An Explainable Climate-Aware Generative and Predictive Modelling Framework for Simulation of “What-if” Plausible Climatic Scenarios across Multiple Crops

Jayati Vijaywargiya, Omer Waqar

School of Computing, University of the Fraser Valley, British Columbia, Canada

Keywords: Climate-Aware Agriculture, “What-if” Scenario Simulation, SF24 dataset, CTGAN, Explainable AI.

Abstract

Climate fluctuations influence many aspects of agriculture including crop growth, soil conditions, and the distribution of fertilizer and water resources. These climatic fluctuations thereby pose significant challenges for agricultural productivity worldwide. However, the availability of agricultural datasets to study the impact of various adverse climatic conditions on different crops remains limited. To address this data availability limitation for agro-climatic impact study, this paper introduces an Explainable Climate-Aware Generative AI framework (XCA-GenAI). The framework combines a Conditional Tabular GAN (CTGAN) to generate realistic synthetic datasets, a Random Forest (RF) regressor to predict crop yield and stress-related parameters, and a SHAP-enabled “what-if” simulation module that evaluates and explains crop responses under varying temperature and rainfall conditions. The proposed framework is employed to generate synthetic representations of ten climatic variations ranging from Very Hot–Dry to Very Cool–Wet using the SF24 dataset. Crop-specific predictive models then estimate how change in climatic condition alters crop density, pest pressure, and frost risk. Further, explainability analysis provides interpretable insights of climate impact across multiple crops represented in the dataset. Comprehensively, this work introduces a climate-aware agricultural decision-support framework to aid farmers and agronomists for informed decision making under varying climatic conditions.

1. Introduction

Climate change is significantly transforming global food systems. Rising temperatures, unpredictable rainfall, prolonged droughts, increased flooding, soil degradation, and frequent extreme weather events are threatening agricultural productivity and food security (Raza et al., 2024). Thus, climate-adaptive agriculture is becoming essential to ensure resilient, sustainable, and future-ready crop production and management.

According to the IPCC Special Report on Climate Change and Land, changing precipitation patterns and more frequent extreme events are already affecting yields of major crops, and these trends are projected to worsen (Mbow et al., 2019). Climate-adaptive agriculture requires data-driven systems. These systems utilize predictive learning of crop-associated parameters to support informed decision-making under climatic uncertainty scenarios (Wen et al., 2025). In agriculture, it uses information such as weather records, soil properties, remote sensing inputs, crop growth indicators, and farm management practices to forecast key variables such as crop density, fertilizer needs, irrigation frequency, pest pressure, and frost risk. By converting climate variability into actionable intelligence, predictive learning enables proactive farm management and protects farmers from climate-related losses.

The application of predictive learning for climate-adaptive agriculture depends on its ability to anticipate outcomes under a wide range of climatic conditions, including normal and unprecedented scenarios. Simulation of climate scenarios allows researchers and farmers to model the impact of temperature and rainfall fluctuations on crop growth and related risks. Simulating multiple “what-if” plausible scenarios helps stakeholders understand sensitivities and agronomic climate-adaptive interventions to reduce losses and improve crop resilience (Dockerty et al., 2006). A “what-if” analysis examines possible outcomes by simulating different scenarios or changes in variables to understand their potential impact.

Agricultural systems are highly complex and influenced by numerous interacting variables, and a major limitation in climate scenario simulation is the explainability of the results. When datasets are sparse or incomplete, models may produce predictions that are difficult to interpret or trust, particularly in extreme scenarios where historical observations are not available. Explainability refers to the ability of a model to clearly show why it made a particular prediction or decision (Allgaier et al., 2023). Without explainable outputs, it becomes challenging for farmers, agronomists, and policymakers to understand the drivers behind predicted crop behavior and risk. Thus, due to lack of explainability, the adoption of climate-adaptive measures is limited.

A key reason for the agronomically accurate predictions and lack of explainability in agricultural models is the scarcity of high-quality datasets, especially for extreme weather events. In many regions, agriculture is dominated by smallholder farms, which are often fragmented and underrepresented in existing datasets. Thus farm-specific studies and recommendations are challenging. Although having a crop-specific and region-specific predictive modelling framework for simulation of “what-if” plausible climatic scenarios addresses this challenge, the data gap still limits accurate prediction of crop parameters in diverse weather events. The absence of comprehensive and well-structured agriculture-related datasets restricts model training, reduces predictive accuracy, and limits the development of robust climate-adaptive agricultural strategies (Cravero et al., 2022a).

The challenge of data limitation under extreme scenario can be addressed using generative AI (Krishna et al., 2025). GenAI efficiently creates synthetic agricultural datasets by simulating crop and environmental parameters under various conditions. The synthetically generated dataset, when augmented on real dataset enables predictive learning models to train on scenarios that would otherwise be underrepresented in the dataset. Additionally, when combined with explainable AI techniques, the predictive model enhances transparency by showing how changes in climate variables impact the crop parameters to influence crop performance and risk. This capability allows farmers and policymakers to make informed proactive decisions and strengthen the resilience of agricultural systems under climate uncertainty.

The existing predictive models for climate impact on agriculture remain constrained by sparse or negligible presence of data from extreme climatic scenarios. This reduces agronomic correctness of the predicted outputs and eventually pragmatic adoption of these models by agronomist and farmers. There exists a need for development of a unified, accurate and explainable framework capable of simulating diverse climatic scenarios and providing interpretable predictions. This study is an attempt to answer how can a multi-crop, region-specific, pragmatically usable model for agronomist, policy makers and farmers can be developed to reduce the climate impact on agriculture.

