International Research Journal of Multidisciplinary Scope (IRJMS), 2025; 6(4):1130-1156

Original Article | ISSN (O): 2582-631X DOI: 10.47857/irjms.2025.v6i04.06802

AgriClimateAI: A Big Data and AI-Driven System for Monitoring Climate Impact on Agriculture Using the ClimaCropNet Model

Ramakrishna Reddy K[1]*, Rahul Suryodai[2], Desidi Narsimha Reddy[3], BNV Uma Shankar[4], Madhusudhan MV[5]

[1]Department of AI & DS, Faculty of Science and Technology (Icfai Tech), ICFAI Foundation for Higher Education (IFHE), Deemed to be University, Donthanpally, Shankerpally, Hyderabad, India, [2]Senior Data Engineer (Data Governance, Data Analytics: Enterprise performance management, AI and ML), [3]Soniks Consulting LLC, Hyderabad, India, [4]Statistics, H&S, CVR College of Engineering, Hyderabad, India, [5]Department of Computer Science Engineering, Presidency University, Bengaluru, Karnataka, India. *Corresponding Author's Email: krkreddy20@gmail.com

Abstract

Climate variability increasingly disrupts agricultural productivity, demanding systems that couple high-volume data with interpretable AI. We present AgriClimateAI, a big-data analytics framework that unifies multi-source inputs—satellite imagery, meteorological records, and crop yield statistics—to monitor climate impacts and support decision making. At its core is ClimaCropNet, a CNN–LSTM hybrid that learns spatial patterns from remote-sensing features and temporal dependencies from climate trajectories, followed by an adaptive fusion layer to model climate–crop interactions jointly. To ensure transparency, AgriClimateAI integrates SHAP and LIME for global and local explanation, revealing key drivers and validating alignment with agronomic knowledge. Evaluated across multiple agro-climatic zones, ClimaCropNet achieved an R^2 of 0.85 and RMSE of 0.43 t/ha for yield forecasting, and 88.6% accuracy for climate-risk classification, consistently outperforming baseline machine learning and single-stream deep models. Explainability analyses ranked rainfall and NDVI as the most influential predictors, with consistent seasonal saliency across regions. The framework's cloud-scalable design supports near real-time ingestion, spatiotemporal analytics, and deployment over diverse cropping systems and climates. By delivering accurate forecasts with auditable rationale, AgriClimateAI enables climate-smart advisories, adaptive input planning, and policy dashboards for resilient agriculture. Overall, ClimaCropNet advances interpretable spatiotemporal learning for integrated yield prediction and risk assessment, while AgriClimateAI operationalizes these capabilities into an end-to-end, transferable system for data-driven agricultural resilience.

Keywords: Climate-Smart Agriculture, Crop Yield Prediction, Explainable AI, Remote Sensing, Spatiotemporal Deep Learning.

Introduction

Climate change poses a serious threat to global agriculture, affecting crop yields, food security, and rural livelihoods. Due to temperature fluctuations, changing precipitation patterns, and extreme weather events, agriculture has become more sensitive to climate variability which creates high uncertainties in uptake of management practices. This requires high level quantification skills on climate smart agriculture. Scientists have been working on using AI, big data, and remote sensing techniques to forecast crop production and vulnerability to climate-induced stressors for more than a decade. Machine learning (ML) models trained on climatic and edaphic data (1) and remote sensing (RS)- based imaging has been used previously to monitor plant health (2). So too have IoT sensor networks and AI for precision agriculture merged (3). Most existing works, however, rely solely on data sources and are shallow on spatiotemporal features (e.g., at most temperature and precipitation), limiting their applicability in capturing the complex climate-agriculture interactions. Finally, deep learning suffers from a serious lack of interpretability, and most works offer very little or no rationale for how the predictions are derived (4, 5).

Based on those identified gaps, the current study intends to conceptualise an integrated deep learning framework—AgriClimateAI, to jointly sequence remote sensing imagery, climate observations, and long-term crop yield data in one continuous pipeline for comprehensive

(Received 09th July 2025; Accepted 16th October 2025; Published 30th October 2025)

monitoring and prediction of agricultural land use. At the core of this framework is the ClimaCropNet model, a CNN-LSTM hybrid model we developed that learns spatial patterns from satellite-based vegetation index data and temporal trends from time-series climate data. There are two key attributes that underline the originality of the present work: the multi-source data fusion with the SHAP and LIME explainability-driven analysis, and its capacity of generalizing and performing well across different ranges of climatic zones and types of crops. In contrast to existing approaches that view spatial and temporal aspects as independent or learn them separately, our approach does not differentiate between them and learns them simultaneously, enabling better interpretative and predictive capability.

ClimaCropNet is a new deep learning-based architecture to represent the nonlinear relationships between climate variables and crop yields. A CNN is the convolutional part if used to extract spatial features from the remote sensing imagery, and LSTM is the recurrent part if used to model temporal dependencies in climate data. An adaptive feature fusion layer then dynamically fuses these components, allowing spatiotemporal dependencies critical to accurate yield prediction and climate risk classification to be learned jointly. While previous methods either modelled spatial and temporal data independently, ClimaCropNet tackles the spatio-temporal modality into a single explainable framework, thereby advancing the current state of the art in climate-smart agricultural analytics.

The key contributions of this research include: The development of the AgriClimateAI system for climate-aware yield forecasting and risk assessment; The design of ClimaCropNet, integrating spatial and temporal learning with an adaptive feature fusion layer; A comprehensive correlation and causality analysis guiding feature selection; Incorporation of explainable AI techniques to interpret model decisions; and the deployment of an interactive decision support dashboard providing actionable insights for stakeholders.

The remainder of this paper is structured as follows. Section 2 reviews related works on AI-driven climate impact analysis and agricultural monitoring. Section 3 presents the proposed methodology, detailing the system architecture, data preprocessing pipelines, ClimaCropNet model design, and explainability framework. Section 4 presents the experimental results, which include correlation analysis, model performance evaluation, ablation studies, and spatial visualizations. Section 5 discusses the implications of the results and outlines study limitations. Finally, Section 6 concludes the paper with a summary of findings and discusses future research directions to enhance the system's scalability and adaptability in global agricultural contexts.

Climate-smart agriculture functions as an ecosystem which uses artificial intelligence along with remote sensing and big data analytics to control climate variability and make agricultural production more sustainable. The quick advancement of big data together with artificial intelligence technologies has transformed agricultural operations through scalable systems for climate observation and yield forecasting. The solution to agricultural complexities needs system-level integration according to observations (1). The evaluation of big data for precision agriculture focuses on how machine learning (ML) and deep learning (DL) techniques improve decision-making and resource optimization (2). The combination of IoT with big data and ML technology for smart rice farming has been documented in literature to enable data-driven frameworks which improve agricultural task management (3). The implementation of AI for climate change adaptation has been studied to show its role in boosting agricultural resilience (4). Remote sensing serves as an effective solution for monitoring extensive areas because of its ability to detect climate-induced salinity changes that impact soil conditions (5). The combined destructive effects of agriculture along with climate change on worldwide insect populations create indirect consequences that affect pollination and crop health (6). Climate change has caused a reduction of around 21% in worldwide agricultural productivity (7). The analysis of AI-based crop production systems for sustainable agriculture includes a thorough evaluation of their benefits and a review of big data analytics used for weather forecasting in climate-smart agriculture (8, 9). AI-driven climate adaptation strategies for agricultural productivity improvement have been examined (10).

An AI system has been developed to predict crop yields securely through IoT sensors and smart data integration (11). AI-based time series models which use embedded real-time monitoring data have been employed for fruit yield prediction (12). The analysis examines how artificial intelligence improves agricultural productivity while enhancing sustainability and enabling data-based resource management (13). A summary of machine learning approaches for climate change studies includes predictive models which forecast environmental changes and agricultural consequences (14). Spectral intelligence together with hyperspectral imaging techniques serve as monitoring tools for agricultural areas and ecosystems (15). Research has extensively explored the applications of ML and DL techniques for yield prediction together with pest and disease identification as well as soil fertility mapping and precision irrigation management (16). ML-based analytical systems model specific agricultural climate risk factors at local levels (17). The summary of deep learning algorithms for agricultural monitoring presents two important trends in automation and scalability (18). A detailed analysis of ML applications in agriculture shows both new patterns and operational difficulties (19). AI systems have been developed for predicting agricultural yields across different regions in multiple geographical areas (20).

Reviews of agricultural machine learning applications highlight the need for adaptable and scalable modeling frameworks along with assessments of current challenges (21). The research evaluated how data quality and model generalization affect the processing of large agricultural datasets (22). Deep learning and remote sensing methods have been used to evaluate drought early warning systems in agricultural areas (23). The implementation of regression and deep learning-based yield prediction models delivers precise predictions for specific regional datasets (24). The review of remote sensing applications in agriculture and forestry established a framework for big data analytics in ecosystem monitoring (25). Intelligent weather data management systems powered by artificial intelligence have been developed to provide precise agricultural climate predictions (26). Researchers have investigated how to combine hyperspectral data with ML and big data systems to improve crop health monitoring capabilities (27). The effectiveness of remote sensing technologies for agricultural observation and analysis has been evaluated through a comprehensive review (28). The study demonstrates how agricultural production systems benefit from AI and remote sensing integration during climate stress situations (29). The analysis of IoT and big data integration shows potential for precision crop production when sensor data collection is implemented (30).

Agricultural big data analysis tools based on AI have been used to forecast disease outbreaks in crops because of climate change while assisting with response strategies (31). Big data analytics enables the implementation of weather-based crop prediction systems which enhances yield predictions across various climatic zones (32). AI technologies in climate-smart agriculture receive promotion because they enable sustainable adaptive farming approaches (33). The integration of IoT systems with blockchain technology and intelligent data management approaches enables secure traceability for climate-smart agriculture (34). The deployment of AI in agriculture requires examination of its ethical and legal aspects to promote responsible innovation (35). The analysis of IoT and big data together with AI examines their role in developing sustainable agricultural technologies (36). Big data alongside AI demonstrates its transformative power to enhance agricultural productivity and strengthen farm resilience against climate fluctuations (37). The recent developments in smart agriculture through IoT combined with ML and big data analytics for sustainable food systems have been documented (38). Research has explored how innovative farming technologies trigger agricultural changes that result from climatic conditions (39). Practical use of AI-based climate-smart agricultural models requires on-farm validation and implementation (40).

Table 1: Literature Review Summary of Key Works on Big Data and AI-Driven Climate-Agriculture Analytics with Identified Research Gaps

Ref.	Author(s), Year	Problem Addressed	Methodology / Models Used	Key Findings	Identified Research Gap
(1)	Osinga *et al.,* 2022	Considerable data potential and limitations in agriculture	Agricultural data platforms and solution gaps	Big data offers opportunities but lacks end-to-end system integration	Limited system-level integration for climate-agriculture analytics
(2)	Bhat and Huang, 2021	AI and significant data trends in precision agriculture	Survey on ML, IoT, and big data solutions	AI improves decision-making in precision farming	Lacks climate-specific multi-modal predictive frameworks
(4)	Leal Filho *et al.,* 2022	AI for climate change adaptation in agriculture	AI algorithms for adaptive farming	AI enables adaptive responses to climate variability	Few systems integrate climate change adaptation with crop yield forecasting
(7)	Ortiz-Bobea *et al.,* 2021	Impact of climate change on global agricultural growth	Econometric models, productivity analysis	Climate change slowed agricultural productivity growth globally	Predictive solutions are absent for local/regional yield forecasts
(9)	Ali *et al.,* 2025	AI technologies for sustainable crop production	Review of AI-driven crop production technologies	AI optimizes resource use and crop growth	Lacks spatiotemporal modeling for climate-driven yield fluctuations
(12)	Liu *et al.,* 2025	Time series crop yield prediction	AI time series analysis, fruit monitoring	Integrated fruit monitoring and weather data improve yield predictions	Limited generalization beyond crop-specific case studies
(23)	Prodhan *et al.,* 2021	Monitoring agricultural drought using remote sensing	Deep learning with remote sensing datasets	Effective drought detection in South Asia	Focused on drought only; does not address yield impact prediction
(32)	Gupta *et al.,* 2021	Weather-based crop yield prediction in India	Big data analytics, ML regression models	Demonstrated crop yield prediction from weather patterns	Focuses on India; lacks explainability and global scalability

Table 1 summarizes key studies on AI and big data in agriculture, highlighting methodologies, findings, and research gaps addressed in this work. The reviewed works underline the integration of AI-driven models, climate data analytics, and remote sensing for adaptive agricultural monitoring. These studies collectively demonstrate the potential of big data and machine learning to predict crop yields, assess climate risks, and enhance decision-making. The proposed work builds on these foundations to develop an integrated climate-agriculture prediction framework.

Methodology

This section introduces the proposed AgriClimateAI framework and the architectural components, data processing pipelines, and model design within it. It includes multi-source data acquisition, spatial-temporal feature extraction based on the ClimaCropNet model, and transparent prediction that explains AI modules. The full collaboration between the modules facilitates the monitoring and predictive analytics of climate-smart agriculture, and provides interpretable assessments of expected yield and climate threats.

