

INTERPRETABLE PADDY YIELD PREDICTION USING DYNAMIC DROPOUT REGULARIZED FCNN WITH SHAP EXPLANATIONS

Ms. S. Abarna

*Assistant Professor, Department of Computer Science and Engineering
Grace College of Engineering, Chandy Nagar, Mullakkadu, Tuticorin, Tamilnadu, India
abarna.kichi@gmail.com*

Dr. D. Kesavaraja

*Professor, Department of Computer Science and Engineering
Dr. Sivanthi Aditanar College of Engineering, Tiruchendur, Tamilnadu, India
dkesavaraja@gmail.com*

DOI: doi.org/10.34293/shanlax.9789361631474.ch007

Abstract

Accurate paddy yield prediction using structured agricultural datasets is essential for planning, resource allocation, and decision support. However, many deep learning approaches for yield forecasting either rely on complex multimodal inputs or provide limited interpretability, making them difficult to trust and deploy in real-world settings. This paper presents an explainable deep learning framework based on a Feature Normalization and Dropout Regularized Fully Connected Neural Network (FN-DR-FCNN) for paddy yield prediction using tabular agronomic and environmental variables. The model integrates z-score feature normalization and a dynamic dropout mechanism to improve generalization and reduce overfitting during training. To enhance transparency, The SHapley Additive exPlanations (SHAP) are also included to measure the contribution of each input feature to the predicted yield to be able to interpret it globally and locally. The proposed approach is validated on two real-world datasets, namely the LDS Rice Dataset and the Vellore Rice Dataset. Experimental results demonstrate that FN-DR-FCNN achieves strong predictive performance while providing consistent and meaningful feature attributions that support agronomic

reasoning. The combination of accuracy, computational efficiency, and explainability makes the proposed framework suitable for practical decision-making in smart agriculture applications.

1. Introduction

Precise yield prediction of crops is a crucial aspect of contemporary agriculture as it assists in the initial planning, effective resource distribution, and making decisions that are resistant to risks. For paddy cultivation in particular, yield is influenced by multiple interacting factors such as soil conditions, irrigation availability, fertilizer application strategies, cultivation practices, and climate variability. As these parameters differ dramatically across regions and season, yield forecasting is a difficult regression problem that needs models with the ability to learn complex non-linear relationships using real world agricultural data.

The use of machine learning and deep learning techniques in yield

prediction is a topic that has been studied intensively over the last few years owing to their features of identifying patterns in the large datasets. Traditional machine learning algorithms (linear regression, decision trees and support vector regression) may be able to give baseline performance and can be partially interpretable; but, predictive accuracy tends to drop when interactions between features are highly non-linear or when the data comprise both categorical and numeric variables. Deep learning models overcome these problems by learning hierarchical representations by data. In certain studies, hybrid models like CNN-LSTM and Transformer-based models have shown great performance especially when satellite images or time-series measurements are provided. However, these architectures generally demand increased computing capabilities, complex input pipelines and specialized data modalities which are not necessarily available in most real world farming scenarios.

The other key issue with deep learning models is that their transparency is limited thus making yield prediction a challenging task. Although neural networks may be accurate, they are black-box models thus it is not easy to understand what makes a given yield value be predicted. This lack of interpretability reduces user trust and restricts real-world adoption, especially in agriculture where decisions based on predictions can directly affect

cost, productivity, and food security. Therefore, along with improving predictive performance, it is equally important to develop models that provide meaningful explanations of the factors influencing yield estimates.

To solve them, the proposed work presents a explainable framework of yield prediction, which is based on Feature Normalization and Dropout Regularized Fully Connected Neural Network (FN-DR-FCNN). The proposed model is designed for structured/tabular agricultural datasets and integrates z-score feature normalization to stabilize training and ensure balanced feature contribution. In addition, a dynamic dropout regularization mechanism is incorporated to improve generalization and reduce overfitting during learning. In addition to accuracy, the proposed framework directly aims at interpretability through the integration of SHapley Additive explanations (SHAP), a method that offers both global and local features prominence and sample-specific explanations to yield predictions. This allows the model to estimate the contribution of each agronomic or environmental parameter to the final yield that is being predicted which enhances transparency and facilitates agronomic arguments.

