International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 13 Issue: 10 | Oct 2026 www.irjet.net p-ISSN: 2395-0072

Satellite-Based Near-Real-Time Rice Yield Estimation using Vegetation Indices and Artificial Neural Network

Kakali Das[1], Madhubanti Sur[2], Suvajit Ghosal[3], Saradindu Sengupta[4]

[1] Senior Scientist, Geoinformatics & Remote Sensing Cell, Department of Science and Technology & Biotechnology, GoWB, West Bengal, India

[23]Scientific Associate, West Bengal State Council of Science and Technology, Geoinformatics & Remote Sensing Cell, Department of Science and Technology & Biotechnology, GoWB, West Bengal, India

[4] Retired Professor, IIT Kharagpur, West Bengal, India

Abstract - *Rice is major staple crops in India, making accurate pre-harvest yield estimation essential for crop insurance, agricultural planning, policy-making and food security.Traditional methodsare labor-intensive, time-consuming, costly and susceptible to sampling errors, limiting their efficiency for large scale assessment. This study investigates Remote Sensingand Artificial Intelligence for near real-time prediction of Kharif rice yield at block level without survey-based data. Sentinel-2A/2Bmultispectral imagerybetween2016-2023was processed using Google Earth Engine to extract four vegetation indices:Normalized Difference Vegetation Index, Green Normalized Difference Vegetation Index, Soil Adjusted Vegetation Index, and Green Chlorophyll Index and correlated with actual yield data from Agriculture Department.Predictor selection employed Leave-One-Out Cross Validation.Multiple Linear Regression (MLR) and Artificial Neural Network (ANN) based models were applied using Statistical Package for Social Sciences. Integrating vegetation indices with linear (MLR) and non-linear (ANN) models estimated yield reliably, achieving a coefficient of Determination (R^2) of 0.82-0.83.*

Key Words: Rice Yield Prediction, Sentinel-2A/2B, Artificial Neural Network, Vegetation Indices, Google Earth Engine, Machine Learning, Leave One Out Cross Validation.

1. INTRODUCTION

Not only is Rice (Oryza sativa L.) the dominant food crop for more than 65% India's population, but India is also one of the world's largest producers and exporters of rice, with an annual production of around 135 million tonnes from covering area of around45–48 million hectares. Among the major rice producing states in India, West Bengal shares the largest rice growing area nearly 5.5-6.0 million hectares because of its favorable agro-climatic conditions and contributes about 13–14% to country's total rice production (Source: DoA&FW, GoI;DoA, GoWB; IREF;ICAR – IIRR,2024-25).Besides being a major food crop, rice contributes significantly to agriculture economy of India and meets the needs of a growing population. Therefore, availability of accurate pre-harvest yield with production data is essential for agricultural management, pricing, planning and strategic decision-making (Dadhwal and Roy 2000; Kosmowskiet.al., 2021; dela Torre et.al., 2021). But till date in most of the regions it is done by traditional methods which are constrained by time delay, limited temporal coverage as well as prone to large errors(Sapkota et. al.,2016; Bhattacharya et.al,2020; Gummadi et al., 2022; Gumma et al., 2022; Wu et.al,2023). Therefore, development of advanced technology-based approaches for timely, reliable, efficient and accurate crop yield estimation along with decision-support systems at multiple spatial scales is essential to ensure sustainable agricultural productivity and food security.

Since 1970s the inception of satellite based Remote Sensing technology emerges as a faster, cost effective , timely and reliable tool for synoptic observations, condition assessment and monitoring of agricultural crops, while the subsequent availability of multi-source spatio-temporal data from active and passive remote sensing systems has further opened up a new dimension in agricultural applications world-wide, embracing the aspects like crop acreage estimation, monitoring crop growth, stress and damage assessment, yield forecasting, precision farming, food security and agricultural decision making etc.(MacDonald et al.,1975;Sahai& Dadwal,1990; Singh et al.,1992;Weiss et.al., 2020;Prasanth Kumar et.al., 2021; Lin et.al., 2022;GuntamukkalaSekhar et.al., 2022;Holecz et al.,2022; Becker-Reshef et al.,2022).

However, in order to get accurate pre-harvest yield estimates across diverse climate regimes to improve market transparency and food security, the integration of high-resolution remote sensing data with advanced algorithm and data-driven approaches like Big Data Analytics, Deep Learning (DL), Artificial Intelligence (AI) and Internet of Things (IoT) etc. along with rigorous statistical validation is essential (Ruben Fernandez-Beltran et al. 2021; Sun and Scanlon 2019; Sobue et al., 2022; Colombo-Mendiza et al.,2022; Zaho et al., 2020; Gomez et al., 2019, Hunt et al., 2019; Prasad et al., 2006; Gandhi et al., 2016; Kayad et al., 2019).

The present study focuses one valuating the potential of integrated satellite data and neural networks for predicting Kharif rice yields in Daspur-I Block of Paschim Medinipur District using multispectral data, specifically Sentinel 1A and

Sentinel 2A/2B from the European Space Agency, combined with Machine Learning (ML) and Deep Learning (DL) based models parallel to statistical analysis. The major objective of this work is to provide, block-level seasonal crop yield information to support effective agricultural planning and decision making.

2. OVER VIEW OF THE STUDY AREA

PaschimMedinipur, a prominent district in the state of West Bengal, is geographically located between 21046′ N to 22057′ N latitudes and 86033′ E to 87044′ E longitude, covering an area of approximately 6270 square kilometers (Figure 1).The district is blessed with fertile agriculturallands, abundant water resources, favorable climateconditions that supports a wide range of crops, including paddy, jute, wheat, maize, mustard, pulses, vegetables and fruits etc., thereby making agriculture an key contributor of the state's agricultural economy and significantly contributing to agricultural production of West Bengal.