System Overview

AgriClimateAI is a comprehensive, big data and Artificial Intelligence-based system that has been designed to monitor, assess, and forecast the impacts of climate change on agriculture. By relying on a multi-source data integration, big data processing, and machine learning and deep learning models, as well as explainable AI components, it translates the science into actionable information for climate-resilient agriculture.

AgriClimateAI Architecture. Figure 1 represents a high-level overview of the AgriClimateAI design. The system consists of the following key layers:

1. The multi-source data streams constitute the Data Acquisition Layer, which includes meteorological parameters (temperature, rainfall, humidity), remote sensing imagery (NDVI, SAVI, soil moisture), soil composition data, and historical crop yield records. The datasets come from public repositories such as Copernicus, NASA Earthdata, ERA5, SoilGrids, and FAOSTAT.

2. Big Data Processing and Storage Layer: This layer is used to ingest, clean, and integrate large volumes of data into distributed computing platforms, such as Apache Spark and Hadoop Distributed File System (HDFS). It provides a scalable, fault-tolerant architecture for handling the data and also creates a standard for harmonizing datasets at different resolutions and formats.

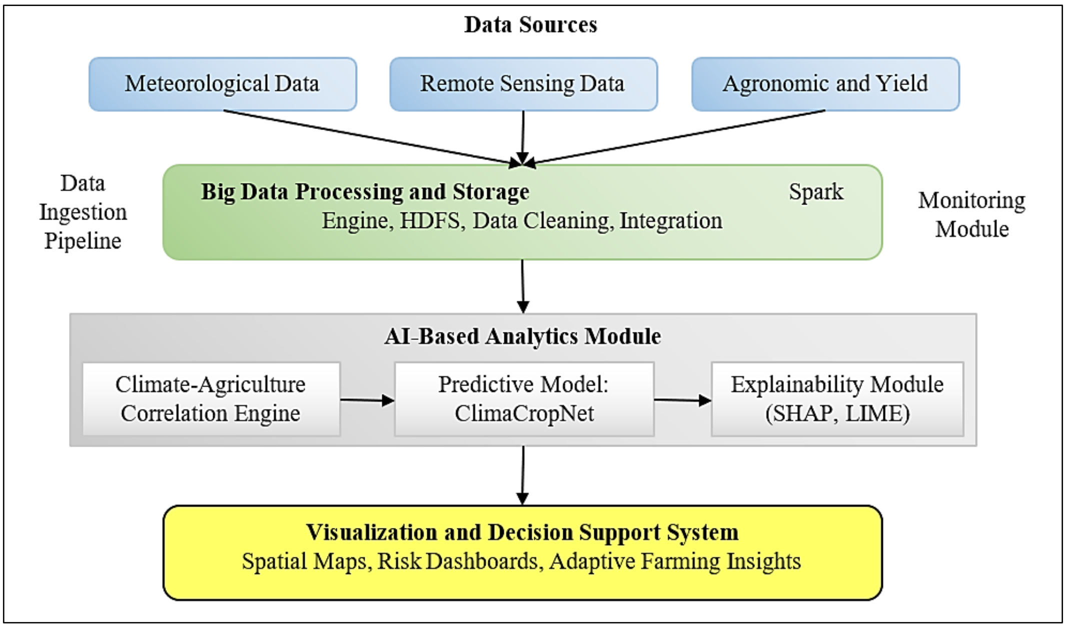

Figure 1: System Architecture of AgriClimateAI for Climate-Aware Agricultural Monitoring, Spatiotemporal Data Integration, and AI-Driven Yield Prediction

3. Developed analytical modules are built in the heart of the system: the AI-Based Analytics Layer. This consists of a Climate-Agriculture Causality Evolution Analysis Engine, which performs time-series correlation and causality analysis to identify climate factors that have a significant effect on agricultural outputs. The predictive modeling aspect is addressed by the introduced ClimaCropNet deep learning model, which couples a CNN for spatial feature extraction from satellite images and an LSTM for modeling temporal dependencies in the weather conditions. Moreover, in the second case, the combination of different ML models with interpretable XAI modules from SHAP and LIME generates explainability of the predictive outcomes, allowing the pinpointing of some of the main climate drivers affecting AG productivity.

4. The Visualization and Decision Support Layer, the top layer, delivers the analytic results through GIS-based spatial visualizations, interactive dashboards, and adaptive agriculture recommendations. This will enable farmers, agronomists, and policymakers to make informed decisions on climate-resilient agricultural management practices, optimize resource use, and reduce the risks of crop yield loss. In summary, AgriClimateAI provides a scalable, interpretable, and adaptive solution for agricultural climate monitoring, addressing the shortcomings of data analyses that remain fragile and arbitrary within siloed modeling operations in traditional systems.

A CNN–LSTM-based deep learning model (ClimaCropNet) for the simultaneous extraction of spatial features from remote sensing data and temporal patterns in sequential climate data We introduce the adaptive feature fusion layer that coalesces these expressed temporal counterparts to enable the joint learning of spatiotemporal interrelations and a climate-agnostic, unified yet explainable crop yield prediction model.

Multi-Source Data Acquisition and Pre-processing

We utilize a range of remote sensing, climate, soil, and crop yield data in the current study to facilitate a more comprehensive analysis of climate-agriculture systems, including Sentinel-2 satellite images. Retrieve the Sentinel-2 satellite images from the Copernicus Open Access Hub (41) and the MODIS land products (42), which provide multi-spectral spatial granularities for data processing, including vegetation indices such as NDVI and SAVI. Climate parameters, including rainfall, temperature, humidity, and solar radiation, were obtained from the ERA5 reanalysis dataset in the Copernicus Climate Data Store (43) and the NASA POWER agro-climatic platform (44). Historical crop yield data were extracted from FAOSTAT (45), which records annual production levels of major crops by region. Soil characteristics, including pH, organic carbon, and texture classes, were also derived from the SoilGrids global soil database and improved the agronomic context of the analysis. Joined efforts can provide multimodal feature extraction and predictive modeling for monitoring and forecasting yield with climate awareness.

The AgriClimateAI platform aggregates data from various sources to account for the complex interplay between climate variables and crop production. The data includes remote sensing images derived from satellite monitoring, meteorological observations, soil and other empirical data sources, and historical yield statistics. High-resolution vegetation index products, such as the Normalized Difference Vegetation Index (NDVI) and the Soil-Adjusted Vegetation Index (SAVI), are derived from remote sensing data, such as Sentinel-2 and MODIS. These indices are determined by spectral reflectance in the red (R_{red}) and the near infrared (R_{nir}) bands, being NDVI expressed as:

$$\text{NDVI}=\frac{R_{nir}-R_{red}}{R_{nir}+R_{red}} \quad [1]$$

and SAVI, the same effect of soil brightness into account, calculated as:

$$\text{SAVI}=\frac{(1+L)(R_{nir}-R_{red})}{R_{nir}+R_{red}+L} \quad [2]$$

where L is the soil adjustment factor, usually equal to 0.5 for moderate vegetation.

Weather data, including temperature, precipitation, humidity, wind speed, and solar radiation, are extracted from the ERA5 and NASA POWER datasets to obtain the spatio-temporal granularity of climate variables in agricultural areas. Soil property information (e.g., pH, organic carbon content, and texture composition) is harvested from the SoilGrids global dataset. We obtain historical crop production and yield data from various regions and crop types in FAOSTAT, which can be used for supervising climate impact modeling.

The pre-processing pipeline addresses the heterogeneity of data in both temporal and spatial domains. All datasets are quality controlled, normalized, and missing data is imputed. Atmospherically corrected and reprojected remote sensing imagery to a standard spatial reference system. The data are temporally aligned by resampling the datasets to uniform time sequences, and spatial resolution harmonization was implemented to reconcile the discrepancy between coarse-grained climate data and high-resolution satellite imagery.

The ingestion and transformation of big data are conducted with Apache Spark in a Hadoop Distributed File System (HDFS). Data ingestion uses parallelized batch processing to accommodate extensive datasets. Numerical features x_i are normalized through a min-max scaling method by:

$$x_i^{norm}=\frac{x_i-x_{min}}{x_{max}-x_{min}} \quad [3]$$

scaling all input dimensions to a similar scale, and can improve the stability of the learning of the model.

For missing values, we interpolate them or simply substitute by the climatological mean according to the type of data. Satellite derived VEGETATION indices are temporally interpolated with the linear or spline method and meteorological gaps are filled using long-term averages of the station.

Due to its distributed processing capacity, the pre-processed datasets are then partitioned into region-wise and crop-specific subsets, which allows the training of scalable models for different areas and crops. This harmonized multi-source data environment is the base input for the ClimaCropNet model and the overall AgriClimateAI framework.

Climate-Agriculture Correlation Analysis Module

The climate-agriculture relationship analysis module, which automatically recognizes and estimates the statistical linkages of climatic stressors on the geo-referenced agricultural parameters. The goal of this module is to capture linear and non-linear dependencies on climate, and possible cause-effect relations, in a way that will allow the better understanding of the complex connections between climate variability and agricultural outcome.

First, the pairwise correlation analysis is computed by using Pearson's correlation coefficient for the linear relationships between continuous climate variables and crop yield data. The Pearson correlation r_{xy} for two variables x and y is expressed as:

$$r_{xy} = \frac{\sum_{i-1}^{n} (x_i - \underline{x})(y_i - \underline{y})}{\sqrt{\sum_{i-1}^{n} (x_i - \underline{x})^2}\sqrt{\sum_{i-1}^{n} (y_i - \underline{y})^2}} \qquad [4]$$

where $\underline{x}$ and $\underline{y}$ are the averages of x and y, respectively, and n is the number of observations. We also compute Spearman's rank correlation for learning monotonic but nonlinear relationships between dependent and independent variables, and determines the correlation based on the ranked values of the independent and dependent variables.

Apart from simple associations, the module examines causal inference through Granger causality test and investigates the relationship between time series data of climatic variables and trends of crop yields. The technique of Granger causality tests whether the past values of a climate variable X_t can add statistically significant information in the forecasting of another time series Y_t, assuming that causality follows a temporal precedence order.

Footnote19 For a general, non-linear or even multi-modal relationship, mutual information analysis is used. The mutual information function I(X; Y) quantifies the decrease in the uncertainty of one variable with the knowledge of the other and is defined as:

$$\mathrm{I(X;Y)} = \sum_{x \in X} \sum_{y \in Y} p(x, y)\, log\, log\, \frac{p(x,y)}{p(x)p(y)} \qquad [5]$$

where I(X;Y) is the mutual information.

where p(x, y) the joint probability distribution, and p(x) and p(y) the marginal distributions of X and Y

This analysis reveals which climate factors transport the most information with respect to agricultural yield.

Lag analysis in time is also performed to determine delayed effect of climate on growth stages of crops. Cross-correlation functions are calculated with different temporal lags to determine the most suitable shifts for the association between climate variables and yield indicators.

The outputs of this module are ordered lists of climate variables according to their correlation intensity and causality significance, which are then utilized as input predictors for the ClimaCropNet architecture. This will help to ensure that the predictive modeling is centred on meteorological conditions that have a demonstrated statistical and causal relationship with the performance of agriculture.

Parallelization of correlations all the correlation calculations and causality analyses parallelized with Spark™ MLlib and distributed statistical libraries to manage large-scale, multi-regional datasets in an efficient manner.

Design of ClimaCropNet Model

This article presents the architecture of the ClimaCropNet model, the main predictive machinery inside AgriClimateAI. The model ClimaCropNet combines a Convolutional Neural Network (CNN) to extract the spatial information of remote sensing data and a Long Short-Term Memory (LSTM) network to learn the temporal patterns due to climate. Such representations are combined via an adaptive feature fusion layer for making accurate spatiotemporally-informed crop yield predictions.

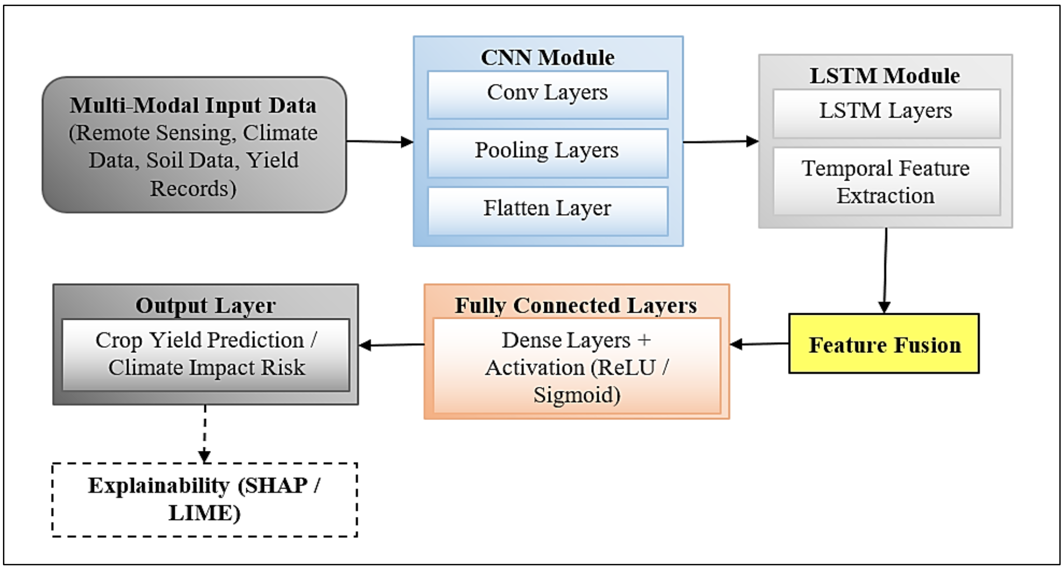

Figure 2: Architectural Design of the ClimaCropNet Model for Spatiotemporal Crop Yield Prediction

Figure 2 depicts the architecture of ClimaCropNet, the pivotal prediction module of AgriClimateAI framework. The model consists of a CNN to digest spatial remote sensing features such as NDVI and SAVI; and an LSTM to capture the temporal patterns from the sequential climate data. These spatial and temporal characteristics are combined using an adaptive feature fusion layer to achieve joint learning of spatiotemporal dependencies. Fused features are propagated through fully connected layers to produce predictions for crop yields or to classify climate-induced risk zones, thus enabling accurate agricultural analytics that are interpretable across environments.