The usefulness of the suggested method is justified with the help of two actual databases Vellore Rice Dataset and LDS Rice Dataset, which comprises large-scale survey-based cultivation and

management variables that have been measured in several rice producing states of India and the dataset Vellore Rice Dataset, with the large-scale measurements of agro-hydrological and climatic factors that should be involved in estimating the paddy yield. By evaluating the model on two diverse datasets, the study demonstrates both predictive robustness and explanation consistency across different cultivation contexts. Overall, the proposed explainable FN-DR-FCNN framework aims to provide an efficient, accurate, and transparent solution for paddy yield prediction that can be deployed as a practical decision-support tool in smart agriculture.

2. Background and Related Work

The usefulness of the suggested method is justified with the help of two actual databases Vellore Rice Dataset and LDS Rice Dataset, which comprises large-scale survey-based cultivation and management variables that have been measured in several rice producing states of India and the dataset Vellore Rice Dataset, with the large-scale measurements of agro-hydrological and climatic factors that should be involved in estimating the paddy yield. However, despite their strong performance, several existing models remain computationally intensive and are difficult to deploy in practical agricultural settings that rely primarily on structured/tabular data. In addition, most deep learning-based yield prediction systems provide limited

interpretability, making it challenging to understand which input variables drive the final prediction. This creates a need for lightweight yet accurate predictive models that also support transparent decision-making through explainable AI techniques. In this context, the following studies are reviewed to highlight current advancements and the remaining gaps addressed by the proposed explainable FN-DR-FCNN framework.

New developments in crop yield prediction have progressively embraced machine learning and deep learning technologies to explain the intricate relationship that exists between yield and agronomic variables. Several works have reported improved accuracy using hybrid and ensemble-based learning, where deep feature extraction is combined with optimization or multi-model fusion strategies. Other studies have utilized satellite imagery, meteorological variables, and soil parameters to strengthen prediction capability through multimodal learning. In addition, recent research has explored reinforcement learning and large language model-assisted frameworks to support adaptive decision-making and interactive agricultural planning tools. Rice-specific studies have also compared deep architectures such as CNN, RNN, and LSTM, and highlighted the benefits of deep learning in learning non-linear patterns from large-scale cultivation datasets. Lightweight deep models and attention-based networks have been proposed to improve efficiency, while

hybrid CNN-LSTM frameworks have been used to handle spatial and temporal dependencies for yield estimation.

Despite these developments, notable limitations remain in existing yield prediction systems. Many state-of-the-art approaches depend on complex architectures and external sensing inputs such as UAV or satellite data, which increases computational cost and reduces feasibility in real-time, low-resource environments. Traditional machine learning models provide partial interpretability but may fail to represent high-dimensional non-linear interactions present in structured agricultural data. Deep learning models trained on tabular datasets can also suffer from unstable learning and overfitting when feature scaling and regularization are not carefully designed. Therefore, there is a strong need for a lightweight and accurate framework that is specifically tailored for structured datasets while also offering transparent decision-making through explainable AI techniques. This motivates the development of an interpretable model that combines systematic feature normalization, controlled regularization, and feature-attribution based explanations for reliable paddy yield prediction.

3. Problem Statement and Motivation

Paddy yield prediction is a challenging regression task due to the complex and non-linear interactions among cultivation practices, soil

properties, irrigation conditions, fertilizer schedules, and climate-related variables. In real-world agricultural datasets, these factors are often recorded in structured/tabular form with a mixture of categorical and numerical attributes. Such datasets typically contain noise, missing values, inconsistent scales across features, and region-specific variations that make generalization difficult. Although deep learning models can capture non-linear relationships effectively, many existing yield prediction frameworks rely on complex hybrid architectures or multimodal inputs such as satellite imagery and high-frequency weather time series. These approaches may offer strong accuracy but often require expensive sensing infrastructure, large computation, and specialized preprocessing pipelines, limiting their applicability in practical and resource-constrained environments.

In addition to accuracy and generalization, interpretability is a major requirement for real-world adoption. Agricultural decision-making involves significant cost and risk, and stakeholders such as farmers, agronomists, and policymakers require not only the predicted yield value but also a clear explanation of why the model produced that estimate. Black-box predictions without reasoning reduce trust and make it difficult to validate whether the model is learning meaningful agronomic patterns or relying on spurious correlations.