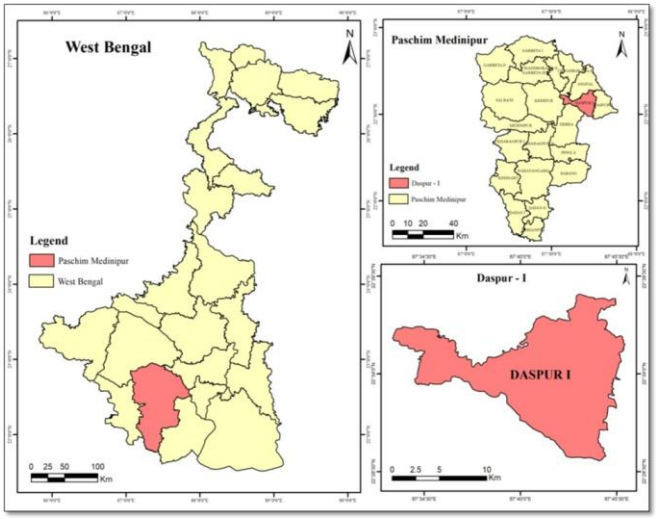


Fig -1: Study Area

3. MATERIALS AND METHODS

3.1 Data Used

Primarily, multispectral data from Sentinel-2 mission with a nominal 5-days revisit frequency, a 290 km wide-swath and 13 numbers of spectral bands at different spatial resolutions including 10 meters for visible and Near Infra Red (NIR), 20 meters for Red Edge and Short Wave Infra Red (SWIR) bands and 60 meters for atmospheric have been used, as it facilitates in assessing crop condition and growth, identification of crop type, detecting changes in vegetation health through various Indices, and predicting yield etc. (Xiong et al., 2017; He et al.,2019; Tripathi et al., 2021; Cao et al., 2021; Berger et al., 2022; Kumar et al., 2022; Chen et al.,2023; Perich et al., 2023; Sun et al.,2024b). Apart from remote sensing data, crop survey data obtained from the Department of Agriculture and Department of Statistics have also been used for model development and validation.

3.2 Methodology

The overall methodology adopted in the present study is illustrated in Fig-: 2.

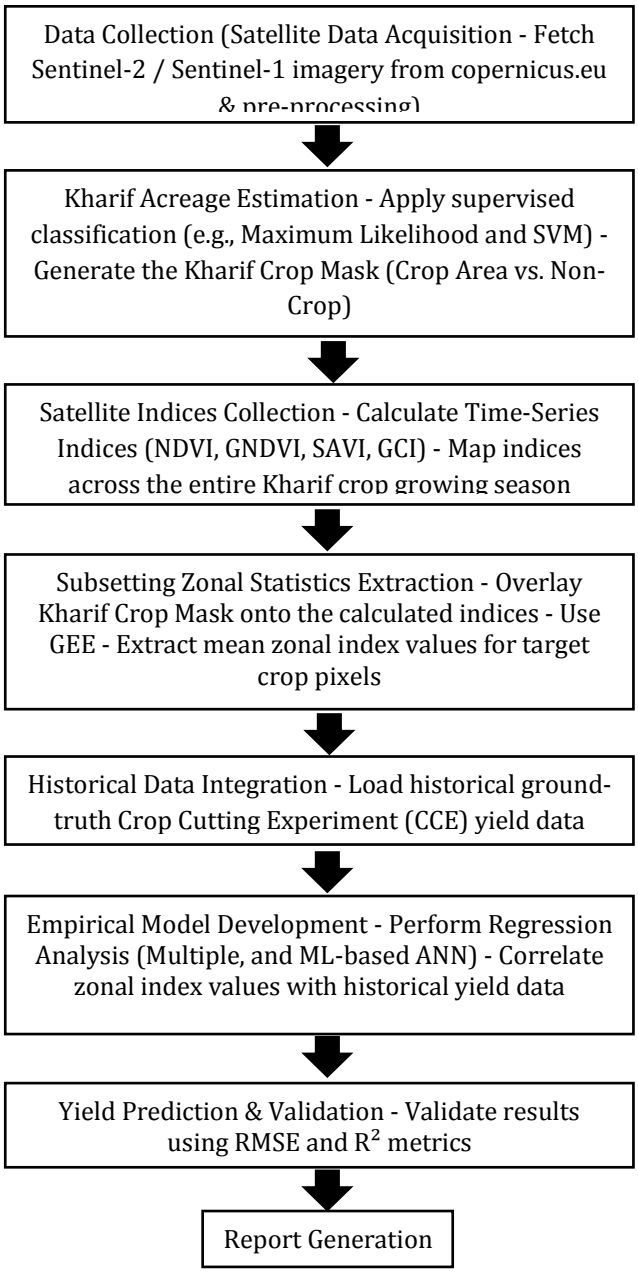


Fig -2: Methodology

3.2.1 Data Preparation

Initially, Sentinel-2 Imagery covering the study area was acquired and pre-processed, followed by the identification of the Kharif paddy growing areas using supervised classification techniques like Maximum Likelihood, Support Vector Machine (SVM) etc. After getting the Kharif paddy coverage areas, Google Earth Engine (GEE) was utilized to get the zonal statistics of a number of indices namely the Normalized Difference Vegetation Index (NDVI), Green Normalized Difference Vegetation Index (GNDVI), Soil Adjusted Vegetation Index (SAVI) and Green Chlorophyll Index (GCI), which were derived using the respective formulas presented in Table-1 and are widely used for assessing vegetation health and vigor, monitoring crop growth and condition, evaluating weather-included changes in vegetation and characterizing crop productivity etc. (Kogan et al. 2003, Singh et al., 2016; Dubey et al., 2017).

To extract the seasonal peak vegetation metrics across the study period (2016–2024), a temporal maximum-value filtering protocol was executed on the tabular comma-separated values (CSV) time-series dataset generated via Google Earth Engine (GEE). For each geographical zone and respective year, the absolute peak vegetation index value observed over the entire Kharif season was computed and isolated within Microsoft Excel. Because non-crop background features maintain consistently low spectral profiles, whereas active crop canopies exhibit a sharp, distinct surge during their peak vegetative phase, this annual maximum value serves as a dynamic, self-correcting agricultural filter that automatically accounts for localized shifts in monsoon sowing schedules. The primary spatial resolution was aggregated at the Block / pixel level of the Kharif crop area.

NDVI (Normalized Difference Vegetation Index)

Spectral Basis: NDVI capitalizes on the structural contrast of leaves. Chlorophyll strongly absorbs Red light (~660 nm) for photosynthesis, while the internal spongy mesophyll structure of healthy leaves reflects Near-Infrared (NIR) light (~860 nm).