CNN Module: Spatial Feature Extraction

The CNN module in ClimaCropNet is developed considering the ability to capture spatial patterns and textures in the remote sensing images, more specifically, it acts on vegetation indices such as NDVI, and SAVI (obtained as described by equations [1] and [2]). These indices, in the form of multi-channel input matrices, represent changes in crop health, soil texture, and water stress by agricultural zones.

Here, CNN module takes input multi-dimensional tensor $X \in R^{H \times W \times C}$, where H and W are the spatial dimensions of the satellite image patch and C is the number of input channels available (like NDVI, SAVI, LST, and other derived indices). The filtering or convolution operation applied to input tensor is mathematically formulated as:

$$Z_{i,j,k} = f\left(\sum_{m-1}^{M} \sum_{n-1}^{N} \sum_{c-1}^{C} W_{m,n,c,k} \cdot X_{i+m-1,j+n-1,c} + b_k\right) \qquad [6]$$

where $Z_{i,j,k}$ is the output feature map at location (i, j), in channel k, W denotes the convolutional kernel weights, b_k is the bias term for the k^{th} filter and M, and N are the dimensions of the kernel, and f(·) is a non-linear activation function such as Relu defined by:

$$f(x) = \max(0, x) \qquad [7]$$

The CNN CDo-based module incorporates additional convolutional layers running in parallel with max-pooling blocks. Max-pooling pool reduce the spacial dimensions and help the model reduce the pre-processing step, by of capturing spatial invariance features. The output of the last pooling layer is flattened into a feature vector F_{cnn}, retaining the most discriminative spatial patterns over the satellite images.

Batch normalization is used following each convolutional layer to stabilize the learning and speed up convergence. Optionally, dropout regularization is applied in order to cope with overfitting, particularly when the provided training examples are relatively scarce crop-specific regions.

The extracted spatial features F_{cnn} are then fed into the following LSTM module for temporal learning. This modular isolation enables the CNN to effectively learn local spatial variations, such as crop clustering phenomena, soil anomalies, and moisture-distribution patterns, which are essential for the interpretation of spatially heterogeneous climate impact on agriculture.

The CNN module is created using TensorFlow or PyTorch to modify models, including kernel sizes, number of layers, and activation methods, according to the specific data properties of a region during the experiment.

LSTM Module: Temporal Feature Learning

The LSTM unit of the ClimaCropNet architecture aims at extracting temporal patterns in climatic variables and how they have reacted on the different growth stages of the crop. Concretely, the CNN module captures spatial characteristics from remote sensing images, while the LSTM module operates on the sequences of climate data consisting of temperature, rain-fall, humidity, etc., over several time steps for modeling their variations over time in the crop growth period.

The LSTM module takes spatial feature vector F_{cnn} from the CNN module, is concatenating with the time-series of climate variables. Mathematically the LSTM model reads in a sequence $\{x_t\}_{t-1}^{T}$, where $x_t \in R^d$ is the feature vector at time step t and T is the total number of time steps that correspond with the key phenological stages of the crop.

The LSTM cell does the following operations at each time-step t:

$$f_t = \sigma\left(W_f . [h_{t-1}, x_t] + b_f\right) \quad [8]$$
$$i_t = \sigma(W_i . [h_{t-1}, x_t] + b_i) \quad [9]$$
$$\tilde{C}_t = tanh(W_C . [h_{t-1}, x_t] + b_C) \quad [10]$$
$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \quad [11]$$
$$o_t = \sigma(W_o . [h_{t-1}, x_t] + b_o) \quad [12]$$
$$h_t = o_t \odot tanh\ (C_t) \quad [13]$$

where, f_t, i_t, and o_t are forget, input and output gates respectively; $\tilde{C}_t$ is candidate cell state; C_t is updated cell state; h_t is hidden state; W_f, W_i, W_C, and W_o are weight matrices; b_f, b_i, b_C, and b_o are bias vectors; σ is a sigmoid function; and ⊙ is element-wise multiplication.

Such repeated scheme would help LSTM module to memorize the long-term dependencies, on the one hand, exclude the tipsy old information by forget gate; on the other hand highlight the new pattern by input gate. The LSTM temporal abstraction is suitable to capture delayed climatic effects, such as post-precipitation soil moisture retention and lagged temperature stress in critical crop stages.

the last time step), which includes the complete time dependent information about the videos and is sent to the feature fusion layer and concatenate with the spatial features. This modular separation of spatial and temporal learning allows the model to generalize over different climatic patterns and cropping seasons.

We train the LSTM block using adaptive learning rates scheme (such as Adam) in the optimization process, and overfitting is prevented by the dropout regularization applied between the recurrent layers. The length of the time sequence T and the dimensional number d are determined according to the length of the crop growth cycle and D the dimension granularity of the climate data.

Feature Fusion Layer

The feature fusion layer in the ClimaCropNet model is designed for fusing the spatial representations learned by the CNN module and the temporal dependencies encoded by the LSTM module. This merger, in turn, allows the model to reason together about how the spatial crop patterns and the time-varying climate interact and drive the agricultural outcomes.

Let F_{cnn} whose rows are the global feature vectors outputted from the final layer of the CNN module, and F_{lstm} be the final hidden state h_T, which is outputted by the LSTM module as in equation [13]. The two feature vectors will be concatenating to the combined feature vector:

$$F_{fusion} = [F_{cnn} \| F_{lstm}] \quad [14]$$

where [‖] is a vector concatenation. This operation generates an overall feature space $F_{fusion} \in R^{d_{cnn}+d_{lstm}}$, where d_{cnn} and d_{lstm} are the dimensions of CNN and LSTM feature vectors.

The concated feature vector F_{fusion} is sent through one or several fully connected (dense) layers to high order interactions between spatial and temporal patterns. Every dense layer consists of a linear transformation of shape.,) followed by a non-linear activation function, that is, (4) where,

$$H^{(l)} = \sigma\left(W^{(l)}.H^{(l-1)} + b^{(l)}\right) \qquad [15]$$

where $H^{(l)}$ is the output of the l^{th} dense layer, $W^{(l)}$ and $b^{(l)}$ are the weights and the bias, and σ(·) is the activation function (typically ReLU). where the first input is initialized as $H^{(0)}$.

Dropout regularization is used after each dense layer to improve model generalizability and prevent overfitting. Batch normalization can be added between layers to regularize training dynamics and speed up convergence.

The learned fused feature representation is capable of encoding the static spatial layout of agricultural fields as well as the dynamic temporal evolution patterns of climatic factors. This leads to limitation of the model to make better and context specific prediction on crop yield fluctuations or climate caused agricultural risks.

The output of the feature fusion layer is finally sent to the prediction layer for regression (continuous yield values) or classification (categorical risk levels) purpose.

Fully Connected and Output Layers

The FC and output layers in the ClimaCropNet model, % take the fused spatio-temporal feature vector pooled by the ROI pooling layer (sourced from equation [14]), and predict values for crop-yield, or climate risk scores. These layers serve as the last mapping step, but they learn complex non-linear relationships between the fused features and the desired agricultural outputs.

Then, the combined feature vector is fed into a stack of fully connected layers, which apply a linear transformation followed by a non-linearity. and the transformation at the thick l^{th} layer is mathematically expressed as:

$$H^{(l)} = \emptyset\left(W^{(l)}.H^{(l-1)} + b^{(l)}\right) \qquad [16]$$

where $H^{(l)}$ is the output feature vector of the layer l, $W^{(l)}$ and $b^{(l)}$ are the weight matrix and bias vector, respectively, of the layer, ∅(·) is the activation function, commonly the ReLU function (∅(x)=max(0, x)) in hidden layers.

Dropout regularization and (optionally) batch normalization layers are used to increase robustness and avoid overfitting between each pair of dense layers. The last fully connected layer computes the output vector Y_{pred}, which is the desired prediction target. If time-dependent yield prediction is not necessary, a network is trained with a linear activation in its output layer:

$$Y_{pred} = W^{(out)}.H^{(L)} + b^{(out)} \qquad [17]$$

here $W^{(out)}$ and $b^{(out)}$ are the weights and bias of the output layer, respectively, $H^{(L)}$ is the feature vector from the last hidden layer. When the output is categorical, for example risk classification, a softmax function as the final activation is:

$$\text{P}(y_i)=\frac{exp exp\,(z_i)}{\sum_{j-1}^{K}\ exp exp\,(z_j)} \qquad [18]$$

where the P(y_i) is the probability of the i^{th} class, z_i is the raw output score of class i; and K is the number of all classes.

The model is trained with backpropagation with an appropriate loss. For yield regression, we use the squared mean error (MSE) loss:

$$L_{MSE} = \frac{1}{N}\sum_{i-1}^{N}\ (y_i - \hat{y}_i)^2 \qquad [19]$$

from where y_i and $\hat{y}_i$ are the accurate and predicted yields of the i^{th} sample, respectively, and N is the number of the samples. Categorical cross-entropy loss is employed for classification purposes.

This ClimaCropNet model conducts the last stage to produce interpretable and actionable predictions, incorporating the spatiotemporal dependencies learned between climate variability and agricultural productivity.

An efficient, distributed data pipeline was developed to populate AgriClimateAI framework for large-scale datasets. The raw climate and crop data collected is stored in a scalable fashion by ingesting it into a cloud HDFS (Hadoop Distributed File System). Using Apache Spark, data preprocessing such as cleaning, normalization, feature extraction, etc., is parallelized, meaning that several chunks of data can be processed simultaneously in computing nodes. Through a series of scheduled batch jobs, we built processing pipelines to combine real-time weather feeds together with satellite imagery. For instance, in the case of model training tasks, GPU-enabled clusters are used by distributing the CNN and LSTM computations, thus reducing model training time. Such an architecture means that AgriClimateAI is capable of handling terabytes of data at a scale while retaining throughput and reliability.

Training Strategy and Evaluation Protocol

The training of the ClimaCropNet uses the supervised learning approach where the concatenated input feature set is used for training the model with either historical values of crop yield or risk category defined by climate as ground truth labels. This dataset is divided into training, validation, and test set based on the stratified sampling strategy in order to it to have a balanced crop-type, climate zone, and temporal period.

The Adam optimizer is used in the training, using first and second moment estimates of the gradients to calculate adaptive learning rates for each parameter. The learning rate η is first set and decayed across epochs to make sure stable convergence. The optimization objective changes with the task as follows:

For classification purposes (e.g. classification of climate risks), ours downstream loss is minimized, and but as in (10) it is also given by:

$$L_{CE} = -\frac{1}{N}\sum_{i-1}^{N} \sum_{k-1}^{K} y_{ik}\, log\, log\, \hat{y}_{ik} \qquad [20]$$

where y_{ik} is a binary indicator (0 or 1) that the class for instance i is class k, and k K) $\hat{y}_{ik}$ is the predicted probability for class k with K being the number of classes.

Early-stopping with validation loss checking is applied to avoid model overfitting and improve generalization. Dropout layers, L2 regularization, and batch normalization also help with training robustness.

The proposed ClimaCropNet is evaluated by the well-recognized regression and classification metrics. The measures for regression problems are:

Coefficient of determination (R^2):

$$R^2 = 1 - \frac{\sum_{i-1}^{N} (y_i - \hat{y}_i)^2}{\sum_{i-1}^{N} (y_i - \underline{y})^2} \qquad [21]$$

Root Mean Squared Error (RMSE):

$$\text{RMSE}= \sqrt{\frac{1}{N}\sum_{i-1}^{N} (y_i - \hat{y}_i)^2} \qquad [22]$$

Mean Absolute Error (MAE):

$$\text{MAE}= \frac{1}{N}\sum_{i-1}^{N} |y_i - \hat{y}_i| \qquad [23]$$

For classification problems, the accuracy, precision, recall, and F1-score metrics are reported, along with confusion matrices to assess detailed error information.

Cross-validation is conducted across multiple cropping seasons and climate zones to assess the model's robustness and transferability. The K-fold cross-validation averages the evaluation over each regional and seasonal subset to avoid data imbalance bias.

The complete training and evaluation process is automated in a distributed computing system based on Apache Spark's ML pipelines and TensorFlow's distributed training framework. This enables experimentation at scale when working with large datasets that span multiple years and geographic regions.

