Journal of Agrometeorology
(A publication of Association of Agrometeorologists)
ISSN : 0972-1665 (print), 2583-2980 (online)
Vol. No. 28 (3) : 275-284 (September - 2026)
https://doi.org/10.54386/jam.v28i3.3399
https://journal.agrimetassociation.org/index.php/jam

Research paper

Field Spectroscopy-Assisted Machine Learning Framework for Assessing Agroclimatic Effects on Paddy

KARTHIK KARUNAKARAN and KARUPPASAMY SUDALAIMUTHU*

Department of Civil Engineering, College of Engineering and Technology, SRM Institute of Science and Technology, Kattankulathur Campus, Chengalpattu, Tamil Nadu, India.
[*]*Corresponding Author Email:* karuppas@srmist.edu.in

ABSTRACT

Assessment of yield-limiting factors in paddy production should be carried out properly in order to improve productivity under varying agro-climatic conditions. This research paper introduces a combined geospatial and machine learning system to estimate stage-based paddy yield stress during *Kharif* 2023 in the Thanjavur region using vegetation indices and agrometeorological variables. Five different satellite-derived vegetation indices were calibrated using spectroradiometer data during the vegetative, reproductive, and ripening stages. The field- and satellite-derived indices showed good agreement in calibration (R^2=0.85-0.92). Correlation analysis showed significant associations between vegetative-stage rainfall, humidity during reproductive stage, and yield variability, whereas the canopy vigor represented by the Enhanced Vegetation Index (EVI) and Soil Adjusted Vegetation Index (SAVI) during the ripening stage was positively correlated with grain filling and final productivity. Yield prediction before and after calibration was performed using machine learning models such as Multiple Linear Regression (MLR), K-Nearest Neighbors (KNN), Random Forest (RF), and eXtreme Gradient Boosting (XGBoost). Among these, XGBoost performed better after calibration (R^2=0.908, RMSE=179.728 kg ha^{-1}, MAE=143.487 kg ha^{-1}). Feature importance revealed that reproductive-stage humidity, rainfall, and ripening-stage soil moisture were the most important yield drivers. This model can be used to provide a powerful decision-support tool for spatial stress mapping, accurate irrigation regulation, and climate-resilient crop strategies in irrigated rice environments.

Keywords: Agroclimatic Stress, Crop Monitoring, Machine Learning, Paddy, Spectral Indices, Spectroradiometer.

Rice (*sativa* L.) is one of the most important staple crops, playing a critical role in ensuring food security in tropical and subtropical regions, especially in South and Southeast Asia. Millions of livelihoods in India depend on paddy farming, which is very delicate to climatic variability, water availability, and land surface processes (IPCC, 2023). The growing climate change, rising temperatures, abnormal rainfalls, and extreme weather conditions are placing an enormous load on rice production systems. These agroclimatic disruptions have a direct impact on crop growth, phenology, crop productivity, and eventual food security in the region (Dineshkumar *et al.,* 2019). This has made constant observation and evaluation of the crop response to the agroclimatic environment an indispensable aspect of sustainable agricultural planning and climate-resilient crop management.

Conventional methods of monitoring crops involve field surveys and agronomic observations, which are, in most cases, labour-intensive, space-restrained, and time-consuming. The development of remote sensing technologies has made the use of satellite data very common in the massive monitoring of agriculture (Dela Torre *et al.,* 2021). However, satellite observations are limited in their spectral resolution for detecting fine-grained crop stress signals at an initial stage. Field-based spectroradiometer data give a very detailed spectral reflectance, although it tends to record fine differences in plant physiology, chlorophyll content, water stress, and canopy structure (Verrelst *et al.,* 2019). This review summarizes the state-of-the-art retrieval methods that have been applied in experimental imaging spectroscopy studies inferring all kinds of vegetation biophysical variables. Identified retrieval method and

Article info - DOI: https://doi.org/10.54386/jam.v28i3.3399
Received:24 February 2026; Accepted: 11 June 2026; Published online : 7 September 2026

improve accuracy in the crop monitoring by allowing a calibration and validation of remotely sensed vegetation indices.

Vegetation indices are useful in the evaluation of the condition of crops. Commonly used indices are Normalized Difference Vegetation Index (NDVI), Green NDVI (GNDVI), Soil Adjusted Vegetation Index (SAVI), Moisture Stress Index (MSI), and Normalized Difference Water Index (NDWI), are widely used to assess vegetation vigor, chlorophyll concentration, biomass, water stress (Vidican *et al.,* 2023), slow adoption of sustainable farming, a need for climate-resilient food systems, resource inequity, and the protection of small-scale farmers' practices. These issues are integral to food security and environmental health. Remote sensing technologies can assist precision agriculture in effectively addressing these complex problems by providing farmers with high-resolution lenses. The use of vegetation indices (VIs), thereby enabling the monitoring of crop growth dynamics and stress response throughout the crop cycle. Another important parameter that is vital and indicates surface energy balance and stress of plants is Land Surface Temperature (LST) (Sajan *et al.,* 2023). Increased surface temperatures usually signify crop water stress, less transpiration, and physiological effects of heat. Satellite derived temperature provides valuable information about the heat stress and evapotranspiration (ET) processes, especially in irrigated agricultural landscapes. A supplementation of spectral vegetation indices with LST data can be used for better understanding of crop response to agroclimatic variability. Agroclimatic factors like rainfall, air temperature, humidity, wind speed, solar radiation, and ET play a great role in the crop phenology, such as vegetative, reproductive, and grain filling stages (Jha *et al.,* 2021). The difference in these parameters may cause a decrease in yield due to drought stress, excess moisture, or heat stress (Hatfield & Prueger, 2015). Thus, the analysis of crop-climatic interaction can be performed through the combination of meteorological records and spectral data in order to obtain a multidimensional framework.

Artificial intelligence and machine learning (ML) have just recently revolutionized the agricultural remote sensing applications. Crop classification, yield forecasting, stress identification, and crop growth modelling (Krithikha Sanju & Bhagavathiappan, 2022) are some of the tasks where ML algorithms have become increasingly popular because they deal with highly nonlinear correlations involving several variables. Random Forest (RF), Support Vector Machines (SVM), Gradient Boosting, and Neural Networks are algorithms that have shown great capabilities in crop prediction using multisource data (Zheng *et al.,* 2023). This paper aims to provide insights into current and future trends in remote sensing for paddy crop monitoring. ML models allow successfully combining spectral, thermal, and agroclimatic data and generate reliable crop condition data and predictive analytics. Irrespective of these developments, there are a number of challenges remain, including data quality and availability, spatial heterogeneity, model transferability across regions and seasons. Numerous developed works utilize only the satellite data and avoid using the field-level spectral measurements as a calibration method (Wang *et al.,* 2021) providing possibilities for flowering monitoring. The proposed framework addresses these gaps by combining multi-source datasets within a unified modelling approach, improving prediction accuracy and robustness.