Relevance to Yield Estimation: NDVI is the historical and operational industry standard for yield forecasting. It correlates linearly with Fractional Vegetation Cover (FVC) and early-to-mid season biomass. By capturing the maximum greenness of the Kharif crop, it acts as a direct proxy for the crop's vegetative capacity right before grain filling.

GNDVI (Green Normalized Difference Vegetation Index)

Spectral Basis: GNDVI replaces the Red channel with the Green spectrum (~550 nm). While Red light is absorbed entirely by the very top layer of a leaf, Green light can penetrate much deeper into the crop canopy.

Relevance to Yield Estimation: GNDVI is highly sensitive to chlorophyll concentration rather than just leaf surface area, as it penetrates deeper into the canopy, it saturates at a much higher leaf density than standard NDVI. It is an excellent choice for forecasting yield in high-biomass Kharif crops where standard NDVI loses its predictive variance mid-season.

SAVI (Soil-Adjusted Vegetation Index)

Spectral Basis: SAVI introduces a mathematically derived soil brightness correction factor (L). In agronomic remote sensing, L is conventionally set to 0.5 to account for most standard agricultural soil variations.

Relevance to Yield Estimation: During the early Kharif season or in areas with wider row spacing (like cotton or low-density pulses), the bare soil between crop rows reflects significant amounts of NIR and Red light. Standard NDVI treats dry or wet soil as an anomaly, skewing the actual vegetation signal. SAVI dynamically suppresses this background soil noise. Choosing SAVI guarantees that your peak seasonal value reflects true plant tissue growth rather than variations in soil moisture or texture.

Table -1: Vegetation indices utilized in the present study along with corresponding formulas

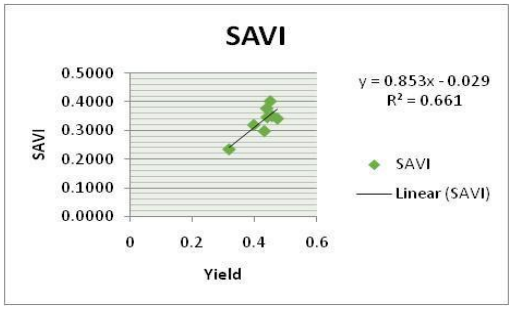
Vegetation Indices	Formula
Normalized Difference Vegetation Index (NDVI)	(NIR – Red) / (NIR + Red)
Green Normalized Difference Vegetation Index (GNDVI)	(NIR-GREEN) /(NIR+GREEN)
Soil Adjusted Vegetation Index (SAVI)	((NIR – R) / (NIR + R + L)) * (1 + L)
Green Coverage Index (GCI)	(NIR) / (Green) – 1

GCI (Green Chlorophyll Index)

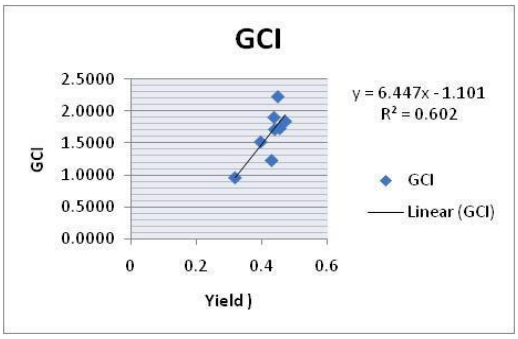
Spectral Basis: GCI is a linear, ratio-based index rather than a normalized difference index. It isolates the direct ratio of NIR scattering to Green light absorption.

Relevance to Yield Estimation: GCI tracks the Canopy Chlorophyll Content with high precision. In grain crops, final yield is directly proportional to the plant's nitrogen status and photosynthetic capacity during the reproductive phase. GCI provides a wider, linear dynamic range across varying levels of plant health. When fed into your Artificial Neural Network (ANN), GCI serves as a highly effective feature because its linear scale prevents the network from underestimating yield variations in premium, high-yielding fields.

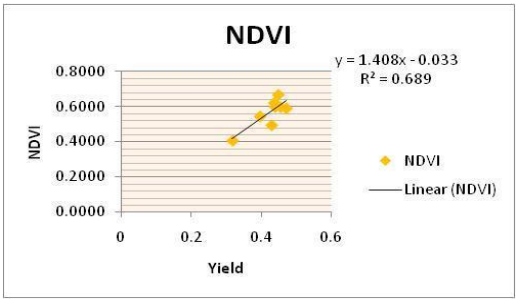
After extracting vegetation index values, correlation analysis was performed between independent/predictor (extracted vegetation index values) and dependent/response (observed crop yield) variables, and the coefficient of determination (R2) was used to identify the most relevant predictor variables based on the strength of their relationship with the dependent variable. The less-correlated outputs were discarded through the Leave-One-Out Cross-Validation (LOOCV) approach to ensure the model's predictive accuracy was thoroughly evaluated, minimize the potential bias and enhanced the reliability of rice yield estimation. Later on, Multiple Linear Regression (MLR) analysis and Artificial Neural Network (ANN) analysis were performed using seasonal data from July to October to predict Kharif paddy yield. The sample yield data collected from the Department of Agriculture served as the response variable, whereas the spectral values derived from Sentinel-2 imagery were used as predictor variables, based on which the yield prediction models were subsequently generated (Fig -3:).

