

Received: 18/06/2026

Accepted: 12/08/2026

Published: 05/09/2026

Citation: Mercy R, Lucia Agnes Beena T (2026) Fertilizer-Focused Stacked Ensemble Model for Explainable Paddy Yield Prediction Using Context-Residualized Nutrient Sensitivity Learning. Indian Journal of Science and Technology 19(32): 2248-2263. <https://doi.org/10.17485/IJST/v19i32.921>

* **Corresponding author.**mercy.loyola@gmail.com**Funding:** None**Competing Interests:** None

Copyright: © 2026 Mercy & Lucia Agnes Beena. This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Published By Indian Society for Education and Environment ([iSee](#))

ISSN

Print: 0974-6846

Electronic: 0974-5645

Fertilizer-Focused Stacked Ensemble Model for Explainable Paddy Yield Prediction Using Context-Residualized Nutrient Sensitivity Learning

R Mercy^{1*}, T Lucia Agnes Beena²

¹ Research Scholar, Department of Computer Science, Holy Cross College (Autonomous), Affiliated to Bharathidasan University, Tiruchirappalli-620002, Tamil Nadu, India

² Research Supervisor, Department of Computer Science, Holy Cross College (Autonomous), Affiliated to Bharathidasan University, Tiruchirappalli-620002, Tamil Nadu, India

Abstract

Objectives: To build an explainable model that predicts paddy yields based on the fertilizer application, removing the influence of the climatic, irrigation and spatial factors. **Method:** The study suggests a new model called Fertilizer-Focused Stacked Ensemble Model (FFSEM). Context residualisation eliminates contextual effects. Mutual information, fertilizer-context synergy, redundancy suppression and stability penalization generate fertilizer sensitivity scores. Softmax weighting forms a Context-Integrated Feature (CIF) space. The Gradient Boosting meta-learning technique is used for stacking regressors like Light GBM, XGBoost and Multilayer Perceptron (MLP). **Findings:** FFSEM gave the best Root Mean Square Error (RMSE of 0.0406), Mean Absolute Error (MAE=0.0287), Mean Absolute Percentage Error (MAPE of 0.202) and R^2 of 0.9512 for paddy yield of Tamil Nadu district level data. The model performed better than the existing ensemble models. Nitrogen was found to be the most important fertilizer component and Theni, Dharmapuri, Thiruvavur and Dindigul districts had been identified as high fertilizer-sensitive districts using SHapley Additive exPlanations (SHAP). **Novelty:** The proposed model provides context-residualized fertilizer sensitivity learning, dependency, synergy, redundancy and stability-aware nutrient weighting.

Keywords: Paddy yield; Fertilizer sensitivity; Stacked ensemble; Context residualization; Explainable artificial intelligence

1 Introduction

Paddy yield prediction is useful for agricultural planning, food security, fertilizer scheduling, irrigation management, storage planning and market supply. Productivity of paddy is affected by rainfall, temperature, solar radiation, irrigation, soil condition, seed rate and fertilizer application⁽¹⁾. These factors are interrelated and may have non-linear interactions. Therefore, traditional statistical models might not be appropriate for complex agricultural relationships.

Machine learning models have been developed to learn hidden patterns from climatic, spatial, remote sensing and agronomic variables for better prediction of crop yields⁽²⁾. The ensemble models like Random Forest, XGBoost, LightGBM and stacked regression models have performed well in agricultural prediction problems. These models are more appropriate than simple regression models in representing nonlinear relationships. But most of the existing studies consider all variables as generic predictors.

Among the various factors which are most under the control of paddy production, the application of fertilizer is one. Nitrogen is involved in vegetative growth and yield formation⁽³⁾. Crop development is also affected by phosphorus, potassium and seed rate. But fertilizer effectiveness varies across the situation. Rainfall and availability of irrigation can affect nitrogen response⁽⁴⁾. Soil and regional management practices may play a role in phosphorus contribution. The effect of potassium influence may vary under water stress and district-level agricultural conditions. Fertilizer variables should not, therefore, be evaluated without considering other factors, such as raw correlation and global feature importance.

In several crop yield prediction studies, the variables of fertilizer are taken as standard input factors⁽⁵⁾,⁽⁶⁾. This practice can lead to more accurate predictions, but sources for nutrient levels may not be clear. Fertilizer effects may be overlaid by the dominant climatic, irrigation and spatial effects. This may lead to skewed fertilizer interpretation and restricted nutrient-based decision-making.