The novelty of the current research is in the creation of a multicomponent field spectroscopy-engineered ML model to examine agroclimatic impacts on paddy crops. The model capitalizes on ML models to take into account non-linear interactions between spectral responses and climatic factors during different periods of crop development. This can be integrated to improve the ability to identify crop stress conditions and measure agroclimatic impacts more realistically. The specific objectives are (i) to acquire multi-temporal satellite-derived vegetation indices and calibrate them using in-situ field spectroradiometer observations for paddy crop across different phenological stages during *Kharif* 2023, (ii) to assess the impact of agro-meteorological and biophysical parameters on stage-wise paddy crop growth and yield variability, (iii) to compare different ML models for accurate prediction of stage-based paddy yield stress and generate spatial stress maps in paddy cultivation. This novel work leads to the development of precision agriculture and climate-resistant crop monitoring strategies.

MATERIALS AND METHODS

Study Area

This research was carried out in Thanjavur district, Tamil Nadu, India, which is one of the largest rice-producing areas in the Cauvery delta. The study region lies between latitudes 10°08′ N to 11°08′ N and longitudes 78°48′ E to 79°57′ E, as shown in Fig. 1, with plain alluvial soils and fertile deltaic soils, which could sustain the intensive paddy farming. Thanjavur has a tropical climate with monsoon effects. The annual rainfall ranges between 900 and 1100 mm, received mainly during the Northeast monsoon (October-December), which is very favourable to the Samba cropping season. It has hot summers of 38°C to 40°C, which are coupled with winter temperatures of between 20°C and 25°C. The irrigation mostly relies on the canal systems of the river Cauvery as well as the groundwater. The region involves paddy cultivation in three seasons, including the *Kharif* or *Kuruvai* (June-September), *Rabi* or *Samba* (September-January), and *Thaladi* (January-March) seasons. This research examines the crop responsiveness during the 2023 *kharif* season, considering three key phenological stages, such as vegetative, reproductive, and ripening, to characterize crop growth dynamics. These phases are important crop growth stages influenced by agroclimatic changes.

Data Sources

The research combines the multisource data, such as satellite data, field spectroscopic measurements, and agroclimatic data, shown in Fig. 2. The LULC map was generated from Resourcesat-2 Advanced Wide Field Sensor (AWiFS) imagery using supervised Maximum Likelihood Estimation approach. A total of 150 training samples for major land-cover classes were selected based on visual interpretation of Google Earth imagery and field observations. The agricultural area covered approximately 1,644 km² (61% of the total study area) and was subsequently used for paddy phenology analysis. The resulting LULC map achieved an Overall Accuracy of 92.8% and a Kappa coefficient of 0.9203.

Landsat 8/9 Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS) Level-1 imagery acquired during the 2023

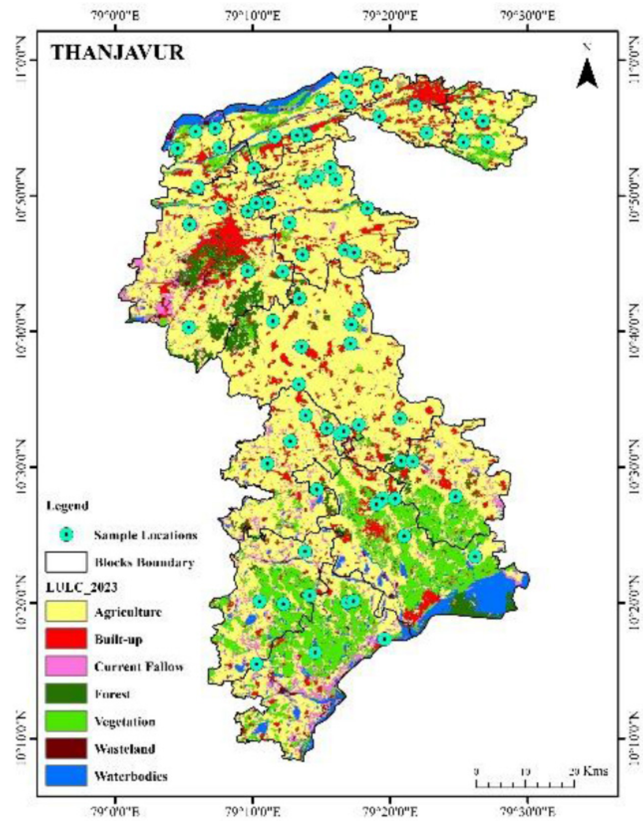

Fig. 1: Spatial representation of major Land Use/Land Cover (LULC) categories within the study area, overlaid with selected sample paddy field locations (n=72).

cropping seasons were used in this study. The imagery provides a spatial resolution of 30 m for optical bands and resampled 30 m for thermal bands, with a temporal resolution of 16 days. The study area, covering approximately 2,627.66 km², falls within Path 142/ Row 53, and cloud cover below 10% were selected to ensure data quality for subsequent analyses all preprocessing and analysis were performed in QGIS.

Spectral signatures were collected from 72 sampling points at 40, 85, and 110 days after sowing (DAS), representing the vegetative, reproductive, and ripening stages of paddy crop respectively (Yoshida, 1981) during the *kharif* season. Measurements were acquired using an ASD Field Spec® Handheld™ 2 spectroradiometer to calibrate the vegetation indices, as shown in Table 1.

Table 1: Landsat and field spectroradiometer data acquisition dates

Paddy Crop Stage	Landsat	Field Data
Vegetative	08-07-2023	07-09 July 2023 (40 DAS)
Reproductive	25-08-2023	21-23 August 2023 (85 DAS)
Ripening	10-09-2023	15-16 September 2023 (110 DAS)

Weather data, including temperature, relative humidity, rainfall, and wind speed were obtained from the India Meteorological Department (IMD). Soil information was obtained from the Tamilnadu Agriculture University. Paddy crop yield data were obtained from the Joint Director of Agriculture, Thanjavur, based on field crop-cutting experiments conducted by government agricultural officials during harvest.

Computation of Biophysical and Agro-Climatic Parameters

The derivation of vegetation indices and LST was done using the Landsat imagery. Digital numbers were converted to Top-of-Atmosphere (TOA) reflectance using the USGS radiometric rescaling coefficients (multiplicative factor=0.00002; additive offset=−0.1) before vegetation index computation. Spectral indices are computed identically to spectral angles, and values range from -1 to +1. The indices aid in the measurement of vegetation vigor, chlorophyll level, and moisture stress.

$$\text{Green Chlorophyll Index (C)} = \frac{NIR}{Green} - 1 \tag{1}$$

$$\text{Green Normalized Difference Vegetation Index } (GNDVI) = \frac{NIR - GREEN}{NIR + GREEN} \tag{2}$$

$$Enhanced\ Vegetation\ Index = G\ x\ \frac{NIR - RED}{NIR + C_1 * RED - C_2 * BLUE + 1} \tag{3}$$

$$\text{Normalized Difference Water Index } (NDWI) = \frac{GREEN - NIR}{GREEN + NIR} \tag{4}$$

$$\text{Soil Adjusted Vegetation Index } (SAVI) = \frac{NIR - RED}{NIR + RED + L} x(1 + L) \tag{5}$$

where GREEN = Green reflectance, NIR = Near Infrared, RED = Red

reflectance, G=Gain Factor (2.5), C1 and C2 = Atmospheric resistance coefficients = 6 and 7.5 respectively, L = Soil adjustment factor (0.5).

