Predictive Modelling of Agricultural Factors to Maximize Crop Yield

Anvesha Nayak, Pramathi Vummadi, Apoorva Raj, Nasam Saimani and Suresh Jamadagni
Department of Computer Science and Engineering, PES University, 100 Feet Ring Road, Bengaluru, India

Keywords: Rice, Paddy, Crop Yield Maximization, Evapotranspiration Forecasting, Surface Runoff Prediction, Irrigation Management, Pest and Disease Detection.

Abstract: Crop yield prediction and factor analysis are methods through which technology can be utilized to improve the quality of current Agricultural practices. This study focuses on improving crop yields based on different factors and ascertaining how climate change affects these factors and their prediction. The aim is to create a tool for farmers to practice precision agriculture and to be made aware of what controllable factors can lead to better yield. The study proposes a three-step methodology for this process. First, we will analyse past years' data and also take into consideration the impact of climate change to know how this relates to these variables as well as crop yield. Secondly, we suggest some spatial variable management practices that could improve the overall agricultural output. Along with that, preventative measures to ensure crop safety are also suggested. Regular updates on these spatial variables will play an important role in helping the farmer make key decisions during the life cycle of the crop. Finally, in the third step of this process, we aim to perform anomaly analysis on pests, weeds, diseases, and climatic anomalies, and suggest relevant countermeasures to the farmer.

1 INTRODUCTION

Agriculture is a fundamental part of human survival and also one of the key pillars of the world economy. However, agricultural economic growth faces an unprecedented challenge in the form of climate change. We take a case study of paddy grown predominantly in Asia, specifically India. Rice is the staple food for more than half the population of India and one of the critical crops for the nation's economy. However, rice cultivation depends heavily on climate variables such as temperature, precipitation, etc. Worsening conditions such as abrupt rises in temperature and irregular rainfall make optimizing crop yield harder. Anomalies such as floods, pests, and diseases also add a layer of unpredictability. This paper aims to use various datasets related to crop growth, agricultural factors, weather etc., to study and understand the relationships between agricultural productivity and climate variables in India. We also segregate the entire crop growth cycle into the different stages and aim to provide suggestions for each stage based on the output of the machine learning model. Image analysis models are used to identify anomalies and diseases while providing suggestions for tackling them. This paper aims to identify effective methods to improve the resilience and sustainability of farming practices through early detection and timely prevention.

2 RELATED WORK

Crop yield prediction particularly concerning the climate impacts on wheat production highly features ensemble models and random forest regression. Ensemble models, particularly Random Forest Regressor (RFR), have been gaining increasing usage in crop yield prediction since they can manage large complex data and can pick up the nonlinear relationship between climate variables and yield. The ensemble model increases the accuracy and minimizes bias through the combination of various machine learning insights, thus becoming robust for high-dimensional, nonlinear climate data (Satpathi et al., 2023). RFR is highly effective for regional-scale predictions and feature selection, which minimizes the risk of overfitting (Breiman, 2001; Pang et al., 2022). Combining RFR with other models, such as artificial neural networks or boosted trees, improves the reliability of predictions under various climate

Nayak, A., Vummadi, P., Raj, A., Saimani, N. and Jamadagni, S.
Predictive Modelling of Agricultural Factors to Maximize Crop Yield.
DOI: 10.5220/0013369400003890

In *Proceedings of the 17th International Conference on Agents and Artificial Intelligence (ICAART 2025) - Volume 3*, pages 1360-1367
ISBN: 978-989-758-737-5; ISSN: 2184-433X

scenarios. These ensemble-based approaches are critical in developing climate-resilient agricultural solutions and managing climate-yield complexities (Iqbal et al., 2024).

An important aspect of optimization of agricultural processes is irrigation management. Irrigation management in turn is heavily dependent on weather prediction models. The citation (Teixeira et al., 2024) proposed a combined approach that integrates Long Short-Term Memory (LSTM) neural networks with Genetic Algorithms (GA) to forecast short- and medium-term weather conditions in the Douro area.

