predict GDD and ET values concerning non-formulated features and create a decision support system(DSS) for the farmer. The DSS system proposed would be able to take real-time inputs from the sensors, analyze and give time-lined predictions for the farmer to prevent frequent manual intervention and for easy and effective management of the farm. Using future predictions of GDD the system can estimate various stages of the crop and its requirements. By utilizing those insights, Farmers can equip themselves with all the necessities required such as irrigation times, fertilizers, pesticides, manpower, warehouse spaces, and transportation.

4. Results and Discussion

The process of prediction considered totally five ML algorithms for analysis, namely Random Forest, Decision Tree, SVR, Voting Regressor and Stacking Regressor. Cross Validation and Hyperparameter optimization were also considered. The growing degree days (GDD) and evapotranspiration (ET) were two critical indicators for smart agriculture decision support systems, the measures set out to evaluate performance of these models after prediction. Among the evaluation metrics used are the coefficient of determination (R^2), mean absolute error (MAE), and root mean square error (RMSE). The accuracy, robustness, and generalizability of each model are revealed by these metrics as shown in Table 1.

Table 1. Model interpretation for proposed dataset

S. No	Model	Interpretation
1	Random Forest Regressor	High accuracy and consistent
2	Decision Tree Regressor	Slightly lower accuracy, more variability
3	SVR (Support Vector Regressor)	High accuracy and consistent
4	Voting Regressor	Best performance and consistent
5	Stacking Regressor	Very good performance and consistent

It was observed that while modelling and evaluating performance parameters Artificial neural network was used in literature, hence ANN was also considered for the proposed work. But the performance of ANN continued to be unsatisfactory even after adjustments, suggesting potential problems with the selection of the model or compatibility with the data. Hyper parameter optimization enhanced the performance of certain models, such as the Decision Tree, but did not yield significant improvements as shown in Table 2.

Table 2. Hyper parameter Tuning of ANN and Decision Tree

Model	Best Parameters	Mean R^2 Score	Standard Deviation (Std Dev)	Interpretation
Decision Tree	max_depth: 30,min_samples_split: 2	0.832	0.078	Slight improvement, but not the best
ANN	alpha: 0.01, hidden layer size:(100,100)	0.319	0.4	No significant improvement

Table 3. Accuracy of evaluated models

Model	Metric	R^2 Score	Accuracy (Explained Variance %)
Random Forest	GDD	0.98	98%
Random Forest	ET	0.94	94%
Decision Tree	GDD	1	100%
Decision Tree	ET	1	100%
SVM	GDD	0.88	88%
SVM	ET	0.58	58%
Voting Regressor	GDD	0.98	98%
Voting Regressor	ET	0.92	92%
Stacking Regressor	GDD	0.95	95%
Stacking Regressor	ET	0.86	86%