The primary objective of this study is to develop a reproducible and explainable framework that integrates generative AI-based data augmentation for predictive modelling of climate impact on crop density and associated risk. The framework simulates “what-if” scenarios across multiple crops to derive insights for climate-adaptive decision-making for agriculture management. The proposed framework was analysed using the SF24 dataset. The sub-objectives of this study are:

1. Generate synthetic agricultural datasets using CTGAN to augment the SF24 dataset.

2. Train predictive models for each crop available in the dataset and forecast key agronomic parameters.

3. Perform “what-if” climate scenario simulations using ten scenarios ranging from Very Hot–Dry to Very Cool–Wet conditions to assess crop sensitivity.

4. Apply SHAP-based explainable AI approaches to interpret model predictions and quantify feature contributions for future climate-adaptive decision-making.

2. Literature Review

Recent research highlights growing role of climate-adaptive approaches in agriculture to manage the increasing risks posed by climate change. Domingues et al. (2022) developed models in which pest dynamics were used to predict crop yield losses. Crane-Droesch (2018) used machine learning with multi-model climate scenarios to predict crop yields and revealed significant climate change induced declines in crops like corn (Crane-Droesch, 2018). Shelia et al. (2019) proposed Climate-Change Agricultural Forecasting Toolbox (CRAFT), which integrates multiple crop models on gridded framework to forecast yields, conduct risk analysis, and assess climate impacts at regional scales (Shelia et al., 2019).

For predictive modelling in agriculture, a variety of machine learning and deep learning algorithms have been utilized to forecast crop yields under diverse agro-climatic and management conditions. Prasad et al. (2021) applied a Random Forest (RF) regression model to predict regional-level cotton yield by capturing non-linear relationships and complex interactions between climatic, soil, and agronomic variables (Prasad et al., 2021). Yamparla et al. (2022) compared several regression algorithms, including Linear, Lasso, Ridge, and RF regressions, for crop yield prediction and reported consistently superior performance across diverse datasets (Yamparla et al., 2022). Dhillon et al. (2023) combined RF with crop modeling frameworks for winter wheat and oilseed rape, showing that RF’s regression capabilities can complement crop growth simulation outputs to improve prediction accuracy (Dhillon et al., 2023). Similarly, Elavarasan and Vincent (2021) proposed a reinforced RF model that integrates agrarian parameters, which achieved significantly higher performance than conventional RF or other regression-based models (Elavarasan and Vincent, 2021).

Recent advances in predictive agriculture increasingly integrate climate and scenario-based modelling to optimize farming practices. Ukhurebor et al. (2022) highlighted the importance of weather monitoring for future farming, emphasizing how dynamic climate conditions across agricultural regions influence decision-making (Ukhurebor et al., 2022). Similarly, Ahmed et al. (2022) reviewed the impacts of climate change on dryland agricultural systems, advocating the use of process-based models to test “what-if” scenarios for adaptation strategies (Ahmed et al., 2022).

Data-driven and digital approaches are also being applied for crop yield prediction and farm management: Chang et al. (2023) developed a maize yield prediction model incorporating weather scenarios to improve planning at the time of seed selection, while Akkem et al. (2024) employed ensemble learning and explainable AI to enhance crop recommendation systems in smart farming (Chang et al., 2023, Akkem et al., 2024). Recent advancements in smart agriculture increasingly leverage explainable AI (XAI) techniques to enhance interpretability, trust, and decision-making in crop and soil management. Hussein et al. (2024) integrated SHapley Additive exPlanations (SHAP) with multiple machine learning models to evaluate irrigation water quality indices, demonstrating improved feature selection and model fine-tuning (Hussein et al., 2024). Cartolano et al. (2024) also applied SHAP to interpret knowledge learned by AI models alongside expert agronomists’ insights in smart agriculture environments (Cartolano et al., 2024). Jothi (2025) used XGBoost combined with SHAP to predict soil fertility rates, providing transparent explanations for model predictions (Jothi, 2025). Kawakura et al. (2022) analyzed diverse agricultural worker data using LightGBM and XAI methods such as SHAP and LIME to explain individual prediction outcomes (Kawakura et al., 2022). Kumar and Kumar (2024) implemented an XAI-based crop recommendation system combining BiLSTM networks with explainable modules, achieving high predictive accuracy for decision support (Kumar and Kumar, 2024).

For predictive modelling and its explainability, dataset play an important role. Agricultural datasets face multiple challenges. Cravero et al. (2022b) identified issues related to lack of standardized data framework that limits interoperability and integration of datasets across different agricultural systems (Cravero et al., 2022b). Barbedo (2022) highlighted that field-collected data often lack variability and often lack variability and may contain ambiguities due to environmental conditions, sensor differences, and collection inconsistencies (Barbedo, 2022). Generative Artificial Intelligence (GenAI) has emerged as a promising tool for addressing data scarcity and improving analytics in agriculture. Bharambe et al. (2024) explored GenAI applications for sustainable soil analytics, highlighting its ability to optimize fertilizer use, generate realistic synthetic data, and support decision-making in agricultural management (Bharambe et al., 2024). Nyambo et al. (2023) applied Conditional Tabular GANs (CTGAN) to synthesize categorical data for peste des petits ruminants (PPR), demonstrating how generative models can supplement limited field datasets for livestock disease prediction (Nyambo et al., 2023). Dos Santos and Vasconcelos (2025) employed Gaussian Copulas combined with GANs to create synthetic data for beet productivity analysis, showing improved sample coverage and statistical variability compared to traditional CTGAN methods (Dos Santos and Vasconcelos, 2024). Wang et al. (2024) proposed TREB, a BERT-based approach, for tabular data imputation, demonstrating generative techniques for filling missing agricultural data while maintaining realistic attribute dependencies (Wang et al., 2024).

