



AGRIENSEMBLE LEVERAGING MULTI-MODEL LEARNING FOR PRECISION CROP SELECTION



M. Shalini¹, G.Dikshitha², A.Sireesha³, B.Asha⁴

Original Article

^{1,2,3}Assistant Professor, IT Department, Stanley College of Engineering & Technology for Women

⁴Assistant Professor, IT Department, Stanley College of Engineering & Technology for Women

*Corresponding Author's Email: shalinibydl@gmail.com, dikshitha2112@gmail.com, luckyreddysireesha1987@gmail.com, asha2372@gmail.com

Abstract

Agriculture is a back bone of economy in certain countries even though it plays an crucial role the sector faces few problems to cover the demands of the high population growth in the world without the proper use of scientific approach by the use of modern technology such as machine learning we can provide suitable remedies to completely transform the conventional farming methods. This smart agriculture model recommends the crops to be planted by studying the soil nutrient condition, area of cultivation and other relevant factors this ensemble model uses the combination of Random Forest, Gradient Boosting, Extra Trees Classification to give accurate crop yield predictions and recommendations.

Keywords: *Machine Learning, Smart Agriculture, Random Forest, Gradient Boosting, Extra Trees Classification*

Introduction

The current problem of food shortage in the has been escalating in a rapid state hence there is a necessity increase of the food production by the process of smart agriculture which offers satisfying results to convert conventional farming practice to sustainable and resource efficient processes. This project uses machine learning methods to monitor the nutrient values, soil conditions, and surrounding environmental factors and provides solution for enhancing the crop yields this approach gives ideal conditions for different types of crops based on Nitrogen content, Potassium content, phosphorus content, Temperature, Humidity, PH and Rainfall. In this project we use an ensemble machine learning model to analyse and categorize the data to obtain more accurate crop recommendations and higher yield productions.

Literature Review

Based on the climatic conditions and to improve the food security and to improve the rural livelihoods the most common approach is the landscape approach which are to be converted into climate smart landscapes[1]. Based on the concept of IoT by the use of wireless sensor technology and networks integration a remote monitoring system real time data is collected[2]. Smart farming by the use of unmanned aerial vehicles is a best technology to make smart agriculture the best than conventional farming methods [3]. The automation of the smart agriculture by the use of IoT, aerial imagery

to tackle problems like pest control, weed management, irrigation can be solved [4] Smart Agriculture by the integration of certain technologies and computational procedures where big data, AI are involved the security is a threat by the use of IoT this problem is coupled [5]. By the use of Nano Technology being explored in the fields which benefits the farms to reduce the losses which helps in the gains of rural farmers[6]. The methodical framework of smart agriculture has seen significant advancements, but the steps to encourage the development of smart agriculture from the perspective of agricultural economic management have not received enough attention. Using the egg price of a city's wholesale market as the research object, analyze the factors influencing and fluctuating egg price by first mining and analyzing the relevant agricultural big data, and then visualizing the big data in order to innovate agricultural economic management, promote the construction and development of smart agriculture, and realize the transformation of agriculture. The findings demonstrate that agricultural big data plays a significant role in building smart agriculture and offers robust data support for innovations in agricultural economic management[7]. The notion of Climate-Smart Agriculture (CSA) is becoming more widely acknowledged as a crucial tool for tackling agricultural issues in the context of climate change. In order to promote food security, CSA combines mitigation and adaptation techniques. Even though several nations have expressed interest in implementing CSA, the idea is still misinterpreted and handled unevenly. This book uses agricultural development and economics to codify the theoretical and methodological underpinnings of CSA. It offers case examples that are country-specific and illustrate the economic advantages of CSA, such as decreased susceptibility and improved adaptability. The policy implications for national and international agricultural and climate change policies are also covered in the book. It provides tried-and-true methods as well as creative ways to implement CSA at the national level, making it a valuable resource for scholars, development organizations, policymakers, and the commercial sector[8].