Studies on paddy yield prediction have recently shifted from the conventional forecasting models to machine learning, deep learning, remote sensing, crop simulation and explainable modelling. Sharma *et al.*⁽⁷⁾ developed the models for rice yield estimation based on climatic variables and multivariate analysis. Sah *et al.*⁽⁸⁾ fused biophysical variables with SAR and optical remote sensing data and obtained good performance with XGBoost, Random Forest, Cubist and neural networks. For the rice production forecasting in China, Mokhtar *et al.*⁽⁹⁾ employed multiple sources of information and combined machine learning approaches. These studies enhanced prediction performance, but primarily the effect of weather, biophysical and production level variables as opposed to the effect of fertilizer.

Rice yield prediction has been enhanced by remote sensing based approaches as well. The authors of Chang *et al.*⁽¹⁰⁾ suggested a hybrid deep neural network based on multi-task learning. Li *et al.*⁽¹¹⁾ used the process clustering method for time-series remote sensing images. Ibrahim *et al.*⁽¹²⁾ applied recursive feature elimination and mutual information regression with PlanetScope imagery. Quille-Mamani *et al.*⁽¹³⁾ demonstrated that the use of spectral and textural indices collected from an UAV was useful in the prediction of rice yield. Liang *et al.*⁽¹⁴⁾ combined the UAV multispectral data, meteorological data and employed SHAP for feature interpretation. These studies enhanced the spatial feature extraction and interpretability, but failed to model fertilizer response as a nutrient-specific representation.

Other works were related to climate impact, phenotyping and pixel-level forecasting. Tasneem *et al.*⁽¹⁵⁾ used machine learning models to determine rice yield prediction under climate change condition. Bak *et al.*⁽¹⁶⁾ estimated yields by segmenting panicle coverage in terms of semantics. For the rice plant, Zhang *et al.*⁽¹⁷⁾ estimated the rice traits based on hyperspectral and texture features. Sentinel-2 data was used to build a pixel level model of rice yield by Simeón *et al.*⁽¹⁸⁾. These studies gave useful predictions but nutrient-context interaction and the weighing of fertilizer-specificity were not sufficiently addressed.

In the case of Tamil Nadu, the importance of spatial and hybrid forecasting also emerges from the studies. In the Cauvery Delta, Manikandan *et al.*⁽¹⁹⁾ introduced rice yield estimation based on Sentinel-1A SAR data and DSSAT simulation. Satheesh *et al.*⁽²⁰⁾ evaluated the accuracy of spatial estimation of rice yield using DSSAT-CERES Rice model and semi-physical model in Tamil Nadu delta districts. Mahalakshmi and Manimekalai⁽²¹⁾ applied fuzzy logic, Monte Carlo simulation and LSTM to predict the location specific paddy yield. Ravikumar *et al.*⁽²²⁾ proposed an ensemble model with LSTM, SVR and XGBoost based on which achieved the RMSE of 70.16, the MAE of 47.93 and the R^2 of 0.9918. Jasmine⁽²³⁾ proposed a crop yield prediction system for the Tamil Nadu State for the crops using Decision Tree and Random Forest model. Although these studies have come a long way in terms of accuracy, most of them have considered the fertilizer a normal input or given fertilizer sensitivity no consideration at all.

Gayen and Dutta developed a machine learning method for paddy yield prediction using MODIS-derived vegetation indices, meteorological parameters, and ancillary crop yield datasets for the lower Ganga Basin. Their study shows that integrating satellite-derived vegetation information with weather variables can improve paddy yield modelling under spatially variable cultivation conditions⁽²⁴⁾. Similarly, Anu *et al.* emphasized interpretable crop yield prediction through stacked regression and explainable artificial intelligence. Their work used ensemble models such as Ridge Regression, Random Forest, XGBoost, CatBoost, LightGBM, and Gradient Boosting, showing that prediction accuracy must be supported by model transparency for precision agriculture⁽²⁵⁾. Ramesh and Kumaresan proposed a stacked ensemble crop yield prediction model using multiple base learners and a decision tree meta-model. Their results showed that stacked learning can outperform individual machine learning models in crop yield prediction⁽²⁶⁾.

The previous study has validated good achievements in paddy yield prediction based on meteorological analysis, remote sensing, UAV image, crop simulation, deep learning and ensemble learning. However, there is limited sensitivity to the fertilizer. Fertilizer variables are typically not included in current models or treatment as they have been grouped with climatic, irrigation, soil and spatial variables with nitrogen, phosphorus, potassium and seed rate. This can lead to some contextual bias because fertilizer effects are influenced by rainfall, temperature, irrigation availability, seasonal changes and district cultivation practices.