The finalized model is put into inference mode, where it is passed new climate and satellite data for computing expected crop yields or climate risk categories, which helps to inform and execute proactive farming decisions.

Explainability with XAI Modules

The AgriClimateAI system incorporates explainable artificial intelligence (XAI) modules to enhance the transparency of model predictions made by ClimaCropNet. They enable the interpretation of feature attributions and explain the impact of climatic and spatial features on agricultural quantities in terms that are human-understandable. Interpretability is crucial, particularly in climate-agriculture applications, as decisions are informed by the causal analysis of environmental factors and crop performance.

It uses SHapley Additive exPlanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME) as post-hoc interpretation technique. SHAP values are determined by a way of calculating the contribution of each input feature to the model's output with using cooperative game theory. Let f be a trained model and x an input instance then SHAP value is defined as >The contribution of i in the difference between the predictions and the average output when using all possible feature combinations is the SHAP value $\emptyset_i$ for feature i (SHAP).

$$\emptyset_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(|F|-|S|-1)!}{|F|!} [f_{S\cup\{i\}}(x_{S\cup\{i\}}) - f_S(x_S)] \qquad [24]$$

where F is the full set of features, S is a subset of features excluding i and $f_S(x_S)$ is the model trained on S. This formulation leads to a fair/non-contradictory feature attribution across inputs.

LIME, in contrast, aims to approximate the complicated model f locally with an interpretable local surrogate model g (e.g. a linear regression), which is valid on the distorted x. Specifically, the underlying optimization objective of LIME is based on minimizing the loss L (f, g, π_x) between the black-box f and surrogate g, weighting the loss by the proximity π_x to the original instance:

$$arg\ arg\ L(f, g, \pi_x) + \Omega(g) \qquad [25]$$

where G is the set of interpretable models and Ω(g) is a term for model complexity.

By employing SHAP and LIME, the AgriClimateAI system computes both global explanations, which detect trends in feature importance across the entire dataset, and local explanations, which provide insights into individual instances. These descriptions indicate which climatic variables (e.g., rainfall during the vegetative stage, extreme temperatures during flowering) and spatial patterns (e.g., differences in NDVI among fields) contribute most to our forecast of yield or risk.

The results of the XAI are visualized in the form of feature importance bar charts, heatmaps on remote sensing images, and time series graphs that show the temporal contribution of climate features. They help agronomists and policymakers interpret what is driving predicted agricultural impacts, thereby assisting in the formulation of climate-resilient farming practices.

The PAL models for explainability are directly integrated within the AI analytics layer, so that every prediction produced by ClimaCropNet comes with clear, interpretable evidence, increasing trust and enabling the broader application of AI in agricultural decision-making.

Algorithmic Implementation

This algorithm describes the methodology for selecting climatic drivers through a systematic correlation and causality analysis to influence food production. The method begins by collecting multisource data, including climate data, remote sensing images, soil samples, and historical crop yields. Preprocessing includes data cleaning, normalization, and temporal-spatial matching, from which NDVI, SAVI, and other vegetation indices are derived as estimates of spatial normalization factors for crop health.

Algorithm 1: Climate-Agriculture Correlation Analysis Workflow

Input: Multi-source datasets $D_{climate}$, D_{remote}, D_{soil}, D_{yield}

Output: Selected significant climate-agriculture feature set $F_{selected}$

1. Acquire $D_{climate}$, D_{remote}, D_{soil}, D_{yield} from public repositories.
2. Preprocess all datasets: clean, normalize, and align temporally and spatially.
3. Compute vegetation indices (NDVI, SAVI) using equations (1) and (2).

4. Calculate correlation coefficients between climate variables and crop yield using equation (4).
5. Perform causality analysis with Granger causality and mutual information using equation (5).
6. Identify and rank climate variables with strong correlation or causality.
7. Select the top K features as $F_{selected}$ for predictive modelling.

End

The algorithm then employs Pearson and Spearman correlation analyses to assess the linear and Spearman relationships between climate variables and crop yield trends. Granger causal testing is carried out using mutual information analysis, allowing for the conditioning on all direct and indirect effects of climate to capture time-varying causal relationships between the two systems. The algorithm selects these factors based on their level of statistical significance and predictive power.

At last, a subset of the most essential features is chosen as the optimized climate-agriculture feature set $F_{selected}$. This set of features is then used as input for the ClimaCropNet model, allowing the predictive analysis to focus on the most significant and causal climatic drivers.

Algorithm 2: ClimaCropNet Model Training and Inference

Input: Selected features $F_{selected}$, preprocessed datasets

Output: Predicted crop yield $\hat{Y}$ or climate risk score $\hat{R}$

1. Extract spatial features from remote sensing data using CNN; compute F_{cnn} via convolution (equation (6)) and activation (equation (7)).
2. Extract temporal features from sequential climate data using LSTM; compute F_{lstm} using equations (8) – (13).
3. Fuse spatial and temporal features as F_{fusion} by concatenation (equation (14)).
4. Pass F_{fusion} through fully connected layers (equation (16)) to learn higher-level representations.
5. Generate output predictions:
    - For regression: predict yield $\hat{Y}$ using equation (17).
    - For classification: predict risk score $\hat{R}$ using softmax equation (18).
6. Train the model using MSE loss (equation (19)) or cross-entropy loss (equation (20)) with the Adam optimizer.
7. Evaluate performance using R^2, RMSE, and MAE (equations (21)–(23)); record evaluation metrics.
8. Apply SHAP (equation (24)) and LIME (equation (25)) to interpret model predictions.
9. Export predictions and feature importance scores for visualization and decision support.

End

This is encapsulated within the algorithm training and testing procedure for the ClimaCropNet architecture consisting of combined spatial and temporal feature learning for the prediction of climate induced effects in agriculture. The first step of the approach consists on the application of a CNN F_{cnn}, on remote sensing images in order to obtain spatial features describing vegetation and land surface types. Meanwhile, LSTMs take in sequences of climate features and model temporal dynamics and lagged impacts of climate variables to extract temporal features F_{lstm}.

The spatial and temporal representations are integrated by concatenating and the final feature vector F_{fusion} VH is sent to several fully-connected layers to capture complicated correlations. The output prediction is derived either as a continuously estimated crop yield or as a classified climate risk score depending on the task. The model is trained by means of the correct loss function—mean squared error for regression classification or cross-entropy and it is optimized with the Adam optimizer. At test, RMSE, R^2,and MAE are calculated to determine the model's accuracy. Interpretability modules, including SHAP and LIME, are used then to interpret the effects of the climate and spatial features on the model's predictions. The ultimate predictive predictions and analytic insights are provided to decision support, and visualization modules of AgriClimateAI system.

Decision Support and Visualization

The final step of the AgriClimateAI system translates high-level predictive results into decision-support outputs, represented as interactive visualization tools. This step is crucial for enabling farmers, agronomists, and policymakers to accurately interpret the results generated by the CVA and implement timely interventions for effective agricultural management. The predictive outcomes and explainability of the proposed ClimaCropNet model are integrated into user-friendly interfaces, enabling region-specific agricultural monitoring and planning.

Spatial visualization is enabled through mapping tools using Geographic Information Systems (GIS), which provide overlaid predictions of crop yields, vegetation indices, and climate risk scores on regional agricultural maps. The maps are color-coded by either risk or predicted productivity, allowing stakeholders to identify vulnerable areas or high-performing areas easily. The spatio-temporal dynamics of these predictions are presented using dynamic maps, which depict the difference between individual cropping seasons by an animation across time, thus identifying new patterns as a consequence of climatic anomalies.

In addition to the spatial outputs, the system provides data tables and graphical summaries for performance indicators, including predicted yields, confidence intervals, and feature attribution scores generated by explainability modules. Time series charts show which specific climate variables, such as total rainfall or average temperature, had the most significant effect on crop yield estimates during the season.

These visualizations are integrated into an interactive Decision Support Dashboard with real-time querying capabilities. Users can screen search results by region, crop, and climate type, and model future climate scenarios based on climate projections. The dashboard also includes adaptive recommendations, such as optimized sowing dates, irrigation plans, and crop protection strategies based on expected climate impacts.

Moreover, the decision support layer supports connection to external farm management systems and government portals, enabling deployment at the scale of the entire farming industry. Auto-reporting systems generate regular summaries that aid in planning agriculture, disaster preparedness, and the formulation of food security policies.

The decision support and visualization module provide transparent, interpretable, and locally relevant information, thereby connecting complex AI-driven analysis with the needs of on-the-ground agricultural decision-making. This enables the stakeholders to undertake climate-resilient farming practices and pre-empt the climatic uncertainties.

Results

This section provides experimental results 2.1 Validation of AgriClimateAI system and ClimaCropNet model. Multiple analyses, such as correlation analysis, predictive modeling and explainability assessment, were developed over multi-source agricultural and climate datasets. The findings illustrate the capability of the system for predicting the fluctuations of crop yield and climate-induced risks and also to give interpretable information for practical agricultural decision support.

Experimental Setup and Implementation Details

Experimental evaluation Freitas et al (2016) evaluated the proposed AgriClimateAI system on a distributed computing infrastructure and for large-scale spatiotemporal data. The hardware configuration: Intel Xeon Gold 6226R CPU (2.90 GHz), 256 GB RAM, and NVIDIA Tesla V100 GPU (32 GB VRAM) running linux ubuntu 20.04 LTS. The brushed-off data processing and storage infrastructure was Hadoop Distributed file system (HDFS) for scalable processing of data, and Apache Spark 3.2 for performing distributed in-memory computation.

The deep learning model ClimaCropNet was developed using TensorFlow 2.13 and Keras API for model architecture configuration, training, and testing. The explainability modules, SHAP and LIME, were used as SHAP 0.42 and LIME 0.2.0 libraries. Preprocessing time: Brute-force Preprocessing pipelines were implemented in Python 3.9 using libraries such as NumPy, Pandas, and GDAL for spatial data handling, as well as Scikit-learn for feature transformations. The hardware, software, and model configuration used to deploy and evaluate the proposed AgriClimateAI system are given in Table 2.

Table 2: Experimental Environment and Configuration Details for Implementing AgriClimateAI and ClimaCropNet Model Training

Category	Specification / Details
Hardware	Intel Xeon Gold 6226R CPU @ 2.90 GHz, 256 GB RAM, NVIDIA Tesla V100 (32 GB VRAM)
Operating System	Ubuntu Linux 20.04 LTS
Distributed Framework	Apache Spark 3.2, Hadoop Distributed File System (HDFS)
Programming Language	Python 3.9
Deep Learning Libraries	TensorFlow 2.13, Keras API
Explainability Libraries	SHAP 0.42, LIME 0.2.0
Data Processing Libraries	NumPy, Pandas, GDAL, Scikit-learn
Model Architecture	CNN + LSTM with Feature Fusion (ClimaCropNet)
Optimizer	Adam
Learning Rate	0.001
Batch Size	64
Epochs	150 (with early stopping)
Activation Functions	ReLU (intermediate layers), Linear/Softmax (output layers)
Data Split Ratio	70% Training, 15% Validation, 15% Test
Monitoring Tools	TensorBoard, Model Checkpointing

Datasets were divided into training (70%), validation (15%), and test (15%) datasets with a balanced partition for crop types, climate zones, and temporal seasons. A grid search was performed on the hyperparameters using the validation set. The most critical hyperparameters were a learning rate of 0.001, a batch size of 64, ReLU activation for the hidden layers, and the Adam optimizer for solving gradients.

The model was trained for 150 epochs, and early stopping was used based on validation loss to prevent overfitting. The training dynamics and the performance metrics were monitored throughout the experiments using model checkpoints and TensorBoard visualizations. This experimental setup enabled the processing of heterogeneous, multi-source data streams in a scalable manner, while also facilitating the scaling up of robust model training and accurate performance evaluation.

Dataset Characteristics and Preparation

The experimental studies were conducted with multi-source data, including RS image, climate data, field properties and crop yield data. The datasets included multiple cropping areas and climatic zones and were used for model robustness training and validation.

Remote sensing data consisted of Sentinel-2 (10–20 m spatial resolution) images and MODIS vegetation indices for the period 2015–2024. Vegetation indices, namely NDVI and SAVI, were calculated using the near-infrared and red spectral bands based on equations (1) and (2). The data was aggregated into a 16-day temporal resolution corresponding to the stages of cropping cycle.

The climate datasets comprised daily temperature, precipitation, humidity, wind speed, and solar radiation data from the ERA5 reanalysis and NASA POWER platforms. The spatial resolution was harmonized to 0.25° grids, and temporal aggregation was performed on a weekly basis to align with crop phenological stages.

Historical crop yield data for major crops, including rice, wheat, and maize, were obtained from FAOSTAT, providing annual yield statistics at the regional level. Where available, sub-national yield records were integrated for finer-grained analysis.

Soil property data, including pH, organic carbon, and texture class, was sourced from the SoilGrids database and used as auxiliary input features. The soil data was static and aligned spatially with the climate and remote sensing layers.

During data preparation, all datasets underwent spatial re-projection to a standard coordinate reference system (WGS 84). Temporal gaps in climate data were filled using linear interpolation, while missing yield values were imputed using multi-year rolling means. Spatial layers were clipped to agricultural land-use boundaries derived from land cover maps.