Therefore, there is a strong need for a yield prediction framework that is (i) effective for structured/tabular agricultural datasets, (ii) computationally efficient for real-time use, (iii) robust against overfitting, and (iv) capable of providing transparent explanations that identify the most influential features driving the predictions.

Motivated by these needs, this work proposes an explainable and lightweight yield prediction framework that integrates feature normalization, controlled regularization, and SHAP-based feature attribution. The proposed approach aims to provide high predictive accuracy while ensuring that the model remains interpretable and deployable for practical agricultural decision-support systems.

4. Proposed Explainable FN-DR-FCNN Framework

This section describes the proposed explainable framework for paddy yield prediction. The model is designed for structured agricultural datasets and focuses on three major goals: stable learning, improved generalization, and transparent reasoning. To achieve these goals, the framework combines (i) feature normalization to ensure consistent scaling of inputs, (ii) dropout-based regularization to reduce overfitting, and (iii) SHAP explainability to interpret predictions at both global and local levels. The overall workflow begins with data preparation and pre-processing, followed

by training the FN-DR-FCNN model, and finally generating feature-attribution explanations using SHAP.

4.1 Data Preparation and Preprocessing

The quality of input data strongly influences the reliability of yield prediction. Hence, a systematic preprocessing pipeline is applied to prepare the datasets before training. The preprocessing stage includes data cleaning, handling missing values, encoding categorical attributes, scaling numerical features, and preventing target leakage.

First, duplicate entries and non-informative attributes are removed to reduce redundancy and improve training efficiency. Columns containing constant values or extremely high missing ratios are excluded since they do not contribute to meaningful learning. For missing values, numerical attributes are imputed using a neutral placeholder such as zero, which is appropriate in agricultural datasets where missing entries can represent non-application or absence of an input (for example, fertilizer not applied during a particular stage). For categorical attributes, missing values are replaced with a dedicated label such as “No” or “Unknown” to preserve category consistency without dropping records.

Since the datasets contain both categorical and numerical variables, categorical fields are converted into numerical representations using label encoding. This is a computationally

efficient encoding that does not require the high dimensionality of one-hot encoding in case of large number of categories. Numerical features are standardized by using the z-score normalization by encoding. To normalize a variable x' use:

$$x' = \frac{x - \mu}{\sigma}$$

where μ and σ are the mean and standard deviation of the feature, respectively. It is done to make all the input variables on the same scale and avoid the influence of the features with significant magnitudes in the learning process. Normalization is also more stable in gradients and it makes the training process faster.

Lastly, the data is divided into a training and testing set with the help of a uniform approach like k-fold cross-validation. This provides a robust evaluation setting by ensuring that model performance is not dependent on a single random split. Cross-validation is particularly useful for agricultural datasets where regional variability and uneven distributions can influence the results.

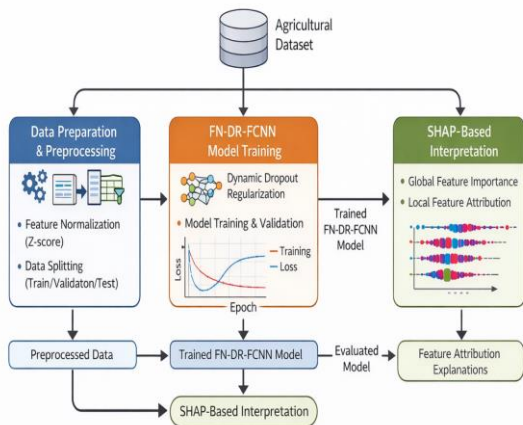


Figure 1: Overall workflow of the proposed FN DR FCNN Framework for paddy yield prediction

The complete pipeline of the proposed explainable FN-DR-FCNN framework, from pre-processing to SHAP-based interpretation, is illustrated in Fig. 1. The overall workflow begins with data preparation and pre-processing, followed by training the FN-DR-FCNN model, and finally generating feature-attribution explanations using SHAP.

4.2 FN-DR-FCNN Model Architecture

The proposed prediction model is a Fully Connected Neural Network (FCNN) enhanced with feature normalization and dropout regularization. Unlike convolutional or recurrent architectures, FCNN models are well suited for structured/tabular datasets where the input features do not follow spatial grids or strict sequential dependencies. The FCNN learns complex non-linear interactions among features by stacking multiple dense layers with non-linear activation functions.