Another weather component that impacts irrigation management is Evapotranspiration (ETo). A number of machine learning (ML) and deep Learning (DL) methods have been proposed to forecast ETo for improving agricultural productivity. One paper (Kadkhodazadeh et al., 2022) proposed a novel approach to predict ETo under climatic changes using a combination of historical ETo data and meteorological data through use of regression models. The study (Granata & Di Nunno, 2021) explored use of ensemble models on two types of recurrent neural networks (RNNs), Long Short-Term Memory (LSTM) and Nonlinear Autoregressive Network with Exogenous Inputs (NARX) and did a comparative study on effectiveness in different climatic conditions (subtropical and semi-arid). The findings indicated that the LSTM models outperformed the NARX models in the subtropical climate, whereas the inverse was true for the semi-arid climate. Another study (Yin et al., 2020) addressed the issue of scarce meteorological data for ETo forecasting through the creation of a hybrid Bi-directional Long Short-Term Memory (Bi-LSTM) model. The study highlights the importance of considering local climate conditions and data availability when selecting the appropriate modelling approach.

Prediction of another component, surface runoff, helps optimize irrigation scheduling and in reducing water waste due to overirrigation. One paper (Gauch et al., 2021) presented a novel MTS-LSTM model for forecasting rainfall-runoff. This method was developed to predict extreme flooding events at various time scales, addressing the challenge of accurately forecasting daily and short-term incidents that occur more frequently than regular daily forecasts.

For anomaly detection, recent studies proposed using CNNs to label rice pests and diseases strategically by using various advanced architectures and methods to achieve high accuracy in detecting and classifying diseases. Key algorithms cover Visual Geometry Group(VGG), ResNet, You Only Look Once Version 3 (YOLOv3), RestNETV2 101, YOLOv5, Inception-V3, DenseNet, AlexNet, GoogLeNet, Faster R–CNN, K-most familiar neighbours, Support vector tool etc. The performance evaluations related to each model showed that ResNet has better accuracy and gave efficient results in differentiating affected and healthy image patterns with a fully connected layer using Softmax Function and cross-validation techniques to improve the potential results.

3 RESEARCH METHODS

This section includes the research methods used in this paper, including the Study Area, Datasets Used and Methodologies.

3.1 Study Area

Telangana, a state present in the southern part of India, spans from latitudes 13’N and 19’N and longitudes 78’E to 81’E. Telangana has a varied agricultural terrain, backed by a tropical climate with warm summers, a moderate monsoon period, and gentle winters. Farming constitutes the bulk of Telangana's economy with major crops including rice, cotton, maize, sorghum, groundnut and soybean. Rice is the major staple crop sown in large areas both in the Kharif and Rabi seasons. Climate and geographical conditions have made Telangana an important region for agricultural innovation and development.

Figure 1: Position of Telangana in India (Minhaz, 2023).

3.2 Datasets Used

Data in Climate Resilient Agriculture (DiCRA): This is a platform with a vast dataset containing multiple factors contributing to agriculture and key indicators in climate resilience in agriculture. Environmental features including land surface temperature (LST), normalized difference vegetation index (NDVI), temperature, and precipitation as well as Socio-economic features such as cropland, crop intensity, etc were used.

Directorate of Economics and Statistics, Ministry of Agriculture and Farmers Welfare, Government of India (DESAgri): It contains the historical yearly yield of all the crops harvested in India. The data is organized by distinct and the metrics used are Tonnes/Hectare.

Open Data Telangana: This contains all the public datasets of Telangana following the open data policy. The datasets used in this research contain monthly information regarding temperature (℃), wind speed (Kmph), humidity (%) and precipitation (mm) with district and date as the row identifier.

Bhuvan is a web-based platform from the Indian Space Research Organisation (ISRO) that provides access to satellite remote sensing data for public use. We use this to get the Evapotranspiration and Surface Runoff data.

Evapotranspiration (mm): The data for evapotranspiration was collected by the National Remote Sensing Centre (NRSC), one of the agencies under the National Hydrology Project (NHP). The evapotranspiration was calculated using the Modified Preistley Taylor (PT) method. (Ai & Yang, 2016) (Priestley & Taylor, 1972) (Parlange & Katul, 1992). There were a few missing data points due to technical errors and weather phenomena such as clouds. Considering the limited meteorological data availability, using the existing average temperature data, crop coefficient (Kc) for paddy at different growth stages and average day length month-wise were used in the Blaney-Criddle (BC) (French, 1950) equation to fill the missing values. The Telangana Weather data was combined with the evapotranspiration data to ensure any dependencies are captured.

Surface Runoff (mm): The data for surface runoff was calculated using the Variable infiltration Capacity (VIC) model, a semi-distributed, physically based hydrological model, adopted to model water balance components.