Spectral band correspondence between Landsat and spectroradiometer data was established using the blue (0.45–0.51µm; 470nm), green (0.53–0.59µm; 550nm), red (0.64–0.67µm; 670nm), and NIR (0.85–0.88µm; 800nm) wavelengths for vegetation index calculations. Vegetation indices (Eq. 1 to 5) derived from satellite imagery were calibrated using field spectroradiometer measurements collected during corresponding satellite overpass periods. Regression analysis was employed to establish correction relationships between field and satellite observations, and calibration coefficients were derived to minimize systematic bias. The resulting calibration equations were applied to satellite datasets to improve the accuracy of vegetation index values used in subsequent crop yield stress modelling. The general linear regression equation used for calibration is given by Eq. 6.

$$Y = aX + b \tag{6}$$

where, Y = Calibrated vegetation index (field-spectroradiometer value), X = Satellite-derived vegetation index, a = Regression slope (scaling factor), b = Intercept (bias correction term). After deriving coefficients, the calibration is applied as:

$$VI_{calibrated} = a \times VI_{satellite} + b \tag{7}$$

Ultimately, the Inverse Distance Weight method (IDW) was employed to generate all-weather variable maps in a GIS

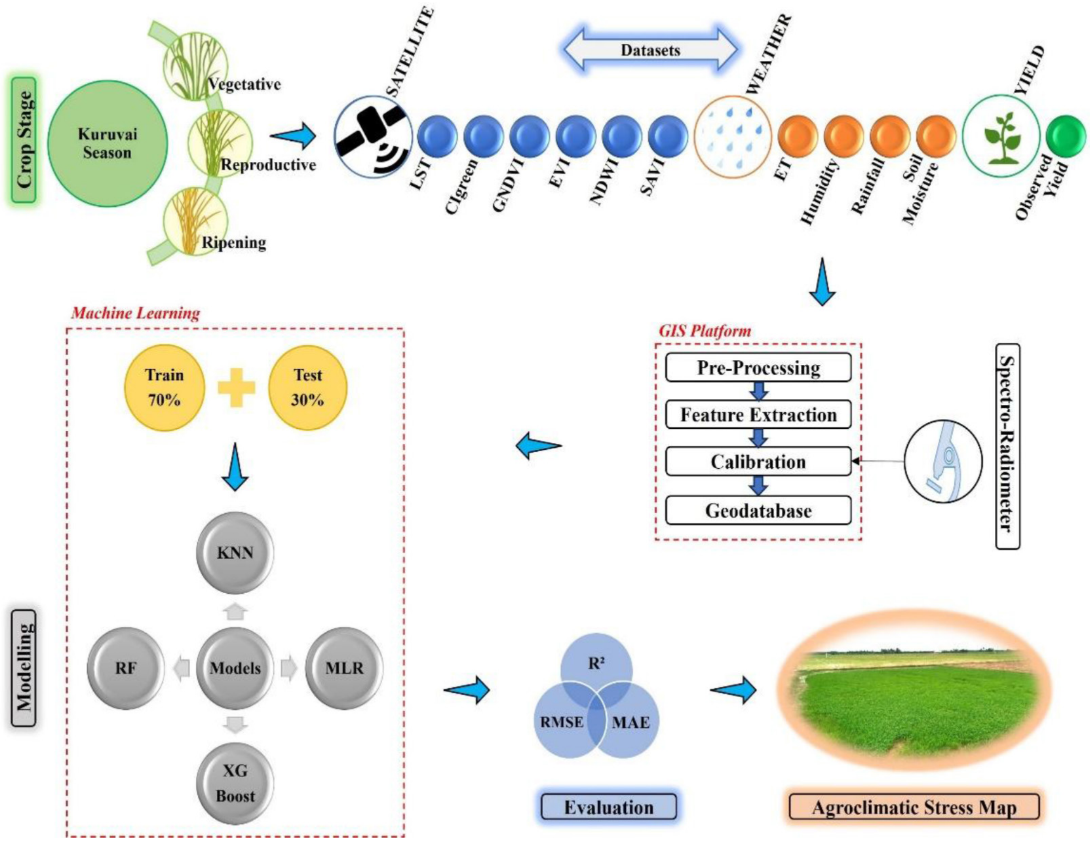

Fig. 2: Methodological framework adopted for paddy yield stress estimation.

platform. The Food and Agriculture Organization (FAO) Penman-Monteith equation, widely known as the standard method for ET_0 computation (Ahmad *et al.,* 2017).

$$ET_0 = \frac{0.408\Delta(R_n\text{-}G)+\gamma\frac{900}{T+273}u_2(e_s\text{-}e_a)}{\Delta+\gamma\,(1+0.34u_2)} \quad (8)$$

Retrieval of LST using Thermal Infrared Bands

LST was derived from satellite thermal infrared imagery (band 10; λ=10.60-11.19) to assess surface thermal conditions influencing crop growth. Thermal band digital numbers were first converted to spectral radiance, followed by conversion to at-sensor brightness temperature. Surface emissivity was estimated using vegetation indices derived from red and near-infrared bands, and emissivity correction was applied to obtain accurate LST values (Eq.9). Stage-wise LST values were extracted for further crop responsiveness and yield modelling analysis.

$$\text{LST (°C)} = \frac{T}{1+\left(\frac{\lambda.T}{\rho}\right)\ln(\varepsilon)} \quad (9)$$

where, T=Effective at-satellite brightness temperature (in kelvin), λ=wavelength of emitted radiance (10.9 µm for Band 10, Landsat8/9), ε=Emissivity (unitless). In physics, ρ is calculated using fundamental constants derived from Planck's Law.

$$\rho = \frac{h.c}{\sigma} = 1.438 \times 10^{-2}\ \text{m.K},$$

Where, h=Planck's constant (6.626×10^{-34} Js), c=speed of light (2.998×10^{8} m/s), σ=Boltzmann constant (1.38×10^{-23} J/K).

Machine Learning approach for evaluating Agro-climatic effects on Paddy yield variability

The objective was to capture complex nonlinear interactions among vegetation indices, climatic factors, and soil parameters across different crop growth stages. This study focuses on identifying and quantifying crop yield stress conditions rather than direct yield prediction. Yield stress refers to the reduction in crop performance caused by adverse environmental and climatic conditions during critical growth stages.