Let the input feature vector be represented as:

$$x = [x_1, x_2, \dots, x_d]$$

Where, d represents the input features post processing. The FCNN changes the input using a series of dense layers. Each hidden layer computes:

$$h^{(l)} = f(W^{(l)}h^{(l-1)} + b^{(l)})$$

where $W^{(l)}$ and $b^{(l)}$ are the weights and biases of the l^{th} layer, $f(\cdot)$ is a non-linear activation function like ReLU, and $h^{(0)} = x$. The output layer will be regressive and approach a linear activation function in order to generate the intended yield y^{\wedge} :

$$\hat{y} = W^{(out)}h^{(L)} + b^{(out)}$$

where L is the number of hidden layers. This output corresponds to the estimated paddy yield in the desired unit (for example, quintals per acre). The internal architecture of the proposed FN-DR-FCNN model, including feature normalization and dynamic dropout regularization, is displayed in Fig. 2.

Overfitting is a common issue in yield prediction tasks because agricultural datasets may contain redundant variables, noisy measurements, and region-specific patterns that do not generalize well. In order to enhance robustness, the dropout regularization is used on the fully connected layers. The random deactivation of a fraction of neurons in training by dropout stops the network that is over-reliant on a high percentage of activations. This encourages the model to learn distributed representations and improves generalization.

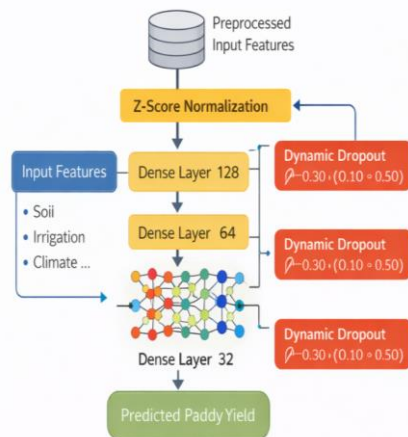


Figure 2. Architecture of the proposed FN-DR-FCNN model with feature normalization and dynamic dropout regularization.

4.3 Dropout Regularization for Generalization

If a dropout rate of p is used, then during training, each neuron is retained with probability (1-p). The dropout operation can be expressed as:

$$h = m \odot h$$

where m is multiplication of the elements and is a binary mask sampled by a Bernoulli distribution. and \odot denotes element-wise multiplication. In inference, the dropout is turned off and the entire network is employed in the generation of the deterministic predictions.

In the proposed FN-DR-FCNN framework, dropout is integrated after key dense layers to control overfitting while maintaining model capacity. This ensures stable training even when the feature space is large and heterogeneous.

5. SHAP-Based Explainability Module

While predictive accuracy is important, interpretability is essential for deploying yield prediction models in real agricultural environments. To address this requirement, the proposed framework integrates SHapley Additive exPlanations (SHAP) to explain the predictions of FN-DR-FCNN. SHAP provides a theoretically grounded method for explaining model outputs by assigning each feature an importance value based on its contribution to the prediction.

5.1 Global Interpretability

Global interpretability refers to understanding which features are most

influential across the entire dataset. SHAP computes the contribution of each feature for every sample, and the mean absolute SHAP value is used to rank features by importance. This helps identify key factors such as irrigation availability, fertilizer usage, soil properties, and climatic conditions that strongly influence yield estimation.

Global feature importance is particularly valuable for agricultural policy and planning, as it highlights the variables that consistently affect yield outcomes across regions and farming practices.

5.2 Local Interpretability

Local interpretability explains why the model produced a specific prediction for an individual sample. For a given test instance, SHAP provides positive and negative contributions of each feature relative to a baseline value. Features with positive SHAP values increase the predicted yield, while features with negative SHAP values reduce it. This sample-level explanation allows stakeholders to validate the reasoning of the model and assess whether the prediction aligns with practical agronomic knowledge.

Local explanations are especially useful for decision support because they provide actionable insights. For example, if the model predicts a low yield and SHAP indicates that limited irrigation or insufficient fertilizer application is the main cause, corrective actions can be planned accordingly.