Maplogs day length Dataset: This dataset gives the day length for the different areas in the world by using Latitude and Longitude as the key.

Rice Pest Dataset: This rice pest dataset, a subset of the IP102 dataset, includes of images categorized into 12 distinct classes that are specific for the purpose of detecting rice pests. The classes include rice leaf roller (605 images), rice leaf caterpillar (475 images), paddy stem maggot (325 images), Asiatic rice borer (745 images), yellow rice borer (455 images), rice gall midge (791 images), brown plant hopper (290 images), rice stem fly (1110 images), rice water weevil (1194 images), rice leaf hopper (686 images), rice shell pest (480 images), and thrips (580 images). These were then augmented with various techniques including vertical flipping, horizontal flipping, multiplication and linear contrast adjustment to enhance the dataset.

Rice Leaf Diseases Detection Dataset: The dataset consists of images showcasing rice leaves in various conditions, including both healthy and unhealthy states. This includes healthy rice leaves (1,085 images), bacterial leaf blight (1,197 images), brown spot (1,546 images), leaf blast (1,748 images), leaf scald (1,332 images), narrow brown leaf spot (954 images), neck blast (1,000 images), rice hispa (1,299 images), and sheath blight (1,629 images). Moreover, the dataset underwent an augmentation procedure that included the use of different techniques like rotation, scaling, flipping etc., to create a larger and more diverse collection of images.

Rice Insect Pest, Disease Crop Weather Calendar: Developed by Telangana State Agricultural University for Nizamabad district is an all-inclusive guide /tool that informs the occurrence of insect-pests and diseases at the district level on a stage-wise basis to take up control measures in time by thus enabling reduction of losses in yield. Information regarding the crop, its stages and week-to-week weather information during the crop season is essential to forewarn the farmers on occurrence/prevalence and recommend management measures against insects, pests and diseases. Farm operations planned in coordination with weather information would likely curtail the cost of inputs as well as other field operations. Rice-insect pest/disease-weather calendars contain the favourable conditions required for the occurrence of key insect pests or diseases and susceptible crop phenological stages.

3.3 Architecture and Methodologies

The process of creating a predictive analysis model consisted of the following steps:

i. Yield Prediction: The methodology proposed to predict Kharif and Rabi crop yields begins by merging historical crop yield data with the corresponding weather data based on district and year, filling in any missing weather values with column mean. Optimal ranges of monthly temperature, humidity, and rainfall are derived from growth-stage data for each month to act as benchmarks.

The preprocessing of the forecasted weather data creates smoothed features, for instance, 30-day averages or sums of temperature, rainfall and humidity. This is then weighted upon in order to determine how close this data is to falling with its monthly optimal ranges using a weather scoring function. The score then combines temperature, rainfall, and humidity into a single, weighted score that reflects the overall favourability for crop growth, adjusted for Kharif and Rabi seasons.

Two models of Random Forests are trained based on historical weather and yield data to find out the Kharif and Rabi yields. Yields are predicted in terms of future weather conditions and adjusted with the calculated weather score that gives an estimate of what is the expected yield based upon favourable or unfavourable conditions. The performance of the model is validated through R^2 scores in making sure predictions are accurate and geared towards district-level conditions.

ii. Irrigation Calculation: Our approach to calculating net irrigation required involves three main models- for the prediction of Precipitation, Evapotranspiration and Surface Runoff respectively. Due to limited access to meteorological data, historical data and weather data were used. Hence Long Short-Term Models (LSTM) were used to capture the temporal data instead of standard regression models with multiple variables used in previous studies. All the models were trained district-wise in order to capture the data more consistently.

Weather Prediction: The Telangana Weather dataset was used to predict weather features like Precipitation (mm), Minimum and Maximum Temperature (°C), Minimum and Maximum Wind Speed (Kmph), Minimum and Maximum Humidity (%). To ensure consistency in scaling, these features were normalized by subtracting the mean and dividing it by the standard deviation. The data was divided into specific time steps and an 80-20 split was used for the train-test division of data. The architecture used consists of two LSTM layers and mean squared error was used as the loss function.