Data Preparation: All satellite-derived vegetation indices, weather variables, soil parameters, and field observations were aggregated stage-wise for each field. Outliers were removed where necessary, and all features were normalized to reduce scale bias among variables. Observed paddy crop yield data as model target variable. The dataset was split into training (70%) and testing (30%) subsets to ensure unbiased performance evaluation.

Model Development: ML algorithms were applied to classify and estimate yield stress intensity using stage-wise environmental and vegetation parameters for before and after calibration. Algorithms tested include RF, Multiple Linear Regression (MLR), eXtreme Gradient Boosting (XGBoost), and K-Nearest Neighbors (KNN) models, which were trained to learn relationships between environmental stress signals and crop performance variability (Palchowdhuri *et al.,* 2018). The ML models were implemented using predefined parameter settings rather than a formal hyperparameter optimization procedure. MLR and XGBoost were executed using their standard settings, KNN configured with five nearest neighbors (k=5), while the RF model employed 200 decision trees (n_estimators=200).

Model Validation: Model performance was assessed using classification and regression metrics, including Coefficient of determination (R^2), Root Mean Square Error (RMSE), and Mean Average Error (MAE) for stress intensity estimation.

Yield Stress Definition: Yield stress was quantified using deviations of vegetation condition and environmental variables from normal crop conditions. Stress indicators were derived from reductions in vegetation indices, increases in temperature, and moisture deficits during critical phenological stages. Stress conditions were classified using an equal-interval approach with a class width of 0.20, resulting in five stress categories: Very Low (0.00–0.20), Low (0.20–0.40), Moderate (0.40–0.60), High (0.60–0.80), and Very High (0.80–1.00).

Stress Mapping: The trained model was applied spatially across the study region to produce yield stress maps using interpolation, identifying zones experiencing heat or moisture stress during crop growth stages. These maps provide insights for targeted agricultural management and climate risk assessment (Yan *et al.,* 2024). The paddy crop yield stress index was given by,

$$\text{Stress Index} = 1 - \text{Normalised Yield } (0 \leq SI \leq 1) \tag{10}$$

Normalized Yield represents the relative yield performance of each field and was derived using min–max normalization (Alibabaei *et al.,* 2021) of the observed field yield data was given by Eq. 11.

where Yi denotes the observed yield of the i^th field, while Y_{min} and Y_{max} represent the minimum and maximum observed yields within the dataset, respectively. The normalization procedure transforms yield values to a dimensionless range between 0 and 1, facilitating comparison across fields and enabling integration with other normalized biophysical indicators in the stress assessment framework.

RESULTS AND DISCUSSION

Calibration of Multi-Spectral Vegetation Indices with In-Situ Spectral Reflectance Measurements

Calibration analysis was conducted for the indices during three phenological stages of paddy showed a high statistical agreement. The results of the linear regression coefficients (a: slope, b: intercept) and R^2 prove that the radiometric correction method has been successfully used to enhance the consistency between the observation of field-based spectroradiometer and satellite-derived vegetation indices shown in Fig. 3. The Chlorophyll green Index during vegetative stage had a high R^2=0.88 and a=1.049, with a small b=0.046 indicating a strong linear relationship and sensitivity to changes in early canopy chlorophyll. GNDVI during vegetative stage was found to have one of the strongest correlations (R^2=0.90), which indicates its strength in modelling the role of nitrogen-associated physiological action in the early growth of the biomass. Vegetative-stage SAVI had a lower slope (a=0.984), yet it had a good fit (R^2=0.87), which is an affirmation of its ability to reduce the effect of the soil background at the thin canopy conditions. NDWI (R^2=0.86) and EVI (R^2=0.85) also proved to be acceptable in terms of crop water content and structural vigor, respectively, during vegetative stage.

EVI (R^2=0.92; a=1.082) indicates its superior ability to describe the peak canopy structure and biomass accumulation during panicle initiation and flowering. CI_{green} (R^2=0.88) and NDWI (R^2=0.87) also showed a consistent sensitivity to the chlorophyll concentration and the water status of plants in dense vegetation conditions in reproductive stage. The performance of GNDVI and SAVI was similar (R^2=0.87), indicating that they were moderately responsive to physiological stress and soil-adjusted vegetation reflectance during the developmental reproductive stage.

During the ripening period, all indices had high calibration performance with slightly higher values of the slope implying greater spectral separability with the changes in canopy maturation and senescence dynamics. The most reliable was proved to be CI_{green} (R^2=0.89; a=1.153), which proves its ability to identify the trends of chlorophyll degradation. NDWI was also highly correlated (R^2=0.89), hence its applicability in determining the decreasing water content of the crop during grain filling. Reliable capture of late-stage canopy variability was found in GNDVI (R^2=0.88) and SAVI (R^2=0.87), with a comparatively lower yet satisfactory correlation in EVI (R^2=0.86), which could have been because of the saturation effect in the senescent canopy conditions. Such results confirm that calibrated vegetation indices are appropriate to be incorporated into ML-based paddy yield stress models. The obtained calibration results are consistent with previously reported spectroradiometer-based vegetation index studies across agricultural crops (Martinez *et al.,* 2018; Haboudane *et al.,* 2004)as well as to external factors affecting canopy reflectance.

Paddy Yield vs Weather Parameters and Vegetation Indices.

Correlation analysis of agrometeorological and paddy yield (Fig. 4) showed that the humidity in the reproductive stage (HUM_Rep) had the strongest positive correlation with the yield (r=0.72), which indicated that the proper availability of atmospheric moisture in the reproductive stage (during flowering and grain formation) had a significant positive effect on the productivity. The moderate positive correlation (r=0.43) between rainfall during the vegetative (RF_Veg) and biomass accumulation demonstrated that an adequate supply of water at an earlier stage of the canopy formation was vital to biomass accumulation. The humidity at the ripening stage (HUM_Rip), on the other hand, was strongly negatively related (r=-0.53) to the idea that too much moisture in the atmosphere at the time of grain maturation can have a negative influence on grain filling and predisposition to pests and diseases. Otherwise, reproductive (RF_Rep; r=-0.47) and ripening (RF_Rip; r=-0.39) phases of rainfall had inversely correlated relationships with yield, suggesting that reproductive success may be impaired in cases of waterlogging or excessive rainfall at these vulnerable stages of phenological development. There was also a negative effect of LST during the reproductive (LST_Rep; r=-0.29) and ET during