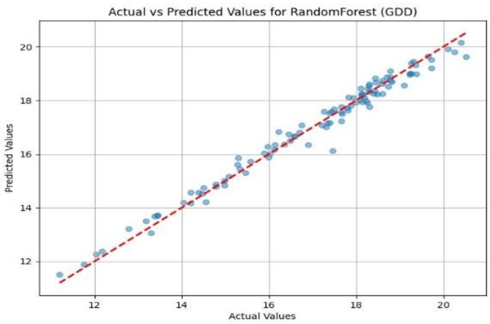

Fig 3. Random Forest Regressor GDD plot

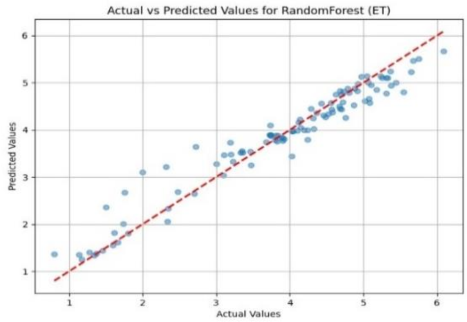

Fig 4. Random Forest Regressor ET plot

Table 3 and table 4 shows the obtained metric values of the evaluated models. Fig 3 to Fig 11 shows the actual and predicted values of different ML algorithms used to analyse the system. The plots clearly show that most of the predicted points are near to the hyper plane line, confirming the models showing good performance for prediction. Among these Random Forest model has performed the best in predicting GDD ($R^2 = 0.98$) and ET ($R^2 = 0.94$), but couldn't model ET ranges properly ($R^2 = 0.508$). The Decision Tree model results were a perfect match with the training data ($R^2 = 1.00$), potentially indicating an issue of overfitting. It also showed a below-average performance for ET ($R^2 = 0.265$).