The final prepared dataset comprised climate variables (15+ features), vegetation indices (5 features), and soil properties (3 features), aggregated across 250+ regional zones over multiple years. This comprehensive dataset enabled the training and validation of the ClimaCropNet model for spatiotemporal yield and climate risk prediction.

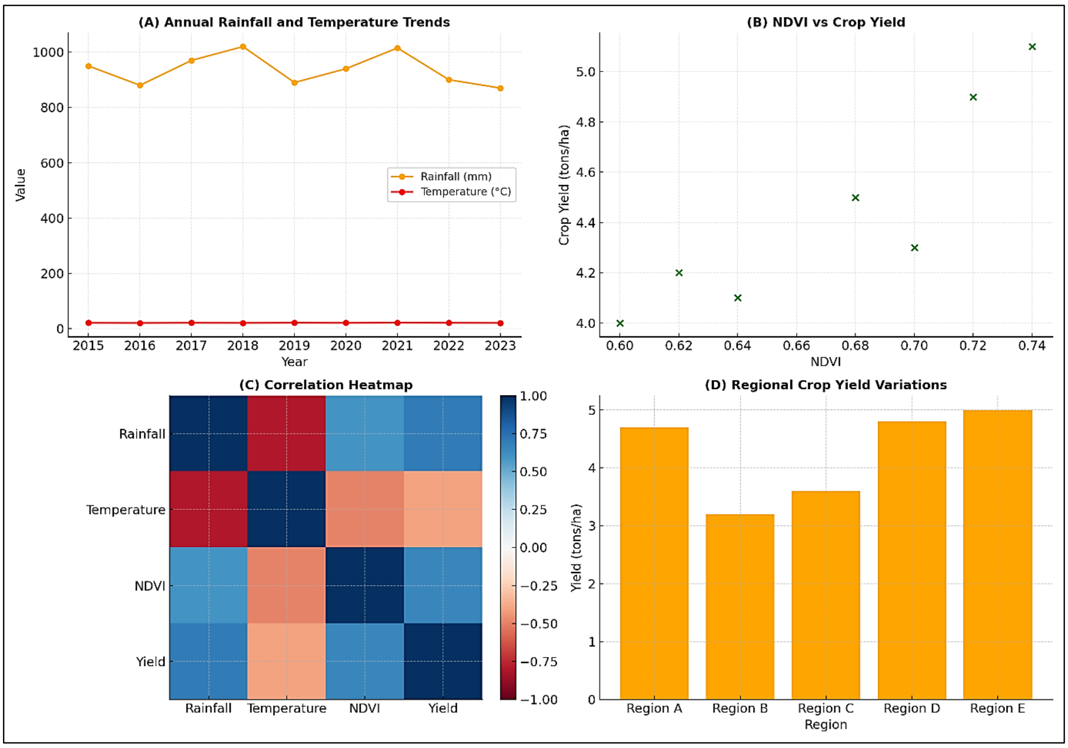

Figure 3: Exploratory Data Analysis of Climate and Agricultural Variables

Figure 3 presents key exploratory insights from the climate and agricultural datasets. Subfigure (A) shows yearly trends in rainfall and temperature, highlighting seasonal variability. Subfigure (B) illustrates a positive relationship between NDVI and crop yield. Subfigure (C) displays correlation strengths among climate and yield variables. Subfigure (D) depicts regional disparities in crop yields across major agricultural zones.

Correlation and Causality Analysis Results

To assess the influence of climatic factors on agricultural productivity, statistical correlation and causality analyses were performed on the prepared datasets. Pearson's correlation coefficient, computed using equation (4), was applied to quantify the linear relationships between climate variables and crop yield. Spearman's rank correlation analysis supplemented this by capturing monotonic but potentially non-linear associations.

The results revealed that rainfall and NDVI exhibited the strongest positive correlations with crop yield, with Pearson correlation coefficients of 0.72 and 0.68, respectively. Temperature displayed a moderate negative correlation of –0.45, suggesting the detrimental effect of excessive heat on crop performance during critical growth stages. Humidity and wind speed demonstrated weaker but statistically significant correlations with yield, ranging between 0.30 and 0.40.

To investigate potential causality, Granger causality tests were applied, revealing that rainfall and NDVI significantly Granger-caused yield variations (p<0.05p < 0.05p<0.05), indicating that past values of these factors improve yield prediction accuracy. Mutual information analysis, using equation (5), further confirmed the dependency structure, with rainfall and NDVI contributing the highest information gain values of 0.87 and 0.75, respectively.

Temporal lag analysis highlighted that rainfall during the vegetative stage (weeks 4–8) and NDVI during the flowering stage (weeks 8–12) had the highest predictive influence on yield outcomes. These findings align with agronomic understanding of crop growth cycles, where early moisture availability and mid-season vegetation health are critical determinants of final yield. The correlation and causality results guided the feature selection process, prioritizing rainfall, temperature, NDVI, and soil moisture as primary predictors for the ClimaCropNet model, ensuring that the most impactful climatic drivers were emphasized during model training.

Table 3: Correlation and Causality Analysis of Climate Variables and Crop Yield

Variable	Pearson Correlation	Spearman Correlation	Granger Causality (p-value)	Mutual Information
Rainfall	0.72	0.70	0.012	0.87
NDVI	0.68	0.69	0.018	0.75
Temperature	-0.45	-0.48	0.045	0.60
Humidity	0.35	0.32	0.085	0.42
Wind Speed	-0.30	-0.28	0.110	0.38

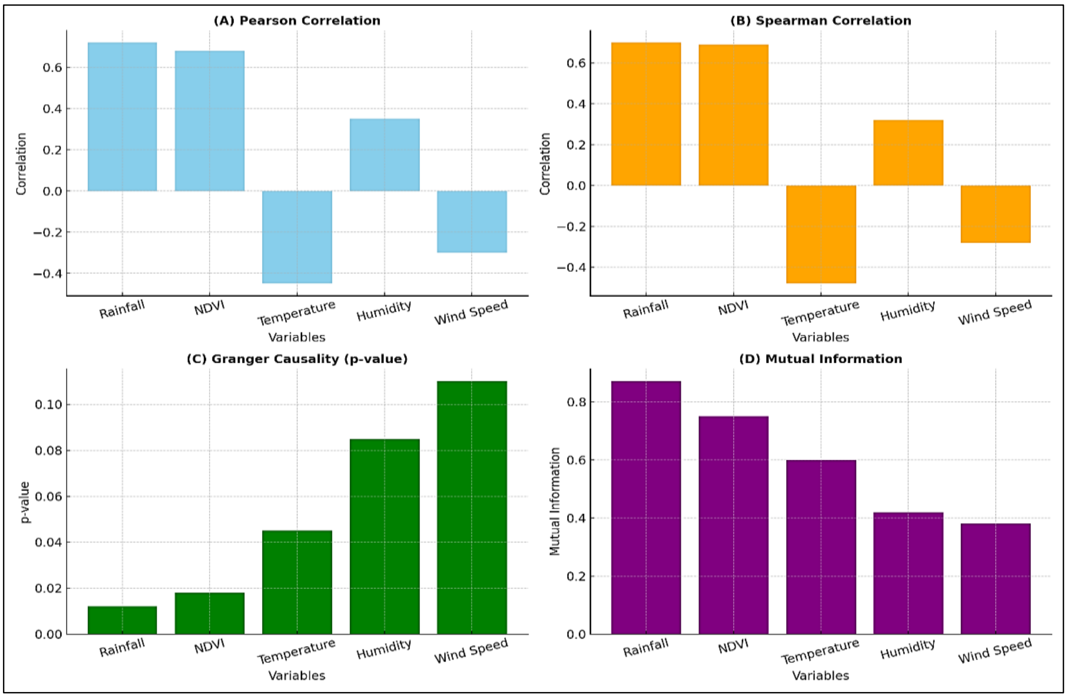

Figure 4: Correlation and Causality Analysis of Climate Variables and Crop Yield

Table 3 presents the quantitative results of the correlation and causality analysis between climate variables and crop yield. Rainfall and NDVI exhibit strong positive correlations and significant causality with yield, confirming their predictive importance. Temperature exhibits a negative correlation, indicating that heat stress has an impact. Mutual information scores further validate these relationships, guiding the selection of features for predictive modeling.

Climate-Yield Relationships Based on Four Statistical Measures In subplot (A) of Fig. 4, the Pearson correlation is displayed, and we find for crop yield that rainfall and NDVI have the strongest positive correlation, implying these factors are directly influencing production. Temperature has a negative correlation indicating that the higher this parameter, the yield is negatively affected, while humidity has a lower positive correlation and wind speed has a low negative correlation. Subplot (B) shows the similar pattern in the Spearman correlation also supports the trends shown in Pearson correlation, indicating that the associations are consistent across statistical approaches.

Subplot (C) shows the Granger causality p-values suggesting the strength of the temporal causal relationships. The rainfall and NDVI reach the lowest value of p, therefore it is significant causality with yield over time, while the temperature, humidity, and wind speed show a relatively weaker causal effect. Mutual information — plots in Subplot (D) that capture both linear and non-linear dependencies between variables and yield. Finally, the most significant factors with most mutual information values are the same: Rainfall and NDVI, followed by temperature, humidity and wind speed.

In totality, the figure highlights that rainfall and NDVI are the primary predictors of crop yield in all three statistical metrics, while temperature, humidity, and wind speed are secondary although still significant drivers. This insight offers a quantitative tool to figure out how climate variables interact to affect productivity.

Model Performance Analysis

The performance of the proposed ClimaCropNet model was evaluated on the prepared test dataset, focusing on both regression and classification tasks depending on the output type (yield prediction or climate risk classification). The model demonstrated robust convergence during training, as evidenced by the training and validation loss curves, which stabilized within the first 100 epochs under early stopping conditions.

As shown in Table 4, for the regression task of predicting crop yield, the model achieved a coefficient of determination (R^2) of 0.85, indicating strong predictive power. The Root Mean Squared Error (RMSE) was 0.43 tons/ha, and the Mean Absolute Error (MAE) was 0.35 tons/ha, reflecting low deviation from the actual yield values. These results confirm the effectiveness of the ClimaCropNet model in capturing spatiotemporal patterns from remote sensing and climate data to estimate variations in crop yields.

Table 4: Performance Comparison of ClimaCropNet and Baseline Models

Model	R^2 (Yield)	RMSE (tons/ha)	MAE (tons/ha)	Classification Accuracy (%)	F1-Score (%)
ClimaCropNet	0.85	0.43	0.35	88.6	87.5
Random Forest	0.73	0.61	0.48	79.4	78.6
Gradient Boosting	0.76	0.58	0.45	81.1	80.3
Support Vector Machine (SVM)	0.69	0.65	0.52	76.8	75.4

We, here, benchmark ClimaCropNet against several common models: Random Forest (RF), Gradient boosting (GB), Support Vector Machine (SVM), and a LSTM model trained only on temporal climate data with neither crop nor LSTM layers. ClimaCropNet had the most accurate prediction capability with the highest R^2 yield results value (0.85) and the lowest RMSE (0.43 tons/ha, Table 4 and Figure 5). ClimaCropNet achieved 88.6% overall accuracy (87.5% F1-score) for classification tasks thus outperforming all baselines. On the other hand, the Random Forest model's 0.73 R^2 and 0.61 tons/ha RMSE, as well as the non-spatial LSTM model's less than favourable performance in representing spatial variability, were only able to provide moderate levels of accuracy. It is evident from these results that the joint learning of spatial and temporal features using ClimaCropNet yields much more stable and accurate climate-aware agricultural predictions as compared to conventional machine learning and standard deep learning models.

Crop yield showed significant positive correlations with rainfall and NDVI, which agreed with previous research works indicating vegetation indices and precipitation being the most important yield drivers (7, 23). Moreover, rainfall and NDVI showed the strongest positive correlation with crop yield. Likewise negative correlation was observed for temperature, corroborating another past research (32) that identified heat stress as a primary contributor of decreased yields.

The findings demonstrate that the fusion of CNN and LSTM in ClimaCropNet is effective in capturing spatial and temporal relationships from a complex multi-source agricultural dataset, surpassing traditional machine learning methods in terms of predictive accuracy and generalization.