5.3 Experimental Methodology

The suggested FN-DR-FCNN model was developed in Python based on TensorFlow and Keras. All the experiments were conducted in a Windows 10 operating system with the Intel Core i5 processor (2.40 GHz) and 16 GB RAM. The datasets were preprocessed by removing redundant columns, imputing missing numerical values with zero, and filling missing categorical entries with a dedicated label. Categorical variables were converted into numerical form using label encoding. To ensure stable learning and balanced feature contribution, all numerical features were standardized using z-score normalization.

The model has been tested on 5-fold cross-validation and four folds were trained and one folded was tested at a time to get the strong performance estimates. FN-DR-FCNN architecture was constructed to be a light network with three dense layers with 128, 64 and 32 neurons respectively and an activation of ReLU. Dropout regularization was applied after each hidden layer to reduce overfitting, and the dropout rate was dynamically adjusted during training. The initial dropout value was set to 0.30 and constrained within bounds of [0.10, 0.50] to prevent under-regularization at very low dropout values and excessive information loss at very high dropout values. The output layer had a linear activation function where it generated continuous yield predictions.

The network was trained on Adam optimizer with learning rate of 0.001 and Mean Squared Error (MSE) was taken as the major loss measure. The training was conducted in 20 epochs with a 32 sized batch, and early stopping was applied to the loss in the validation with a patience of 5 epochs to prevent overfitting and unneeded calculation.

Standard regression measures, such as RMSE, MAE, MSE, MAPE, and coefficient of determination (R^2) were used to measure the model performance. Besides the accuracy evaluation, the explainability analysis was conducted using SHapley Additive exPlanations (SHAP). Given that the proposed model is a deep neural network that works with structured data, SHAP Deep Explainer was used to compute feature attributions.

The interpretability of data worldwide was shown by SHAP summary plots ranking the most significant variables throughout the dataset, whereas the interpretability of data at an individual predictor was shown by SHAP waterfall plots. This conglomerated appraisal guarantees that the framework proposed has a good prediction performance as well as that the decision-maker can have a clear picture when making decisions based on paddy yield estimations.

6. Results and Explainability Analysis

6.1 Quantitative Performance Assessment

The proposed FN-DR-FCNN model was tested on two real world structured

data sets i.e. the LDS Rice Dataset and Vellore Rice Dataset and through 5-fold cross-validation. The measurement was aimed at error measures and goodness-of-fit measures, such as RMSE, MSE, MAE, MAPE, and R^2 . These measures are all useful in giving a sound evaluation of model accuracy, robustness, and stability on datasets with varying distributions and feature characteristics.

The obtained results confirm that the FN-DR-FCNN architecture effectively learns non-linear relationships from tabular agronomic and environmental inputs. The integration of feature normalization improved training stability by ensuring consistent gradient updates across variables with different scales. In addition, the dropout regularization strategy reduced overfitting and enhanced generalization, particularly for the high-dimensional and heterogeneous feature space present in the LDS dataset. Across both datasets, the proposed model consistently achieved low prediction error and high R^2 , indicating that it can accurately approximate yield variations while maintaining reliable generalization performance.

The training and validation loss curves were used to monitor the learning behavior at the epochs in order to further confirm stability in training. The model had a consistent convergence and few oscillations and the model did not continue to train unnecessarily when the validation loss ceased to decrease.

The findings obtained in this observation suggest that FN-DR-FCNN is a stable and efficient learning model that is applicable to structured agriculture data. The training and validation loss curves as in Fig. 3 approach each other gradually, which is a sign of stable learning and regularized overfitting.

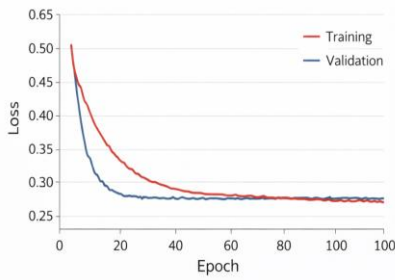


Fig. 3. Training and validation loss curves of FN-DR-FCNN across epochs, demonstrating convergence and stable learning behavior.