To fill this gap, a stacked ensemble model with the concept of fertilizer use (Fertilizer-Focused Stacked Ensemble Model (FFSEM)) for explainable paddy yield prediction is proposed. FFSEM separates contextual variables and distinction from fertilizer variables. Contextual features include rainfall, temperatures, solar radiation, indicators of irrigation and spatial characteristics. Fertilizer features include nitrogen, phosphorus, potassium and seed rate. The climatic effect, irrigation effect and spatial effect are eliminated from the yield and fertilizer variables by context residualization. The proposed method also leverages fertilizer dependency, fertilizer synergy, fertilizer redundancy, fertilizer stability scoring, softmax weighting, stacked ensemble prediction and SHAP explainability to inform fertilizer-informed agricultural decision making.

2 Methodology

2.1 FFSEM Architecture

The FFSEM architecture is proposed to enhance paddy yield prediction by learning the fertilizer-specific sensitivity of yield under different climatic, irrigation and spatial conditions. Figure 1 illustrates that the architecture will start with the input crop yield dataset, which will contain contextual features and features related to fertilizers. The contextual features are rainfall, temperature and solar radiation, irrigation infrastructure and spatial features. The characteristics of fertilizer are nitrogen, phosphorus, potassium and seed rate. This disentangling enables the model to interpret the effects of fertilizers in isolation rather than considering all the input variables as common prediction terms.

The preprocessing of data is done in the first stage of the architecture. Missing values are imputed with median imputation and numerical features are normalized with MinMax scaling. Variables that are likely to leakage like the production and area-derived variables are eliminated to prevent biased prediction. The architecture uses context residualization after preprocessing. During this step, the effect of the dominating climatic, irrigation and spatial factors is eliminated in both yield and fertilizer variables. This is done to help isolate the actual contribution to paddy yield due to fertilizer.

The second step involves calculation of fertilizer sensitivity. Mutual information between residual of fertilizers and yield residuals are used to estimate conditional dependency. The fertilizer-context synergy is then computed to incorporate the interactions between nutrient application and environmental conditions. Redundancy penalty is used to minimize redundant information between fertilizer variables. There is also a stability penalty to enhance consistency within groups of district-years. The scores are added to generate fertilizer-specific values of importance.

Following sensitivity scoring, the architecture uses softmax normalization to produce FFSEM fertilizer weights. The representation of the weighted fertilizer is combined with the contextual variables to create the Context-Integrated Feature (CIF) space. Such CIF representation is forwarded to the stacked ensemble learning module. LightGBM, XGBoost and MLP Regressor are base learners, whereas Gradient Boosting Regressor is the meta-learner. Lastly, explainability through SHAP is applied to understand the role of fertilizers and produce district-level rankings of the fertilizer influence.

2.2 Mathematical Formulation of FFSEM

2.2.1. Context Residualization

Context residualization is to remove the dominant influence of non-fertilizer variables before estimating fertilizer-specific yield sensitivity. In paddy yield prediction, fertilizer effects are often mixed with climatic, irrigation and spatial effects. Therefore, direct correlation between fertilizer inputs and yield may produce biased nutrient importance. To avoid this issue, FFSEM first separates the input variables into contextual features and fertilizer features.

Let the complete crop yield dataset be represented as $D = C, F, y$ where (C) denotes the contextual feature set, (F) denotes the fertilizer feature set and (y) denotes the observed paddy yield. The contextual feature set includes rainfall, temperature, solar radiation, irrigation variables and spatial attributes. The fertilizer feature set is defined as:

$$F = F_1, F_2, \dots, F_m \quad (1)$$

where each F_j represents one fertilizer-related variable such as nitrogen, phosphorus, potassium, or seed rate.