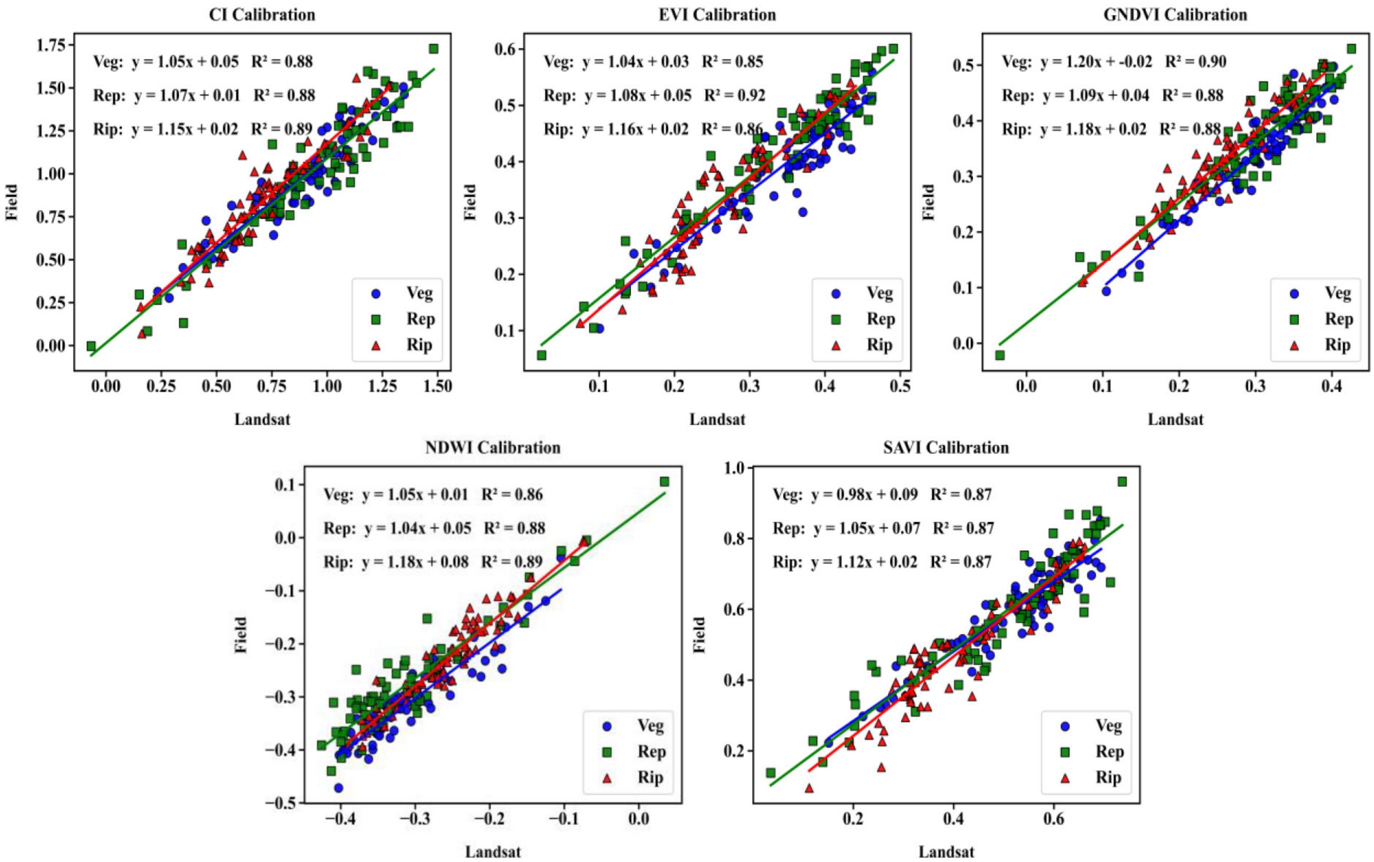

Fig. 3: Comparative analysis of spectral vegetation indices before and after field-based calibration to enhance the reliability of satellite-derived crop condition assessment.

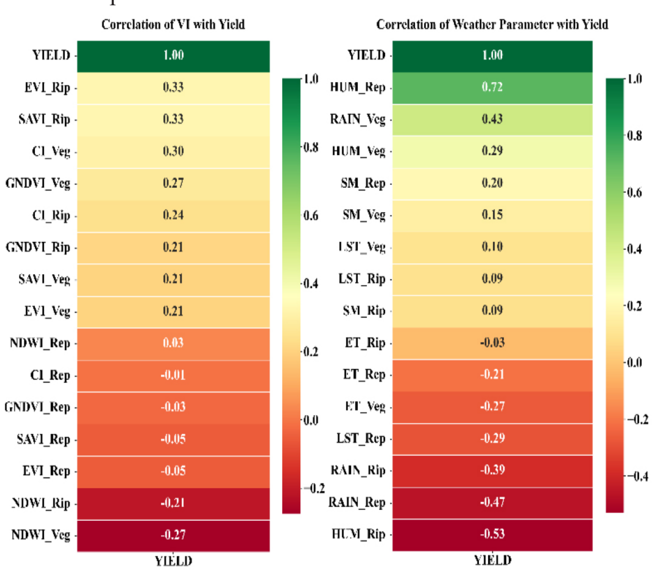

Fig. 4: Stage-wise correlation of paddy yield with spectral vegetation indices and agro-climatic parameters during the *Kharif* 2023 season.

both vegetative (ET_Veg; r=-0.27) and reproductive (ET_Rep; r=-0.21) stages on yield, as a manifestation of the effects of thermal stress and heightened evaporative demand on crop physiological performance. The soil moisture was also weak and positively associated with the vegetative (SM_Veg; r=0.15) and reproductive (SM_Rep; r=0.20) stages, which showed that it contributed to but did not dominate green matter variation.

The positive correlations were highest at EVI and SAVI in the ripening stage (EVI_Rip and SAVI_Rip), and this suggests that vigorous canopies at the end of the season contribute significantly to grain filling and the final realisation of yield. Correspondingly, CI_{green} at the vegetative growth (CI_{green}Veg), was moderately positively correlated (r=0.30), indicating the effect of the initial stage of chlorophyll concentration on the formation of yields. The GNDVI at the vegetative (r=0.27) and ripening (r=0.21) phases also showed significant relationships with yield, as it is sensitive to nitrogen processes and photosynthetic efficiency. Nonetheless, the NDWI indices of all stages, especially vegetative (NDWI_Veg; r=-0.27) and ripening (NDWI_Rip; r=-0.21), had a negative correlation with yield, and therefore, high water content or surface wetness of the canopies could be associated with stressful conditions of waterlogging. There were weak or negative correlations with indices in the reproductive stage (EVI_Rep, SAVI_Rep, GNDVI_Rep), possibly due to spectral saturation effects when the canopy is dense. This shows that spectral mid-season variability may not be as highly sensitive to yield differences as other growth stages (early and late). These stage-specific correlations can be useful in incorporating the climatic stress indicators (Sidhu *et al.,* 2021) and vegetation indices (Torgbor *et al.,* 2023) thus optimizing productivity and reducing food waste. More recent evaluations of technological solutions such as remote (ML models to better map paddy yield stresses and predictive models.