Despite advances in climate-adaptive crop modelling, current studies are limited by scarce and inconsistent datasets. They majorly rely on historical records and lack data augmentation. This restricts model generalization across diverse agro-climatic conditions (Crane-Droesch, 2018, Shelia et al., 2019, Domingues et al., 2022, Cravero et al., 2022b, Barbedo, 2022). Ensemble and advanced models such as RF and reinforced RF achieve strong predictive accuracy (Prasad et al., 2021, Yamparla et al., 2022, Dhillon et al., 2023). However, their outputs often lack interpretability. Integrated explainable AI approaches for multi-crop climate scenario analysis remain underexplored (Hussein et al., 2024, Kumar and Kumar, 2024). Generative AI techniques can generate realistic data and mitigate dataset limitations (Bharambe et al., 2024, Wang et al., 2024). Yet, their integration with predictive, climate-sensitive modelling pipelines is limited. This highlights the need for a reproducible, explainable framework combining data augmentation, predictive modelling, and “what-if” scenario analysis for climate-adaptive multi-crop management. using RF, and (4) SHAP-based explainability for understanding feature relevance in prediction.

3. Methodology and Implementation

The overall workflow of the proposed XCA-GenAI framework is presented in Figure 1. It is designed to provide a reproducible and explainable system that integrates generative AI-based data augmentation with predictive crop-specific climate-sensitivity modeling for “what-if” scenario analysis using the SF24 dataset. The framework comprises four stages: (1) “What-if” climate scenario formulation simulation, (2) synthetic agro-climatic data generation using CTGAN, (3) predictive modelling of crop-sensitivity across different scenarios using RF, and (4) SHAP based explainability for understanding feature relevance in prediction.

3.1 SF24 Dataset

The Smart Farming 2024 (SF24) dataset provides a comprehensive foundation for analysing interactions between crop growth variables and environmental conditions in smart agriculture systems (Aldhahri et al., 2025). The dataset contains 4,800 records of 28 sensor-driven features for various crops. The dataset properties are summarized in Table 1. Several recent studies have employed the SF24 dataset for developing data-driven agricultural solutions (Aldhahri et al., 2025, Srinivas and Monica, 2025, Rani and Banu, 2025, Aasha and Sugumar, 2025, Sasmal et al., 2025).

Table 1. Feature available in the SF24 Dataset

Category	**Features**
Soil Properties	N, P, K, pH, soil moisture, soil type, organic matter
Climatic Variables	temperature, humidity, rainfall, sunlight exposure, wind speed, CO_2 concentration, frost risk
Crop Management Factors	Crop label, irrigation frequency, crop density, pest pressure, fertilizer usage, growth stage, water usage efficiency
Other information	urban area proximity, water source type

3.2 “What-if” Climate Scenarios for Simulation

In this work, ten distinct climate scenarios, as shown in Table 2, were designed to systematically explore the effects of temperature (ΔT) and rainfall (ΔR) variations on crop-related parameters. These scenarios range from Very Hot–Dry to Very Cool–Wet and were specifically chosen due to their high relevance and frequency in real-world contexts. They reflect frequent climatic events, such as heatwaves, droughts, and unusually wet periods, making them realistic for risk assessment. The scenarios strongly influence crop parameters such as soil moisture, pest pressure, etc., and thus directly affect crop growth. These scenarios aim to capture the most prominent aspects of agricultural risk to aid farmers and agronomists in developing climate-adaptive mitigation strategies. Including these extremes also ensures that predictive models are evaluated not only under mean climatic conditions but also under plausible climate scenarios. With changes in temperature and rainfall, the dependent variables humidity and soil moisture are recalculated as follows:

$$\text{humidity}' = h \times (1 - 0.015 \cdot \Delta T) + 0.02 \cdot \Delta R, \quad (1)$$

$$\text{soil\_moisture}' = s \times (1 - 0.02 \cdot \Delta T) + 0.05 \cdot \Delta R, \quad (2)$$

where h and s represent baseline humidity and soil moisture, respectively.

Table 2. Climate scenarios for simulation

Scenario	ΔT **(°C)**	ΔR **(mm)**	**Description**
Baseline	0	0	Reference synthetic data
Hot-Dry	+2	-10	Mild warming, reduced rainfall
Cool-Wet	-2	+10	Mild cooling, increased rainfall
Hot-Wet	+3	+20	Higher temperature with wet conditions
Cool-Dry	-3	-20	Cooler and drier conditions
Very Hot-Dry	+5	-50	Extreme drought stress
Very Cool-Wet	-5	+50	Flood-prone and cool conditions
Moderate Hot-Dry	+1	-25	Light drought stress
Moderate Cool-Wet	-1	+25	Slightly wetter conditions
Extreme Hot	+5	0	Heat stress with unchanged rainfall