Climate-smart agriculture (CSA) is gaining recognition as a vital strategy to feed the growing global population amid climate change. A review of 137 publications using an institutional analysis framework found that only 55.5% specifically addressed institutional dimensions. The CSA concept includes three pillars: productivity, adaptation, and mitigation, yet these are rarely integrated in the literature. The focus on these pillars varies by region, with high-income countries emphasizing mitigation, while middle and low-income countries prioritize productivity and adaptation. Institutional aspects in CSA have gradually gained attention, mainly focusing on knowledge infrastructure and market structure, but less on how historical, political, and social contexts influence CSA adoption. A more integrated approach, combining technology with institutional factors, could enhance the scaling of CSA practices[9]. In at least nine nations, agriculture is the most important industry and contributes 6.4% of the world's economic output. Not only does it sustain billions of people, but it also creates a sizable amount of jobs. In order to increase crop yields, the agricultural sector is increasingly turning to AI, or "Agriculture Intelligence," in response to the challenges posed by population expansion, unpredictable climate change, and food security. This study explores the different uses of AI in agriculture, utilizing deep learning, machine learning [10], IoT, robots, computer vision, and precision farming as well as crop phenotyping and disease detection [11]. These developments improve soil fertility, cut expenses, increase productivity, and minimize the need for chemicals [12][13].

Proposed Methodology

The main purpose of our paper is to predict the suitable crop to be planted based on soil nutrient levels and climatic conditions. The system is designed using an ensemble machine learning techniques and trained using the historical data which is capable in predicting the crop suitable to plant in the region. The system must be able to fetch the Nitrogen content, Potassium content, Phosphorous content, Humidity, Temperature, PH, Rainfall and decide the crop.

The Algorithms used in the paper for ensemble machine learning model are Random Forest, Gradient Boosting, Extra Trees Classification which are used to classify, study and analyse the data to train the model and get the required crop plantation recommendation for higher yields.

Step 1: The field data collected which contains the soil sample data such as Nitrogen content, Potassium content, phosphorus content, Temperature, Humidity, PH and Rainfall as shown in figures 1,2 and 3.