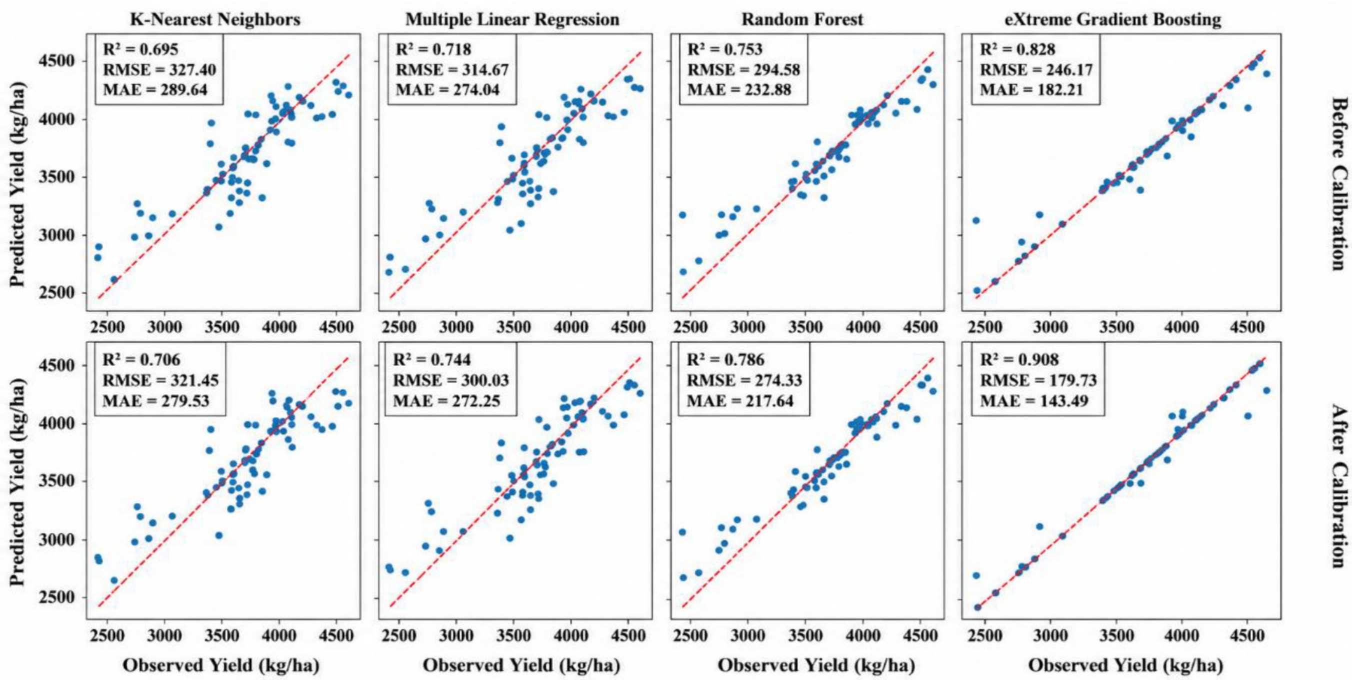

Fig. 5: Observed versus predicted paddy yield values derived from KNN, MLR, RF, and XGBoost models before and after calibration. improvement in prediction accuracy during the *Kharif* 2023 season.

Table 2: Performance metrics of selected ML models for paddy yield prediction

Models	Before Calibration			After Calibration		
	R^2	RMSE	MAE	R^2	RMSE	MAE
KNN	0.695	327.399	289.635	0.706	321.454	279.526
MLR	0.718	314.668	274.037	0.744	300.030	272.247
RF	0.753	294.577	232.879	0.786	274.329	217.636
XG Boost	0.828	246.167	182.207	0.908	179.728	143.487

Pre- and Post-Calibration ML Model Performance for Paddy Yield Prediction.

ML model performance was compared before and after calibrating satellite-based inputs with field spectroradiometer measurements (Fig. 5), with outliers removed using the Interquartile Range method to reduce anomalous influence on predictions. Before calibration, XGBoost achieved the highest accuracy (R^2=0.828; RMSE=246.17 kg ha^{-1}; MAE=182.21 kg ha^{-1}), followed by RF (R^2=0.753), while MLE and KNN showed lower predictive ability (R^2=0.718 and 0.695). After calibration, XGBoost showed marked improvement (R^2=0.908; RMSE=179.73 kg ha^{-1}; MAE=143.49 kg ha^{-1}), whereas RF improved moderately (R^2=0.786). MLE exhibited slight gains (R^2=0.744), and KNN demonstrated marginal improvement, reflecting sensitivity to local data distribution and high-dimensional geospatial variability. The accuracy of the yield prediction, as well as the spectral consistency of the input variables given in Table 2. Eventually, ensemble learning models, capture more variability in the stage-wise stress, as well as canopy dynamics that affect the paddy productivity.

Stage-Specific Yield Stress assessment based on XGBoost Model

The stage-based stresses classification of the paddy yield prediction model estimated by the use of the XGBoost model can show the critical information on the relative importance of the biophysical and agrometeorological parameters in the development and growth of paddy at various growth stages, as shown in Fig. 6. The findings have shown that the majority of vegetation indices, i.e., CI_{green}EVI, GNDVI, NDWI, and SAVI demonstrated a low level of stress impact across all stages, leading to a low contribution to yield variation. This indicates that spectral indices are a good parameter for capturing canopy vigor and chlorophyll dynamics, but may have a low direct sensitivity to yield-limiting stress conditions compared to climatic parameters, especially in dense canopies where spectral saturation can be observed.

ET and LST also had a consistent and consistently low stress influence at each developmental stage, and this is indicative of the fact that either the stress factor associated with thermal and atmospheric needs was reasonably reduced by irrigation practices or that they had a relatively uniform spatial variance within the study area. But humidity (HUM) showed a different stage-specific stress response, low influence on the vegetative, and strong influence on the reproductive as well as the ripening stage, which results in an overall high stress category. This points out how vulnerable flowering and grain filling activities are to changes in moisture content in the atmosphere, which may greatly impact the pollen viability, spikelet fertility, and grain development. Rainfall (RF) was one of the most influential parameters of stress and exhibited

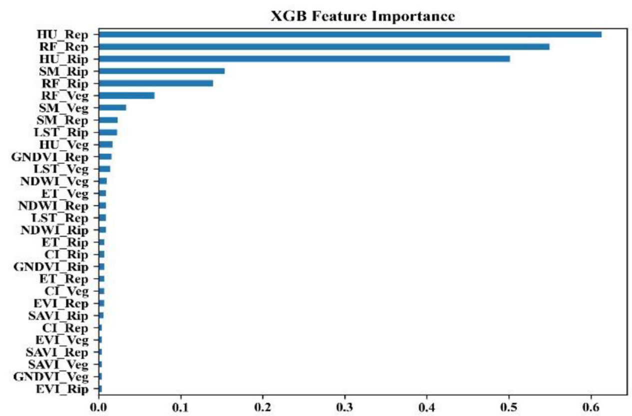

Fig. 6: Feature importance analysis from the XGBoost model showing the key predictors influencing paddy yield at different phenological stages.

high levels of stress effect at vegetative and reproductive stages and less at the ripening stage, which led to very high classification. This means that either too much or too little precipitation in the early and mid-growth phases can cause waterlogging or water stress, hence breaking down the biomass accumulation and panicle formation. Conversely, soil moisture (SM) was highly influential at the vegetative stage, but lowly influential at the reproductive stage and ripening stage, with a mean moderate classification of stress. Such a trend highlights the fact that sufficient soil water supply is needed in the initial stages of canopy formation, and the regulated drainage is more demanded in the later stages of growth. The findings also underscore the significance of integrating phenology-based climatic stress measures into ML models to map the regions under spatial yield stress and the control of precision irrigation practices, which is consistent with the findings of previous studies (Shahhosseini *et al.,* 2021).