Evapotranspiration Prediction: Evapotranspiration plays a vital role in Irrigation Management. Historical data calculated by the Modified Preistley Taylor (PT) (Ai & Yang, 2016) method as recorded by the Bhuvan dataset and the Blaney-Criddle (BC) equation (French, 1950) (Choudhary, 2018) were used to forecast the missing Evapotranspiration values.

The Blaney-Criddle equation is as follows:

$$ET = (0.0173\ Ta - 0.314)\ Kc * Ta(D/4465.6) * 25.4 \quad (1)$$

Where, Ta is mean air temperature (in °F), Kc is crop coefficient and D is day length (in hours).

Day length was taken from the Maplogs day length Dataset and the crop coefficients were considered as per growth stage of the crop (Mote et al., 2018).

Feature engineering led us to add four new features, 1-day lag, 2-day lag, 3-day average and 7-day average, to the data in order to capture the real-time changes in the weather patterns. The model architecture consists of two LSTM layers with dropout layers to prevent overfitting. The mean squared error was used as the loss function.

Surface Runoff Prediction: For the prediction of surface runoff, we considered different models including Linear Regression, Decision Trees, Random Forest Regression, Support Vector Regression (SVR), ensemble models like Gradient Boosting Regression, XGBoost and LightGBM. The data was split into training, validation and testing subsets in a 60-20-20 ratio for each district separately. We incorporated a custom evaluation function to identify the best performing model. The custom evaluation function calculated the mean and standard deviation for the metrics R^2. The best model was identified by assessing which model had the highest mean R^2 while having the least standard deviation. This was identified to be the Decision Tree model as shown in Table 3.

The decision tree was trained on the various weather parameters such as precipitation, humidity, temperature etc. which were given as the input along with historical surface runoff data. A 5-fold cross validation approach, the GridSearchCV technique from the scikit-learn python library, was used to finetune the hyperparameters to find the best bias-variance balance. These hyperparameters were tuned

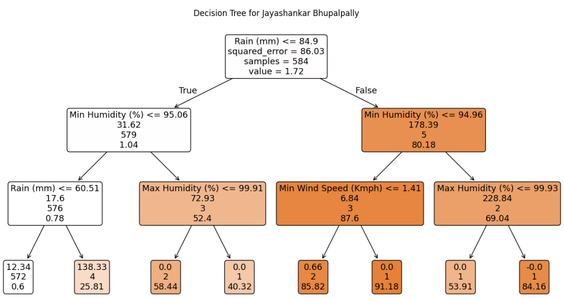

Figure 2: Surface Runoff Decision Tree for Jayashankar Bhupalpally district.

to maximise the R^2 and capture the maximum variance in the data. The best hyperparameters were max_dept=3, min_samples_leaf=1 and min_samples_split=2.

To calculate the net irrigation required the following formula was used:

$$NI = I - (P - Sr - ETa) \quad (2)$$

Where, NI is Net Irrigation (mm), I is Ideal Irrigation (mm), P is Precipitation (mm), Sr is Surface Runoff (mm) and ETa is Actual Evapotranspiration (mm).

Net Irrigation was calculated as irrigation required per hectare of land. So based on the farmers' input of farm area the Irrigation required in total can be calculated.

iii. Preventative Measures Recommendation: Using the Rice Insect Pest, Disease Crop Weather Calendar based on the standard week in the year, crop season and the growth stage of the plant, the possible pest and disease attacks are indicated. The probability of them occurring as well as preventative methods for the pests and diseases are given in a tabular format.

iv. Pest Detection: The images in the dataset were resized to be 244x244 pixels to match our base model ResNet18's input size. It was then converted to tensors and normalized using means and standard deviations. PyTorch's random split function was used to slit the dataset into 80-0 train test datasets. Using transfer learning the base ResNet18 model was modified by changing the fully connected layer to fit the number of classes in the dataset. Cross entropy loss was used to train the model and the Adam optimiser was used at a learning rate of 0.001. A progress tracker (tqdm) was used to monitor the training and validation losses. Early stopping was employed to ensure the model doesn't overfit. The model was evaluated against the training, validation, and test subsets. Accuracy measures the percentage of correct predictions compared with true labels. The model was tested against a final test subset to ensure generalisation.

v. Plant Disease Detection: ResNet18 is a commonly used base neural network for many image-based machine learning applications where extraction of specific and meaningful features is required. It allows for easy learning throughout the network as well as overcoming the degradation because of the vanishing gradient problem. Therefore, this model was used in the disease detection component of our research. We used a modified and augmented dataset to improve the quality and classification capability of the model. For the training and testing stages we utilized the Pytorch Lightning method and the Tensorflow preprocess function.

