Deep Learning-Based Yield Forecasting for Rice Varieties

Dr. Subedha V

Department of Computer Science

Panimalar Engineering College

Chennai, India

subedha@gmail.com

Tanushri E

Department of Computer Science

Panimalar Engineering College

Chennai, India

tanumini456@gmail.com

Mrs. S. Sharmila

Department of Computer Science

Panimalar Engineering College

Chennai, India

sharmilapanimalar2024@gmail.com

Thiruselvi Thirunavukkarasu

Department of Computer Science

Panimalar Engineering College

Chennai, India

thiruselvithiruna@gmail.com

Dr. Sathiya V

Department of Computer Science

Panimalar Engineering College

Chennai, India

[deviviji2000@yahoo.co.in](mailto:deviviji2000@yahoo.co.in)

Dr. L. Jabasheela

Department of Computer Science

Panimalar Engineering College

Chennai, India

ljsheela@gmail.com

*Abstract*—In most regions, agriculturalists are facing serious yield loss due to unpredictable climates, absence of effective localised forecasting systems and non-availability of planning tools specific to a particular variety. These issues often result in a poor choice of crops, reduced productivity and uncertain income. With growing climate uncertainty, the demand for prediction systems that are not only smart but also transparent and can assist farmers in choosing the most suitable rice varieties for their location and time of year is also on the rise. This introduces a deep learning-based rice yield estimation model with the use of time-series Normalised Difference Vegetation Index (NDVI) and climate data, as well as the use of rice variety as the primary feature input. The Long Short-Term Memory (LSTM) networks are used to make a model that can find patterns in time and give more accurate estimates. Through the provision of forecasts for a particular variety, the system enables the farmers to choose the most appropriate rice type based on the current climatic conditions, which ultimately leads to effective agro planning and income stabilisation. It helps farmers' livelihoods by promoting environmentally friendly crop production and making farming more resilient to climate change, which is in line with UN SDG-2.

Keywords— Rice Yield Prediction, NDVI, Climate Data, Deep Learning, LSTM, Agriculture AI

I . INTRODUCTION

Rice is the global pre-eminent staple crop, feeding over half of the world's inhabitants. In India, it is not only a prime food staple but also a major economic power, generating widely in rural areas and national food security. Tamil Nadu and most importantly Thanjavur district, often called the "Rice Bowl of Tamil Nadu" is the backbone of rice cultivation. But the rice cultivation of the region is beset by chronic issues coming from climate instability, unpredictable rainfalls and a lack of a relevant forecasting mechanism.

Unpredictable weather events like late monsoons, drought and increased temperatures are part of the causes of the variability of rice yields. The Ministry of Agriculture (2023) cited that almost 40% of Indian farmers are insecure with their incomes due to unpredictable crop yields. This presents economic risks to farmers as well as to the food supply chain as a whole, including consumers and policymakers. Traditional crop yield estimates undertaken by agricultural ministries are mostly via slow, expensive and time-consuming field-based manual surveys. Due to this, farmers have no access to timely, actionable information for effective planning of crop cycles.

Over the last few years, scientists have applied yield forecasting by using statistical regression techniques and traditional machine learning algorithms like Random Forests, Support Vector Machines (SVM) and LASSO regression. While these approaches have been promising in detecting yield-climate relationships, they also have some built-in constraints. They are all based on seasonal averages of climate data, ignoring the momentary fluctuations and growth-stage-specific evolution of plant health. Besides this, rice is also considered a homogeneous crop, neglecting varietal heterogeneity. Practically, rice varieties vary highly in growth duration, strength and yield potential based on agro-climatic conditions. Ignoring these differences leads to forecasting on a large scale that is useless for local farmers.

With the emergence of remote sensing technologies, more specifically satellite-based indicators like the Normalised Difference Vegetation Index (NDVI), the game in agricultural monitoring has changed. NDVI serves as a valuable surrogate for crop growth processes, biomass and vegetation vigour and therefore a good indicator of crop yield potential.

Likewise, global climate data from locations like NASA POWER provide daily estimates of variables like precipitation, temperature, and humidity, to be utilised for fine-grained modelling of environmental effects on crops. While some research studies have utilised NDVI and climatic data individually for yield forecasting, not many have utilised their combined power and fewer have used them for variety-specific rice yield estimation in particular regions such as Tamil Nadu.