Paddy yield Classification based on stress index levels

The results of the spatial distribution of paddy yield stress (Fig. 7) according to the use of the XGBoost-based stress index demonstrate high intra-district variation of the stress in various administrative blocks. Table 3 presents the severity of yield-limiting environmental situations affecting crop productivity. The stress index was divided into five classes. Blocks like Ammappettai had very low levels of stress, which showed a favourable condition of the hydro-meteorological environment and good crop growth conditions during the critical stages of phenology. Likewise, Thiruvaiyaru, Thanjavur, Orathanadu, and Thiruvidaimarudur were largely in the low stress category, indicating fairly consistent climatic and soil moisture conditions favouring yield development. Conversely, Papanasam, Kumbakonam, Thiruvonam, Madukkur, and Pattukkottai were allied to the medium scale of stress, which showed signs of transitional conditions in the environment that could cause partial physiological stress levels because of the variation in the rainfall patterns, humidity, or soil water. It is important to note that the southern blocks like Sethubhavachathram and some areas in Peravurani had high to very high stress levels, which means that there is unfavourable growth condition, and this may be due to the factors of reproductive stage hydro-meteorological stress, such as an excessive variation of rainfall or an imbalance of moisture in the

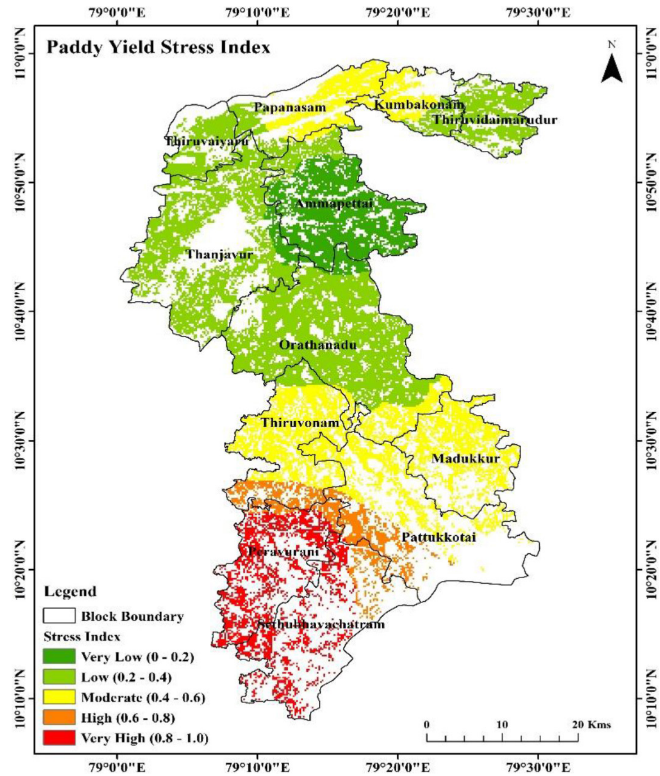

Fig. 7: Spatial distribution of paddy yield stress across the study area during the *Kharif* 2023 season.

atmosphere.

The existence of a spatial gradient of stress intensity between the north and south blocks depicts the effect of climatic parameters during a certain stage-specific humidity and rainfall, on variability in yield, as had been previously found in the feature importance analysis. High levels of stress are associated with areas where the moisture conditions of reproductive stages are likely to affect the grain filling process and can affect the translocation procedures, thus yield potential is minimized. These results are consistent with previous research level (Krithikha Sanju & Bhagavathiappan, 2022) that shows that spatial stress indices simulated by ML are effective to enhance the intricate relationships between climatic variability and crop productivity at the regional.

CONCLUSION

This paper established a stage-specific geospatial-ML model of estimating yield-limiting stress in *Kharif* paddy in Thanjavur Delta by combining field spectroradiometer-calibrated vegetation indices and agrometeorological variables. The framework was able to effectively characterize the nonlinear interactions between crops and weather that affect paddy productivity at various stages. Out of the assessed methods of ML, ensemble-based models have proven to have the highest performance in detecting variability in stage-specific stress, especially in the reproductive period, which was the most sensitive to climatic changes. Using field spectroscopy enhanced the spectral consistency and uncertainty in satellite-based stress estimation over spatially heterogeneous fields. The suggested novel framework offers considerable information to agricultural planners, policymakers, and farmers as it allows

Table 3: Paddy crop yield Stress Index classification with associated stress levels

Stress Index	Stress Level	Yield Reduction (%)	Expected Yield State
0.00 - 0.20	Very Low	0 - 20	Very High
0.21 - 0.40	Low	20 - 40	High
0.41 - 0.60	Moderate	40 - 60	Moderate
0.61 - 0.80	High	60 - 80	Low
0.81 - 1.00	Very High	80 - 100	Very Low

them to detect the conditions of crop stress on time and make better decisions concerning crop management to make the use of crops more sustainable. It can help in the optimization of irrigation water distribution, creation of stage-based agro-advisory, and the empowerment of climate-resilient rice production plans in other semi-arid regions. Additionally, by allowing transparent yield stress measurement and post-event damage measurements, which are satellite-based, the framework will be able to support evidence-based crop insurance programs by enhancing the accuracy of indemnity and minimizing the time to settle claims. The study is constrained by limited soil characteristics, crop farming activities, and irrigation systems, which can cause localized bias during model predictions. Future research can focus on the integration of farmer surveys, paddy types, high-resolution UAV images, real-time networks of soil and microclimate sensors using the IoT, and more sophisticated hybrid deep learning structures can be used in future studies to improve the monitoring of crop stress over time.

ACKNOWLEDGMENT

We would like to express our sincere gratitude to the U.S. Geological Survey (USGS), National Remote Sensing Centre (NRSC), India Meteorological Department (IMD), Geological Survey of India (GSI), Tamil Nadu Agricultural University (TNAU), and the Office of the Joint Director of Agriculture, Thanjavur, for providing the essential datasets required for this research. We also extend our gratitude to the management of SRM University, Kattankulathur Campus, for their continuous support and encouragement throughout the study. Furthermore, we acknowledge the contributions of all individuals involved in data collection and archiving. The authors are also thankful to the anonymous reviewers for their valuable comments and constructive suggestions that helped improve the quality of this work.