The ResNet18 model consists of 18 layers from the input to the output layer. Based on the parameter values it was observed that the ResNet18 model outperformed the other networks in identifying rice leaf disease. The model achieved a test accuracy of 77% and test loss of 1.32 when tested on the dataset. These metrics indicate a fairly efficient model, although there could be opportunities for additional optimization to enhance accuracy and decrease loss.

4 EXPERIMENTAL RESULTS

4.1 Crop Yield

The Random Forest model demonstrated strong performance in predicting crop yields for both Kharif and Rabi seasons.

Overall, the Random Forest model proved to be a reliable tool for predicting crop yields in both Kharif and Rabi seasons. While the model performed slightly better for Kharif, it still provided accurate predictions for Rabi as well. The results are indicated in Table 1.

Table 1: Crop Yield Maximisation Performance Metrics.

Season	Performance Metric			
	R^2	RMSE	MAE	Average Difference (Predicted – Actual)
Kharif	0.7525	0.2192	0.1678	0.5335
Rabi	0.7206	0.2319	0.1786	0.5709

4.2 Irrigation Calculation

Irrigation Calculation was split into predictive analysis of three components- weather prediction, evapotranspiration forecasting and surface runoff prediction.

As the evapotranspiration data is combined with the weather predictive data, it gives a good indication of the performance of the irrigation calculation model. The order of measurement is relatively negligible when compared to evapotranspiration and hence, not as impactful on the irrigation model.

For Evapotranspiration, prediction was carried out over the various districts with separate models for each one. The models were trained for each district to ensure that the minute changes in weather and other conditions are captured as accurately as possible. The mean and standard deviation (SD) of the performance metrics R^2, Accuracy Percentage and Root Mean Square Error (RMSE) is presented in Table 2.

The outcome value shows that the model has fairly taken into consideration the different aspects relating to the prediction of outcomes - the R^2 values range around 0.79 to 0.98 for the different regions and depict a high value of variance explanation in most cases. The mean accuracy level of 81% showcases the effectiveness of the model in all the geographical locations.

Table 2: Evapotranspiration Performance Metrics.

Statistical Measure	Performance Metric		
	R^2	Accuracy	RMSE
Mean	0.9035	81.0616	0.2778
SD	0.0448	5.0134	0.0534

In the case of Surface Runoff, our evaluation focused on the R^2 and RMSE for each model for all the districts. We determined the best models by finding the mean and standard deviation for each model across the districts.

The results are detailed in Table 3. It was found that the Decision tree model (DT) had the highest mean R^2 value and maintained low RMSE values. The highest R^2 was achieved for the district Khammam with the value of 0.8404.

Table 3: Surface Runoff Models Performance Metrics.

Model	Performance Metric			
	R^2		RMSE	
	Mean	SD	Mean	SD
Decision Trees	0.7360	0.2056	7.7135	5.7883
Gradient Boosting Regression	0.5711	0.1457	12.9112	8.3090
LightGBM	0.6867	0.1453	18.9772	14.9264
Linear Regression	0.7129	0.1110	26.3868	14.9429
Random Forest Regression	0.5430	0.3390	27.7005	29.6897
Support Vector Regression	0.2888	0.0778	4.3581	5.5488
XGBoost	0.6711	0.1910	14.0355	17.7094

4.3 Anomaly Detection

For anomaly detection the models used were classification models and hence, the metrics used were modified as needed. Accuracy along with loss were used as criteria to measure how well the model is performing.

Our best performing model on the test dataset for pest detection showed good generalization as indicated by a test accuracy of 98.63% and a low test loss of 0.0437. This implies that the model correctly classified unseen pest images with precision and minimum errors, which indicates that the model has identified the most relevant features without overfitting to the training data. The close agreement of test accuracy with training performance confirms the robustness of the model in pest detection on new data, thereby reinforcing its appropriateness for practical deployment in the identification of pests infesting a crop of rice. The detailed results are shown in Table 6. The performance metrics are shown in Table 4.

Table 4: Pest Detection Performance Metrics.