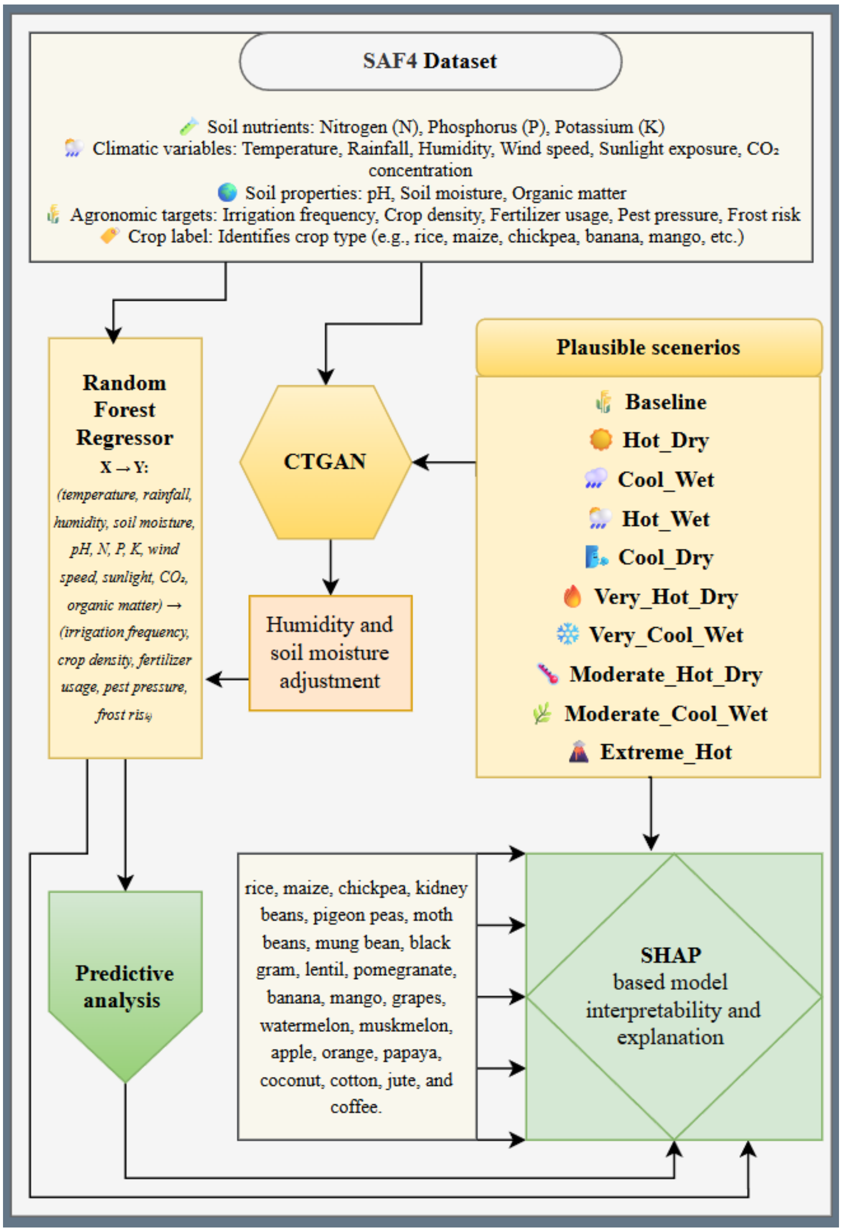

Figure 1. Workflow of the proposed XCA-GenAI framework for agricultural crop “what-if” scenario simulation

3.3 Augmenting SF24 Dataset with Synthetic Data

In similarity with other available agricultural datasets, the SF24 dataset has limitations in the availability of data from extreme climatic scenarios. Generating synthetic data addresses this limitation by expanding the dataset to a wider range of conditions while preserving the statistical properties of the original dataset. In this work, a Conditional Tabular Generative Adversarial Network (CTGAN) is used for synthetic data generation (Nyambo et al., 2023).

CTGAN is a generative model designed for creating synthetic tabular data containing both continuous and categorical features. CTGAN captures the complex distributions and dependencies found in tabular datasets. Its generator produces synthetic records conditioned on discrete attributes, such as crop type, while the discriminator learns to differentiate real from synthetic data. Through adversarial training, the generator iteratively improves until the synthetic data closely matches the statistical patterns of the original dataset. CTGAN explicitly models multi-modal distributions, preserving correlations among variables and maintains relationships between continuous and categorical features. Mode-specific normalization ensures that even infrequently observed categories are adequately represented, while conditional sampling enables targeted generation for specific scenarios, such as specific range of climate conditions in this work.

Mathematically, the CTGAN optimization process consists of four main components: (i) the discriminator’s real-data term, (ii) the discriminator’s fake-data term, (iii) the total discriminator loss, and (iv) the generator’s objective to fool the discriminator. Firstly, the real-data term of the discriminator measures how well the real samples from the original dataset are correctly identified. Secondly, the discriminator’s fake-data term evaluates how well the discriminator detects synthetic samples generated by the generator. Thirdly, the total discriminator loss is obtained by combining the real-data and fake-data terms. This combined loss guides the discriminator to simultaneously maximizes the ability of CTGAN to recognize real samples and reject synthetic ones. Finally, the

generator minimizes its objective by attempting to fool the discriminator. It generates synthetic samples conditioned on the specified attribute and adjusts the parameters so that the discriminator increasingly classifies the generated samples as real.
The CTGAN model was trained for 300 epochs with a learning rate of 0.0001 and a batch size of 128 to produce synthetic data that reflects the variability of climatic parameters and maintains the distribution of the original dataset.

3.4 Predictive Modeling for the SF24 Dataset

In this work RF regressor (Elavarasan and Vincent, 2021) was used for generating the predictive models for different crops using GenAI data augmented SF24 dataset. The predictive model is developed for each crop with input features comprising of nutrient concentrations (N, P, K) and environmental parameters (temperature, humidity, rainfall, soil moisture, wind speed, sunlight exposure, CO_2 concentration, soil pH, and organic matter).
The predicted crop management and stress-related variables include crop density, pest pressure, and frost risk. For each crop and using respective model, predictions were generated for all ten formulated "What-if" climate scenarios.

3.5 SHAP-based Explainability

To interpret the predictions of the RF regressor models and analyze the contribution of input features under each scenario, SHapley Additive exPlanations (SHAP) is used (Kawakura et al., 2022). SHAP provides an interpretable measure of feature influence for each input feature by considering its effect across all possible combinations and additive explanation model. For the explainability of each prediction, SHAP value is calculated for every input feature.
This value reflects the contribution of that input feature in the prediction of the output variable. To reduce the computational overhead in calculation of SHAP Values, we have used TreeExplainer algorithm. TreeExplainer leverages the structure of decision trees to efficiently compute SHAP values for tree-based models such as RF.

