

EXPLAINABLE ARTIFICIAL INTELLIGENCE FOR SMART INDOOR AGRICULTURE: A FRAMEWORK FOR GREENHOUSE CLIMATE CONTROL AND SUSTAINABLE DECISION-MAKING

Ms. D. Radha

Assistant Professor, Shri Krishnaswamy College for Women, Chennai, India

Research Scholar, Vels Institute of Science, Technology and Advanced Studies (VISTAS), Chennai, India

Email: radha.aasc@gmail.com

ABSTRACT

Agriculture has been the primary and major source of income in our country. Environmental conditions change due to rising global warming and other environmental conditions. Several environmental factors like. Temperature, humidity, soil moisture and light conditions must be checked to take appropriate control actions. Manually checking may not always be possible and efficient. . Hence, I used machine learning to predict suitable greenhouse control actions. Traditional irrigation practices may not provide transparency regarding the environmental conditions for generating control decisions. This research proposes an Explainable Artificial Intelligence (XAI) framework for intelligent greenhouse climate control and proper decision-making.

The study used a greenhouse climate control dataset containing 3,000 records and and seven environmental input parameters, namely temperature, humidity, carbon dioxide concentration, soil moisture, light intensity, external weather, and time of day. The target variable consists of five greenhouse control actions: No Action, Turn on Heater, Open Vent, Start Irrigation, and Turn on Light.

Random Forest algorithm is used to detect the correct greenhouse control action. Dataset is divided into training and testing using 80: 20 ratios. The l results show an accuracy of 99.83%, precision of 99.83%, recall of 99.83%, and F1-score of 99.83%. SHAP-based explainability is implemented to predict the contribution of individual environmental parameters to the prediction process.

The feature importance analysis shows that temperature is the most influential parameter. This is followed by soil moisture, light intensity, humidity, external weather, carbon dioxide concentration, and time of day. The proposed research shows that explainable machine learning can support transparent and intelligent greenhouse climate control. The research contributes to sustainable indoor agriculture by combining accurate prediction with interpretable decision-making.

Keywords: Explainable Artificial Intelligence, Smart Agriculture, Indoor Agriculture, Greenhouse Climate Control, Random Forest, SHAP, Sustainable Agriculture, Machine Learning

I. INTRODUCTION

Today agriculture is facing a lot of changes with advanced technologies like Artificial Intelligence (AI), Machine Learning (ML), Internet of Things (IoT), and other sensor-based technologies. Traditional farming practices depended on manual monitoring and environmental parameters. Plant productivity can be affected due to the changing environmental conditions. Hence, regular checking and on time decision making is necessary in today's modern agricultural era.

Indoor agriculture provides a more closed environment than traditional open field agricultural practices. The environmental parameters need to be checked on a regular basis. Environmental factors like light, humidity, temperature can be controlled as per the need of the environment. Hence proper analysis of environmental parameters and appropriate selection of control actions is needed.

Machine learning algorithms will aid in analyzing agricultural data and help in proper decision making. ML algorithms study the trends, patterns and relationships between environmental factors and climate factors, thereby it predicts corresponding control actions. For example, if the temperature is low, then the system may suggest switching on the heater, opening a vent when ventilation is required, pumping out water when moisture of soil is low or turning on artificial lights when the available light is not sufficient.

Even though ML models can provide high accuracy, but their decisions are not always correct. Sometimes, they provide predictions without stating the proper reason for the decision made. This may lead of a cause of concern In agricultural applications. Hence, an automated system shall be needed enable to take proper control decisions which shall help the farmers and other decision makers to understand the system.

This issue is addressed by Explainable Artificial Intelligence (XAI) which makes the predictions of ML models more understandable. It helps in understanding the importance of input parameters individually in order to make a prediction. In agricultural applications, such predicted action is directly related to varying environmental factors. So XAI explains why the model performs a specific action and helps in proper decision making.