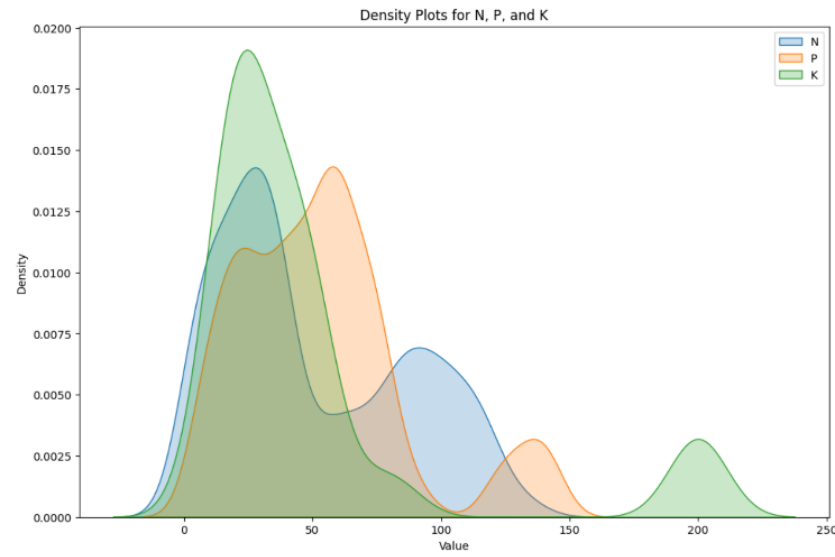


Figure 1 Density of Nutrition values of N,P,K present in the soil

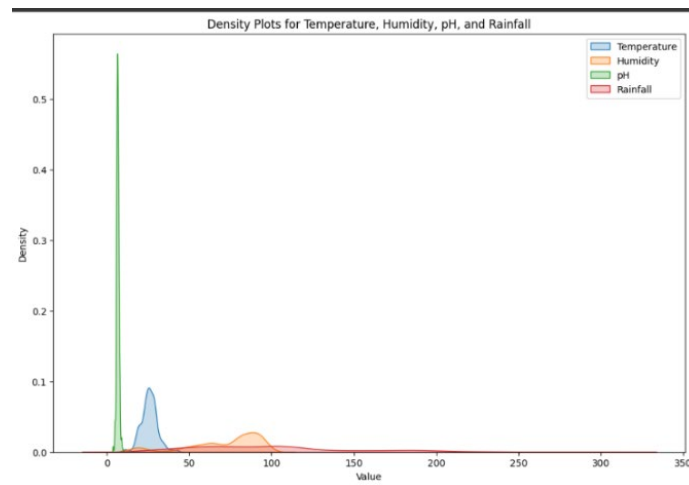


Figure 2 Density of the surrounding climatic values

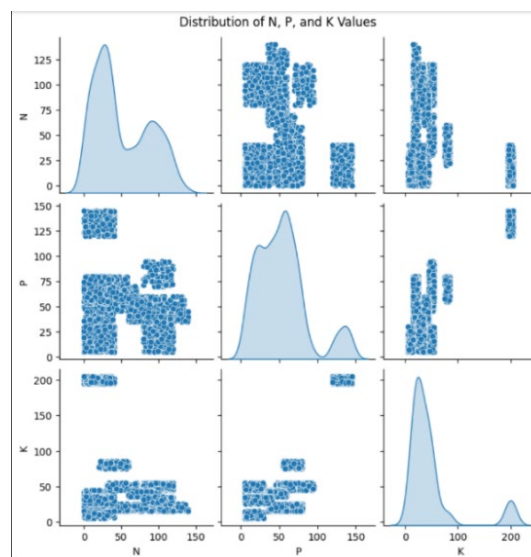


Figure 3 Distribution of the N, P and K values in the soil

Step 2: The Loaded data is pre-processed by fetching out the Null values and duplicate values present in the data set.

a) Checking for Duplicate values :-

The duplicate values in the data set is been checked based on rows in the data set where the present rows set is considered as ' x_p ' the data set as 'A' and the previous row set is considered as x_q where $p < q$. We consider $A = \{x_1, x_2, x_3, \dots, x_n\}$ with n rows for each row x_p where $p=2, \dots, n$ we check if the row x_q is present such that $x_p = x_q$ and $q < p$. The formula checks if $x_p \in \{x_1, x_2, x_3, \dots, x_{p-1}\}$ if this condition returns true any I value then ' x_p ' is considered as duplicate. If any duplicate found in the data set then the row is removed form the data set which reduces the duplicated values.

b) Checking for Null Values :-

The null values which have been present in the data set are represented as nan

In the data set A with m columns y_1, y_2, \dots, y_m and n rows now take an function $f(q)$ that is equal to 1 if 'q' is an null value or represent 0. If null values are found then we fill the null values of the data set by the mean value.

Step 3: Now we divide the data set in to train and test set where the 80% of the is been used to train the model and the rest 20% is used to test the model to improve the model even better the training set is divided in validation which is of 20% and the rest 60% is used for training the final model this makes sure the model has been trained, validated and tested generalizing the performance. Hence let the data set A be divided into features and labels where the features are represented in terms of ' α ' and labels in terms of ' β '. If the total sample proportion is considered as 'w' the training set is calculated by Train set $= (w \times n)$ samples, the test set is calculated as Test set $= (1-w) \times n$ samples. Where in the model 'w' is considered as 0.7 hence Test set $= 0.7n$ and train set $= 0.3n$.

Step 4: Now we prep the ensemble model buy using three different classifiers Random Forest Classifiers $\hat{y} = \text{mode}\{h_1(x), h_2(x), h_3(x) \dots, h_T(x)\}$ where T represents the decision trees trained on a random subset for sample x and $h_i(x)$ is the prediction of the ith tree.

Gradient Boosting Classifiers $\hat{y} = \sum_{t=1}^T \alpha_t h_t(x)$ where $h_t(x)$ is the Tth tree and α_t is the learning rate by this the model minimizes the loss function present in the model.

Extra Trees Classifiers involve more randomness to the model by the process of node splitting $\hat{y} = \text{mode}\{h_1(x), h_2(x), h_3(x) \dots, h_T(x)\}$ here the prediction of model is based on the majority of vote.

The three classifiers are configured using different hyperparameters like number of trees, the depth of trees, nutrition value consideration. The above three models are used as they have the ability to capture non-linear relationships which are present in the complex dataset.