In this study, SHAP is applied to interpret the predictions generated by crop-specific RF regression models under different "What-if" climate scenarios. For each crop and scenario, the model predicts management- and stress-related variables such as irrigation frequency, fertilizer usage, crop density, pest pressure, and frost risk based on nutrient (N, P, K) and environmental inputs (temperature, humidity, rainfall, soil moisture, wind speed, sunlight exposure, CO_2 concentration, soil pH, and organic matter). SHAP decomposes each prediction into individual feature contributions by starting from a baseline value and quantify how much each input variable increases or decreases with respect to the baseline value to reach the final predicted value.
The resulting SHAP values provide both the direction and magnitude of each feature's influence on model predictions, which allows explainable interpretation of the predictions. For example, a positive SHAP value for temperature on irrigation frequency indicates that higher temperatures increase the predicted number of irrigation events. Similarly, a negative SHAP value for soil moisture on frost risk implies that higher soil moisture reduces the predicted likelihood of frost damage. Nutrient levels such as nitrogen or potassium may show strong positive or negative contributions to crop density or fertilizer usage depending on the crop type and climate scenario.
By analyzing SHAP values for individual crop and scenario predictions, researchers can understand which features were most influential for a specific crop under a given climate condition, thereby supporting climate-adaptive decision-making and targeted interventions in crop management.

4. Results and Discussion

This section presents the analysis of all the components of the proposed XCA-GenAI framework.

4.1 CTGAN Generated Synthetic Data Analysis

Distribution pattern of synthetic data generated using CTGAN and distribution of real dataset is shown in Figure 2. The

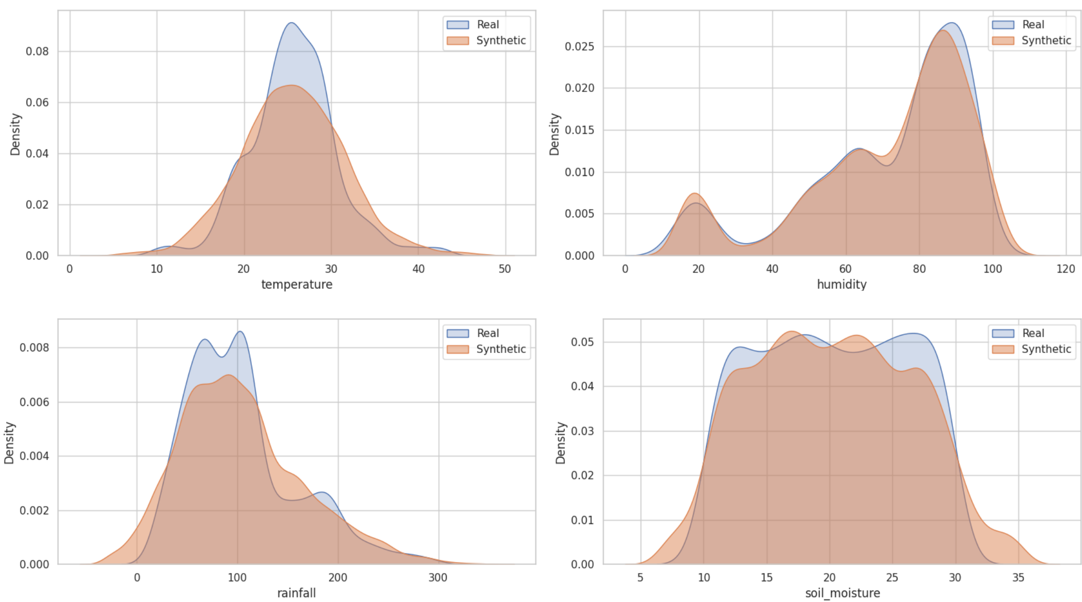

Figure 2. Distribution patterns of real and simulated temperature, humidity, rainfall and soil moisture

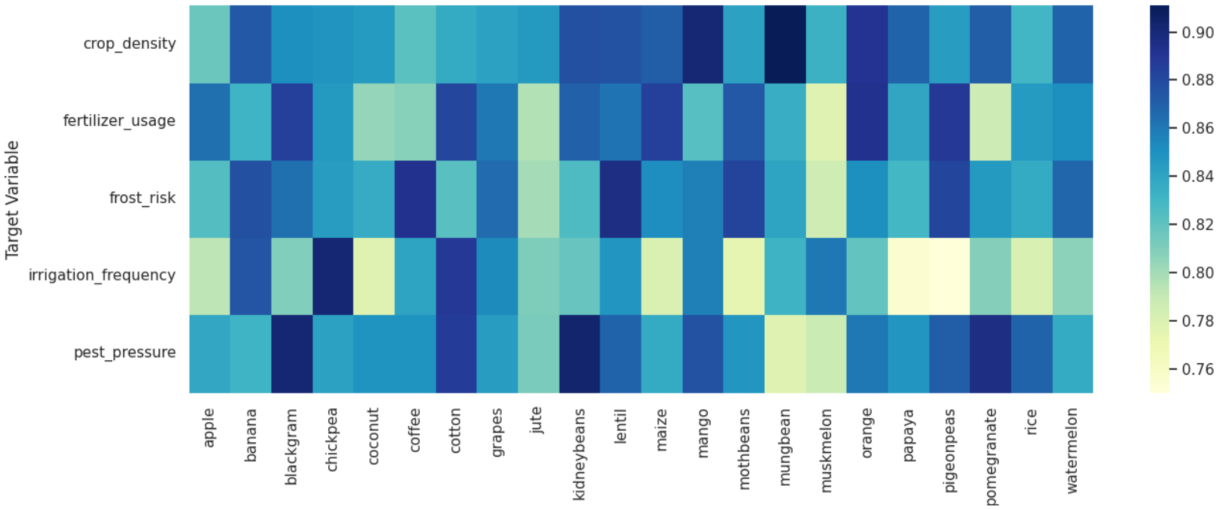

Figure 3. Random Forest prediction accuracy (F1-score) per crop and target variable

temperature of synthetic dataset shows normal distribution, while humidity shows random trend. Rainfall follows a right-skewed pattern and soil moisture is uniformly distributed.