The paper proposes an explainable machine learning framework for greenhouse climate control. A Random Forest classifier is used to predict the most appropriate greenhouse control action based on environmental parameters, while SHAP (SHapley Additive exPlanations) is used to show the importance and contribution of features towards the prediction of model.

SHAP is mainly used to provide a deeper view of the environmental factors and its influence on greenhouse control actions.

This paper produces an Explainable ML framework for indoor agriculture climate control. We use Random Forest to predict the control action based on the environmental conditions. We use SHAP (SHapley Additive explanations) to understand the contribution of individual features to a model's prediction. The proposed framework shall provide a clear, smart and proper decision making in indoor agriculture.

II. RELATED WORK

To address the problems faced in indoor agriculture, nowadays AI and ML are being used rapidly. It is being used in many areas like disease detection, yield prediction, monitoring soil & moisture and climate control. Usage of sensors and automated collection of data has led to the increased usage of advanced technologies in today's modern agricultural fields.

Sensors and IoT devices are being used rapidly in smart agriculture for collecting data about environmental factors. It monitors environmental factors like CO₂, moisture of soil, temperature, humidity etc. ML techniques are used to analyse the dataset collected to predict the trends and patterns and thereby support decision making.

Random Forest is mostly used amongst ML algorithms because of its ability to handle complex relationships between input data and provide better classification. It combines the predicted results of several decision trees and gives the final prediction. So, the model can give better prediction accuracy and reduce overfitting compared to other algorithms.

ML can handle the selection of particular control action based on environmental conditions. Based on the observed environmental condition, the model may suggest operating the heater on, open ventilation, start irrigation, switching on the artificial light etc. So such automated control actions can help the greenhouse to react more quickly as per the changes in the environmental conditions.

We cannot rely on a particular ML algorithm only seeing its high accuracy. We should understand why the model selected the control action. This is really important in agriculture since environmental conditions and control actions can affect the plant growth and usage of water and energy.

We use Explainable AI as it helps study the ML models in a better way. We used SHAP as it tells us the importance of individual features and its role in model prediction and hence identifies the environmental parameters that have the greatest influence on a specific decision.

Explainable AI can provide more support in automated control systems in indoor agriculture. We not only get to know about the control action but also understanding of the environmental factors that played a major role for that prediction.

The research paper used Random Forest with SHAP to provide a framework for indoor agriculture climate control. This framework emphasizes on prediction of control action and importance of the feature that played a greater role in indoor agriculture control decisions.