Step 5: A Voting Classifier is used to merge the initialized classifiers into a single ensemble model here in soft voting each classifier gives us the probability of p_i for each of the class and the final prediction by process of highest average probability as in equation 1.

$$\hat{y} = \text{argmax}(\frac{1}{M} \sum_{i=1}^M p_i(y)) \quad (1)$$

where M is the total count of models present in the ensemble model. By pooling the predictions of each individual model, the ensemble technique makes use of its strengths, potentially improving total performance. The final forecast is made by averaging the anticipated probability from each model, a technique known as "soft voting." Because this approach lessens the volatility and bias that could affect individual models, accuracy is frequently increased. The model

is now trained by using the training data set by utilizing the grouped approach features and target data after the model is trained the accuracy and predictions of the model is calculated and a detailed report is been generated.

Step 6 : Now to further to improve the model performance we perform hyperparameter tuning by the process of Grid Search with cross validation where ' γ ' is the set of hyperparameters and the grid search improves the model performance $S(\gamma)$ across the various hyperparameters with the help of cross validation as specified in equation 2.

$$S(\gamma) = \frac{1}{K} \sum_{i=1}^K \text{Accuracy}(\gamma, D_i) \quad (2)$$

which systematically removes all different combinations of hyperparameters used across the different models and by the cross-validation technique we assess the performance and move with the optimal condition that give high yield accuracy. The final model trained by the optimal hyperparameters γ^* is evaluated by the use of the test set the accuracy is calculated as in equation 3.

$$\frac{1}{n_{test}} \sum_{i=1}^{n_{test}} 1(\hat{y}_i = y_i) \quad (3)$$

where n_{test} gives the total number of samples, \hat{y}_i represents true predicted label and y_i represents the true label hence predictions based on the test set are cross verified and a report is drafted based on the final accuracy and classification. Hyperparameter table with optimal values is specified in table 1.

Table1 Hyperparameter table with the best hyper parameter values

Hyperparameter	Best value
et max depth	10
et n estimators	50
gb learning rate	0.01
gb n estimators	50
rf max depth	10
rf n estimators	100

Results and Discussion

The paper is aimed to give the best crop which is suitable for the external climatic factors and soil nutrient values which can gradually reduce the crop wastage and look over the irrigating resources the dataset used has been cleared of the duplicate and null values and to evaluate the model accuracy, precision, recall and f1 score was used by 'True positive', 'True negative', 'False positive' and 'False negative' values. The model is enhanced by the use of various hyperparameter values and the best were taken for maximum model benefit as shown in table 1. Different types of classifiers were used individually and an ensemble model was also used in the combination of those classifiers as shown in the table 2 where Random Forest obtained an accuracy of 98%, Gradient Boosting obtained an accuracy of 97%, Extra Trees obtained an accuracy of 99% and the ensemble model obtained the accuracy of 99.5% as shown in figure 4.