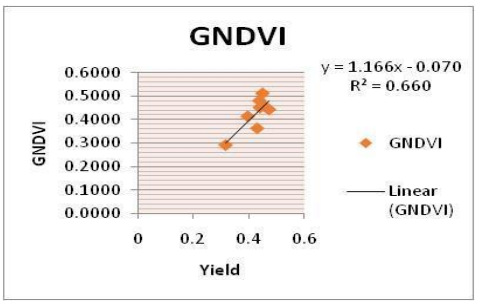


(a)



(b)





(d)

Fig -3: Correlation between Vegetation Indices with Rice Yield

3.2.1.1 Correlation Analysis and Multi Regression Modeling

For yield estimation, the field-survey data and the data obtained from the Department of Agriculture were used as the response variable, while vegetation indices derived from Sentinel 2 imagery namely NDVI, GNDVI, SAVI, and GCI are served as predictor variables. Preliminary, correlation analysis was performed to assess the strength and direction of the relationship between each index and observed rice yields, with Pearson correlation coefficients computed to quantify linear associations. Based on the correlation analysis results, Multiple Linear Regression model was developed and expressed as:

$$Y=\beta0+\beta1\cdot NDVI+\beta2\cdot GNDVI+\beta3\cdot SAVI+\beta4\cdot GCI+\epsilon$$

Where, Y represents predicted yield, β1 denotes regression coefficients, and ϵ represents the error term. Model robustness was evaluated using Leave One Out Cross Validation (LOOCV), and predictive performance was assessed through R2 and Root Mean Square error (RMSE) values. The analysis revealed that vegetation indices with stronger correlations with observed yield contributed more significantly to yield prediction, while the multi regression framework enabled the integration of information from multiple vegetation signals, thereby improving prediction reliability compared to single index models (Fig-4:).

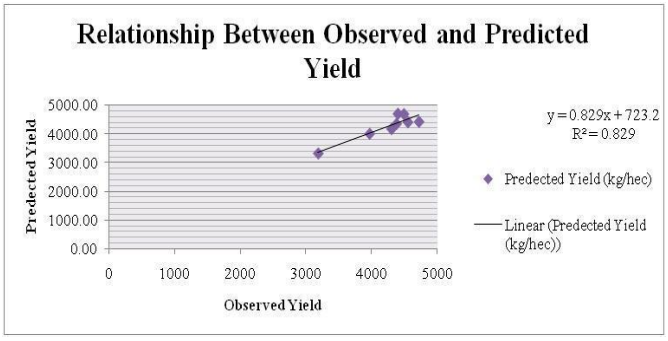


Fig -4: Accuracy of Rice Yield Estimation using Vegetation Indices

3.2.1.2 Artificial Neural Network (ANN) Architecture and Specifications

The Artificial Neural Network (ANN) is a powerful deep learning framework that can be applied to both regression and classification tasks, making it highly suitable for agricultural yield prediction (Jiang et al., 2004; Johnson et al., 2016; Khaki et al., 2019; Kim et al., 2019). In this study, ANN was utilized to estimate Kharif rice yields of the study area by integrating spatial features, vegetation indices including-NDVI, GNDVI, SAVI, and GCI (Refer to Table 1), and historical yield data. The ANN architecture consists of multiple hidden layers connected through nodes, enabling the model to capture complex, non-linear relationships between crop growth indicators and yield. These hidden layers allow the network to learn intricate patterns across multi-temporal satellite data, ultimately producing more accurate yield forecasts compared to traditional statistical methods model complex, non-linear relationships between the engineered Indices max value inputs, and final crop yields, an Artificial Neural Network (ANN) Radial Basis Function framework was constructed. The network was executed using the Multilayer Perception (MLP) engine within IBM SPSS Statistics 27.To comply with open science frameworks and allow exact replication by independent researchers, the network's structural, functional, and algorithmic configurations are detailed below. To ensure deterministic reproducibility of the data partition and weight initializations, the SPSS internal pseudorandom number generator was fixed to a global seed value of [2000000].

3.2.1.2.1 Network Topology and Structure

The predictive modeling framework was implemented using an ANN Multilayer Perceptron (MLP) with a feed-forward architecture organized into sequential layers. The input layer comprised neurons, each corresponding to standardized covariates derived from satellite observations and ancillary data. Specifically, the engineered features included Index max, AUC, and βgreen, supplemented by additional inputs as required. These variables were rescaled to a standardized scale to ensure comparability across predictors. The hidden structure consisted of a single intermediate layer, with the number of processing units either specified or automatically determined by the SPSS algorithm within a range of 1 to 50 nodes. Signal transformation within this layer was governed by the non-linear hyperbolic tangent activation function, which facilitated the capture of complex, non-linear relationships between spectral indices and crop yield outcomes. The output layer contained a single continuous node representing the target parameter, crop yield (kg/hectare). This layer employed the identity activation function to produce direct numerical predictions, optimized using a sum-of-squares error metric appropriate for regression tasks.

Model training was conducted in SPSS using batch optimization with the scaled conjugate gradient algorithm. Initial parameters were set to Lambda = 0.0000005 and Sigma = 0.00005, with interval center and offset values of 0 and 0.5, respectively. Data partitioning allocated 70% of the dataset to training and 30% to testing, with no holdout sample. Automatic stopping rules were applied, including a maximum training time of 15 minutes, an error change threshold of 1.0×10−4, and an error ratio of 0.001. The training process was monitored through error steps and epoch progression, with termination occurring once convergence criteria were satisfied. Model outputs included network information, variable importance measures, predicted versus residual plots, and saved predicted values. The final trained model was exported as an XML file to ensure reproducibility and facilitate subsequent application in yield estimation studies.

3.2.1.2.2 Verification and Replication Syntax

For absolute replication of the neural network workflow, the exact IBM SPSS command syntax utilized to build the network architecture is provided below:

Execute the Multilayer Perceptron (MLP) Crop Yield Network

```
MLP [Insert_Your_Yield_Variable_Name] WITH Max_NDVI
Seasonal_Integral GreenUp_Slope [Insert_Other_Variables]

 /PARTITION TRAINING=70 TESTING=30 HOLDOUT=0

 /ARCHITECTURE AUTOMATIC=NO

  HIDDENLAYERS=1 (NUMUNITS=[Insert_Hidden_Count]
ACTIVATION=TANH)

  OUTPUT=IDENTITY ERROR=SUMOFSQUARES

 /TRAINING                    TYPE=BATCH
OPTIMIZATION=SCALEDCONJUGATE

 /STOPPINGRULES ERRORSTEPS=1 (DATA=AUTO)
MINRELATIVECHANGE=0.0001 MAXEPOCHS=200

 /PRINT SUMMARY DIAGRAM
```

3.2.1.2.3 Training Configuration and Hyper parameters

The Multilayer Perceptron (MLP) model was trained using the SPSS Neural Networks module, with covariates rescaled to standardized values to ensure comparability across predictors. The dataset was partitioned into training (70%) and testing (30%) subsets, with no holdout sample. The architecture was configured for automatic determination of hidden units, constrained between a minimum of 1 and a