III. PROPOSED METHODOLOGY

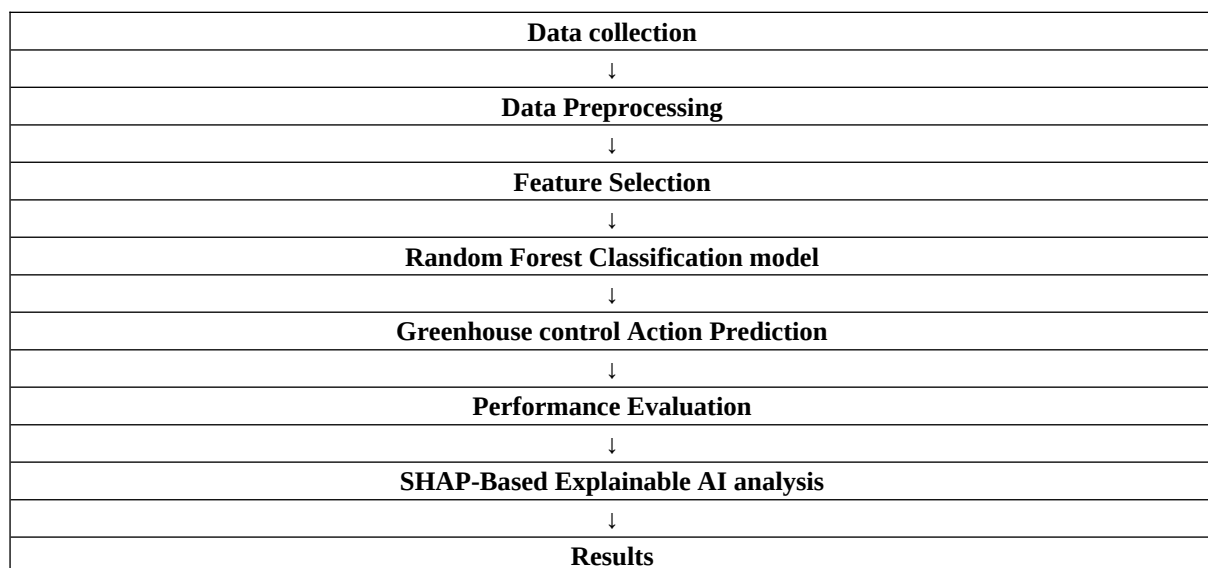
3.1 Overview of the Proposed Framework

The research paper is aimed to identify the climate control action based on the environmental condition. The methodology begins with data collection, preprocessing and ends with model interpretation using SHAP.

The main steps of the proposed framework is as below:

1. Collection of data and data loading
2. Preprocessing of data
3. Environmental parameters selection
4. Training and testing data
5. Random forest model training
6. Greenhouse control actions prediction
7. Performance evaluation
8. Feature analysis
9. SHAP based AI analysis

The overall workflow of the proposed system can be represented as follows:



The environmental factors are analyzed in the model. Data is preprocessed and the selected environmental features are given to the Random forest model. The trained model identifies the possible control action. The model is evaluated using various performance metrics. Feature importance and SHAP based AI analysis are done to study the role of different environmental features in the model's prediction process.

3.2 Dataset Description

The dataset has been downloaded from Kaggle comprising 3000 records and 8 attributes. Seven features represent the environmental features and the remaining attribute represents the greenhouse control action.

Input features in the dataset are:

- Temperature
- Humidity
- CO₂ concentration
- Soil Moisture
- Light Intensity
- External Weather
- Time of Day

Action represents the target variable which represents the greenhouse control action associated with the environmental conditions.

Table 1: Dataset attributes

Attribute	Description
Temperature	Represents temperature inside the indoor farm
Humidity	Humidity level
CO ₂	Carbon dioxide
Soil_Moisture	Soil Moisture level
Light_Intensity	Represents the light available
External_Weather	Information about weather condition outside
Time_of_Day	Represents the time
Action	Required greenhouse climate control action

3.3 Greenhouse Control Actions

The proposed model is aimed to predict the correct control actions based on the environmental conditions.

The dataset contains five control actions:

- No_Action
- Turn_On_Heater
- Open_Vent
- Start_Irrigation
- Turn_On_Light

These 5 actions represent the output variable of the model. The control action is predicted based on the environmental conditions. The model may predict when ventilation, irrigation, heating is required.

3.4 Data Preprocessing

After data collection, the data is preprocessed in ML model. Data is analyzed and its data types, missing values, null values etc. are determined and handled. 7 environmental parameters are used as input features and Action variable is used as target output. The dataset is checked for null values, missing values, incorrect values etc.

X represents the input features

- **Temperature**
- **Humidity**
- **CO₂**
- **Soil_Moisture**
- **Light_Intensity**
- **External_Weather**
- **Time_of_Day**

Y = Action , which represents the target variable

3.5 Random Forest Classification

The Random Forest is selected for this research due to its ability to handle multiple input variables and complex relationships amongst environmental features and control actions. The model was built with 200 estimators and fixed random state so that results can be reproduced.

3.6 Training and Testing

The dataset was divided into training and testing sets in 80:20 ratio. We chose stratified sampling to maintain the proper class distribution across training and testing datasets. The model was built using the training data and later was evaluated using testing data.