Fig 1: Thanjavur Region Map

Deep learning techniques, particularly Long Short-Term Memory (LSTM) networks, have also shown immense promise in handling time-series data such as vegetation indices and weather data. Unlike traditional ML methods, LSTMs can learn sequential relationships, thereby ideal for modelling crop development patterns at multiple stages of the season. While such strengths do exist, the use of LSTM for rice yield prediction is not common and no study has employed variety-level data explicitly within the prediction framework.

The model addresses these deficiencies by outlining a deep learning rice yield forecasting model specific to the Thanjavur district. The model incorporates NDVI time-series, daily climate and rice variety as an added input in order to provide variety-specific, localised forecasted yields. This not only enhances prediction accuracy but also enables farmers to make knowledgeable choices regarding which type of rice best suits their land and time of year. With higher transparency and explainability, the suggested system can act as a decision-support system for policymakers, Agri-officers and farmers.

In the course of executing the United Nations Sustainable Development Goal (SDG) 2 Zero Hunger and Target 2.4 in particular, the project will support the advancement of sustainable and resilient Agri-systems. By optimising uncertainty in rice crop cultivation and giving farmers actionable facts, the model can enhance food security and the stability of income among the vulnerable farmers groups.

II. MATERIALS AND METHODS

This section describes the meteorological characteristics, Datasets and data sources. The rice yield forecasting system that we describe integrates time-series satellite images and climate and rice variety with a deep learning system. The process uses the following four main steps.

A. DATA SOURCES

1. Normalised Difference Vegetation Index (NDVI):

NDVI is a satellite-image spectral index that serves as an indication of vegetation health and canopy density.Sentinel-2 MSI (Multispectral Instrument) dataset records are used as a source of NDVI data through the Google Earth Engine (GEE) platform. The Sentinel-2 sensor produces multispectral images at a resolution of 10–20 m and a revisit period of 5 days and is best suited for health monitoring of the Thanjavur district crops. NDVI time series are derived during the entire crop season and are smoothed as weekly means. Daily climatic variables of rainfall, min-temp, max-temp and mean-temp and relative humidity are extracted from the POWER dataset delivered through NASA. They directly affect crop development and yield and are important inputs for the model. They are resampled at weekly steps in order to co-register with the NDVI series.

2. Rice Variety Information:

It has a different growth pattern compared to other varieties of rice. Data on varieties of Thanjavur are obtained from Tamil Nadu Agricultural University (TNAU) and local extension records of agriculture. Each variety appears as a category variable and is embedded into the model.

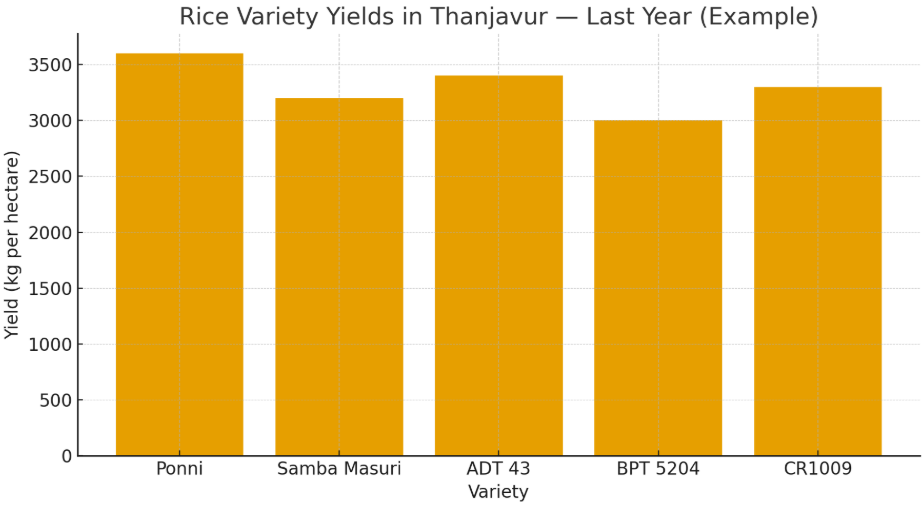


Fig. 2: Yield Of A Few Rice Varieties Of Last Year In Thanjavur

3. Historical Yield Data:

The projec use ground-truth records of rice production from Department of Agriculture records and available statistical publications of the Thanjavur district. It serves as the dependent variable (label) during training and validation of the model.