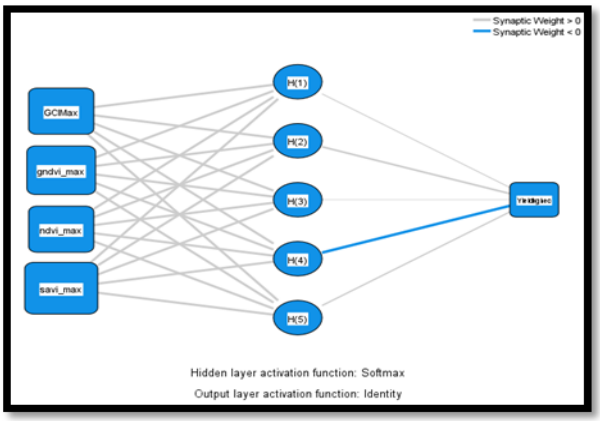
maximum of 50 nodes. Training was conducted in batch mode using the Scaled Conjugate Gradient optimization algorithm, which is well suited for non-linear regression tasks. Initial hyper parameters were set to Lambda = 0.0000005 and Sigma = 0.00005, with interval center and offset values of 0 and 0.5, respectively, and a memory size allocation of 1000.

Stopping rules were applied to prevent over-fitting and ensure computational efficiency. These included a maximum training time of 15 minutes, an automatic determination of epochs, an error change threshold of 1.0×10−4, and an error ratio of 0.001. The training process was monitored through error steps, with termination occurring once convergence criteria were satisfied. Model evaluation outputs included network information, variable importance measures, predicted versus residual plots, and saved predicted values. The final trained model was exported in XML format to facilitate reproducibility and subsequent application in crop yield estimation studies.

For crop yield estimation, three fundamental parameters define the ANN framework:

1. Network architecture – the arrangement and connectivity of neurons across input, hidden, and output layers.
2. Learning procedure – the optimization algorithm used to iteratively update interconnection weights during training.
3. Activation function – the mathematical transformation that converts a neuron's weighted input into its output signal and introduces non-linearity into the model.

By integrating these elements, ANN can effectively combine information from multi-source and multi-temporal satellite data, offering a robust framework for crop yield prediction (**Fig-** 5-8:). This methodology demonstrates the potential of Neural Network based approaches to support agricultural planning and food security by providing timely and reliable yield estimates at the block level.

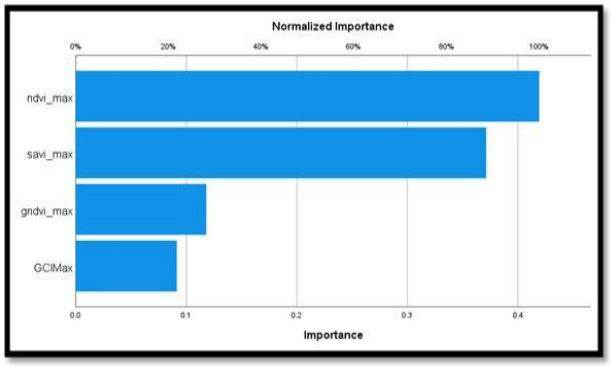


Network Information

Input Layer	Covariates	1	GCI Max
		2	gndvi_max
		3	ndvi_max
		4	savi_max
	Number of Units[a]		4
	Rescaling Method for Covariates		Standardized
Hidden Layer(s)	Number of Hidden Layers		1
	Number of Units in Hidden Layer 1[a]		200
	Activation Function		Hyperbolic tangent
Output Layer	Dependent Variables	1	Yield (kg/hec)
	Number of Units		1
	Rescaling Method for Scale Dependents		Standardized
	Activation Function		Identity
	Error Function		Sum of Squares

a. Excluding the bias unit

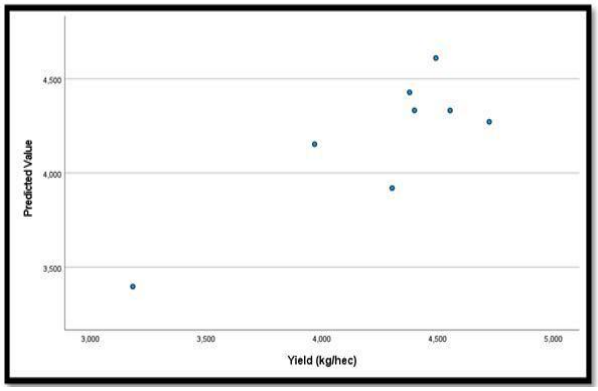
Fig -5: Process Flow of Artificial Neural Network in Yield Estimation



Independent Variable Importance

	Importance	Normalized Importance
GCI Max	.091	21.8%
gndvi_max	.118	28.1%
ndvi_max	.419	100.0%
savi_max	.371	88.6%