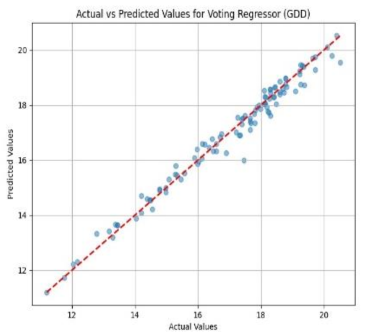

Fig 5. Voting Regressor GDD plot

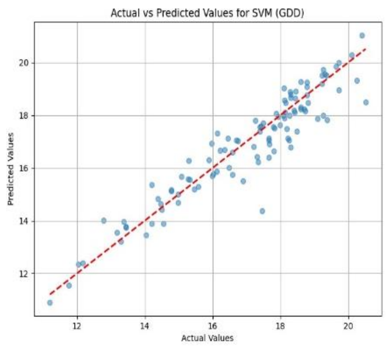

Fig 6. SVM GDD plot

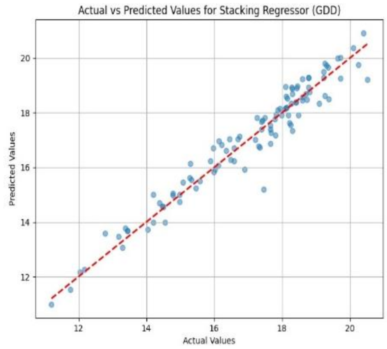

Fig 7. Stacking Regressor GDD plot

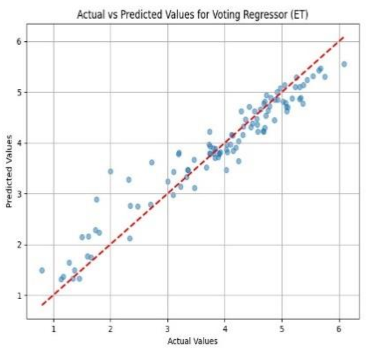

Fig 8. Voting Regressor ET plot

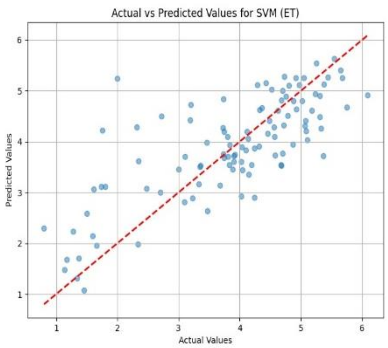

Fig 9. SVM ET plot

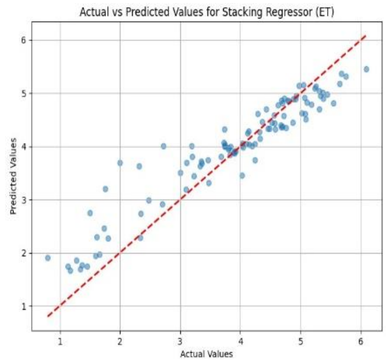

Fig 10. Stacking Regressor ET plot

Table 4. Model Performance with respect to target values

Model	Metric	GDD (Mean)	GDD (Std Dev)	ET (Mean)	ET (Std Dev)
RandomForest	R^2 Score	0.876	0.038	0.508	0.266
	RMSE	0.29	-	0.31	-
	MAE	0.22	-	0.22	-
	R^2	0.98	-	0.94	-
DecisionTree	R^2 Score	0.773	0.077	0.265	0.531
	RMSE	0	-	0	-
	MAE	0	-	0	-
	R^2	1	-	1	-
SVM	R^2 Score	0.859	0.046	0.46	0.173
	RMSE	0.71	-	0.82	-
	MAE	0.53	-	0.61	-
	R^2	0.88	-	0.58	-
Voting Regressor	R^2 Score	0.866	0.045	0.517	0.268
	RMSE	0.32	-	0.36	-
	MAE	0.23	-	0.27	-
	R^2	0.98	-	0.92	-
Stacking Regressor	R^2 Score	0.879	0.031	0.533	0.188
	RMSE	0.46	-	0.47	-
	MAE	0.35	-	0.34	-
	R^2	0.95	-	0.86	-

Table 5. Sample Predicted results

Timestamp	Temperature (°C)	Humidity (%)	Wind Speed (km/h)	Rain fall (mm)	Soil temperature (°C)	Soil moisture primary	Soil moisture secondary	ET	GDD	Cumulative GDD	Predicted GDD	Predicted ET	Cumulative Predicted GDD	Growth Stage	Irrigation Recommendation	Pest/Disease Risk	Fertilization Timing	Weather Alerts
2022-07-09 00:00:00	21.7616	87.0048	8.32	1.442	24.1876	1.0796	1.094	1.172439	12.045	12.045	12.2646	1.247743955	12.2646	Germination stage	Irrigation needed	Low risk of pests/diseases	No immediate fertilization needed	['High humidity warning']
2022-07-10 00:00:00	21.8892	88.1428	9.4	0.7496	23.5956	1.1128	1.0872	1.371003	12.175	24.22	12.37705	1.381244537	24.64165	Germination stage	Irrigation needed	Low risk of pests/diseases	No immediate fertilization needed	['High humidity warning']

The Support Vector Machine (SVM) model with an R-squared value of 0.88 showed good performance for GDD. However, the performance for ET was subpar, with a R-squared value of 0.58. The Voting Regressor and Stacking Regressor have estimated GDD and ET with R^2 values for GDD (0.98 and 0.95) and ET (0.92 and 0.86) which shown high performance, but the Voting Regressor displayed instability in its ET predictions. The predicted sample results are shown in Table 5 for Random Forest predicted model, showing various recommendations for the system.

5. Conclusion

The examination of several regression models demonstrates clear advantages and disadvantages in forecasting Growing Degree Days (GDD) and Evapotranspiration (ET). The Random Forest model demonstrates exceptional performance, attaining a R^2 score of 0.98 for GDD and 0.94 for ET, while also reducing the RMSE. It faced difficulties in accurately representing the range of values in ET. The Decision Tree model demonstrates perfect alignment with the training data which could potentially be a sign of overfitting, and showed poor performance for ET. The Support Vector Regressor (SVR) demonstrates a good level of prediction accuracy for GDD but underperformed for ET. Both the Voting Regressor and Stacking Regressor displayed high R^2 values, indicating their potential for reliable predictions, though both showed volatility in ET predictions. Hyperparameter optimization improved the performance of some models, it did not significantly enhance the Artificial Neural Network (ANN), pointing to potential issues with model selection or data compatibility. Random Forest Regressor showing superior performance among individual models. These findings underscore the importance of careful model selection and optimization to improve prediction accuracy in precision agriculture, ultimately aiding farmers in better managing their resources and enhancing crop yields.

References

[1] C. R. K. Ghosh, M. A. Rahman, and S. K. Biswas, "Smart Agriculture: A Review of Recent Trends and Applications in Precision Farming," *IEEE Access*, vol. 9, pp. 113549–113569, 2021.

[2] D. J. Lee, H. J. Choi, and J. K. Park, "Advancements in Precision Agriculture Technologies for Enhanced Crop Management," *Computers and Electronics in Agriculture*, vol. 183, pp. 106133, 2021.