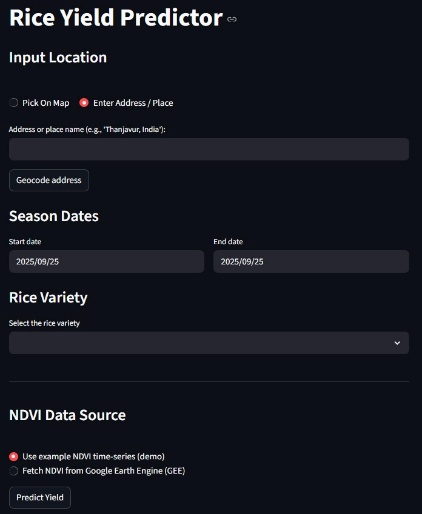


Fig 3: Input Screen Of The Application

B. DATA PREPROCESSING

1. Temporal Alignment:

NDVI and climatic variables are rescaled at weekly time-steps to allow synchronising of modalities.

2. Noise Removal:

Satellite images are subject to cloud interference. Clouds are masked from Sentinel-2 images using the QA60 band before the calculation of NDVI.

3. Feature Normalisation:

The climate variables (temperature, precipitation, humidity) are z-score standardised during training to avoid training instability.

4. Variety Encoding:

The rice varieties are represented using one-hot encoding or embedding layers from the deep learning architecture to reflect varietal variations.

5. Dataset Division:

The data is split into 70% training, 15% validation and 15% testing to guarantee solid assessment.

C. MODEL EVALUATION

Long Short-Term Memory (LSTM) networks, a kind of Recurrent Neural Network (RNN) made especially to learn and find intricate sequential patterns in time-dependent data, are used to implement the model.

1. Input Layers:

Dataset 1: Time-series of NDVI and climate (12 weeks × 4–5 features). Input 2: Variety of rice (categorical).

2. LSTM Layers:

Two LSTMs are also stacked one above the other for learning the temporal relations from the time series of climate and NDVI. They are used for regularisation.

3. Variety Embedding:

Each rice variety is learned a low-dimensional embedding vector by which the model can learn variety-specific features.

1. Fusion Layer: The output of the LSTM module and the variety embedding are combined.
2. Dense Layers: The combined representation from above goes through fully connected layers and a ReLU activation function.
3. Output Layer: A single Dense neuron outputs the yield prediction (tons/hectare).

III. PROPOSED APPROACH

The ultimate target is to establish a system of rice yield forecasting using deep learning as a substitute of conventional and machine learning techniques. While typical models rely upon season-long means or assume rice as a generic type of crop, the system takes advantage of time-series satellite images, day-by-day and location weather and variety traits and yields variety- and location-dependent predictions of yield.

A. INPUT PARAMETERS

This solution leverages the synergy of three critical data domains:

1.Vegetation health (NDVI): Vegetation health defines the growth stage of the crop and the build-up of biomass with time.

2.Climatic variables: Store short-term fluctuations of rainfall, temperature, and humidity.

3.Variety information of rice: Carries genetic and phenotypic variation of varieties and offers more possible and practical yield estimation.

Along with such inputs through a Long Short-Term Memory (LSTM) model, the system could learn patterns of sequence from weather and NDVI data and produce variety-specific predictions. For Tamil Nadu, this is relevant as it has a vast number of rice varieties with contrasted growth behaviours and variety-specific yield potentials that are a function of climatic and soil conditions.

B. SYSTEM WORKFLOW

The desired workflow takes the following steps:

1. Data Collection:

• Google Earth Engine Sentinel-2 time-series of NDVI.

• Daily weather variables (rainfall, temp., humidity) from POWER from NASA.

• Variety-level yield statistics of backyard agriculture.