Fig -6: Relative Contribution of Predictors to Rice Yield Estimation



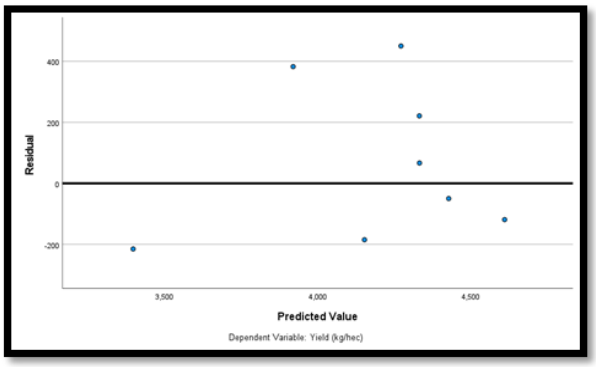


Fig -7: Scatter Plot of Predicted Yield against Actual Yield

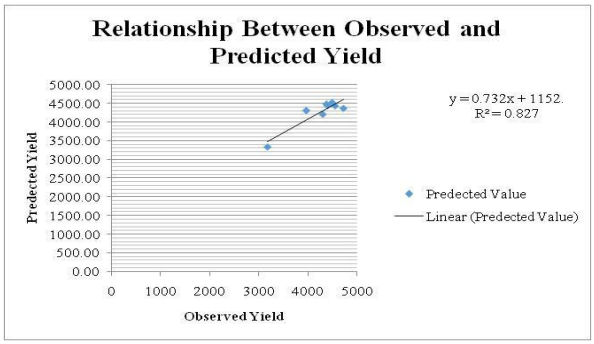


Fig -8: Accuracy of Rice Yield Estimation Using Vegetation Indices

4. Accuracy assessment

To compare the performance of different Machine Learning and Deep Learning models, Cross Validation (CV) was employed as a robust evaluation technique. CV is widely used for its simplicity, general applicability, and effectiveness in reducing the risk of overfitting. Model accuracy was assessed using multiple evaluation metrics including the Root Mean Square Error (RMSE), defined in Equation (2) which measures the average magnitude of prediction errors, indicates the degree of agreement between predicted yields and observed values. Additionally, the coefficient of determination (R^2), defined in Equation (1), was calculated to quantify the proportion of variance in observed yield explained by the model. Models exhibiting lower RMSE values and higher R2 value were considered to have better predictive performance, while cross validation results provided an assessment of their robustness and generalization capability.

$$R^2 = \frac{(\sum_{i=1}^{n}(y_i-\bar{y}_i)(f_i-\bar{f}_i))^2}{\sum_{i=1}^{n}(y_i-\bar{y}_i)^2\sum_{i=1}^{n}(f_i-\bar{f}_i)^2} \quad \text{..... Equation (1)}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - f_i)^2} \quad \text{..... Equation (2)}$$

Where:

- n = number of observations
- yi = actual yield
- y^i = predicted yield

The Artificial Neural Network (ANN) and Multiple Linear Regression (MLR) approaches exhibited different strengths and limitations in crop yield prediction MLR primarily characterized the linear relationships between vegetation indices and yield and ANN was capable of capturing non-linear relationships between vegetation indices (NDVI, GNDVI, SAVI, and GCI), weather variables, and yield data. The ANN achieved strong predictive performance with R2 values exceeding 0.80, demonstrating the potential for modelling complex relationships within multi temporal satellite data.

In contrast, Multiple Linear Regression provided a useful baseline model, explaining approximately 83% of yield variability (R2=0.83) in rice yield. However, its reliance on linear relationships limits its expressive power, especially for complex agricultural tasks, where crop growth is influenced by non linear and multi factor interactions. Consequently, MLR produced higher residual errors compared to ANN model.

Overall, ANN outperformed MLR by leveraging hidden layers and activation functions to capture intricate patterns in the data, while MLR offered greater interpretability and provided a framework for examining the contributions of individual vegetation indices. The success of ANN highlights its flexibility in integrating diverse predictors, whereas MLR serves as a simpler but powerful alternative for yield estimation.

Error-1 Calculation Processes using MLR:

Year	Yield (kg/ha)	Predicted Yield (kg/ha)	Error $(y - \hat{y})$	Error^2
2016	4553.14	4397.38	155.76	24,259.18
2017	4722.26	4411.55	310.71	96,534.65
2018	4378.02	4338.65	39.37	1,550.00
2019	4302.54	4165.27	137.27	18,844.00
2020	4399.08	4685.69	-286.61	82,152.00
2021	3182.08	3323.05	-140.97	19,880.00
2022	4491.56	4676.05	-184.49	34,030.00
2023	3967.80	3998.84	-31.04	963.00
2024	-	4039.25	-	-

Data Provided

1. Sum of squared errors

SSE = (0.000024 + 0.000014 + 0.000086 + 0.000110 + 0.000033 + 0.000047 + 0.000087 + 0.000019 + 0.071437) = 0.071857

2. Mean squared error

MSE = 0.0718579 ≈ 0.007984

3. Root mean square error

RMSE = 0.007984 ≈ 0.0894

MLR explained approximately 82% of yield variability (R^2 = 0.82), with a Root Mean Square Error (RMSE) of 0.089. While this provided a useful baseline, the model's reliance on linear relationships limited its ability to capture complex, non linear interactions among vegetation indices (NDVI, GNDVI, SAVI, GCI) and yield. Consequently, prediction errors were higher compared to deep learning approaches.

Error-2 Calculation Processes using ANN:

Year	Yield (kg/ha)	Predicted (kg/ha)	Error (y – ŷ)	Error^2
2016	4553.14	4513.27	39.87	1589.6
2017	4722.26	4513.10	209.16	43755.5
2018	4378.02	4513.30	-135.28	18302.7
2019	4302.54	3967.82	334.72	112039.1
2020	4399.08	4513.01	-113.93	12980.6
2021	3182.08	3182.08	0.00	0.0
2022	4491.56	4491.35	0.21	0.04
2023	3967.80	3967.83	-0.03	0.001
2024	-	3182.08	-	-

Data Provided

1. Sum of squared errors

SSE = (0.000016 + 0.000437 + 0.000182 + 0.001122 + 0.000130) = 0.001887

2. Mean squared error

MSE = 0.0018878 = 0.000236

3. Root mean square error

RMSE = 0.000236 ≈ 0.0154

Artificial Neural Network (ANN) achieved stronger predictive performance, with R^2 values exceeding 0.89 and a lower RMSE of 0.0154. The hidden layers and activation functions enabled ANN to learn intricate non linear patterns across spatial features and historical yield data, resulting in more accurate predictions.

5. RESULTS AND DISCUSSION

Two modeling approaches Multiple Linear Regression (MLR) and Artificial Neural Network (ANN) were applied to estimate rice yields using vegetation indices and historical yield data. Their performance was evaluated using five-fold Cross-Validation, with accuracy assessed through Root Mean Square Error (RMSE) and the Coefficient of Determination (R^2).

- The Multiple Linear Regression Model explained approximately 83% of variability in rice yield (R^2 = 0.83), with an RMSE of 0.089 across different datasets. While this provided a solid baseline for yield estimation, however, the model's linear relationships limit to capture complex and non-linear interactions among vegetation indices and yield. As a result, prediction errors were relatively higher, particularly in years with more variability in crop growth conditions.
- Artificial Neural Network (ANN) achieved stronger predictive performance, with R^2 values 0.89 and a RMSE of 0.0154, indicating agreement between predicted and observed yields. The hidden layers and activation functions allowed ANN to learn intricate non-linear patterns across spatial features and historical yield data, resulting in highly accurate predictions.