4.2 Prediction Accuracy

The prediction accuracy based on F1-score for crop density, risk and other parameters for different crops is shown in Figure 3. The heatmap reflects the variability and robustness of model in prediction across different crops.

4.3 Climate Scenario Impact

The impact of ten distinct "what-if" climate scenarios is different for different crop, and is illustrated Figure 4. The heatmap visualizes percentage changes (positive or negative) in crop-density, frost-risk, and pest-pressure with reference to baseline values, across ten simulated climate scenarios across different crops.
The heatmap shows that each crop responds differently to changing climate conditions. Some crops have shown strong sensitivity while others remained comparatively stable. Hot–Dry and Hot–Wet scenarios generally lead to increases in crop density and elevated pest pressure across many crops. In contrast, Very Cool scenarios have shown higher frost risk, and reduced crop density for most of the crops. The findings show that cereal crops (rice, maize) are highly sensitive to temperature fluctuations and show pronounced variation in pest pressure and density under thermal extremes. The pulses (mungbean, lentil, pigeonpea) remain relatively stable under moderate climate regimes, suggesting their adaptability in climate fluctuations. The heatmap also reflects that the fruit crops (banana, pomegranate, mango) are more vulnerable to heat and moisture stress. These observations align with the agronomic facts and thus showcase the robustness of proposed XCA-GenAI framework.

4.4 SHAP-Based Feature Importance

The influence of distinct variables in crop growth and risk prediction is different for each crop, and is explained in this framework using SHAP based explainability of the prediction model. The Figure 5 illustrates a bubble representation in