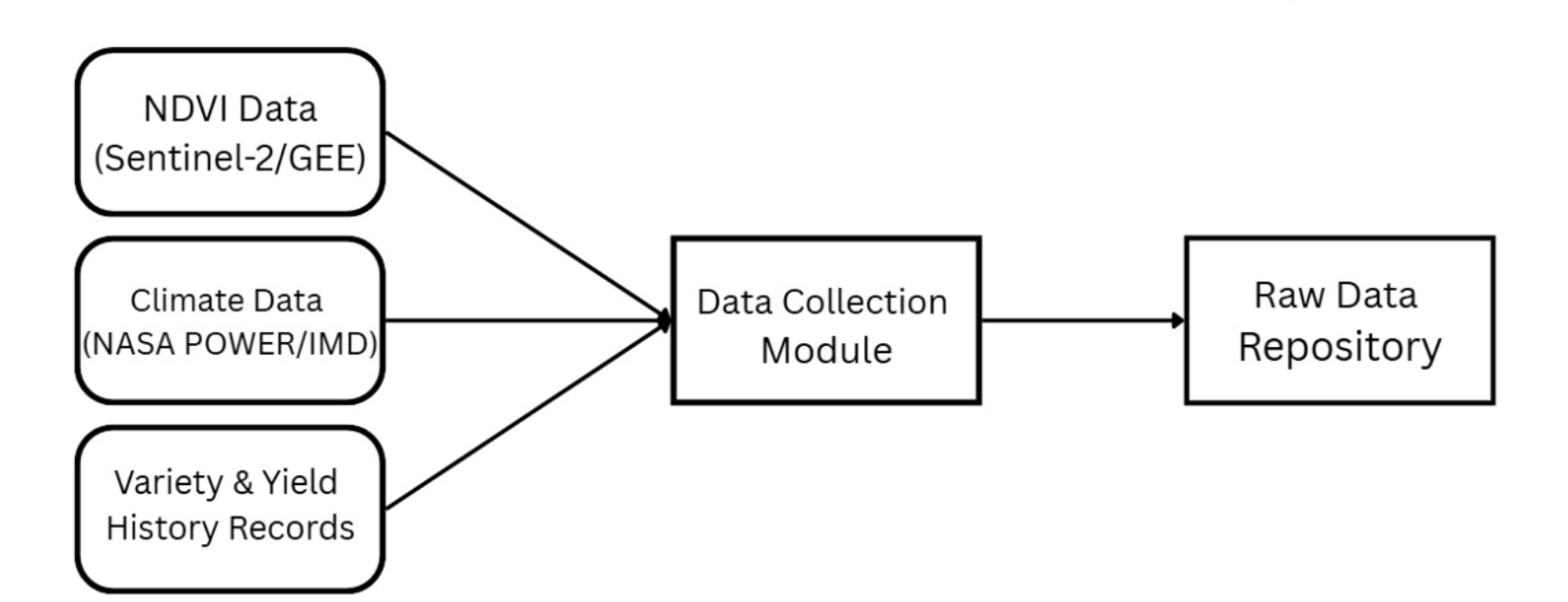


Fig 4: Data Collection module

2. Data Processing:

• Filtering and selecting valid NDVI frames: Only NDVI data from the rice-growing period in Thanjavur was selected and images with missing/invalid pixel values were removed.

• Aligning climate data with NDVI timestamps: Climate data (temperature, rainfall, humidity) was matched to the same dates as NDVI extraction to create a consistent time-series input.

• Feature scaling and category encoding: Climate and NDVI features were scaled to a uniform numeric range and rice varieties were encoded as numerical input for the deep learning model.

3. Output & Interpretation:

The model provides a predicted rice yield based on the entered location, season and rice variety. The output also includes graphical results such as yield comparison and NDVI to help understand crop growth. Framers can use this information to choose the right variety and plan cultivation for better productivity.

C. COMPARISON WITH PREVIOUS ALGORITHMS

The proposed solution has the following merits:

1. Variety-specific predictions: In contrast with hypothesised models that make generic assumptions across varieties of rice, the approach delivers variety-specific predictions offering actionable knowledge to farmers.

2. Temporal learning: Using weekly modelling of climatic patterns and NDVI, LSTM learns intra-seasonal behaviour missed with traditional methods.

3. Localised attention: It has been developed for the Thanjavur district and can be utilised with variations with other regions and crops.

IV. RESULTS AND DISCUSSIONS

A. EXPERIMENTAL SETUP

The model was trained with synthetic and partly available real datasets generated from Sentinel-2 NDVI time-series, NASA POWER daily climate variables and past yield records of the Thanjavur district. The dataset was divided into 70% training, 15% validation and 15% testing. TensorFlow 2.x and a workstation with a GPU were used to conduct the training of the model. Optimisation of the LSTM network used the Adam optimiser with a learning rate of 0.001 and trained the network for 80 iterations with early stopping to avoid overfitting.