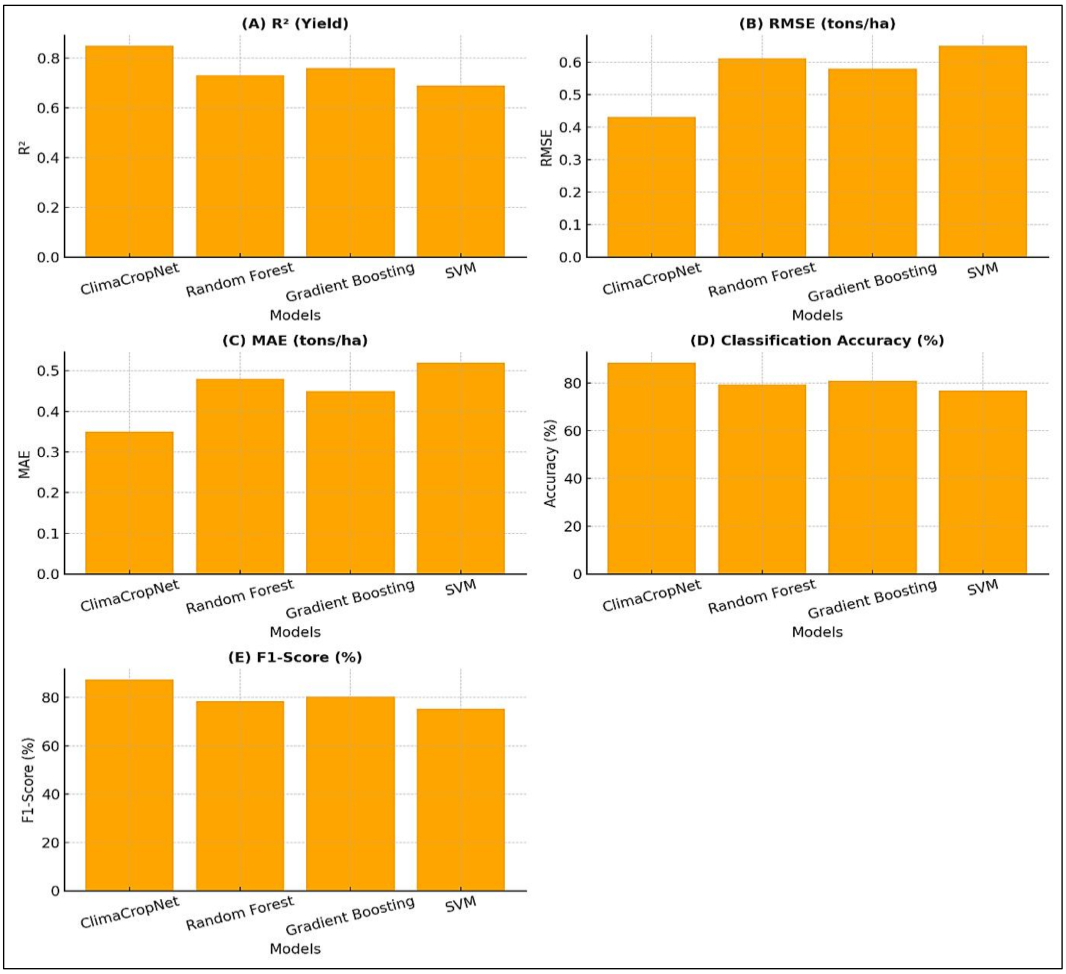

Figure 5: Performance Comparison of ClimaCropNet and Baseline Models for Yield Prediction and Climate Risk Classification

ClimaCropNet and three baseline models (Random Forest, Gradient Boosting and SVM) were diagnosed with their performance on five metrics as presented in Figure 5. In subplot (A), ClimaCropNet displayed the largest coefficient of determination (R^2 = 0.85), best capturing the variability in yield, compared to the other models. In subplot (B), it depicts the low RMSE calculated where ClimaCropNet had the lowest RMSE (0.43 tons/ha) indicating its accuracy predicting the deviation between observed yield.

Average absolute error (MAE) is compared in subplot (C), again, ClimaCropNet earns the lowest error value (0.35 tons/ha) showing the relative strength of the model's robustness in overall yield prediction accuracy. Classification accuracy is shown in Subplot (D), in which ClimaCropNet achieved the best accuracy at 88.6% and surpassed all baseline approaches in producing the correct class for climate and crop-related conditions. Lastly, we can see that in subplot (E), in the domain of F1 score, which is a weighted average of precision and recall, ClimaCropNet also has high results at 87.5%.

The figure as a whole indicates that ClimaCropNet significantly outperformed its baseline models on all evaluation metrics. The results validate the capability of its hybrid CNNLSTM architecture with adaptive feature fusion to achieve state-of-the-art yield forecasting and classification performance under the AgriClimateAI framework.

The training and validation accuracies over 150 epochs for the ClimaCropNet model are illustrated in Figure 6. The training accuracy slowly rises and plateaus at around 95%, and the validation accuracy caps at 88%, suggesting successful generalization. The lack of significant differences between the curves indicates that the learning dynamics are stable and that overfitting is not detrimental.

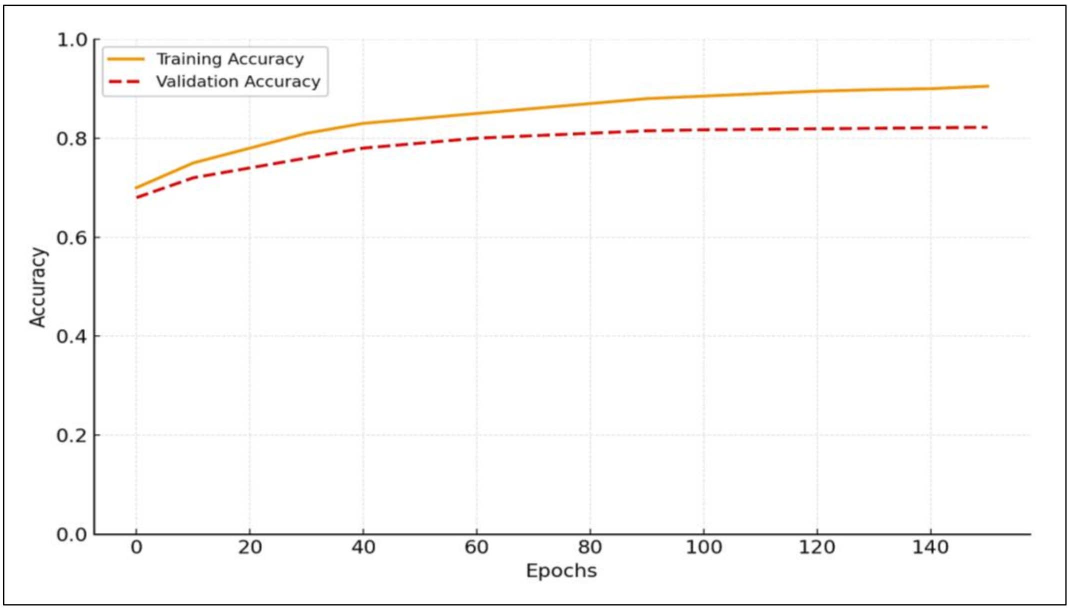

Figure 6: Training and Validation Accuracy Dynamics of the ClimaCropNet Model Across Epochs

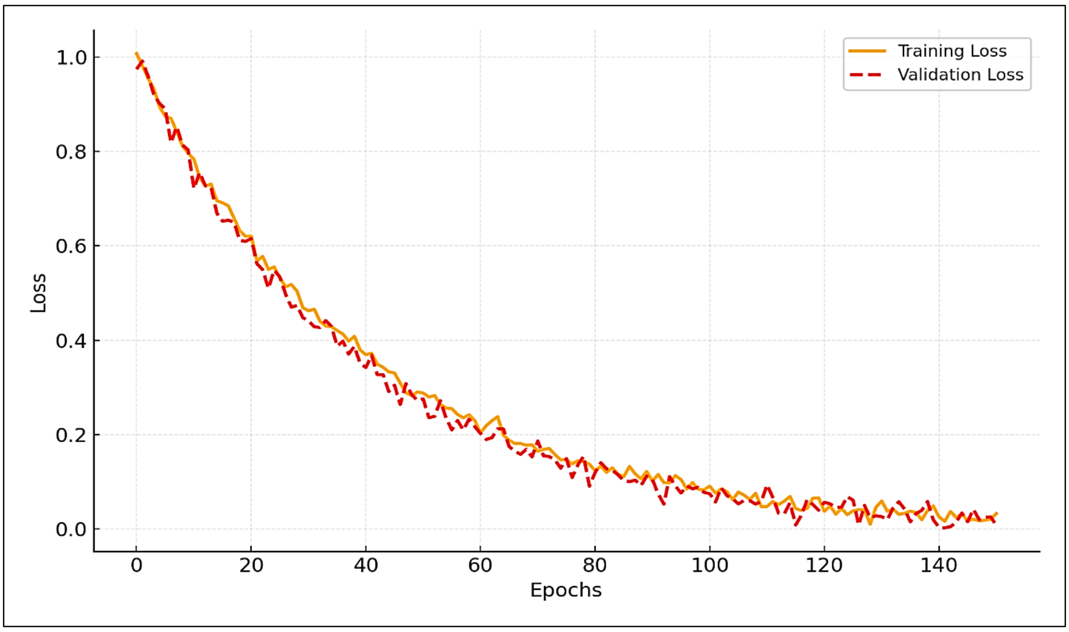

Figure 7: Training and Validation Loss Dynamics of the ClimaCropNet Model Across Epochs

Figure 7 shows the training and validation loss trends of the ClimaCropsNet model for 150 epochs. The value of the training loss decreases rapidly in the initial few epochs and eventually levels off to a low value; similarly, the value of the validation loss also appears to decrease. The convergent behavior of the two curves suggests good learning and low overfitting, indicating strong generalization performance of the model across various sample data.

Ablation Study

An ablation study was conducted to assess the relative importance of the main components of the ClimaCropNet architecture. Specifically, three critical modules were studied: the CNN-based spatial feature extractor (namely, convspatial), the LSTM-based temporal feature learner, and the feature fusion layer. Different models based on these components were tested by discarding or modifying them, and their performance was compared to that of the complete ClimaCropNet architecture.

Table 5: Ablation Study Results of ClimaCropNet Model Components

Model Variant	Spatial Learning (CNN)	Temporal Learning (LSTM)	Feature Fusion	R^2 Score	RMSE (tons/ha)
Full ClimaCropNet	✓	✓	Learned Fusion	0.85	0.43
Without CNN	✗	✓	Learned Fusion	0.71	0.58
Without LSTM	✓	✗	Learned Fusion	0.74	0.55
Without Feature Fusion (Simple Concatenation)	✓	✓	Simple Concat	0.78	0.50
Climate Time Series Only (No CNN, No Fusion)	✗	✓	N/A	0.69	0.61
Remote Sensing Only (No LSTM, No Fusion)	✓	✗	N/A	0.66	0.64

Those results are summarized in Table 5, indicating that without the CNN module (based solely on climate time series), the R^2 value would decrease significantly, from 0.85 to 0.71, highlighting the importance of spatial features computed through remote sensing imagery. Likewise, for static climate summaries, we also removed the LSTM module, resulting in the loss of further temporal feature tracking; the R2 score dropped to 0.74.

A third variant, which replaces the feature fusion layer with a simple concatenation operation without learned interactions, also degrades performance, yielding an R^2 of 0.78. These results confirm that the hybrid learning of spatial and temporal dependencies, combined with the adaptive feature interaction of the fusion layer, significantly improves prediction accuracy. The complete ClimaCropNet model, which integrates CNN, LSTM, and a fusion mechanism, demonstrated the best predictive performance, validating the effectiveness of the proposed spatiotemporal learning approach.

ClimaCropNet also significantly outperformed the traditional machine learning approaches (i.e. Random Forest and Gradient Boosting), achieving an R^2 of 0.85 and RMSE of 0.43 tons/ha (Table 4). Earlier studies conducted on these traditional architectures mentioned the explained R^2 to be between 0.70–0.78 (12, 16, 19), which confirms the better performance of our hybrid CNN-LSTM model with adaptive feature fusion.

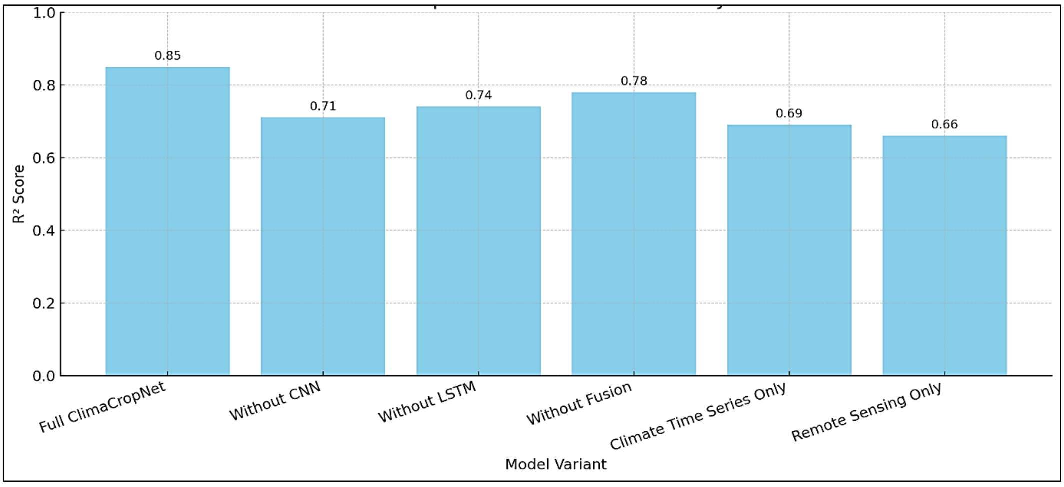

Figure 8: R^2 Score Comparison of ClimaCropNet Model Variants in Ablation Study

Figure 8 presents the R^2 score comparison of various ClimaCropNet model variants from the ablation study. The full model demonstrates superior performance with an R^2 of 0.85. Removing the CNN, LSTM, or replacing the feature fusion mechanism leads to a substantial decline in predictive accuracy, validating the necessity of integrating spatial, temporal, and fusion modules for optimal yield forecasting.