[3] A.P. Patel, S. M. Dey, and A. A. Patel, "Integration of IoT and Machine Learning for Smart Agriculture: A Review," *Journal of Network and Computer Applications*, vol. 183, pp. 103185, 2021.

[4] R. M. Li, C. Y. Zhang, and H. Y. Liu, "Precision Agriculture Using Remote Sensing and Machine Learning for Crop Yield Prediction," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 2, pp. 1264–1277, 2021.

[5] Abba, S.; Wadumi Namkusong, J.; Lee, J.-A.; Liz Crespo, M. Design and Performance Evaluation of a Low-Cost Autonomous Sensor Interface for a Smart IoT-Based Irrigation Monitoring and Control System. Sensors 2019, 19, 3643. https://doi.org/10.3390/s19173643

[6] Farzad Kiani and Amir Seyyedabbasi, "Wireless Sensor Network and Internet of Things in Precision Agriculture" International Journal of Advanced Computer Science and Applications(ijacsa), 9(6), 2018. http://dx.doi.org/10.14569/IJACSA.2018.090614

[7] Saha, G. C. .; Islam, M. R. .; Billah, M. M. .; Khan, H. I. .; Mat, R. C. .; Hossain, M. M. .; Hoque, M. R. .; Pramanik, E. S. C. .; Saha, H. . IoT Based Smart Agricultural Crop Monitoring in Terms of Temperature and Moisture. Int J Intell Syst Appl Eng 2024, 12, 234-245.

[8] Senthil Kumar Swami Durai, Mary Divya Shamili, Smart farming using Machine Learning and Deep Learning techniques, Decision Analytics Journal, Volume 3, 2022, 100041, ISSN 2772-6622, https://doi.org/10.1016/j.dajour.2022.100041.

[9] Z. Hong, Z. Kalbarczyk and R. K. Iyer, "A Data-Driven Approach to Soil Moisture Collection and Prediction," 2016 IEEE International Conference on Smart Computing (SMARTCOMP), St. Louis, MO, USA, 2016, pp. 1-6, doi: 10.1109/SMARTCOMP.2016.7501673.

[10] Zhai, Zhaoyu, José-Fernán Martínez, Victoria Beltran and Néstor Lucas Martínez. "Decision support systems for agriculture 4.0: Survey and challenges." Comput. Electron. Agric. 170 (2020): 105256.

[11] P. S., Maya Gopal, and Bhargavi R. 2019. "Performance Evaluation of Best Feature Subsets for Crop Yield Prediction Using Machine Learning

Algorithms." Applied Artificial Intelligence 33 (7): 621–42. doi:10.1080/08839514.2019.1592343.

[12] P.S. Maya Gopal, R. Bhargavi, A novel approach for efficient crop yield prediction, Computers and Electronics in Agriculture, Volume 165, 2019, 104968, ISSN 0168-1699, https://doi.org/10.1016/j.compag.2019.104968.

[13] Food and Agriculture Organization of the United Nations (FAO), Drainage Principles and Applications, FAO Drainage Paper 56, Rome, Italy, 1986.

[14] L. Y. Lin, H. M. Chen, and J. Z. Chang, "Smart Farming: Data-Driven Solutions for Agricultural Management," *Agricultural Systems*, vol. 198, pp. 103344, 2022.

[15] T. X. Wu, Z. C. Zhao, and Y. F. Chen, "Recent Advances in Precision Agriculture Technologies: A Comprehensive Review," *Sensors*, vol. 22, no. 4, pp. 1227, 2022.

[16] M. S. Ahmed, M. A. Ahmed, and S. K. Shaw, "Machine Learning and Data Analytics in Precision Agriculture: Techniques and Applications," *IEEE Transactions on Automation Science and Engineering*, vol. 20, no. 1, pp. 11–25, 2023.

[17] J. S. Lee, H. J. Kim, and D. K. Kim, "Optimization of Smart Agriculture Systems Using IoT and AI Technologies," *Journal of Agricultural Engineering*, vol. 52, no. 1, pp. 20–32, 2023.