3.7 Explainable Artificial Intelligence Using SHAP

SHAP based AI explained the importance of individual features and thereby aided in the prediction process. It helped to study the importance of environmental features that played a major role in the decision-making process of the model. SHAP also explains the reasons behind specific control action in the decision-making process.

IV. RESULTS AND DISCUSSION

4.1 Performance Evaluation

The model was evaluated using the metrics like accuracy, precision, recall, and F1-score. It helped the model to predict the proper control action based on the environmental conditions.

Table 2: Performance metrics

Metric	Value
Accuracy	0.998333
Precision	0.998347
Recall	0.998333
F1-Score	0.998333

Accuracy of 99.83% was achieved. 99.83% was the value for precision, recall, and F1-score

4.2 Class-Wise Performance

The model was evaluated separately for each greenhouse control action.

Table 3: Performance metrics for control actions

Indoor control Action	Precision	Recall	F1-Score	Support
No_Action	0.99	1.00	1.00	120
Open_Vent	1.00	1.00	1.00	120
Start_Irrigation	1.00	0.99	1.00	120
Turn_On_Heater	1.00	1.00	1.00	120
Turn_On_Light	1.00	1.00	1.00	120

The model performed well across all the control actions. Perfect values close to 1.00 were achieved by the classes Open_Vent, Turn_On_Heater, and Turn_On_Light .

Compared to other classes, No_Action and Start_Irrigation classes showed lesser precision or recall values. But their F1-scores showed values close to 1.00

4.3 Feature Importance Analysis

Feature analysis showed that temperature was the most influential feature in the model prediction process amongst all other input features.

Table 4: Feature analysis

Feature	Importance
Temperature	0.410451
Soil_Moisture	0.206601
Light_Intensity	0.178164
Humidity	0.127104
External_Weather	0.048610
CO ₂	0.017146
Time_of_Day	0.011924

Results (Output from Google Colab)

```

UPLOAD THE FILE: greenhouse_climate_control_dataset.csv
greenhouse_climate_control_dataset.csv(text/csv) - 327309 bytes, last modified:
4/10/2025 - 100% done
Saving greenhouse_climate_control_dataset.csv to greenhouse_climate_control_dataset
(1).csv
  
```

DATASET NAME: greenhouse_climate_control_dataset.csv

DATASET LOADED SUCCESSFULLY

FIRST FIVE ROWS:

	Temperature	Humidity	CO2	Soil_Moisture	Light_Intensity \
0	24.363503	87.042858	958.794548	51.906094	2404.167764
1	16.452090	81.970569	841.003511	59.565080	1185.260449
2	35.811066	42.740347	463.642470	22.838316	3738.180187
3	15.576561	61.486480	659.874875	13.266596	9763.799670
4	30.459650	52.947719	1184.907797	42.673403	8739.463661

	External_Weather	Time_of_Day	Action
0	2	22	No_Action
1	1	23	Turn_On_Heater
2	1	11	Open_Vent
3	2	14	Turn_On_Heater
4	2	20	No_Action

DATASET SHAPE:

(3000, 8)

COLUMN NAMES:

['Temperature', 'Humidity', 'CO2', 'Soil_Moisture', 'Light_Intensity', 'External_Weather', 'Time_of_Day', 'Action']

DATASET INFORMATION

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 3000 entries, 0 to 2999

Data columns (total 8 columns):

#	Column	Non-Null Count	Dtype
0	Temperature	3000 non-null	float64
1	Humidity	3000 non-null	float64
2	CO2	3000 non-null	float64
3	Soil_Moisture	3000 non-null	float64
4	Light_Intensity	3000 non-null	float64
5	External_Weather	3000 non-null	int64
6	Time_of_Day	3000 non-null	int64
7	Action	3000 non-null	object

dtypes: float64(5), int64(2), object(1)

memory usage: 187.6+ KB

None

MISSING VALUES

Temperature	0
Humidity	0
CO2	0
Soil_Moisture	0
Light_Intensity	0
External_Weather	0
Time_of_Day	0
Action	0

dtype: int64

DATASET SHAPE AFTER REMOVING MISSING VALUES:

(3000, 8)

INPUT FEATURES

['Temperature', 'Humidity', 'CO2', 'Soil_Moisture', 'Light_Intensity', 'External_Weather', 'Time_of_Day']

TARGET VARIABLE

Action

AVAILABLE GREENHOUSE ACTIONS:

['No_Action', 'Turn_On_Heater', 'Open_Vent', 'Start_Irrigation', 'Turn_On_Light']

DATA SPLITTING COMPLETED

TRAINING DATA SIZE: (2400, 7)

TESTING DATA SIZE: (600, 7)

TRAINING RANDOM FOREST MODEL
MODEL TRAINING COMPLETED SUCESSFULLY
GREENHOUSE ACTION PREDICTION COMPLETED
MODEL PERFORMANCE

Accuracy : 0.9983333333333333
Precision: 0.9983471074380165
Recall : 0.9983333333333333
F1 Score : 0.9983333043976457

PERFORMANCE TABLE

Evaluation Metric	Value
0 Accuracy	0.998333
1 Precision	0.998347
2 Recall	0.998333
3 F1 Score	0.998333

CLASSIFICATION REPORT

	precision	recall	f1-score	support
No_Action	0.99	1.00	1.00	120
Open_Vent	1.00	1.00	1.00	120
Start_Irrigation	1.00	0.99	1.00	120
Turn_On_Heater	1.00	1.00	1.00	120
Turn_On_Light	1.00	1.00	1.00	120
accuracy			1.00	600
macro avg	1.00	1.00	1.00	600
weighted avg	1.00	1.00	1.00	600

CONFUSION MATRIX

<Figure size 1000x800 with 0 Axes>

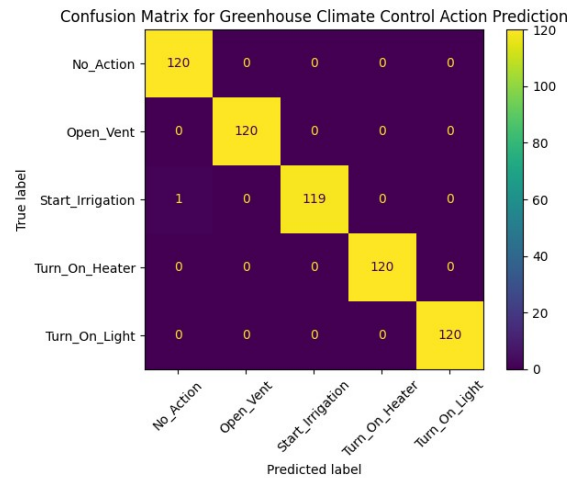


Figure 1: Confusion matrix

ACTUAL VS PREDICTED ACTIONS

	Actual Action	Predicted Action
0	Start_Irrigation	Start_Irrigation
1	No_Action	No_Action
2	No_Action	No_Action
3	Start_Irrigation	Start_Irrigation
4	Turn_On_Light	Turn_On_Light
5	Open_Vent	Open_Vent
6	Start_Irrigation	Start_Irrigation
7	Turn_On_Light	Turn_On_Light
8	Turn_On_Heater	Turn_On_Heater
9	Open_Vent	Open_Vent
10	Start_Irrigation	Start_Irrigation
11	Start_Irrigation	Start_Irrigation
12	Start_Irrigation	Start_Irrigation

13	No_Action	No_Action
14	Open_Vent	Open_Vent
15	Turn_On_Light	Turn_On_Light
16	Open_Vent	Open_Vent
17	Turn_On_Light	Turn_On_Light
18	Open_Vent	Open_Vent
19	Turn_On_Light	Turn_On_Light