Author's certificate: The manuscript or its part is not under consideration for publication elsewhere and the same has been approved by all co-authors.

Conflict of Interests: The authors declare that there is no conflict of interest regarding the publication of this article.

Funding: No funding

Data Availability: The Data used in this study are available upon reasonable request from the corresponding author.

Author's contribution: **Karthik Karunakaran:** Conceptualization, Methodology, Data Collection & Analysis, Original Draft Preparation, Machine Learning; **Karuppasamy Sudalaimuthu:** Data Collection, Machine Learning, Supervision, Reviewing, Validation.

Disclaimer: The contents, opinions, and views expressed in the research article published in the Journal of Agrometeorology are the views of the authors and do not necessarily reflect the views of the organisations they belong to.

Publisher's Note: The periodical remains neutral concerning jurisdictional claims in published maps and institutional affiliations.

REFERENCES

IPCC (2023). Climate Change 2022 – Impacts, Adaptation and Vulnerability: Working Group II Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. *Climate Change 2022 – Impacts, Adaptation and Vulnerability*. https://doi.org/10.1017/9781009325844

Ahmad, L., Parvaze, S., Parvaze, S., & Kanth, R. H. (2017). FAO Reference evapotranspiration and crop water requirement of apple (Malus Pumila) in Kashmir Valley. *Journal of Agrometeorology*, *19*(3), 262–264. https://doi.org/10.54386/jam.v19i3.668

Alibabaei, K., Gaspar, P. D., & Lima, T. M. (2021). Crop Yield Estimation Using Deep Learning Based on Climate Big Data and Irrigation Scheduling. *Energies 14*(11), 3004. https://doi.org/10.3390/EN14113004

Dela Torre, D. M. G., Gao, J., & Macinnis-Ng, C. (2021). Remote sensing-based estimation of rice yields using various models: A critical review. *Geo-Spatial Information Science*, *24*(4), 580–603. https://doi.org/10.1080/10095020.2021.1936656

Dineshkumar, C., Nitheshnirmal, S., Bhardwaj, A., & Priyadarshini, K. N. (2019). *Phenological Monitoring of Paddy Crop Using Time Series MODIS Data*. 19. https://doi.org/10.3390/iecg2019-06205

Haboudane, D., Miller, J. R., Pattey, E., Zarco-Tejada, P. J., & Strachan, I. B. (2004). Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture. *Remote Sensing of Environment*, *90*(3), 337–352. https://doi.org/10.1016/j.rse.2003.12.013

Hatfield, J. L., & Prueger, J. H. (2015). Temperature extremes: Effect on plant growth and development. *Weather and Climate Extremes*, *10*, 4–10. https://doi.org/10.1016/j.wace.2015.08.001

Jha, R. K., Kalita, P. K., & Cooke, R. A. (2021). Assessment of climatic parameters for future climate change in a major agricultural state in India. *Climate*, *9*(7). https://doi.org/10.3390/cli9070111

Krithikha Sanju, S., & Bhagavathiappan, V. (2022). A comprehensive approach on predicting the crop yield using hybrid machine learning algorithms. *Journal of Agrometeorology*, *24*(2), 179–185. https://doi.org/10.54386/jam.v24i2.1561

Martinez, N. E., Sharp, J. L., Johnson, T. E., Kuhne, W. W., Stafford, C. T., & Duff, M. C. (2018). Reflectance-based vegetation index assessment of four plant species exposed to lithium chloride. *Sensors (Switzerland)*, *18*(9), 1–24. https://doi.org/10.3390/s18092750

Palchowdhuri, Y., Valcarce-Diñeiro, R., King, P., & Sanabria-Soto, M. (2018). Classification of multi-temporal spectral indices for crop type mapping: A case study in Coalville, UK. *Journal of Agricultural Science*, *156*(1), 24–36. https://doi.org/10.1017/s0021859617000879

Sajan, B., Kanga, S., Singh, S. K., Mishra, V. N., & Durin, B. (2023). Spatial variations of LST and NDVI in Muzaffarpur district, Bihar using Google earth engine (GEE) during 1990-2020. *Journal of Agrometeorology*, *25*(2), 262–267. https://doi.org/10.54386/jam.v25i2.2155

Shahhosseini, M., Hu, G., Huber, I., & Archontoulis, S. V. (2021). Coupling machine learning and crop modeling improves crop yield prediction in the US Corn Belt. *Scientific Reports*, *11*(1), 1–15. https://doi.org/10.1038/s41598-020-80820-1

Sidhu, B. S., Sharda, R., & Singh, S. (2021). An assessment of water footprint for irrigated rice in Punjab. *Journal of Agrometeorology*, *23*(1), 21–29. https://doi.org/10.54386/jam.v23i1.84

Torgbor, B. A., Rahman, M. M., Brinkhoff, J., Sinha, P., & Robson, A. (2023). Integrating Remote Sensing and Weather Variables for Mango Yield Prediction Using a Machine Learning Approach. *Remote Sensing*, *15*(12). https://doi.org/10.3390/rs15123075

Verrelst, J., Malenovský, Z., Van der Tol, C., Camps-Valls, G., Gastellu-Etchegorry, J. P., Lewis, P., North, P., & Moreno, J. (2019). Quantifying Vegetation Biophysical Variables from Imaging Spectroscopy Data: A Review on Retrieval Methods. *Surveys in Geophysics*, *40*(3), 589–629. https://doi.org/10.1007/S10712-018-9478-Y

Vidican, R., Mălinaș, A., Ranta, O., Moldovan, C., Marian, O., Ghețe, A., Ghișe, C. R., Popovici, F., & Cătunescu, G. M. (2023). Using Remote Sensing Vegetation Indices for the Discrimination and Monitoring of Agricultural Crops: A Critical Review. *Agronomy*, *13*(12), 1–27. https://doi.org/10.3390/agronomy13123040

Wang, F., Yao, X., Xie, L., Zheng, J., & Xu, T. (2021). Rice yield estimation based on vegetation index and florescence spectral information from UAV hyperspectral remote sensing. *Remote Sensing*, *13*(17), 1–19. https://doi.org/10.3390/rs13173390

Yan, X., Guo, Y., Ma, B., Zhao, Y., Guga, S., Zhang, J., Liu, X., Tong, Z., & Zhao, C. (2024). Hazard assessment of rice cold damage based on energy balance in paddy field. *Agricultural and Forest Meteorology*, *358*, 110233. https://doi.org/10.1016/J.AGRFORMET.2024.110233

Yoshida, S. (1981). Fundamentals of Rice Crop Science. *Fundamentals of Rice Crop Science*, 65–109.

Zheng, Q., Huang, W., Xia, Q., Dong, Y., Ye, H., Jiang, H., Chen, S., & Huang, S. (2023). Remote Sensing Monitoring of Rice Diseases and Pests from Different Data Sources: A Review. *Agronomy*, *13*(7). https://doi.org/10.3390/agronomy13071851