6.2 Comparative Analysis with Existing Approaches

To showcase the effectiveness of the proposed procedure, FN-DR-FCNN was correlated with representative yield prediction models reported in prior studies. Existing methods include sequential models such as LSTM-based forecasting approaches, hybrid deep learning models combining temporal and spatial learning, and regularized hybrid models that integrate classical regression techniques. In comparison, FN-DR-FCNN is specifically designed for structured/tabular datasets and avoids dependence on specialized data modalities such as satellite imagery or dense time-series measurements.

The relative outcomes indicate that the proposed FN-DR-FCNN has a better prediction accuracy and less computational resources. The proposed architecture is much simpler and offers more practical in terms of real-time decision support because compared to complex hybrid architectures, it offers an efficient model structure with fewer parameters, less memory utilization, and a reduced inference time. These advancements demonstrate the advantage of lightweight dense learning with effective regularization and constant feature scaling to allow the model to learn meaningful interactions between features without being overly complex. Fig. 4 summarizes the comparative performance of FN-DR-FCNN against baseline models with respect to RMSE, MAE, and R².

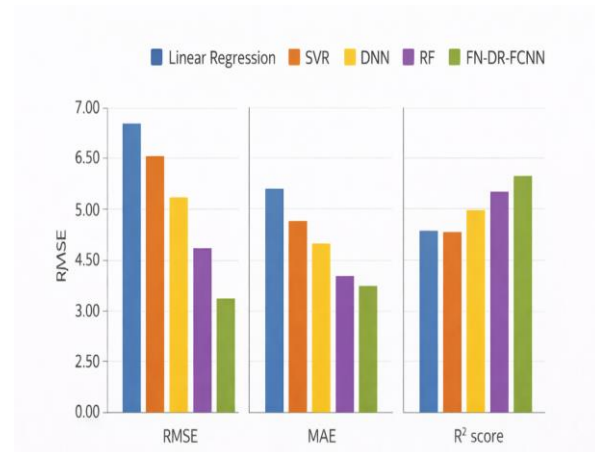


Fig. 4. Comparative performance of FN-DR-FCNN against baseline models in terms of RMSE, MAE, and R².

6.3 Explainability Analysis using SHAP

While strong predictive performance is essential, the practical deployment of yield prediction models also requires transparency and trust. Deep neural

networks often behave as black-box predictors, making it difficult to understand which variables drive the predicted yield. To address this limitation, SHapley Additive exPlanations (SHAP) concept was applied to explore the FN-DR-FCNN predictions. SHAP provides additive feature attributions grounded in cooperative game theory, where each input feature is assigned a contribution score representing its impact on the final output relative to a baseline expectation. For a given input sample, the predicted yield can be expressed as:

$$\hat{y} = \phi_0 + \sum_{i=1}^d \phi_i$$

where ϕ_0 denotes the baseline prediction, d is the count of input features, ϕ_i represents the SHAP value of the i -th feature. When SHAP is positive, it implies that the given feature enhances the predicted yield, and when SHAP is negative, it implies that it diminishes the prediction yield.

6.3.1 Global Interpretability

Global interpretability aims to identify which variables consistently influence yield prediction across the dataset. This was achieved using SHAP summary plots, where features are ranked by the mean absolute SHAP magnitude. The global SHAP analysis reveals the dominant agronomic and environmental factors that the proposed model relies on for yield estimation. Such insights are valuable because they provide an interpretable ranking of yield drivers and enable validation of model reasoning against agricultural domain knowledge.

In general, features related to irrigation availability, fertilizer application practices, soil properties, and climatic conditions tend to contribute significantly to yield prediction. The SHAP ranking also helps distinguish between features that have a strong and consistent effect across samples and features whose influence is localized to specific cultivation conditions. This supports transparent decision-making by allowing stakeholders to focus on high-impact variables during crop planning and resource allocation.

Table 1 reports the top-ranked features obtained from SHAP analysis using the mean absolute SHAP magnitude as the importance score. For the LDS Rice dataset, irrigation availability (0.084) and fertilizer-related inputs such as basal (0.072), tillering (0.061), and panicle-stage application (0.056) appear as the strongest contributors, indicating that management practices have a dominant influence on yield estimation in the survey-based dataset. In contrast, the Vellore Rice dataset shows higher importance for climatic and soil-related factors, where rainfall (0.091), temperature (0.083), soil pH (0.074), and nitrogen content (0.067) contribute most to the predicted yield. Overall, the feature ranking demonstrates that the proposed FN-DR-FCNN model learns meaningful agronomic relationships, while SHAP provides transparent evidence of how key cultivation and environmental parameters drive the prediction process across different datasets.