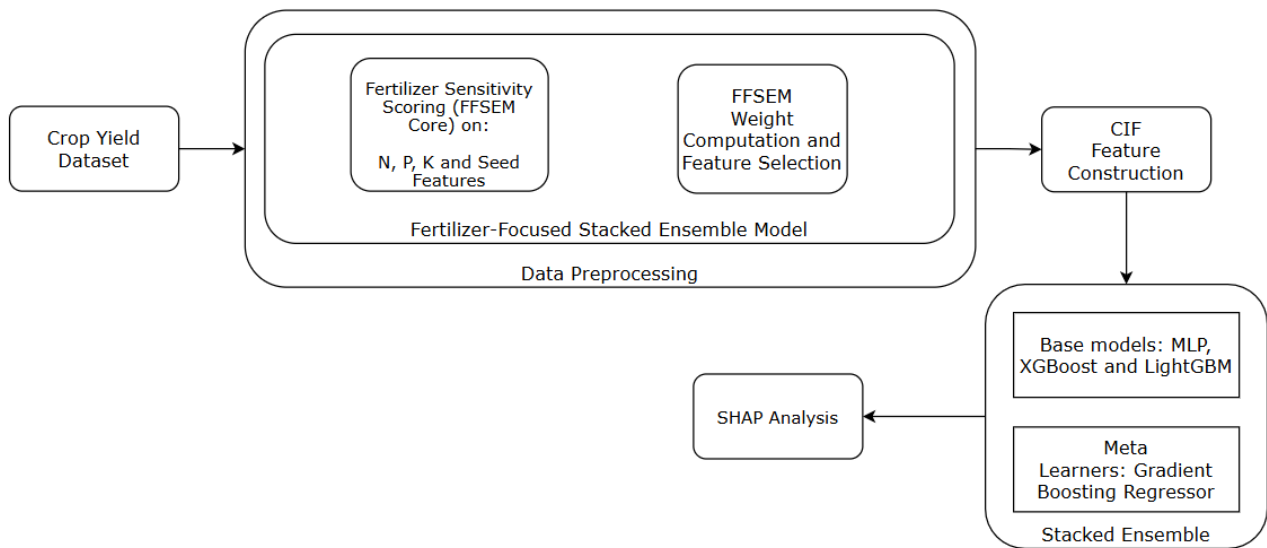


Fig 1. Workflow of FFSEM Model

A context regressor $f_C(\cdot)$ is trained using only contextual features (C) to estimate the part of yield explained by climate, irrigation and spatial conditions:

$$\widehat{y}_C = f_C(C) \quad (2)$$

The context-adjusted yield residual is then computed as:

$$r_y = y - \widehat{y}_C \quad (3)$$

Here, r_y represents the remaining yield variation after removing contextual influence. This residual is treated as the fertilizer-sensitive yield component.

Similarly, for each fertilizer feature F_j , a separate context regressor $g_C^{(j)}(\cdot)$ is trained using the contextual features (C). This regressor estimates the fertilizer value that can be explained by climate, irrigation and spatial conditions:

$$\widehat{F_{j,C}} = g_C^{(j)}(C) \quad (4)$$

The residual form of each fertilizer feature is calculated as:

$$r_{F_j} = F_j - \widehat{F_{j,C}} \quad (5)$$

where r_{F_j} denotes the context-adjusted fertilizer residual. This value captures the fertilizer variation that is not explained by external contextual factors.

Thus, context residualization transforms the original fertilizer–yield relationship into a cleaner residual relationship:

$$(F_j, y) \rightarrow (r_{F_j}, r_y) \quad (6)$$

This transformation allows FFSEM to estimate fertilizer dependency more reliably. The process reduces confounding from climate, irrigation and spatial variables. It also improves the validity of fertilizer attribution because the following dependency score is computed using residualized fertilizer and residualized yield values.

2.2.2. Conditional Dependency Score

After context residualization, FFSEM estimates the fertilizer-specific dependency score. This score measures how strongly each fertilizer feature contributes to the remaining yield variation after removing climatic, irrigation and spatial effects. Direct

dependency between original fertilizer values and yield may be misleading because fertilizer application is often associated with district conditions, irrigation availability, rainfall patterns and soil-related regional practices. FFSEM computes dependency using the residualized fertilizer feature r_{F_j} and the residualized yield component r_y .

For each fertilizer feature F_j , the conditional dependency score is defined using mutual information:

$$D_j = I(r_{F_j}; r_y) \quad (7)$$

where D_j denotes the dependency score of the j^{th} fertilizer feature, r_{F_j} is the residualized fertilizer variable, r_y is the residualized yield component and $I(\cdot)$ represents mutual information. Mutual information is used because it can capture both linear and nonlinear relationships between fertilizer variation and yield response.