Performance Metric	Value
Accuracy	0.9692
Precision	0.9703
Recall	0.9692
F1 Score	0.9690

For disease detection, the model achieved a test accuracy of 83.96% when tested on the dataset. These metrics indicate a fairly efficient model, although there could be opportunities for additional optimization to enhance accuracy and decrease loss. The detailed metrics are given in Table 5.

Table 5: Disease Detection Performance Metrics.

Performance Metric (Rice Disease)	Value
Accuracy	0.8396
Precision	0.8471
Recall	0.8396
F1 Score	0.8414

Table 6: Pest Detection Results.

Pest	TPR	FPR
Rice Leaf Roller	97.69%	0.07%
Rice Leaf Caterpillar	98.53%	0.28%
Paddy Stem Maggot	97.54%	0.43%
Asiatic Rice Borer	97.58%	0.17%
Yellow Rice Borer	98.02%	0.04%
Rice Gall Midge	99.75%	0.48%
Brown Plant Hopper	89.66%	0.11%
Rice Stem Fly	99.46%	1.24%
Rice Water Weevil	98.32%	0.43%
Rice Leaf Hopper	85.01%	0.01%
Rice Shell Pest	97.29%	0.06%
Thrips	97.76%	0.14%

Table 7: Disease Detection Results.

Disease	TPR	FPR
Bacterial Leaf Blight	89.36%	0.13%
Brown Spot	80.26%	3.16%
Leaf Blast	70.44%	5.23%
Leaf Scald	95.34%	2.57%
Narrow Brown Spot	79.84%	2.27%
Neck Blast	97.20%	0.10%
Rice Hispa	87.56%	1.16%
Sheath Blight	75.69%	0.83%
Tungro	88.39%	0.39%
Healthy	77.24%	2.21%

5 DISCUSSIONS AND CONCLUSIONS

In crop yield maximisation, the ensemble model demonstrated strength in both generalizability and robustness, by minimizing overfitting through a combination of multiple models. While the proposed Random Forest model achieves acceptable error margins for Kharif and Rabi models respectively, employing similar ensemble techniques as those used for the benchmark could improve the accuracy even more. By adding an ensemble technique, the proposed model could potentially achieve a lower RMSE and increase resilience to variance, potentially outperforming the benchmark ensemble model in terms of robustness and precision for seasonal and district-level applications. Thus, while the benchmark model excels in overall error reduction, the proposed model's tailored approach provides certain practical benefits in forecasting agricultural yield for localized and season-specific contexts.

For irrigation calculation, the values for surface runoff were observed to be several degrees lower than that of evapotranspiration and precipitation. So, the effectiveness of our study is indicated by the performance of our weather and specifically our evapotranspiration model.

We have also taken into account the lack of data availability in certain areas by making our model less dependent on meteorological data, which is harder to capture. By displaying similar or better performance than benchmark models, despite having limited data, our model is more robust in areas where data is not widely available. This efficiency along with the high confidence score, makes the model a good fit for broader applications.

For pest detection, the model used, ResNet 18 can process large amounts of data using data augmentation and transfer learning, resulting in higher accuracy using various classifiers. Unlike traditional methods, deep learning models can automatically extract relevant features from images improving accuracy, CNNs are capable of multi-spectral and hyperspectral processing, enabling the detection of disease symptoms. Therefore, they can detect rice diseases accurately and timely. The results are shown in Table 6 and the performance metrics are detailed in Table 4.

In case of disease detection, the algorithms use different data to identify and classify various diseases, which is important for timely intervention and disease control. Machine learning models can be adapted to different locations and environments to address emerging field conditions. Although this has been the outcome, there are still some barriers in the general understanding of the data and the different machine learning models. Therefore, future study should be developed more accurately and more efficient models should be deployed while improving data collection and documentation. Hence, using machine learning, agricultural sustainability can be improved and utilization of pesticides, fertilizers, and other materials easier and also reduces the environmental impact besides enhancing efficiency. The detailed results for different diseases are shown in Table 7.

The different components of this study come together to create a comprehensive tool for farmers to use to ensure that their crops are climate resilient, on a day-to-day basis. By using predictive analysis, the effects of climate change can be captured effectively to provide real-time insights into maximizing crop yield. Future work includes expanding our models to other crops as well as to other areas in the country and beyond. With more data availability a standard, but area-specific, model can be developed for climate resilient agriculture in India.

ACKNOWLEDGMENT

At the outset, we would like to thank our Professors at PES University for their unwavering guidance in ensuring the quality of our research. We are also extremely indebted to our friends and family for their help and support.