6. CONCLUSION

- Multiple Linear Regression provided greater interpretability, offering a clear framework for examining the relationships between predictor variables and rice yield, with NDVI emerging as the strongest predictor.

- However, ANN demonstrated greater flexibility in modelling complex and non-linear relationships that may not be adequately represented by a linear regression framework.
- The results highlight that, while regression provides a transparent baseline, ANN offers a more reliable and robust approach for block-level yield estimation, especially when dealing with multi-temporal satellite data and non-linear crop growth dynamics.

The ANN significantly demonstrated flexibility than MLR in terms of predictive accuracy, making it the preferred model for rice yield estimation in this study. MLR remains useful for explanatory analysis, but ANN is better suited for operational forecasting and agricultural planning.

ACKNOWLEDGEMENT

The authors are thankful to the Additional Chief Secretary, Special Secretary, Department of Science & Technology and Biotechnology, Government of West Bengal for providing administrative help to carry out the work and to extend the computational facilities of the Geoinformatics & Remote Sensing Cell of the Department.

REFERENCES

[1] Dadhwal, V. K., & Ray, S. S. (2000). Crop assessment using remote sensing-part-II: Crop condition and yield assessment. Indian Journal of Agricultural Economics, 55(2), 55–67.

[2] Kosmowski, F., Chamberlin, J., Ayalew, H., Sida, T., Abay, K., &Craufurd, P. (2021). How accurate are yield estimates from crop cuts? Evidence from smallholder maize farms in Ethiopia. Food policy, 102, 102122.

[3] Dela Torre, D.M.G., Gao, J., Macinnis-Ng, C., 2021. Remote sensing-based estimation ofrice yields using various models: a critical review. Geo-spat. Inf. Sci. 24 (4),580–603.

[4] Sapkota, T. B., Jat, M. L., Jat, R. K., Kapoor, P., & Stirling, C. (2016). Yield estimation of food and non-food crops in smallholderproduction systems. In T. Rosenstock, M. Rufino, K. Butterbach-Bahl, L. Wollenberg, & M. Richards (Eds.), Methods for measuring greenhouse gas balances and evaluating mitigation options in smallholder agriculture. Cham: Springer.

[5] Bhattacharyay, D., Maitra, S., Pine, S., Shankar, T. and PeddaGhousePeera S.K. (2020). Future of Precision Agriculture in India. In: Protected Cultivation and Smart Agriculture edited by Sagar Maitra, Dinkar J Gaikwad and Tanmoy Shankar, pp. 289-299.

[6] Gummadi, S., Dinku, T., Shirsath, P. B. and Kadiyala, M. D. M. (2022). Evaluation of multiple satellite precipitation products for rainfed maize production systems over Vietnam. Scientific reports, 12(1), 1-18.

[7] Gumma, M.K.; Thenkabail, P.S.; Panjala, P.; Teluguntla, P.; Yamano, T.; Mohammed, I. Multiple agricultural cropland products of south asia developed using landsat-8 30 m and modis 250 m data using machine learning on the google earth engine (gee) cloud and spectral matching techniques (SMTs) in support of food and water security. GISci. Remote Sens. 2022, 59, 1048–1077.

[8] Wu, B., et al., 2023. Challenges and opportunities in remote sensing-based crop monitoring: a review. Natl. Sci. Rev. 10 (4), nwac290.

[9] MacDonald, R. B.; Hall, F. G.; and Erb, R. B., "The Use of LANDSAT Data in a Large Area Crop Inventory Experiment (LACIE)" (1975). LARS Symposia. Paper 46.

[10] SahaiB, and DadhwalVk (1990) Remote Sensing in Agriculture. In 'Technology Blending and Agrarian Prosperity' (JP Verma, A Varma, eds.) p 83-98, Malhotra Publishing House.

[11] Singh, R., Goyal,R.C., Saha and Chhikara ,R.S.(1992) Use of satellite spectral data in crop yield estimation surveys. International Journal of Remote Sensing. .13 No.14 2583-2592.

[12] Weiss ,M., Jacob,F.,&Duveiller, G.(2020). Remote Sensing for agricultural applications: A meta-review. Remote Sens. Environ., 236,111402.

[13] Prasanth Kumar, P., Sairam, M., Shankar, T., Praharaj, S. and Maitra, S. (2021). Application of Remote Sensing in Agriculture. Indian Journal of Natural Sciences. 12(69): 37475-37479.

[14] Lin, L., Di, L., Zhang, C. Guo, L., Di, Y., Li, H. and Yang, A. (2022). Validation and refinement of cropland data layer using a spatial-temporal decision tree algorithm. Scientific Data 9, 63. https://doi.org/10.1038/s41597-022- 01169-w.

[15] GuntamukkalaSekhar, JogeshGoswami,Jami Naveen (2022).Crop Acreage and Yield Estimation using Remote Sensing and GIS and Various Related Schemes.Indian Journal of Natural Sciences, Vol.13 / Issue 72 / June / 2022.

[16] Holecz, F., Setiyono, T. D., Barbieri, M., Collivignarelli, F., Gatti, L., Anthony, M. and Zaugg, B. (2022). RIICE: The Service and Its Operational Use. In Remote Sensing of Agriculture and Land Cover/Land Use Changes in South and Southeast Asian Countries. pp. 105-131. Springer, Cham.

[17] Becker-Reshef, I., Bandaru, V., Barker, B., Coutu, S., Deines, J. M., Doorn, B. and Justice, C. (2022). The NASAProgram on and. In Remote Sensing of Agriculture and Land Cover/Land Use Changes in South and Southeast Asian Countries (pp. 53-80). Springer, Cham.

[18] Fernandez-Beltran, R.;Baidar, T.; Kang, J.; Pla, F.(2021). Rice-Yield Prediction with Multi-Temporal Sentinel-2 Data and 3D CNN: A Case Study in Nepal. Remote Sens. 2021, 13, 1391.https://doi.org/10.3390/rs13071391.