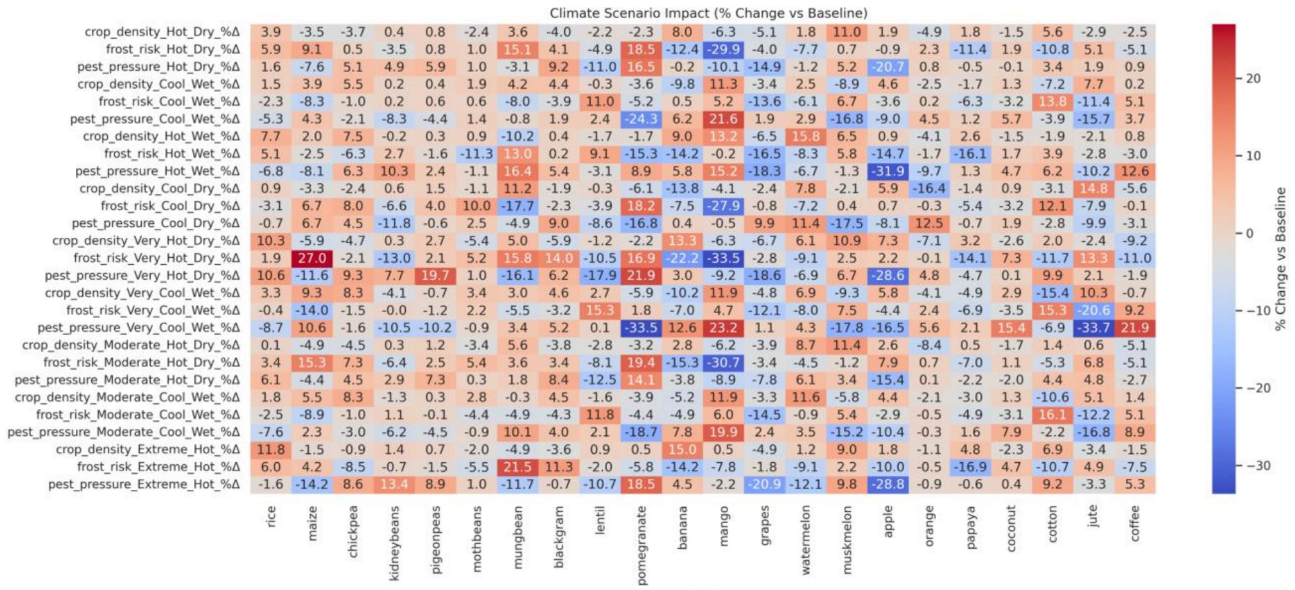

Figure 4. Heatmap illustrating crop responses under ten climate scenarios

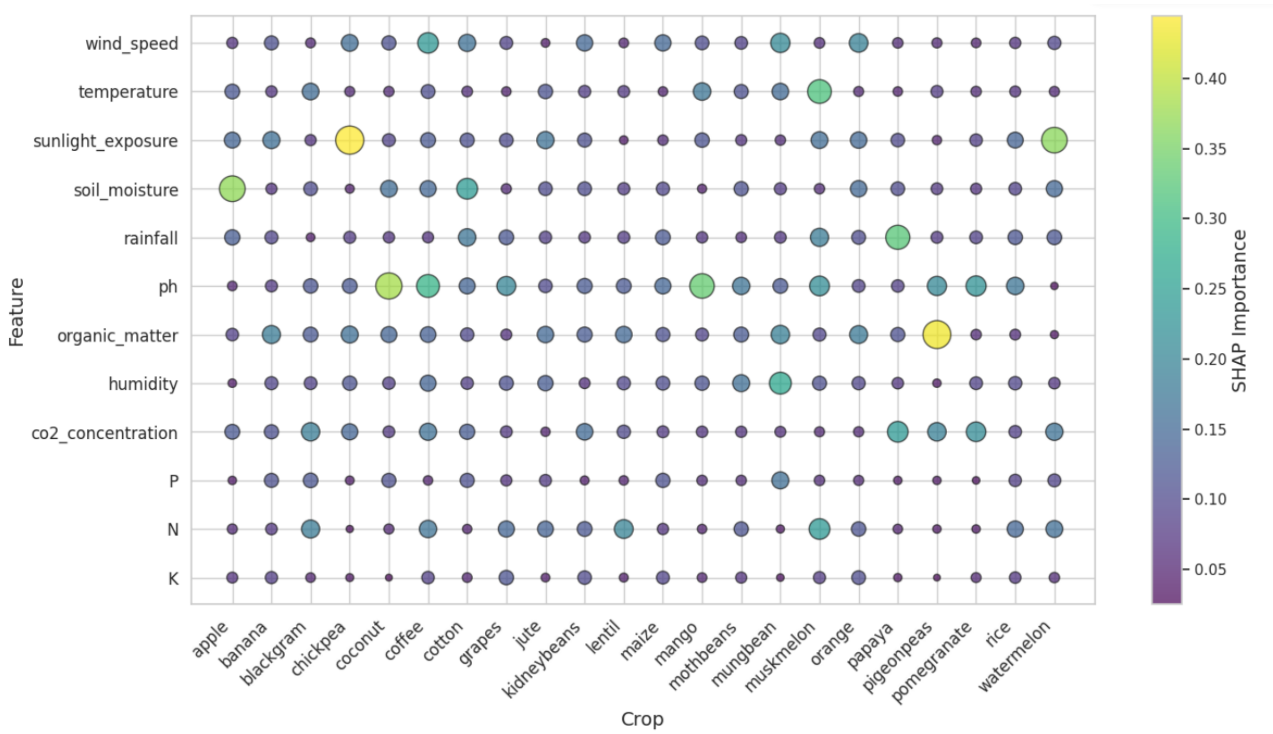

Figure 5. Matrix-style SHAP summary visualization showing crop-specific feature importance.

matrix style to showcase the importance of each variable in prediction of crop density and risk across different crops. The variations in the bubble size and color intensity demonstrate that the predictive model does not rely on a uniform subset of variables for all crops. This heterogeneity in SHAP values suggests that the classifier captured the agro-ecological variability across different crops.

5. Conclusion

This study proposed an integrated and explainable agro-climatic scenario simulation framework that combines generative AI–based synthetic data generation (CTGAN), machine learning prediction, and SHAP-based interpretability to assess climate impacts on crop-related productivity and risk parameters. The agricultural data limitation in extreme and underrepresented climatic conditions is addressed by augmenting the real dataset with synthetically generated climatic and agro-ecological dataset. Further, in this study, ten "what-if" climate scenarios were simulated to evaluate crop density, frost risk, and pest pressure to analyse climate-induced variability across multiple crops. The approach demonstrates how generative modelling and explainable results can enhance agricultural datasets and support adoption of scenario-driven predictive analysis for climate-resilient agriculture.

The results indicate that model performance varies across different crops. With SF24 dataset, the scenario-based simulations revealed heterogeneous crop sensitivity under extreme climatic conditions, where hot–dry and very hot scenarios increased pest pressure and altered crop density patterns, while very cool scenarios elevated frost risk. The results were well aligned with the observed agronomic patterns. SHAP-based analysis further demonstrated that climatic variables influence predictions for temperature-sensitive crops, and soil properties influence leguminous and perennial crops.

This framework provides transparency and enhances interpretability for the usability of proposed XCA-GenAI framework by agronomists, policymakers, and farmers by translating model outputs into actionable insights. The scenario-based sensitivity analysis supports adaptive strategies by identifying crops that are vulnerable to thermal and rainfall extremes. This analysis can be used as data-driven insight for empirical targeted interventions such as irrigation management, soil improvement, and pest control.

This framework can be advanced by incorporating multi-source remote sensing data such as satellite-derived vegetation indices, land surface temperature, soil moisture products, and high-resolution spatial observations to improve spatial transferability and real-time monitoring. Integration of near-real-time weather forecasts would further enhance predictive capability for operational decision support. Additionally, the framework can be expanded to include spatially explicit risk mapping and temporal dynamics for region and crop specific agricultural monitoring. The combination of generative AI, explainable machine learning, and remote sensing has potential to offer a scalable pathway for climate-smart agriculture under increasing environmental uncertainty.

Acknowledgement

This work is supported by Discovery Grant of the Natural Sciences and Engineering Research Council (NSERC) of Canada.

References

Aasha, S., Sugumar, R., 2025. Sensor malfunction simulation and data imputation using DL for precision agriculture. *Indian Journal of Science and Technology*.

Ahmed, M., Hayat, R., Ahmad, M., Ul-Hassan, M., 2022. Impact of climate change on dryland agricultural systems: A review of current status, potentials, and further work need.

Akkem, Y., Biswas, S. K., Varanasi, A., 2024. Streamlit-based crop recommendation with explainable AI. *Neural Computing and Applications*.

Aldhahri, E. A., Almazroi, A. A., Alkinani, M. H., Ayub, N., Alghamdi, E. A., Janbi, N. F., 2025. Smart farming: Enhancing urban agriculture through predictive analytics and resource optimization. *IEEE Access*, 13, 72375–72388.

Allgaier, J., Mulansky, L., Draelos, R. L., Pryss, R., 2023. How does the model make predictions? A systematic literature review on explainability in machine learning for healthcare. *Artificial Intelligence in Medicine*, 143, 102616.

Barbedo, J. G. A., 2022. Data fusion in agriculture: Resolving ambiguities and closing gaps. *Sensors*.

Bharambe, U., Mahato, M., Durbha, S., 2024. Generative ai for sustainable soil analytics. *Geospatial Technology and Applications*, Springer.

Cartolano, A., Cuzzocrea, A., Pilato, G., 2024. Assessing explainable AI models for smart agriculture. *Multimedia Tools and Applications*.

Chang, Y., Latham, J., Licht, M., Wang, L., 2023. A data-driven crop model for maize yield prediction. *Communications Biology*.