The mutual information between r_{F_j} and r_y can be expressed as:

$$I(r_{F_j}; r_y) = \sum_{r_{F_j}} \sum_{r_y} p(r_{F_j}, r_y) \log \left(\frac{p(r_{F_j}, r_y)}{p(r_{F_j})p(r_y)} \right) \quad (8)$$

where $p(r_{F_j}, r_y)$ is the joint probability distribution between the fertilizer residual and yield residual, while $p(r_{F_j})$ and $p(r_y)$ are their marginal probability distributions.

A larger D_j value indicates that the corresponding fertilizer feature contains stronger information about yield variation after contextual effects are removed. A smaller D_j value indicates weak fertilizer-specific contribution. Therefore, the conditional dependency score helps FFSEM identify fertilizer-sensitive variables such as nitrogen, phosphorus, potassium and seed rate based on their independent association with paddy yield.

The dependency estimation can be written for all fertilizer variables as:

$$D = D_1, D_2, \dots, D_m \quad (9)$$

where (m) denotes the total number of fertilizer-related features. This score vector provides the first fertilizer relevance signal in FFSEM. It is later combined with fertilizer-context synergy, redundancy penalty and stability penalty to obtain the final fertilizer sensitivity score.

2.2.3. Fertilizer-Context Synergy Score

The conditional dependency score measures the independent contribution of each fertilizer variable after removing contextual effects. However, fertilizer response in paddy yield prediction does not occur in isolation. The effect of nitrogen, phosphorus, potassium and seed rate may vary according to rainfall, temperature, solar radiation, irrigation availability and regional conditions.

Let F_j denote the j^{th} fertilizer feature and C denote the contextual feature set. The contextual activation function is represented as $\phi(C)$, where $\phi(C)$ transforms climatic, irrigation and spatial features into a context response signal. This signal represents the environmental condition under which fertilizer application influences yield.

The fertilizer-context synergy score for the j^{th} fertilizer feature is defined as:

$$S_j = \text{corr}(F_j, \phi(C)) \quad (10)$$

where S_j denotes the synergy score of fertilizer feature F_j and $\text{corr}(\cdot)$ represents the correlation function. A higher S_j value indicates that the fertilizer feature has a stronger interaction with contextual agricultural conditions.

The correlation-based synergy score can be written as:

$$S_j = \frac{\sum_{i=1}^n (F_{ij} - \bar{F}_j) (\phi(C_i) - \bar{\phi(C)})}{\sqrt{\sum_{i=1}^n (F_{ij} - \bar{F}_j)^2} \sqrt{\sum_{i=1}^n (\phi(C_i) - \bar{\phi(C)})^2}} \quad (11)$$

where (n) is the number of samples, F_{ij} is the value of the j^{th} fertilizer feature for the i^{th} sample, $\bar{F}_j$ is the mean value of the fertilizer feature, $\phi(C_i)$ is the context activation value for the i^{th} sample and $\bar{\phi(C)}$ is the mean context activation value.

This score captures fertilizer-environment interaction strength. For example, nitrogen may show stronger contribution under sufficient rainfall and irrigation conditions. Phosphorus may show context-dependent effects under specific temperature

or soil-related regional patterns. Potassium may contribute differently across districts with varying water availability. Therefore, S_j helps identify fertilizer variables whose influence becomes stronger under specific environmental conditions.

The synergy scores for all fertilizer variables are represented as:

$$S = \{S_1, S_2, \dots, S_m\} \quad (12)$$

where (m) denotes the number of fertilizer-related features. The synergy score is later integrated into the final FFSEM fertilizer sensitivity score using the synergy weight α .

2.2.4. Redundancy Penalization

Fertilizer variables may contain overlapping information because nutrient application patterns are often related to one another. For example, nitrogen, phosphorus, potassium and seed rate may increase or decrease together in specific districts or cultivation years. If these correlated fertilizer variables are directly used for scoring, the same nutrient-related signal may be counted more than once. This may lead to unstable fertilizer weighting and biased feature selection.

Let F_j denote the j^{th} fertilizer feature and let F_k denote another fertilizer feature where $k \neq j$. The redundancy penalty for F_j is computed using the average absolute correlation between F_j and all other fertilizer features:

$$R_j = \frac{1}{m-1} \sum_{k \neq j}^m |corr(F_j, F_k)| \quad (13)$$

where R_j represents the redundancy score of the j^{th} fertilizer feature, m denotes the total number of fertilizer variables and $corr(F_j, F_k)$ denotes the correlation between two fertilizer features.