Table 1: Top-ranked input features based on mean absolute SHAP values for FN-DR-FCNN on LDS &Vellore rice datasets

Rank	LDS Rice Dataset - Feature	Mean SHAP	Vellore Rice Dataset - Feature	Mean SHAP
1	Irrigation availability	0.084	Rainfall	0.091
2	Fertilizer (Basal)	0.072	Temperature	0.083
3	Fertilizer (Tillering)	0.061	Soil pH	0.074
4	Fertilizer (Panicle stage)	0.056	Nitrogen (N)	0.067
5	Crop variety	0.049	Phosphorus (P)	0.047
6	Sowing method	0.043	Potassium (K)	0.041
7	Soil texture	0.038	Humidity	0.059
8	Drainage type	0.031	Evapotranspiration	0.052
9	Pest severity level	0.026	Groundwater salinity / chloride	0.035
10	Land preparation method	0.021	Groundwater permeability	0.029

6.3.2 Local Interpretability

In addition to global trends, local interpretability explains why the model predicted a specific yield value for an individual sample. This was visualized using SHAP waterfall plots, where the baseline prediction is incrementally adjusted by the contributions of the most influential features. Features with positive SHAP values push the prediction upward, while features with negative values reduce it.

Local explanations are particularly useful in agriculture because they provide actionable insights. As an example, when the model estimates a lower yield of a given sample, SHAP can identify whether the primary causes are related to limited irrigation, insufficient fertilizer input, unfavorable soil conditions, or adverse climatic indicators. Similarly, for high-yield

predictions, SHAP can highlight the key positive drivers such as optimal cultivation practices or supportive environmental conditions. This level of transparency improves trust and enables the model to serve not only as a forecasting tool but also as a decision-support mechanism.

6.4 Discussion

The combined quantitative and explainability results demonstrate that FN-DR-FCNN provides both strong prediction performance and meaningful interpretability for structured paddy yield prediction. Feature normalization improves learning stability and prevents scale-dominance effects, while dropout regularization enhances generalization by reducing co-adaptation among neurons. The SHAP-based interpretation complements the predictive model by revealing the feature-level contributions

that drive yield estimates, thereby addressing a key limitation of deep learning systems in agriculture.

Compared to conventional black-box deep models, the proposed framework improves transparency without sacrificing accuracy. The ability to provide both global feature ranking and sample-specific explanations makes the proposed approach more reliable for real-world adoption, where users require justifiable predictions for planning and intervention. Overall, FN-DR-FCNN with SHAP explanations offers an efficient and interpretable solution for paddy yield prediction, with strong potential for deployment in smart agriculture and related structured-data forecasting applications.

7. Conclusion

This source discussed a deep learning prediction model of paddy yield with structured/tabular agricultural data that is explainable. The designed FN-DR-FCNN model combines the z-score feature normalization with dropout regularization to enhance the stability of learning and minimize overfitting and therefore increase the generalization of various cultivation conditions. The model was tested on LDS Rice Dataset and Vellore Rice Dataset with 5 fold cross-validation and results showed that the proposed method gives high yield estimation performance with high regression accuracy.

In addition to predictive capability, the proposed framework improves transparency through SHapley Additive exPlanations (SHAP). The SHAP-based global feature ranking identified the most influential variables contributing to yield prediction, while local explanations provided sample-level reasoning by showing how individual features shift the prediction from the baseline to the final estimated yield. This combination of accuracy and interpretability makes the proposed FN-DR-FCNN framework suitable for practical agricultural decision-support applications where trust and justification of predictions are essential.

One way of extending this work in the future is to use more sophisticated feature selection and dimensionality reduction methods to make it even more robust when applied on high-dimensional agricultural data. The model can also be expanded to include uncertainty-aware prediction to quantify confidence intervals for yield estimates, which would support risk-aware planning. In addition, integrating real-time weather streams and IoT-based field sensor measurements can enable dynamic yield forecasting under changing environmental conditions. Finally, lightweight deployment strategies such as model pruning and quantization can be explored to implement the proposed framework efficiently on edge devices for scalable smart agriculture applications.