Crane-Droesch, A., 2018. Machine learning methods for crop yield prediction and climate change impact assessment in agriculture. *Environmental Research Letters*, 13(5), 053001.

Cravero, A., Pardo, S., Galeas, P., López Fenner, J., Caniupán, M., 2022a. Data type and data sources for agricultural big data and machine learning. *Sustainability*, 14(23), 16131.

Cravero, A., Pardo, S., Sepúlveda, S., Muñoz, L., 2022b. Challenges to machine learning in agricultural big data. *Agronomy*.

Dhillon, M. S., Dahms, T., Kuebert-Flock, C. et al., 2023. Integrating random forest and crop modeling improves yield prediction for wheat and oilseed rape. *Frontiers in Remote Sensing*, 4, 102345.

Dockerty, T., Lovett, A., Appleton, K., Bone, A., Sünnenberg, G., 2006. Developing scenarios and visualisations to illustrate policy and climatic influences on future agricultural landscapes. *Agriculture, Ecosystems & Environment*, 114(1), 103–120.

Domingues, T., Brandão, T., Ferreira, J. C., 2022. Machine learning for detection and prediction of crop diseases and pests: A comprehensive survey. *Agriculture*, 12(4), 567.

Dos Santos, D., Vasconcelos, J., 2024. Using Gaussian Copulas and Generative Adversarial Networks for Generating Synthetic Data in Beet Productivity Analysis. *Sugar Tech*, 27.

Elavarasan, D., Vincent, P. M. D. R., 2021. A reinforced random forest model for enhanced crop yield prediction. *Journal of Ambient Intelligence and Humanized Computing*, 12(9), 8797–8812.

Hussein, E., Zerouali, B., Bailek, N., Abdessamed, D., Ghoneim, S., Santos, C., Hashim, M., 2024. Harnessing Explainable AI for Sustainable Agriculture: SHAP-Based Feature Selection in Multi-Model Evaluation of Irrigation Water Quality Indices. *Water*, 17, 34.

Jothi, R., 2025. Soil fertility rate prediction using xgboost and shap. *Farming and Smart Agriculture for Sustainable Future*, Taylor & Francis.

Kawakura, S., Hirafuji, M., Ninomiya, S., Shibasaki, R., 2022. Analyses of Diverse Agricultural Worker Data with Explainable Artificial Intelligence: XAI based on SHAP, LIME, and LightGBM. *European Journal of Agriculture and Food Sciences*, 4, 11-19.

Krishna, C. S., Sai, S. P., Roy, P. D., Sathya, V., 2025. Agroinsight pro: Sustainable agriculture analytics & decision support using gen-ai. *Proceedings of the 4th International Conference on Smart Technologies, Communication and Robotics (STCR)*, IEEE, 1–5.

Kumar, S., Kumar, M., 2024. Explainable ai - based crop recommendation system. *International Conference on Signal Processing*, IEEE.

Mbow, C., Rosenzweig, C., Barioni, L. G., Benton, T. G., Herrero, M., Krishnapillai, E., Liwenga, P., Pradhan, M. G., Sapkota, F. N., Tubiello, Y. X., 2019. *Climate Change and Land: IPCC Special Report on Climate Change, Desertification, Land Degradation, Sustainable Land Management, Food Security, and Greenhouse Gas Fluxes*. IPCC.

Nyambo, D. G., Ngulumbi, N., Mduma, N., 2023. Data synthesis technique for categorical peste des petits ruminants (ppr) data using ctgan model. *2023 First International Conference on the Advancements of Artificial Intelligence in African Context (AAIAC)*, 1–5.

Prasad, N. R., Patel, N. R., Danodia, A., 2021. Crop yield prediction in cotton using random forest. *Spatial Information Research*, 29(4), 573–585.

Rani, N. S., Banu, N. S., 2025. Iot-enabled smart irrigation with stacked ml algorithms. *3rd International Conference on Data*, IEEE.

Raza, A., Safdar, M., Adnan Shahid, M., Shabir, G., Khil, A., Hussain, S., Khan, M., Aziz, S. U., Azam, S., Sattar, J., Muhammad, N. E., 2024. Climate change impacts on crop productivity and food security: an overview. *Transforming Agricultural Management for a Sustainable Future: Climate Change and Machine Learning Perspectives*, —, 163–186.

Sasmal, B., Das, S. C., Barai, S., Biswas, S., 2025. Hybrid ml and xai approach for sustainable crop recommendation. SSRN. Available: https://papers.ssrn.com/abstract$_id = 5404609$.

Shelia, V., Hansen, J., Sharda, V., Porter, C. et al., 2019. A multi-scale and multi-model gridded framework for forecasting crop production and climate risk. *Environmental Modelling & Software*, 114, 83–97.

Srinivas, A. C. M. V., Monica, C. L., 2025. Optimizing smart irrigation with fuzzy logic and genetic algorithms. *5th International Conference (IEEE)*.

Ukhurebor, K. E., Adetunji, C. O., Olugbemi, O. T., Nwankwo, W. et al., 2022. Precision agriculture: Weather forecasting for future farming. *AI, Edge and IoT-based Smart Agriculture*, Academic Press, 101–121.

Wang, S., Zhou, W., Wang, S., Zheng, R., 2024. Treb: A bert attempt for imputing tabular data. arXiv preprint arXiv:2410.00022.

Wen, G., Cao, Y., Wei, X., 2025. The data-driven analysis of soil health and crop adaptability: Technologies, impacts, and optimization strategies. *Advances in Resources Research*, 5(1), 350–368.

Yamparla, R., Shaik, H. S., Guntaka, N. S. P., 2022. Crop yield prediction using random forest algorithm. *Proceedings of the 7th International Conference on Smart Computing and Communications (ICSCC)*, IEEE, 101–106.

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLIX-B3-2026-1147-2026