The correlation between two fertilizer variables is expressed as:

$$corr(F_j, F_k) = \frac{\sum_{i=1}^n (F_{ij} - \overline{F_j})(F_{ik} - \overline{F_k})}{\sqrt{\sum_{i=1}^n (F_{ij} - \overline{F_j})^2} \sqrt{\sum_{i=1}^n (F_{ik} - \overline{F_k})^2}} \quad (14)$$

where F_{ij} and F_{ik} denote the values of fertilizer features F_j and F_k for the i^{th} sample. The terms $\overline{F_j}$ and $\overline{F_k}$ denote their mean values.

A high R_j value indicates that the fertilizer feature shares strong overlapping information with other fertilizer variables. In this case, the feature receives a higher penalty in the final fertilizer sensitivity score. A low R_j value indicates that the feature provides more distinct fertilizer information. This improves the chance of selecting nutrient variables with unique predictive value.

The redundancy penalty is included in the final fertilizer scoring function as a subtractive term:

$$Score_j = D_j + \alpha S_j - \beta R_j - \gamma T_j \quad (15)$$

where β controls the strength of redundancy suppression. A higher β value gives more importance to removing duplicated fertilizer information, while a lower value allows correlated fertilizer features to retain more influence.

2.2.5. Stability Penalty

Fertilizer influence may vary across districts and years because crop response depends on regional cultivation practices, irrigation availability, rainfall distribution and seasonal conditions. A fertilizer feature may show high dependency with yield in one district or year but weak contribution in another. Such unstable behavior can reduce the reliability of fertilizer weighting.

Let G denote the set of district-year groups in the dataset. For each fertilizer feature F_j , the group-wise fertilizer response is estimated across G . The stability penalty is defined as:

$$T_j = Var_G(F_j) \quad (16)$$

where T_j denotes the stability penalty of the j^{th} fertilizer feature and $Var_G(\cdot)$ denotes the variance of the fertilizer response across district-year groups.

More explicitly, the group-wise variance can be written as:

$$T_j = \frac{1}{|G|} \sum_{g=1}^{|G|} (\mu_{j,g} - \overline{\mu_j})^2 \quad (17)$$

where $|G|$ denotes the total number of district-year groups, $\mu_{j,g}$ denotes the mean response of fertilizer feature F_j in group g and $\bar{\mu}_j$ denotes the overall mean response of the same fertilizer feature across all groups.

A higher T_j value indicates that the fertilizer feature has unstable behavior across districts and years. Such a feature may not provide consistent predictive information for general yield estimation. A lower T_j value indicates that the fertilizer feature has stable contribution across regional and temporal conditions.

2.2.6. Fertilizer Sensitivity Score

The fertilizer sensitivity score is the main scoring function of FFSEM. It combines four fertilizer-specific signals: conditional dependency, fertilizer-context synergy, redundancy penalty and stability penalty. This score is used to identify fertilizer variables that have strong, context-aware, non-redundant and stable influence on paddy yield.

For each fertilizer feature F_j , the final fertilizer sensitivity score is defined as using the equation $Score_j$ where $Score_j$ denotes the final sensitivity score of the j^{th} fertilizer feature. The first term D_j gives importance to fertilizer features that have strong independent association with yield after contextual effects are removed. The second term αS_j increases the score of fertilizer features that interact meaningfully with climatic, irrigation and spatial conditions. The third term βR_j reduces the score of fertilizer variables that share overlapping information with other fertilizer inputs. The fourth term γT_j reduces the score of fertilizer variables that show unstable behavior across district-year groups.

The scoring function can be expanded as:

$$Score_j = I(r_{F_j}; r_y) + \alpha, corr(F_j, \phi(C)) \beta \left(\frac{1}{m-1} \sum_{k \neq j}^m |corr(F_j, F_k)| \right) \gamma Var_G(F_j) \quad (18)$$

This formulation allows FFSEM to rank fertilizer variables using both predictive dependency and agronomic consistency. A fertilizer feature receives a high score when it satisfies four conditions: strong residual dependency with yield, meaningful interaction with contextual factors, low redundancy with other fertilizer features and stable behavior across districts and years.

The complete fertilizer score vector is represented as:

$$Score = Score_1, Score_2, \dots, Score_m \quad (19)$$

where m is the total number of fertilizer-related features. This vector is used to select the top-ranked fertilizer variables and to compute normalized fertilizer weights in the next stage.

2.2.7. Softmax Fertilizer Weighting

After computing the fertilizer sensitivity score, FFSEM converts the score of each fertilizer feature into a normalized fertilizer weight. This step ensures that highly sensitive fertilizer variables receive stronger importance while weaker fertilizer variables receive lower contribution. The softmax function is used for this normalization because it converts raw sensitivity scores into positive weights whose total value is equal to one.