REFERENCES

Satpathi, A., Setiya, P., Das, B., Nain, A. S., Jha, P. K., Singh, S., & Singh, S. (2023). Comparative Analysis of Statistical and Machine Learning Techniques for Rice Yield Forecasting for Chhattisgarh, India. *Sustainability*, *15*(3), 2786. https://doi.org/10.3390/su15032786

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/a:1010933404324

Pang, A., Chang, M. W. L., & Chen, Y. (2022). Evaluation of Random Forests (RF) for Regional and Local-Scale Wheat Yield Prediction in Southeast Australia. *Sensors*, *22*(3), 717. https://doi.org/10.3390/s22030717

Iqbal, N., Shahzad, M. U., Sherif, E.-S. M., Tariq, M. U., Rashid, J., Le, T.-V., & Ghani, A. (2024). Analysis of Wheat-Yield Prediction Using Machine Learning Models under Climate Change Scenarios. *Sustainability*, *16*(16), 6976–6976. https://doi.org/10.3390/su16166976

Teixeira, R., Cerveira, A., Solteiro, E. J., & Baptista, J. (2024). Enhancing Weather Forecasting Integrating LSTM and GA. *Applied Sciences*, *14*(13), 5769–5769. https://doi.org/10.3390/app14135769

Kadkhodazadeh, M., Valikhan Anaraki, M., Morshed-Bozorgdel, A., & Farzin, S. (2022). A New Methodology for Reference Evapotranspiration Prediction and Uncertainty Analysis under Climate Change Conditions Based on Machine Learning, Multi Criteria Decision Making and Monte Carlo Methods. *Sustainability*, *14*(5), 2601. https://doi.org/10.3390/su14052601

Granata, F., & Di Nunno, F. (2021). Forecasting evapotranspiration in different climates using ensembles of recurrent neural networks. *Agricultural Water Management*, *255*, 107040. https://doi.org/10.1016/j.agwat.2021.107040

Yin, J., Deng, Z., Ines, A. V. M., Wu, J., & Rasu, E. (2020). Forecast of short-term daily reference evapotranspiration under limited meteorological variables using a hybrid bi-directional long short-term memory model (Bi-LSTM). *Agricultural Water Management*, *242*, 106386. https://doi.org/10.1016/j.agwat.2020.106386

Gauch, M., Kratzert, F., Klotz, D., Nearing, G., Lin, J., & Hochreiter, S. (2021). Rainfall–runoff prediction at multiple timescales with a single Long Short-Term Memory network. *Hydrology and Earth System Sciences*, *25*(4), 2045–2062. https://doi.org/10.5194/hess-25-2045-2021

Minhaz, A. (2023, June 15). *Telangana's erratic summer showers*. Frontline. https://frontline.thehindu.com/the-nation/agriculture/telangana-erratic-summer-showers/article66941025.ece?utm_source=relatedstories&utm_medium=article&utm_campaign=trackRelArt

Ai, Z., & Yang, Y. (2016). Modification and Validation of Priestley–Taylor Model for Estimating Cotton Evapotranspiration under Plastic Mulch Condition. *Journal of Hydrometeorology*, *17*(4), 1281–1293. https://doi.org/10.1175/jhm-d-15-0151.1

Priestley, C., & Taylor, R. (1972). On the Assessment of Surface Heat Flux and Evaporation Using Large-Scale Parameters. *Semantic Scholar*. https://doi.org/10.1175/1520-0493(1972)100%3C0081:OTAOSH%3E2.3.CO;2

Parlange, M. B., & Katul, G. G. (1992). An advection-aridity evaporation model. *Water Resources Research*, *28*(1), 127–132. https://doi.org/10.1029/91wr02482

French, H. (2021). *Determining Water Requirements in Irrigated Areas From Climatological and Irrigation Data; TP-96*. Hassell Street Press.

Choudhary, D. (2018, March 22). *Methods of Evapotranspiration*. https://doi.org/10.13140/RG.2.2.14533.76007

Mote, K., V. Praveen Rao, Kumar, K., & V. Ramulu. (2018). Estimation of crop evapotranspiration and crop coefficients of rice (Oryza sativa L.) under low land condition. *Journal of Agrometeorology*, *20*(2), 117–121. https://doi.org/10.54386/jam.v20i2.521.