[19] Sun, Y. A. and Scanlon, B. R. (2019). How can Big Data and machine learning benefit environment and watermanagement: a survey of methods, applications, and future directions? Environmental Research Letters. 14 (2019)073001.

[20] Sobue, S., Samarakoon, L., Steventon, M., Oyoshi, K., & Le Toan, T. (2022). Asia-RiCE: Collaborative Frameworkfor Rice Crop Management in Asia. Remote Sensing of Agriculture and Land Cover/Land Use Changes in South andSoutheast Asian Countries, 81.

[21] Colombo-Mendoza, L. O., Paredes-Valverde, M. A., Salas-Zárate, M. D. P., & Valencia-García, R. (2022). Internetof Things-Driven Data Mining for Smart Crop Production Prediction in the Peasant Farming Domain. AppliedSciences, 12(4), 1940.

[22] Zhao, Y.; Potgieter, A.B.; Zhang, M.; Wu, B.; Hammer, G.L. Predicting wheat yield at the field scale by combining high-resolution Sentinel-2 satellite imagery and crop modelling. Remote Sens. 2020, 12, 1024.

[23] Gómez, D.; Salvador, P.; Sanz, J.; Casanova, J.L. Potato yield prediction using machine learning techniques and sentinel 2 data. Remote Sens. 2019, 11, 1745.

[24] Hunt, M.L.; Blackburn, G.A.; Carrasco, L.; Redhead, J.W.; Rowland, C.S. High resolution wheat yield mapping using Sentinel-2. Remote Sens. Environ. 2019, 233, 111410.

[25] Prasad, A.K.; Chai, L.; Singh, R.P.; Kafatos, M. Crop yield estimation model for Iowa using remote sensing and surface parameters. Int. J. Appl. Earth Obs. Geoinf. 2006, 8, 26–33.

[26] Gandhi, N.; Armstrong, L.J.; Petkar, O.; Tripathy, A.K. Rice crop yield prediction in India using support vector machines. In Proceedings of the 2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE), KhonKaen,Thailand, 13–15 July 2016.

[27] Kayad, A.; Sozzi, M.; Gatto, S.; Marinello, F.; Pirotti, F. MonitoringWithin-Field Variability of Corn Yield using Sentinel-2 andMachine Learning Techniques. Remote Sens. 2019, 11, 2873.

[28] Xiong, J., et al., 2017. Nominal 30-m cropland extent map of continental africaby integrating pixel-based and object-based algorithms using Sentinel-2 and Landsat-8data on Google Earth Engine. Remote Sens.

[29] He, L.; Mostovoy, G. Cotton Yield Estimate Using Sentinel-2 Data and an Ecosystem Model over the Southern US. Remote Sens.2019, 11, 2000.

[30] Tripathy, R., Chaudhari, K.N., Bairagi, G.D., Pal,O.,Das,R., &bhattacharya,B.K.(2021). Towards fine scale yield prediction of three major crops of India using datafrom multiple satellite. J.Indian Soc. Remote Sens., 1-14.

[31] Cao, D., et al., 2021. Delineating the rice crop activities in Northeast China throughregional parametric synthesis using satellite remote sensing time-series data from2000 to 2015. J. Integr. Agric. 20 (2), 424–437.

[32] Berger, K., et al., 2022. Multi-sensor spectral synergies for crop stress detection andmonitoring in the optical domain: a review. Remote Sens. Environ. 280, 113198.

[33] Kumar, D.A., Neelima,T.L., Srikanth,P., Devi, M.U.,Suresh,K., &Murthy,C.S. (2022). Maize Yield Prediction using NDVI derived from Sentinel 2 data in Siddipet district of telengana state. J. Agrometeorol., 24(2), 165-168.

[34] Chen, H., et al., 2023. A novel Greenness and Water Content Composite Index (GWCCI) for soybean mapping from single remotely sensed multispectral images. RemoteSens. Environ. 295, 113679.

[35] Perich, G., et al., 2023. Pixel-based yield mapping and prediction from Sentinel-2 usingspectral indices and neural networks. Field Crops Res. 292, 108824.

[36] Sun, X., Zhang, P., Wang, Z., Yijia, W., 2024b. Potential of multi-seasonal vegetationindices to predict rice yield from UAV multispectral observations. Precis. Agric. 25(3), 1235–1261.

[37] Kogan, F. N., Gitelson, A., Zakarin, E., Spivak, L., & Lebed, L.(2003). AVHRR-based spectral vegetation index for quantitative assessment of vegetation state and productivity. : Calibration and validation. Photogrammetric Engineering & Remote Sensing, 69(8), 899–906.

[38] Singh, U. K., Rani Kumudini, I., Rajmohan, V., Kumar, R., Kumar, S., Prakash, S., et al. (2016). Manual for drought management (p. 154). New Delhi: DAC&FW, MoA&FW, Government of India.

[39] Dubey, S.K., Gavli, Ashutosh, S.K. Diwakar, Neetu& Ray, S.S. (2017). Use of vegetation condition index for rice yieldforecasting. In 38th Asian conference on remote sensing—Spaceapplications: Touching human lives, ACRS 2017, October 2017, New Delhi.

[40] Jiang, D., X. Yang, N. Clinton, and N. Wang. 2004. "An Artificial Neural Network Model for EstimatingCrop Yields Using Remotely Sensed Information." International Journal of Remote Sensing 25 (9):1723–1732.

[41] Johnson, M. D., W. W. Hsieh, A. J. Cannon, A. Davidson, and B. Frédéric. 2016. "Crop Yield Forecastingon the Canadian Prairies by Remotely Sensed Vegetation Indices and Machine Learning Methods." Agricultural and Forest Meteorology 218-219: 74–84.

[42] Khaki, S., and L. Wang. 2019. "Crop Yield Prediction Using Deep Neural Networks." Frontiers in PlantScience 10: 621.

[43] Kim, N., H. Kyung-Ja, N.-W.Park, J. Cho, S. Hong, and Y.-W. Lee. 2019. "A Comparison between MajorArtificial Intelligence Models for Crop Yield Prediction: Case Study of the Midwestern UnitedStates, 2006–2015." ISPRS International Journal of Geo-Information 8: 240.