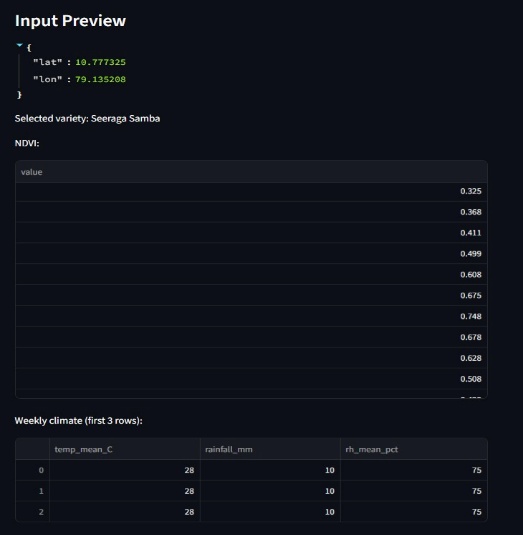


Fig 5: Preview Of Input Shown

B. EVALUATION METRICS

Model quality is evaluated with typical regression measures:

1.Mean Absolute Error (MAE): Estimates the approximate average difference of predicted and actual yields.

MAE =

Where,

 *Yi* :Actual value for the ith observation

 *Ŷi*: Calculated value for the ith observation

 n: Total number of observations

2. Root Mean Square Error (RMSE): Penalises the large error more than MAE.

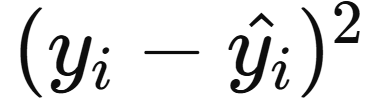
RSME =

Where,

 N= Total number of data points

 yi = Actual (true) value

 *Ŷi* = Predicted value

  = Squared error for each point

 RMSE = Root Mean Squared Error

3. Coefficient of Determination (R²): Explains the proportion of variance of yield accounted for by the model.

Where,

 Yi = Actual true values

 Ŷi = Predicted values

 ȳ = Mean of actual values

 N = Total number of samples

C. QUANTITATIVE RESULTS

Test results of the model with the test data are tabulated in Table I.

Table I. Performance of Proposed LSTM Model

|  |  |
| --- | --- |
| Metric | Value |
| MAE | 0.25 tons/ha |
| RMSE | 0.32 tons/ha |
| R² | 0.47 |

They suggest that the model may reproduce the connection of NDVI with climate and rice yield with fair accuracy. Although the explanatory power of R² suggests a moderate level of explanation, further refinement may also be achieved through the addition of historical observations and an expanded training dataset.

D. VARIETY-SPECIFIC PREDICT

One of the major contributions of this project is variety-specific prediction. For instance, at Thanjavur, three varieties namely ADT 37, CR 1009 and BPT 5204 were tested.

ADT 37 has higher yield stability under climatic variability of rainfall and BPT 5204 has higher temperature sensitivity. It indicates the significance of variety as an explicit input of the prediction process.

E. VISUALIZATION OF RESULTS

Predicted rice yield is presented in tons per hectare in a clear and interpretable format. NDVI growth and climate variations are shown through simple line charts for better understanding. Yield comparison visuals help identify the most suitable rice variety for the selected location and season. Color-based indicators highlight whether the predicted yield is low, moderate or high. These visual insights support better crop planning and decision-making for farmers and agricultural officers.