FEATURE IMPORTANCE

	Feature	Importance
0	Temperature	0.410451
3	Soil_Moisture	0.206601
4	Light_Intensity	0.178164
1	Humidity	0.127104
5	External_Weather	0.048610
2	CO2	0.017146
6	Time_of_Day	0.011924

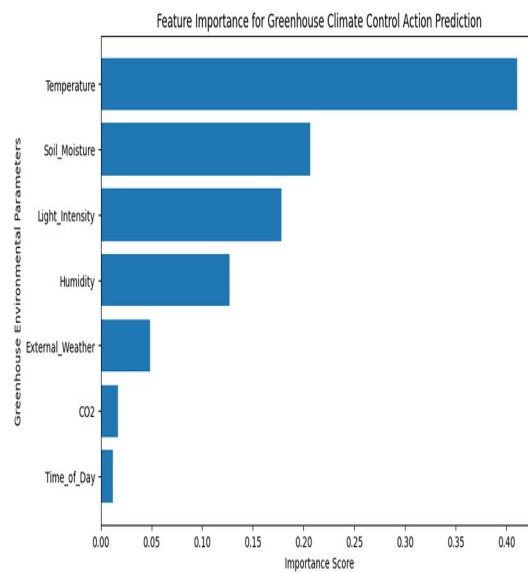


Figure 2: Feature importance

GENERATING SHAP EXPLANATIONS

SHAP VALUES GENERATED SUCCESSFULLY

SHAP SUMMARY PLOT

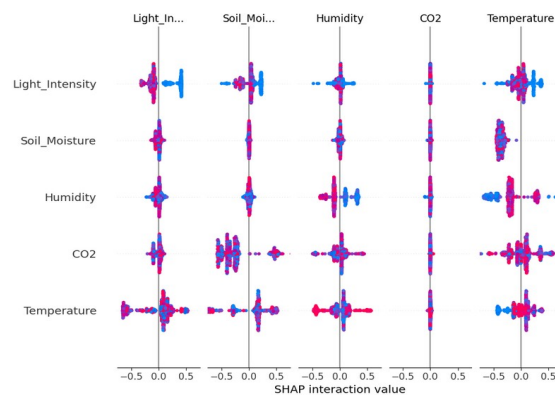


Figure 3: SHAP summary plot

SHAP FEATURE IMPORTANCE

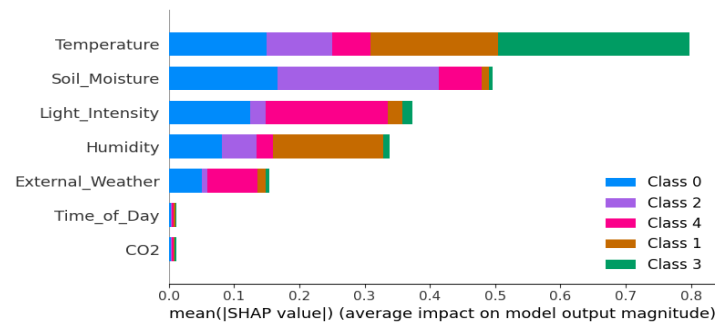


Figure 4: SHAP feature importance

RESULT FILES SAVED SUCCESSFULLY

1. greenhouse_action_predictions.csv
2. greenhouse_model_performance.csv
3. greenhouse_feature_importance.csv

PROJECT COMPLETED SUCCESSFULLY

V. SUSTAINABILITY IMPLICATIONS

Efficient usage of resources like water, energy can contribute to sustainable agriculture in smart indoor systems. The research paper gives a framework that uses environmental parameters to predict proper control actions, which shall optimize the proper usage of resources.

5.1 Efficient Water Management

In the proposed model, soil moisture is one among the most important features as it can aid in decisions on irrigation by providing information about the water requirements of the medium.

We use ML techniques to predict the situation when irrigation is needed. By giving proper decisions related to irrigation, it shall aid in reducing water consumption and thereby contribute to efficient usage of water.