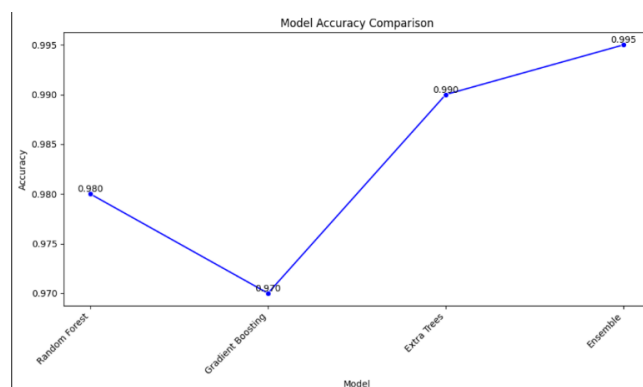


Figure 4 Model Accuracy Comparing

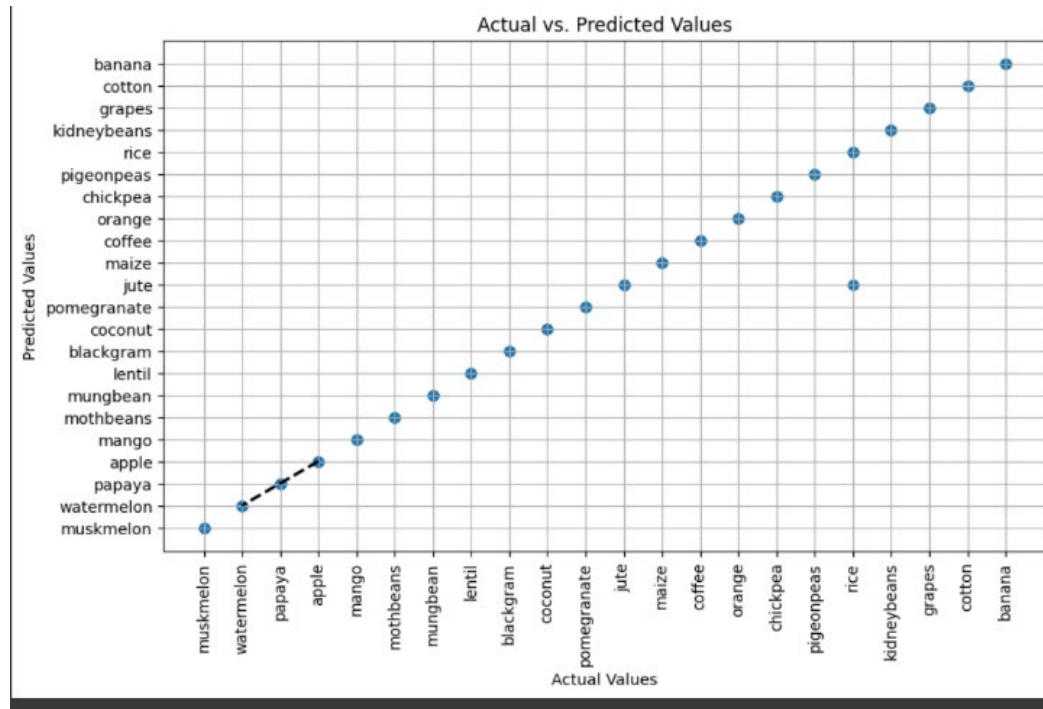


Figure 5 Actual vs Predicted values

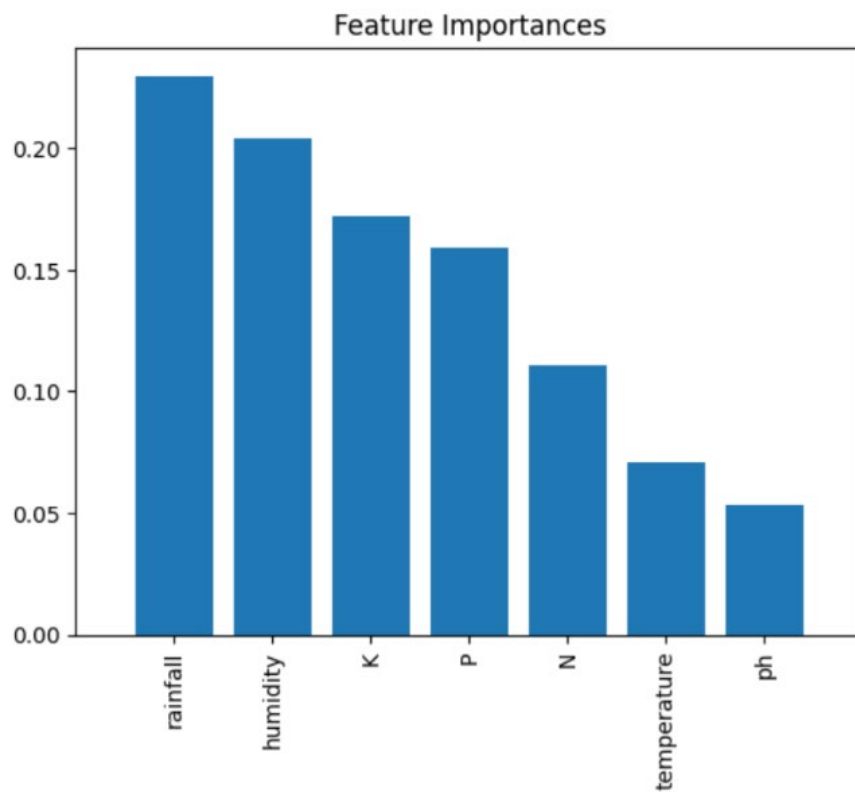


Figure 6 Feature Importance graph

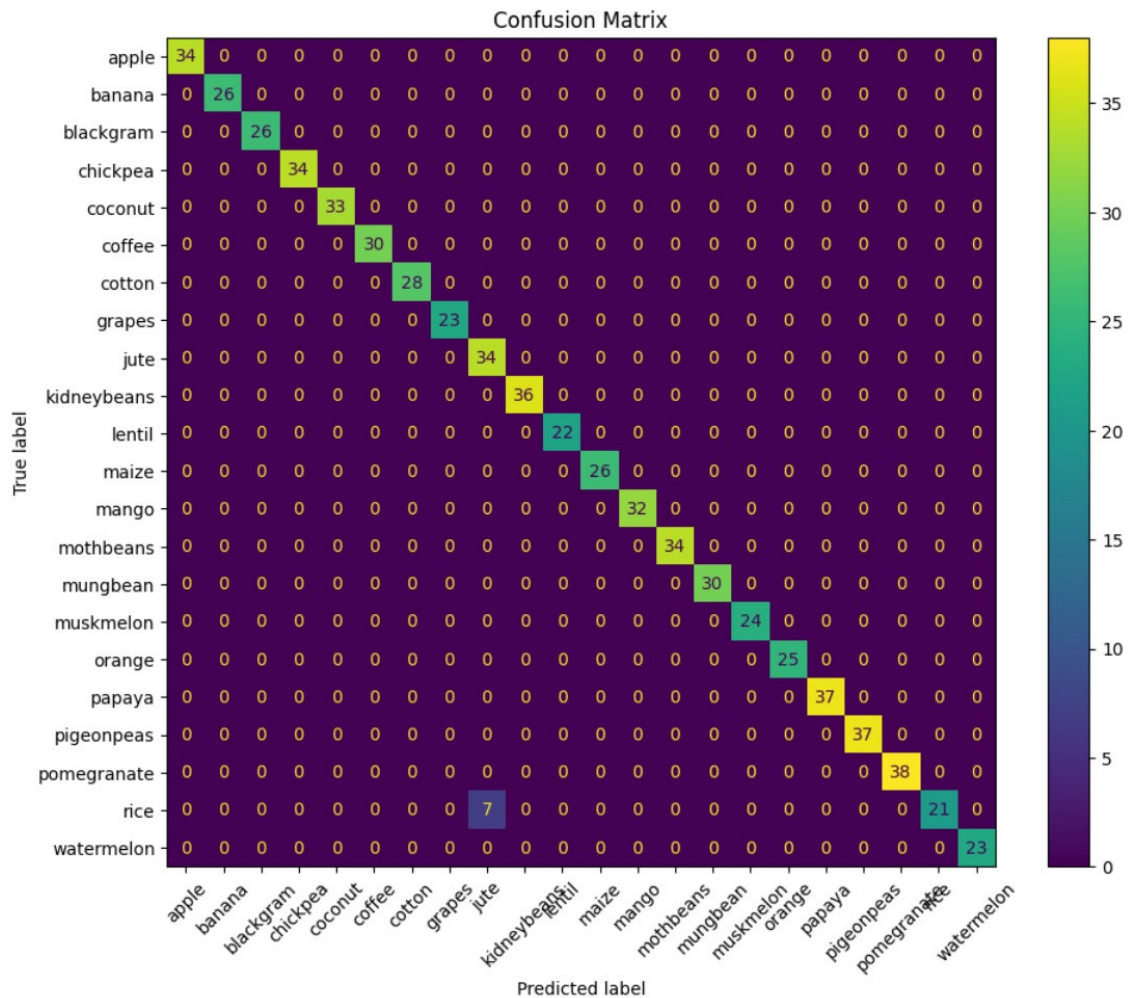


Figure 7 Confusion Matrix between the True and Predicted Labels

Table 2 comparing of the accuracy, percision, Recall and F1 score between the models

S.no	Model	Accuracy	Percision	Recall	F1 score
1	Random Forest	98%	98.1%	97.5%	97%
2	Gradient Boosting	97%	96.4%	97%	96.5%
3	Extra Trees	99%	98%	98.3%	98%
4	Ensemble model (Random forest+ Gradient Boosting+ Extra Trees)	99.5%	99.12%	98.94%	98.91%