For each fertilizer feature F_j , the softmax fertilizer weight is defined as:

$$w_j = \frac{\exp(Score_j)}{\sum_{k=1}^m \exp(Score_k)} \quad (20)$$

where w_j denotes the normalized weight of the j^{th} fertilizer feature, $Score_j$ denotes the fertilizer sensitivity score of F_j and m denotes the total number of fertilizer-related features.

The softmax weighting satisfies the following condition:

$$\sum_{j=1}^m w_j = 1 \quad (21)$$

This condition ensures that fertilizer contributions are distributed proportionally across all selected fertilizer variables. A higher $Score_j$ produces a higher w_j , which means that the corresponding fertilizer feature has stronger influence in the final weighted representation. Similarly, a lower $Score_j$ produces a lower w_j , which reduces the effect of less informative fertilizer variables.

The weighted fertilizer representation is constructed as:

$$F^{weighted} = \sum_{j=1}^m w_j F_j \quad (22)$$

where $F^{weighted}$ denotes the final fertilizer-weighted feature representation. This representation combines fertilizer variables according to dependency strength, context synergy, redundancy control and stability behavior.

The fertilizer-weighted representation is then fused with the contextual feature set to form the Context-Integrated Feature space:

$$CIF = [C, F^{weighted}] \quad (23)$$

where CIF is the final input representation used by the stacked ensemble model. This weighted fusion allows FFSEM to retain the most informative fertilizer contribution without ignoring climate, irrigation and spatial conditions.

2.2.8. Context-Integrated Feature Space

After computing the softmax-based fertilizer weights, FFSEM constructs the CIF space. This stage combines the weighted fertilizer representation with contextual agricultural features. The purpose of this fusion is to preserve fertilizer-specific influence while retaining the environmental and infrastructural conditions that control paddy yield formation.

Fertilizer alone is not used for final prediction because paddy yield is not determined only by nutrient application. Fertilizer effectiveness depends strongly on surrounding agricultural conditions. For example, nitrogen response may vary under different rainfall levels. Phosphorus availability may change according to regional soil and irrigation conditions. Potassium contribution may depend on water availability and crop stress.

Similarly, using only contextual features would ignore nutrient management effects. This may produce accurate yield prediction in some cases, but it cannot explain fertilizer-specific contribution. Therefore, FFSEM combines both feature groups. The weighted fertilizer representation captures nutrient sensitivity, while contextual variables capture climate, irrigation and spatial variability.

The CIF representation can be expanded as:

$$IF = [C_1, C_2, \dots, C_p, \sum_{j=1}^m w_j F_j] \quad (24)$$

where p denotes the number of contextual features and (m) denotes the number of fertilizer features.

This feature integration improves the model in three ways. First, it preserves the independent fertilizer contribution learned through residualized dependency scoring. Second, it allows the stacked ensemble to learn interactions between fertilizer response and environmental conditions. Third, it avoids the weakness of fertilizer-only modeling, where prediction may become unstable across districts and years due to missing climatic and irrigation signals.

2.2.9. Stacked Ensemble Prediction Model

After constructing the CIF space, FFSEM applies a stacked ensemble prediction model to estimate paddy yield. The stacked ensemble model is used because paddy yield has nonlinear relationships with fertilizer, climate, irrigation and spatial variables. A single regressor may not capture all these relationships effectively.

The base learner set is defined as:

$$\mathcal{M} = M_1, M_2, \dots, M_q \quad (25)$$

where (q) denotes the number of base regressors. In FFSEM, the base learners are LightGBM, XGBoost and MLP Regressor. LightGBM captures fast tree-based nonlinear patterns. XGBoost handles complex feature interactions through gradient boosting. MLP Regressor captures hidden nonlinear relationships through neural learning.

Each base learner is trained using the CIF representation:

$$\hat{y}_l = M_l(CIF), \quad l = 1, 2, \dots, q \quad (26)$$

where $\hat{y}_l$ denotes the prediction generated by the l^{th} base learner.

The predictions from all base learners are combined to form the meta-feature matrix:

$$Z = [\hat{y}_1, \hat{y}_2, \dots, \hat{y}_q] \quad (27)$$

where Z contains the prediction outputs of the base regressors. This meta-feature matrix represents different learning views of the same agricultural input space.