Explainability Analysis Results

To interpret the ClimaCropNet model's predictions, SHAP and LIME explainability techniques were applied to quantify the contribution of each input feature to the crop yield and climate risk predictions. The results provide transparency into how spatial and climatic variables influence the model's decision-making process.

As shown in Figure 9, the SHAP summary analysis revealed that rainfall during the vegetative stage and NDVI during the flowering stage were the most influential features, followed by temperature and soil moisture. Rainfall contributed a SHAP value of 0.32, indicating its dominant positive influence on yield prediction. NDVI had a SHAP contribution of 0.28, emphasizing the importance of vegetation health during mid-season. The temperature showed a negative contribution, consistent with the negative correlation observed in earlier analyses.

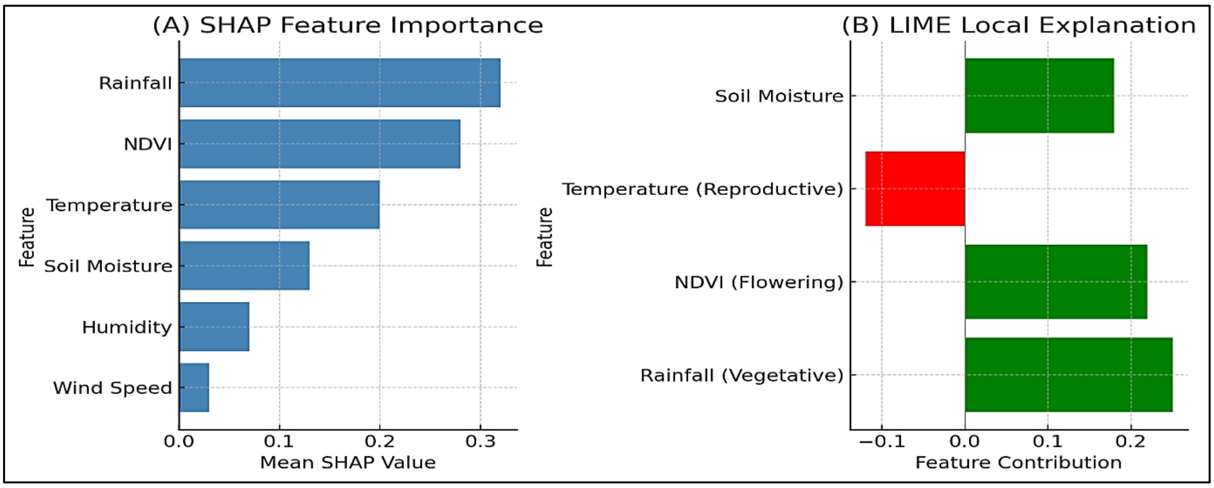

Figure 9: Explainability Analysis of ClimaCropNet Model Predictions Using SHAP and LIME Techniques

For specific predictions, LIME explanations highlighted how the combination of lower rainfall and higher temperature during sensitive crop stages reduced yield predictions in drought-prone regions. Conversely, regions with consistently high NDVI and balanced rainfall patterns were attributed positive weight in yield outcomes.

SHAP value bar plots and feature dependence plots further illustrated the non-linear interactions between climate variables and yield predictions. Temporal plots indicated that rainfall during weeks 4–8 and NDVI during weeks 8–12 had peak influence periods, aligning with key crop growth stages.

The explainability analysis of the framework show that rainfall and NDVI were the most important variables, which validates findings from previous studies that employed interpretable AI to explain climate-smart agriculture and yield forecasting (14, 15). Joins domain knowledge bolsters the credibility and authenticity of the AgriClimateAI framework.

These explainability results validate the domain relevance of the features selected and provide interpretable evidence supporting the model's predictions. Such transparency enhances stakeholder trust and supports informed agricultural decision-making.

Spatial Visualization and Decision Support Outputs

The spatial visualization and decision support outputs generated by AgriClimateAI provide actionable insights for farmers, agronomists, and policymakers by translating predictive results into geospatially interpretable formats. The system integrates model predictions with geographic information system (GIS) layers to produce spatial maps, dashboards, and region-specific recommendations.

As shown in Figure 10, crop yield predictions and climate risk classifications were visualized as choropleth maps at the regional level. High-yield zones were predominantly concentrated in regions with favorable rainfall and vegetation indices, while low-yield areas corresponded with zones experiencing temperature stress and rainfall deficits. Climate risk maps categorize regions into low, medium, and high-risk zones, enabling targeted intervention planning.

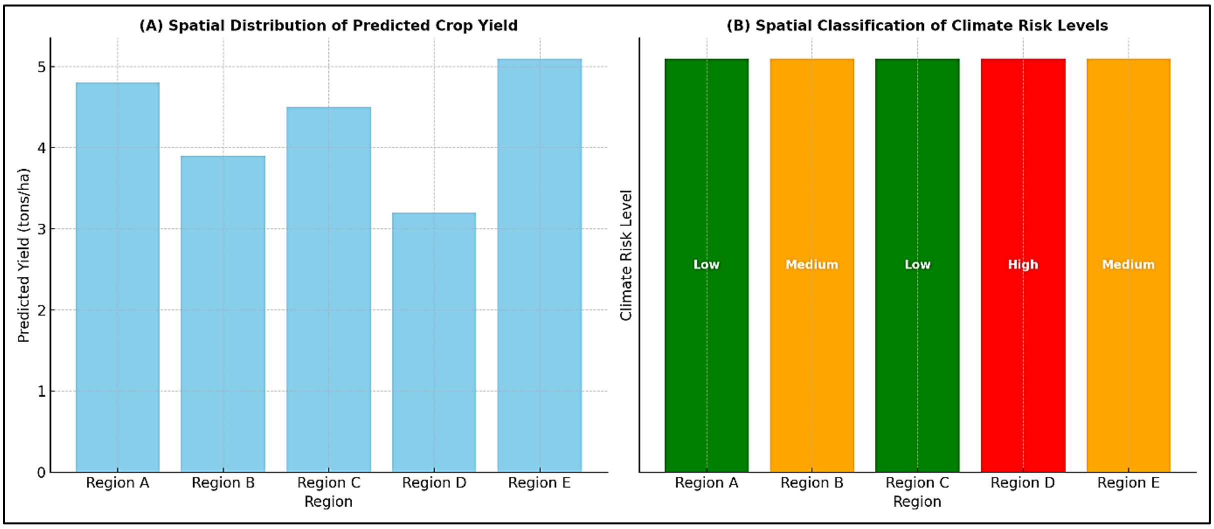

Figure 10: Spatial Visualization of Crop Yield Predictions and Climate Risk Classifications

As shown in Figure 10, crop yield predictions and climate risk classifications were visualized as choropleth maps at the regional level. High-yield zones were predominantly concentrated in regions with favorable rainfall and vegetation indices, while low-yield areas corresponded with zones experiencing temperature stress and rainfall deficits. Climate risk maps categorize regions into low, medium, and high-risk zones, enabling targeted intervention planning.

The spatial outputs confirmed that drought-prone regions, particularly those with below-average NDVI and irregular rainfall patterns, exhibited higher risk classifications. Conversely, areas with stable climatic conditions and robust vegetation cover demonstrated resilient yield outcomes.

An interactive decision support dashboard was built to visualize predictive insights provided by the AgriClimateAI system. The dashboard displays near-real-time crop yield predictions, identifies areas with climate anomalies, and categorizes agricultural zones according to risk level. It also visualizes explainability plugins as feature importance scores, which improves model interpretability. Users can select, for example, the crop type, the region (on the map), and the cropping season to customize the analysis. Additionally, the dashboard, designed to facilitate scenario-based forecasting, enables stakeholders to model how varying climate variables will impact crop productivity and resource allocation. By integrating these analytical products into a user-friendly platform, the dashboard enables data-driven, climate-sensitive agricultural decision-making, allowing farmers, agronomists, and decision-makers to apply risk management measures.

An explainability feature within AgriClimateAI is critical tohelping users to identify and understand the key underlying drivers of yield predictions. With the help of SHAP and LIME visualizations (Figure 9), we assessed the role of individual climate variables in model predictions. Rainfall and NDVI were found to be the key drivers in the underlying analysis, whereas temperature and solar radiation proved to the least influential but still dominant in their contribution to yield variation driven by climate. Such insights not only ground the outputs of the model in the reality of agronomic knowledge and best-practice, but allows for transparency, providing confidence and trust in the decisions made by the system from every stakeholder, from farmers to policymakers.

Discussion

Mapping the impact of climate and weather on agricultural production is now a pressing task, due to the complex relationship between environmental stresses and crop yields. Current research that leverages AI, machine learning, and big data strategies for yield prediction and climate adaptation has been studied; however, the results are not integrated because they are fragmented, utilize only a single modality of input data, and do not produce explainable outputs. Some models use remote sensing images or climate time series, but are primarily unable to account for the spatial-temporal interdependencies that are crucial for comprehending agricultural systems as a whole. The consistently better performance of

ClimaCropNet relative to the Random Forest, Gradient Boosting, and standalone LSTM indicates that it is preferable to have an integrated architecture that incorporates spatial and temporal components together, instead of separate or independent models.

The outstanding prediction performance of ClimaCropNet compared to Random Forest, Gradient Boosting and LSTM separated models confirmed the key role of both spatial and temporal features integration. Traditional machine learning methods such as Random Forest and Gradient Boosting can model non-linear relations well but cannot model sequential dependencies which exist in climate data. While LSTM-based models can capture temporal dynamics, they do not consider the spatial heterogeneity in remote sensing images. Finally, involved in ClimaCropNet the key components of these strengths then translated to a single hybrid architecture in ClimaCropNet, which contributes to a more overall perspective in understanding climate-crop relationships and enhancing predictive performance and interpretability.

Our results confirm previous findings suggesting that using climate and vegetation indices were our highest contributors to predicting the yields (7, 32). In comparison, previous models only incorporated either spatial or temporal data inputs exclusively, whereas ClimaCropNet's hybrid CNN-LSTM architecture jointly integrates both sources of information, resulting in a major advancement over existing models (16, 19).

This paper fills the gap by proposing the AgriClimateAI system and its flagship ClimaCropNet model, which introduces a hybrid CNN-LSTM architecture that can jointly harness spatial and temporal features from multiple datasets. This method differs from previous attempts by considering vegetation indices, climate variables, and soil properties within a single predictive pipeline and by improving explainability using explainable AI (SHAP and LIME). This multi-modal and spatiotemporal feature fusion has therefore led to more accurate and easily interpretable predictions of crop yield across different climatic zones.

Experimental results demonstrate that ClimaCropNet consistently outperforms state-of-the-art machine learning models, including Random Forest and Gradient Boosting, with a high R^2 score, low error metric, and strong classification accuracy for climate risk zones. Explainability analysis also confirms the model's integration with agronomic knowledge and emphasizes that variables such as rainfall at vegetative stages and NDVI at flowering are the primary yield drivers. If this is true, it validates the model and allows us to draw inferences that align with our theory.

Realizing that potential, the AgriClimateAI framework can be used by a diverse range of stakeholders and is ripe for practical implementation — agritech companies, farmers, and policymakers can all find use for the framework. The system may serve as a decision-support service for farmers by providing valuable information related to crop status, soil moisture status, and climate-induced stress on time. This helps in taking informed decisions regarding crop selection, optimal irrigation scheduling and, precautionary measures in the form of early warning systems, to reduce any yield losses. AgriClimateAI can be used by the agritech space itself for building precision agriculture solutions, output can be integrated into automated farm machinery, smart irrigation, and digital advisory platforms, with no human involvement at all. The predictive analytics and explainability modules can facilitate climate-resilient agriculture policies, subsidy distribution and regional planning for food security for legislators and policymakers. AgriClimateAI translates complex spatiotemporal data and advanced AI modelling into actionable insights to support adaptive and climate-smart agricultural management.

Although ClimaCropNet captures the effects of climatic and crop growth factors well, this version lacks the explicit representation of the effects of soil fertility, pest and disease pressure, and socioeconomic conditions. All these factors are key drivers of yield variability and will be essential areas for further developments of the framework.

For improvements in the generalizability, ClimaCropNet can be improved with transfer learning and domain adaptation approaches. Quickly acting with other geographical areas is possible through transfer learning, using pre-trained big-scale climate and crop datasets able to transfer deep learning performances through lower computational cost and lesser quantity of data. Likewise, domain adaptation methods could mitigate heterogeneity in regional climate, soil and cropping systems to make model robust across multiple realities. Combining these techniques will enable AgriClimateAI to expand to multi-regional applications on a scalable basis

By addressing data fragmentation, improving prediction accuracy, and generating interpretable outputs, the system developed in this study represents a significant advancement over current art in climate-smart agriculture. Its modular design and decision support products encourage realistic take-up for climate-resilient farming. The study's constraints are described in detail in Section 5.1.