References

1. Agarwal, N., Choudhry, N. and Tripathi, K.C., et. al., 2025. A novel hybrid time series deep learning model for forecasting of cotton yield in India. *International Journal of Information Technology*, pp.1-8.
2. Ajay, A., Craufurd, P., Kumar, V., Samaddar, A., Malik, R.K., Sharma, S., Ranjan, H., Singh, A.K., Paudel, G., Pundir, A.K. and Poonia, S., 2022. Large survey dataset of rice production practices applied by farmers on their largest farm plot during 2018 in India. *Data in Brief*, 45, p.108625
3. Bharathi, S.V., 2025 and Manikandan, A., 2025. A novel hybrid meta-heuristic-enabled ensemble learning model with deep feature extraction for crop yield prediction with heuristic ensemble yield. *International Journal of Information and Decision Sciences*, 17(1), pp.1-31.
4. Chen, D. and Huang, Y., (2025). Integrating reinforcement learning and large language models for crop production process management, optimization, and control through a new knowledge-based deep learning paradigm. *Computers and Electronics in Agriculture*, 232, p.110028.
5. Chu, Z. and Yu, J., 2024. An end-to-end model for rice yield prediction using deep learning fusion. *Computers and Electronics in Agriculture*, 174, p.105471.
6. Elavarasan, D. and Vincent, P.D., 2020. Crop yield prediction using deep reinforcement learning model for sustainable agrarian applications. *IEEEaccess*, 8, pp.86886-86901.
7. Gaso, D.V., La Rosa, L.E.C., Puntel, L.A., Edreira, J.I.R., de Wit, A. and Kooistra, L., (2025). Predicting spatial and temporal variability in soybean yield using deep learning and open source data. *European Journal of Agronomy*, 164, p.127498.
8. Geetha, M., Suganthe, R.C., Latha, R.S., Anju, R., Sastimalar, K., and Shobana, P., 2022, January. Deep Learning Based Yield Prediction Model to Predict the Yield of Paddy in Cauvery Delta Region. In *2022 International Conference on Computer Communication and Informatics (ICCCI)* (pp. 1-6). IEEE.
9. Halim, V.J., Ryanlie, A., Djong, K., Setiawan, K.E. and Hasani, M.F., et 2023, August. Rice Yield Production Forecasting using Deep Learning Models. In *2023 International Conference on Networking, Electrical Engineering, Computer Science, and Technology (IConNECT)* (pp. 54-59). IEEE.
10. Islam, A., Sultana, T., Hasan, M.M. and Nadim, M.I., 2024, June. Bangladeshi Paddy Yield Estimation by Deep Learning Approach. In *2024, 15th International Conference on Computing, Communication and Networking Technologies (ICCCNT)* (pp. 1-4). IEEE.
11. Krupavath, K., Raghu Babu, M., and Mani, A., 2022. Comparative

- evaluation of neural networks in crop yield prediction of paddy and sugarcane crops. The digital agricultural revolution: innovations and challenges in agriculture through technology disruptions, pp.25-55.
12. Kumar, R.M., 2025. Crop yield prediction by Mestrial Environ Net (MENN). *International Journal of Advanced Science and Computer Applications*, 4(1).
 13. Mohan, R.J., Rayanoothala, P.S., and Sree, R.P., (2025). Next-gen agriculture: integrating AI and XAI for precision crop yield predictions. *Frontiers in Plant Science*, 15, p.1451607.
 14. Yang, G., Jin, N., Ai, W., Zheng, Z., He, Y. and He, Y., (2025). Integrating remote sensing data assimilation, deep learning, and large language models for interactive wheat breeding yield prediction. arXiv preprint arXiv:2501.04487.
 15. Yewle, A.D., Mirzayeva, L. and Karakuş, O., 2025. Multi-modal Data Fusion and Deep Ensemble Learning for Accurate Crop Yield Prediction. arXiv preprint arXiv:2502.06062.
 16. Zulkifley, M.A., Aziz, N.A.A. and Akhyar, A., 2024, December. Gross Paddy Yield Estimation Using a Lightweight Deep Learning Model. In *2024 International Conference on Decision Aid Sciences and Applications (DASA)* (pp. 1-5). IEEE.