5.2 Energy Management

In our proposed system, temperature and humidity are considered the most essential environmental features. These features are directly contributing to the operation of equipment like heating systems and artificial lighting.

Energy can be managed more efficiently by predicting correctly when these actions are required in indoor systems. This can prevent continuous operation of equipment and also avoid manual monitoring. It can provide recommendations based on environmental conditions.

5.3 Controlled Resource Utilization

Proper and automated decision making can lead to efficient usage of the resources available. When the system monitors the environmental conditions, then equipment like irrigation systems, heaters, ventilation systems, lighting systems can be made functional depending upon

the environmental conditions. This technique shall lead to better and efficient usage of water and other energy resources.

5.4 Transparent Decision-Making

When advanced technologies are incorporated into the model, then it is essential to understand the automated decisions. Efficient usage of resources alone does not matter. SHAP based AI shall enhance transparency of the system which shall help the users to understand the environmental features and respective control actions. So it becomes easier to interpret the recommendations and enhances the confidence of the users in ML based models.

5.5 Support for Sustainable Indoor Agriculture

The research paper integrates ML model-based prediction, environmental parameters monitoring and SHAP based AI. These all together provide a smart indoor agriculture system. Predicting control decisions in a timely manner along with explanations provide a better prediction. This led to the development of an efficient indoor agriculture system.

VI. CONCLUSION

The research provided a framework using Explainable AI for smart indoor agriculture with emphasis on automatic green house control actions. The research paper has used 7 environmental conditions like temperature, humidity, carbon dioxide concentration, soil moisture, light intensity, external weather, and time of day to identify the greenhouse control actions.

We have used Random Forest to predict the 5 possible control actions: No_Action, Turn_On_Heater, Open_Vent, Start_Irrigation, and Turn_On_Light. We divided the dataset into training and testing sets and we evaluated the performance using accuracy, precision, recall and F1 score.

The results from the research showed overall good performance with accuracy of 99.83%. The performance metrics like precision, recall, and F1-score values were around 99.83%. These results showed that model performed consistently across all the evaluation metrics.

Temperature was found to be the most dominating factor by feature analysis in the training model. Soil moisture, light, humidity also contributed likely in the prediction process. The results show that the environmental factors involved in the research helped to predict the correct control actions depending upon the environmental conditions.

Along with these performance metrics, SHAP based AI also contributed to enhance the prediction process of the Random forest model. SHAP helped in identifying the feature that played a major role in the prediction process among other features. This resulted in making the framework more transparent in identifying the greenhouse control action.

We have combined Random forest model with SHAP based AI for greenhouse control action prediction. This approach aided in decision making and provided insight into the factor that played significantly in the prediction process. These technologies can be used for developing smart indoor agricultural systems in today's modern era.

In the future, we can evaluate the framework using real time sensor data collected from real agriculture fields. IoT based monitoring systems can be integrated and we can compare the performance of the Random forest model with other ML, deep learning algorithms. We can also test the research for different crops, other greenhouse conditions and real-world environmental conditions.