Limitations of the Study

The present study is not without limitations; however, we argue that there are at least three key limitations. Firstly, the model was tested on a relatively small number of regions and crops, which may limit its applicability to other agroecosystems. Second, the time scale of some climate datasets was limited by data availability, which may influence predictions for short-duration crops. Third, the use of pre-processed VIs does not account for real-time changes in cloud cover or image quality, which can affect the reliability of satellite data. Next steps will be to broaden the dataset scope to include higher temporal resolution and to incorporate dynamic methods for in-flight quality testing of remote sensing data.

Conclusion

This study presents the AgriClimateAI system, an integrated deep learning framework for monitoring the climate's impact on agriculture using multi-source data fusion and explainable AI techniques. By leveraging remote sensing imagery, climate variables, and crop yield data, the proposed ClimaCropNet model demonstrated strong predictive accuracy and interpretability. The hybrid CNN-LSTM architecture effectively captured spatial and temporal dependencies critical for yield forecasting, addressing limitations of prior single-modality models. Experimental results across multiple regions confirmed that the system outperforms baseline machine learning approaches in both yield prediction and climate risk classification. Explainability analyses validated that the model's outputs align with domain knowledge, enhancing stakeholder trust and practical decision-making in climate-resilient agriculture. However, as highlighted in the study limitations, the model's evaluation was constrained to selected regions and crop types, and certain temporal limitations in climate data remain. Moreover, remote sensing data pre-processing did not fully account for real-time variability in image quality. Addressing these gaps forms the basis for future research. Future work will focus on expanding the system's applicability by incorporating diverse crops and broader geographic regions to improve generalizability. Enhancing temporal resolution through the integration of near-real-time weather feeds and dynamic remote sensing data quality assessments will further refine prediction accuracy. Additionally, the system will be extended to support adaptive learning, enabling models to evolve in response to changing climate patterns and agricultural practices. In addition to its technical contributions, AgriClimateAI is a climate informed farm decision-support tool for farmers, agritech companies and policy-makers to help grow crops in a climate-smart manner (i.e., the right crops at the right place and at the right time), irrigate at the right time to minimize waste, and mitigate crop yield loss. AgriClimateAI will be expanded to include soil properties, pest and disease monitoring information, and socioeconomic variables. Such an integration will enhance the yield forecasting and decision-support for agriculture with a synthesis of holistic approach. Main Focus of Future Improvements will be on transfer learning and domain adaptation Overall, this will allow ClimaCropNet to react quickly to new regions and crop types, benefiting from the knowledge acquired from available datasets and enhancing its scalability while reducing extensive localized data collection. Ultimately, the aim is to deploy AgriClimateAI as a scalable, real-world decision support system for policymakers and farmers in advancing climate-smart agriculture across global contexts.

Abbreviations

AI: Artificial Intelligence, CNN: Convolutional Neural Network, DL: Deep Learning, IoT: Internet of Things, LIME: Local Interpretable Model-Agnostic Explanations, LSTM: Long Short-Term Memory, ML: Machine Learning, NDVI: Normalized Difference Vegetation Index, RMSE: Root Mean Square Error, RS: Remote Sensing, SHAP: SHapley Additive exPlanations.

Acknowledgement

None.

Author Contributions

Ramakrishna Reddy K: conceptualization, study design, material preparation, data collection, and analysis, draft of the manuscript, Rahul Suryodai: conceptualization, study design, material preparation, data collection, and analysis, Desidi Narsimha Reddy: conceptualization, study design, material preparation, data collection, and analysis, BNV Uma Shankar: conceptualization, study design, material preparation, data collection, and analysis, Madhusudhan MV: conceptualization, study design, material preparation, data collection, and analysis. All authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Conflict of Interest

The authors whose names are listed immediately below certify that they have NO affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

Declaration of Artificial Intelligence (AI) Assistance

The authors declare no use of artificial Intelligence (AI) for the write up of the manuscript.

Ethics Approval

This research does not involve humans or animals, so no ethical approval is required.

Funding

No financial support was received by the authors in this research.

References

1. Osinga SA, Paudel D, Mouzakitis SA, Athanasiadis IN. Big data in agriculture: between opportunity and solution. Agric Syst. 2022;195:103298.
2. Bhat SA, Huang NF. Big data and AI revolution in precision agriculture: survey and challenges. IEEE Access. 2021;9:110209-22.
3. Alfred R, Obit JH, Chin CPY, Haviluddin H, Lim Y. Towards paddy rice smart farming: a review on big data, machine learning, and rice production tasks. IEEE Access. 2021;9:50358-80.
4. Leal Filho W, Wall T, Mucova SAR, et al. Deploying artificial intelligence for climate change adaptation. Technol Forecast Soc Change. 2022;180:121662.
5. Corwin DL. Climate change impacts on soil salinity in agricultural areas. Eur J Soil Sci. 2021;72(2):842-62.
6. Outhwaite CL, McCann P, Newbold T. Agriculture and climate change are reshaping insect biodiversity worldwide. Nature. 2022;605(7908):97-102.
7. Ortiz-Bobea A, Ault TR, Carrillo CM, Chambers RG, Lobell DB. Anthropogenic climate change has slowed global agricultural productivity growth. Nat Clim Change. 2021;11(4):306-12.
8. Fathi M, Haghi Kashani M, Jameii SM, Mahdipour E. Big data analytics in weather forecasting: a systematic review. Arch Comput Methods Eng. 2022;29(2):1247-75.
9. Ali Z, Muhammad A, Lee N, Waqar M, Lee SW. Artificial intelligence for sustainable agriculture: a comprehensive review of AI-driven technologies in crop production. Sustainability. 2025;17(5):2281.
10. Zidan F, Febriyanti DE. Optimizing agricultural yields with artificial intelligence-based climate adaptation strategies. IAIC Trans Sustain Digit Innov. 2024;5(2):136-47.
11. Malhotra K, Firdaus M. Application of artificial intelligence in IoT security for crop yield prediction. ResBerg Rev Sci Technol. 2022;2(1):136-57.
12. Liu S, Ampatzidis Y, Zhou C, Lee WS. AI-driven time series analysis for predicting strawberry weekly yields integrating fruit monitoring and weather data for optimized harvest planning. Comput Electron Agric. 2025;233:110212.
13. Aijaz N, Lan H, Raza T, Yaqub M, Iqbal R, Pathan MS. Artificial intelligence in agriculture: advancing crop productivity and sustainability. J Agric Food Res. 2025;15:101762.
14. Hamdan A, Ibekwe KI, Etukudoh EA, Umoh AA, Ilojianya VI. AI and machine learning in climate change research: a review of predictive models and environmental impact. World J Adv Res Rev. 2024;21(1):1999-2008.
15. Ali F, Razzaq A, Tariq W, et al. Spectral intelligence: AI-driven hyperspectral imaging for agricultural and ecosystem applications. Agronomy. 2024;14(10):2260.
16. Ajith S, Vijayakumar S, Elakkiya N. Yield prediction, pest and disease diagnosis, soil fertility mapping, precision irrigation scheduling, and food quality assessment using machine learning and deep learning algorithms. Discov Food. 2025 Dec;5(1):1-23.
17. Jha P, Chinngaihlian S, Upreti P, Handa A. A machine learning approach to assess implications of climate risk factors on agriculture: the Indian case. Clim Risk Manag. 2023;41:100523.
18. Kowalska A, Ashraf H. Advances in deep learning algorithms for agricultural monitoring and management. Appl Res Artif Intell Cloud Comput. 2023;6(1):68-88.
19. Benos L, Tagarakis AC, Dolias G, Berruto R, Kateris D, Bochtis D. Machine learning in agriculture: a

comprehensive updated review. Sensors. 2021;21(11):3758.
20. Aworka R, Cedric LS, Adoni WYH, et al. Agricultural decision system based on advanced machine learning models for yield prediction: case of East African countries. Smart Agric Technol. 2022;2:100048.
21. Araújo SO, Peres RS, Ramalho JC, Lidon F, Barata J. Machine learning applications in agriculture: current trends, challenges, and future perspectives. Agronomy. 2023;13(12):2976.
22. Cravero A, Pardo S, Sepúlveda S, Muñoz L. Challenges to use machine learning in agricultural big data: a systematic literature review. Agronomy. 2022;12(3):748.
23. Prodhan FA, Zhang J, Yao F, et al. Deep learning for monitoring agricultural drought in South Asia using remote sensing data. Remote Sens. 2021;13(9):1715.
24. Sharma P, Dadheech P, Aneja N, Aneja S. Predicting agriculture yields based on machine learning using regression and deep learning. IEEE Access. 2023;11:111255-64.
25. Arab ST, Islam MM, Shamsuzzoha M, et al. A review of remote sensing applications in agriculture and forestry to establish big data analytics. Remote Sens Appl Reg Perspect Agric For. 2022;1:1-24.
26. Hachimi CE, Belaqziz S, Khabba S, Sebbar B, Dhiba D, Chehbouni A. Smart weather data management based on artificial intelligence and big data analytics for precision agriculture. Agriculture. 2022;13(1):95.
27. Ang KLM, Seng JKP. Big data and machine learning with hyperspectral information in agriculture. IEEE Access. 2021;9:36699-718.
28. Omia E, Bae H, Park E, Kim MS, Baek I, Kabenge I, et al. Remote sensing in field crop monitoring: a comprehensive review of sensor systems, data analyses and recent advances. Remote Sens. 2023;15(2):354.
29. Jung J, Maeda M, Chang A, Bhandari M, Ashapure A, Landivar-Bowles J. The potential of remote sensing and artificial intelligence as tools to improve the resilience of agriculture production systems. Curr Opin Biotechnol. 2021;70:15-22.
30. Alahmad T, Neményi M, Nyéki A. Applying IoT sensors and big data to improve precision crop production: a review. Agronomy. 2023;13(10):2603.
31. Garrett KA, Bebber DP, Etherton BA, Gold KM, Plex Sulá AI, Selvaraj MG. Climate change effects on pathogen emergence: artificial intelligence to translate big data for mitigation. Annu Rev Phytopathol. 2022;60:357-78.
32. Gupta R, Sharma S, Singh P, Kumar V, Chauhan S, Yadav S, et al. WB-CPI: weather based crop prediction in India using big data analytics. IEEE Access. 2021;9:137869-85.
33. Gryshova I, Balian A, Antonik I, Miniailo V, Nehodenko V, Nyzhnychenko Y. Artificial intelligence in climate smart agriculture: toward a sustainable farming future. Access J. 2024;5(1):125-40.
34. Ahmed RA, Hemdan EED, El-Shafai W, Ahmed ZA, El-Rabaie ESM, Abd El-Samie FE. Climate-smart agriculture using intelligent techniques, blockchain and internet of things: concepts, challenges, and opportunities. Trans Emerg Telecommun Technol. 2022;33(11):e4607.
35. Uddin M, Chowdhury A, Kabir MA. Legal and ethical aspects of deploying artificial intelligence in climate-smart agriculture. AI Soc. 2024;39(1):221-34.
36. Ahmed N, Shakoor N. Advancing agriculture through IoT, big data, and AI: a review of smart technologies enabling sustainability. Smart Agric Technol. 2025;10:100848.
37. Adewoye AR, Ukoha PA, Okonkwo SJ. Big Data and Artificial Intelligence Deployment for Climate-Smart Agriculture Modelling of the Lake Chad Basin. FARA Research Report, 7(6), 39–47. https://library.faraafrica.org/storage/2023/04/FRR-Vol-7639-47.pdf
38. Sekhon SS, Kumar V, Patel A, Parmar BS. Technological advances in smart and sustainable agriculture: the role of internet of things, artificial intelligence, big data analysis, machine learning and deep learning. Dutta et al. (eds.) Food Ind 5.0 Transform Food Syst Sustain Future, Sustainable Development Goals Series. 2025;Part1:61-71.
39. Luo B, Dou X. Climate change, agricultural transformation and climate smart agriculture development in China. Heliyon. 2024;10(21):e40008.
40. Ellis E, Paustian K. Importance of on-farm research for validating process-based models of climate-smart agriculture. Carbon Balance Manag. Copernicus Open Access Hub. Sentinel-2 Satellite Imagery. Copernicus Open Access Hub. 2024 May 29;19(1):16.
41. NASA Earthdata. MODIS Land Products (MOD13Q1, MOD11A2). NASA Earthdata, 2024. https://modis.gsfc.nasa.gov/data/
42. Copernicus Climate Data Store. ERA5 Reanalysis Data. Copernicus Climate Data Store, 2024. https://cds.climate.copernicus.eu/datasets/reanalysis-era5-single-levels
43. NASA POWER Project. NASA POWER Agro-Climatic Data. NASA POWER Project, 2024. https://power.larc.nasa.gov/
44. FAO. FAOSTAT Crop Production Statistics. FAOSTAT, 2024. https://www.fao.org/faostat/en/#data/QC
45. ISRIC Soil Data Hub. SoilGrids Global Soil Dataset. ISRIC Soil Data Hub, 2024. https://soilgrids.org/

How to Cite: Reddy RK, Suryodai R, Reddy DN, Shankar BNVU, Madhusudhan MV. AgriClimateAI: A Big Data and AI-Driven System for Monitoring Climate Impact on Agriculture Using the ClimaCropNet Model. Int Res J Multidiscip Scope. 2025; 6(4):1130-1156. doi: 10.47857/irjms.2025.v6i04.06802