Conclusion and Fature Works

In conclusion we have shown an data-driven approach in the recommendation of the suitable crops to be planted in the certain locality based on the soil nutrient value and the environmental factors such as humidity, temperature, PH and Rain Fall this model has the potential to reduce the crop wastage and farmer income loss with the optimization of

irrigation resources therefore promoting the smart agricultural methods by the help modernised computer techniques such as machine learning. In the near future focus looking ahead to expanding the data in terms of various species of crops and cover more diverse conditions integrating real time data by the help of live sensors with advanced IoT techniques and an user friendly interface for the farmers to increase the models practical ability.

References

- [1] Scherr, S. J., Shames, S., & Friedman, R. (2012). From climate-smart agriculture to climate-smart landscapes. *Agriculture & Food Security*, 1, 1-15.
- [2] Patil, K. A., & Kale, N. R. (2016, December). A model for smart agriculture using IoT. In *2016 international conference on global trends in signal processing, information computing and communication (ICGTSPICC)* (pp. 543-545). IEEE.
- [3] Maddikunta, P. K. R., Hakak, S., Alazab, M., Bhattacharya, S., Gadekallu, T. R., Khan, W. Z., & Pham, Q. V. (2021). Unmanned aerial vehicles in smart agriculture: Applications, requirements, and challenges. *IEEE Sensors Journal*, 21(16), 17608-17619.
- [4] Hassan, S. I., Alam, M. M., Illahi, U., Al Ghamdi, M. A., Almotiri, S. H., & Su'ud, M. M. (2021). A systematic review on monitoring and advanced control strategies in smart agriculture. *Ieee Access*, 9, 32517-32548.
- [5] de Araujo Zanella, A. R., da Silva, E., & Albin, L. C. P. (2020). Security challenges to smart agriculture: Current state, key issues, and future directions. *Array*, 8, 100048.
- [6] Rameshaiah, G. N., Pallavi, J., & Shabnam, S. (2015). Nano fertilizers and nano sensors—an attempt for developing smart agriculture. *Int J Eng Res Gen Sci*, 3(1), 314-320.
- [7] Su, Y., & Wang, X. (2021). Innovation of agricultural economic management in the process of constructing smart agriculture by big data. *Sustainable Computing: Informatics and Systems*, 31, 100579.
- [8] McCarthy, N., Lipper, L., & Zilberman, D. (2018). Economics of climate smart agriculture: An overview. *Climate smart agriculture: Building resilience to climate change*, 31-47.
- [9] Totin, E., Segnon, A. C., Schut, M., Affognon, H., Zougmore, R. B., Rosenstock, T., & Thornton, P. K. (2018). Institutional perspectives of climate-smart agriculture: A systematic literature review. *Sustainability*, 10(6), 1990.
- [10] Sreelatha, G. (2024). Transfer Learning Based Bi-GRU for Intrusion Detection System in Cloud Computing. In: Satheeskumaran, S., Zhang, Y., Balas, V.E., Hong, Tp., Pelusi, D. (eds) Intelligent Computing for Sustainable Development. ICICSD 2023. Communications in Computer and Information Science, vol 2121. Springer, Cham.

- [11] PR Anisha, Kishor Kumar Reddy C, NG Nguyen, G Sreelatha, A Text Mining using Web Scraping for Meaningful Insights, Journal of Physics: Conference Series 2089 (1), 012048, 2021
- [12] Kishor Kumar Reddy, Advaita Daduvy, R Madana Mohana, Basem Assiri, Mohammed Shauib, Sadaf Alam and Abdullah Sheneamer, “Enhancing Precision Agriculture and Land Cover Classification: A Self –Attention 3D Convolutional Neural Network Approach for Hyperspectral Image Analysis, IEEE Access, 2024, DOI: 10.1109/ACCESS.2024.3420089
- [13] Pathan, M., Patel, N., Yagnik, H., & Shah, M. (2020). Artificial cognition for applications in smart agriculture: A comprehensive review. *Artificial Intelligence in Agriculture*, 4, 81-95.