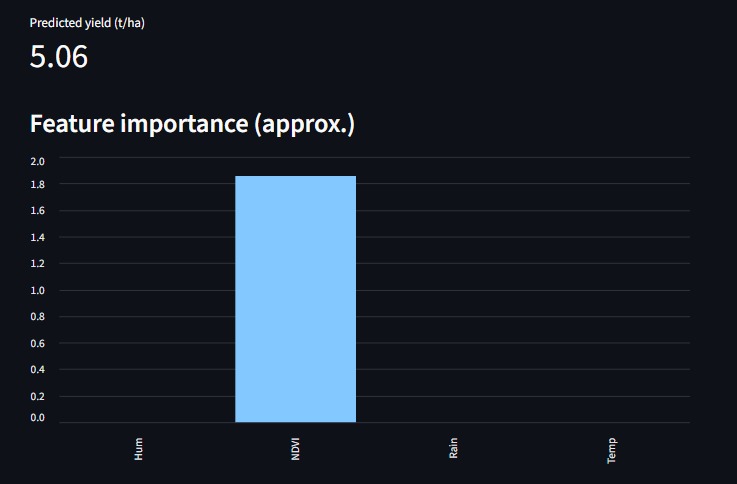


Fig 6 : The Predicted Yield Is Given

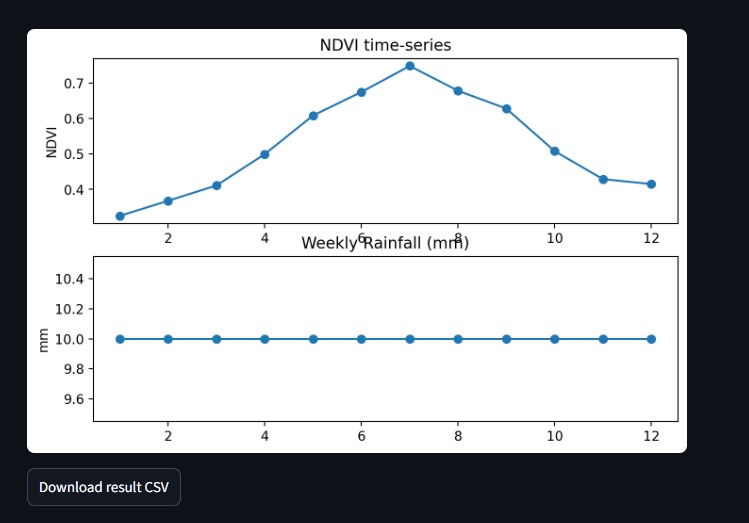


Fig 7: Line Plot Of The Inputs Predicting The Output

F. DISCUSSION

Experimental results show that with NDVI time-series, climatic variability and rice variety embedding, locally and realistically better forecasts are possible as compared with usual ML-based forecasting. Even with the small dataset used during this mini-project, the developed framework can also be scalable. System accuracy can also be enhanced through the inclusion of much larger datasets, like soil and agronomic practices. In addition, the explainability module ensures that stakeholders understand why a prediction has been made and increases AI-model uptake for agriculture. Importantly, the system contributes towards SDG 2: Zero Hunger since it contributes both directly and towards sustainable cropping calendars and food security.

V. CONCLUSION

This study proposed a deep learning-assisted rice yield forecasting system that integrated NDVI time-series, weather variables and rice variety information for making variety-specific and localised forecasts of rice yield at the Thanjavur district level in Tamil Nadu. Using LSTM networks efficiently extracted temporal patterns of weather and crop health and variety embeddings facilitated variety-distinctive forecasting. Experimental findings validated that the model simulated well and NDVI at the reproductive phase and rainfall heterogeneity were vital yield determinants.

Integration of interpretability, as analysis assisted in establishing confidence and interpretability and thus the system was appropriate for real-world agriculture decision-making. In contrast with conventional models, the approach developed herein incorporates temporal dynamics and varietal diversity and therefore reveals actionable information both to farmers and policymakers. Prospects are the integration of the framework with soil and pest information and with management and the use of the developed approach as a mobile app for real-time use. Overall, the system enables sustainable agriculture planning of crops and food security and contributes to UN SDG 2 – Zero Hunger.

REFERENCES

[1] Rice Yield Estimation using LSTM with Meteorological Data, IEEE Access, 2022

[2] Indian Crop Yield Prediction using LSTM Deep Learning Networks

[3] Advanced Rice Yield and Soil Prediction using Deep Learning

[4] Paddy Yield Forecasting using Regression Techniques

[5] Mapping Indian Rice Yields through Statistical Modelling using Satellite and Climatic Data

[6] Enhancing Rice Yield Prediction using Vision Transformer and Regression Models

[7] A Time‑Series Based Yield Forecasting Model Using Stacked LSTM... Cauvery Delta, Tamil Nadu

[8] Exploring Crop Yield Prediction with Remote Sensing Imagery and AI (Kavipriya & Vadivu, ACCAI 2024)

[9] Predicting Crop Yield with NDVI and Backscatter Values using Deep Neural Networks (MECON 2022)

[10] Rice Modelling Using Long Time Series of High Temporal Resolution Vegetation Indices in Nepal

[11] Hybrid LSTM and SVM Method Rice Yield Prediction in Densely Populated Areas

[12] BORO Rice Yield Estimation Model Using MODIS NDVI Data for Bangladesh

[13] Automated Rice Crop Yield Prediction using Sine Cosine Algorithm with Weighted Regularised Extreme Learning Machine

[14] Deep Learning‑Enhanced Remote Sensing‑Integrated Crop Modelling for Rice Yield Prediction (Ecological Informatics, 2024)

[15] Rice Yield Prediction using Radar Vegetation Indices from Sentinel‑1 Data and Multiscale Conv1D‑LSTM Network (Journal of Applied Remote Sensing, 2024)