VII. REFERENCES

1. Anubhove, M. S. T., Ashrafi, N., Saleque, A. M., Akter, M., & Saif, S. U. (2020, July). Machine learning algorithm based disease detection in tomato with automated image telemetry for vertical farming. In 2020 International Conference on Computational Performance Evaluation (ComPE) (pp. 250-254). IEEE.
2. Chen, Z., Wu, R., Lin, Y., Li, C., Chen, S., Yuan, Z., ... & Zou, X. (2022). Plant disease recognition model based on improved YOLOv5. *Agronomy*, 12(2), 365.
3. Zhang, P., Yang, L., & Li, D. (2020). EfficientNet-B4-Ranger: A novel method for greenhouse cucumber disease recognition under natural complex environment. *Computers and Electronics in Agriculture*, 176, 105652.
4. Liu, J., & Wang, X. (2020). Early recognition of tomato gray leaf spot disease based on MobileNetv2-YOLOv3 model. *Plant Methods*, 16, 1-16.
5. Wang, X., Liu, J., & Zhu, X. (2021). Early real-time detection algorithm of tomato diseases and pests in the natural environment. *Plant Methods*, 17(1), 1-17.
6. Krithika, N., & Selvarani, A. G. (2017, March). An individual grape leaf disease identification using leaf skeletons and KNN classification. In 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIECS) (pp. 1-5). IEEE.
7. Sabrol, H., & Kumar, S. (2016). Intensity based feature extraction for tomato plant disease recognition by classification using decision tree. *International Journal of Computer Science and Information Security*, 14(9), 622.
8. Nanehkaran, Y. A., Zhang, D., Chen, J., Tian, Y., & Al-Nabhan, N. (2020). Recognition of plant leaf diseases based on computer vision. *Journal of Ambient Intelligence and Humanized Computing*, 1-18.
9. Swetha, V., & Jayaram, R. (2019, June). A novel method for plant leaf malady recognition using machine learning classifiers. In 2019 3rd International conference on

- Electronics, Communication and Aerospace Technology (ICECA) (pp. 1360-1365). IEEE.
10. Harakannanavar, S. S., Rudagi, J. M., Puranikmath, V. I., Siddiqua, A., & Pramodhini, R. (2022). Plant leaf disease detection using computer vision and machine learning algorithms. *Global Transitions Proceedings*, 3(1), 305-310.
 11. Singh, A. K., Sreenivasu, S. V. N., Mahalaxmi, U. S. B. K., Sharma, H., Patil, D. D., & Asenso, E. (2022). Hybrid feature-based disease detection in plant leaf using convolutional neural network, bayesian optimized SVM, and random forest classifier. *Journal of Food Quality*, 2022, 1-16.
 12. Cruz, A. C., Luvisi, A., De Bellis, L., & Ampatzidis, Y. (2017). Vision-based plant disease detection system using transfer and deep learning. In 2017 asabe annual international meeting (p. 1). American Society of Agricultural and Biological Engineers.
 13. Mekha, P., & Teeyasuksaet, N. (2021, March). Image classification of rice leaf diseases using random forest algorithm. In 2021 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunication Engineering (pp. 165-169). IEEE.
 14. Jaisakthi, S. M., Mirunalini, P., & Thenmozhi, D. (2019, February). Grape leaf disease identification using machine learning techniques. In 2019 International Conference on Computational Intelligence in Data Science (ICCIDS) (pp. 1-6). IEEE.
 15. Ramesh, S., Hebbar, R., Niveditha, M., Pooja, R., Shashank, N., & Vinod, P. V. (2018, April). Plant disease detection using machine learning. In 2018 International conference on design innovations for 3Cs compute communicate control (ICDI3C) (pp. 41-45). IEEE.
 16. Zhao, Y., He, Y., & Xu, X. (2012). A novel algorithm for damage recognition on pest-infested oilseed rape leaves. *Computers and electronics in agriculture*, 89, 41-50.
 17. Hassanien, A. E., Gaber, T., Mokhtar, U., & Hefny, H. (2017). An improved moth flame optimization algorithm based on rough sets for tomato diseases detection. *Computers and Electronics in Agriculture*, 136, 86-96.
 18. Ngugi, L. C., Abelwahab, M., & Abo-Zahhad, M. (2021). Recent advances in image processing techniques for automated leaf pest and disease recognition—A review. *Information processing in agriculture*, 8(1), 27-51.
 19. Rangarajan, A. K., Purushothaman, R., & Ramesh, A. (2018). Tomato crop disease classification using pre-trained deep learning algorithm. *Procedia computer science*, 133, 1040-1047.
 20. Dhaya, R. (2020). Flawless identification of fusarium oxysporum in tomato plant leaves by machine learning algorithm. *Journal of Innovative Image Processing (JIIP)*, 2(04), 194-201.