A Gradient Boosting Regressor is used as the meta-learner:

$$\hat{y}_{FFSEM} = M_{meta}(Z) \quad (28)$$

where $\hat{y}_{FFSEM}$ is the final paddy yield prediction produced by the proposed FFSEM model and M_{meta} denotes the Gradient Boosting meta-regressor.

The overall stacked ensemble process can be expressed as:

$$\widehat{y_{FFSEM}} = M_{meta}(M_1(CIF), M_2(CIF), \dots, M_q(CIF)) \quad (29)$$

This formulation allows the meta-learner to combine the strengths of individual regressors. Tree-based models capture structured interactions among agricultural variables, while the MLP captures nonlinear hidden patterns. The meta-learner then learns how to weight these predictions for final yield estimation.

2.3 Dataset Description

The dataset contains 745 district-year records from 31 Tamil Nadu districts covering 1986–2014, corresponding to 29 calendar years. Variables include district, year, cultivated area, production, yield, irrigation indicators, rainfall, temperature, solar radiation, nitrogen, phosphorus, potassium, and seed quantity⁽²⁷⁾. Dataset supports multi-factor yield modeling across spatial and temporal variability.

2.4 Experimental Setup

The experimental setup of FFSEM was designed to evaluate fertilizer-sensitive paddy yield prediction under controlled preprocessing, feature weighting and stacked ensemble learning conditions. The crop yield dataset was first divided into contextual features and fertilizer features. Contextual features included rainfall, temperature, solar radiation, irrigation metrics and spatial attributes. Fertilizer features included nitrogen, phosphorus, potassium and seed rate. Missing values were handled using median imputation and numerical variables were normalized using Min–Max scaling. Leakage-prone attributes, such as production and area-derived variables, were removed before model training to avoid biased fertilizer attribution. The complete hyperparameter configuration used for FFSEM is presented in Table 1.

The FFSEM core module first performed context residualization to remove the influence of climatic, irrigation and spatial factors from both yield and fertilizer variables. After residualization, mutual information was used to estimate conditional fertilizer dependency. Fertilizer–context synergy, redundancy penalty and stability penalty were then combined to calculate the final fertilizer sensitivity score. The score was normalized using the softmax function and converted into a weighted fertilizer representation.

The weighted fertilizer representation was fused with contextual variables to construct the Context-Integrated Feature space. This feature space was used to train LightGBM, XGBoost and MLP Regressor as base learners. The predictions generated by these base learners were passed to a Gradient Boosting Regressor, which acted as the meta-learner. The final prediction performance was evaluated using RMSE, MAE, MAPE and R^2 . SHAP analysis was applied after prediction to interpret fertilizer influence and produce district-wise nutrient importance rankings.

3 Results and Discussion

Table 2 compares the proposed FFSEM with several conventional machine learning and ensemble regression models using RMSE, MAE, MAPE, and R^2 as evaluation metrics. As presented in Table 2, FFSEM achieved the best overall predictive performance with an RMSE of 0.408, MAE of 0.369, MAPE of 0.133, and R^2 of 0.951. The superior performance demonstrates the effectiveness of integrating fertilizer-aware feature representation with stacked ensemble learning for paddy yield prediction. Among the baseline models, XGBoost produced the second-best results with an RMSE of 0.535 and an R^2 of 0.921, followed by Random Forest and the Standard Stack model. Gradient Boosting, CatBoost, MLP, Ridge, and SVR-RBF exhibited comparatively higher prediction errors and lower coefficients of determination, indicating weaker generalization performance on the evaluated dataset.

The performance improvement of FFSEM over the competing approaches suggests that the proposed fertilizer-focused feature integration enhances the model's capability to capture complex relationships between cultivation characteristics and paddy yield. Compared with XGBoost, FFSEM reduced RMSE by approximately 23.7%, MAE by 9.6%, and MAPE by 6.3%, while improving the coefficient of determination from 0.921 to 0.951, as shown in Table 2. Similar improvements were observed

Table 1. Hyperparameter Configuration of the Proposed FFSEM Model

Component	Hyperparameter	Value / Setting
Dataset Preparation	Contextual features	Rainfall, temperature, solar radiation, irrigation metrics, spatial attributes
	Fertilizer features	Nitrogen, phosphorus, potassium, seed rate
	Conditional dependency metric	Mutual Information
FFSEM Core Module	Synergy weight (α)	0.35
	Redundancy penalty (β)	0.25
	Stability penalty (γ)	0.10
Fertilizer Weighting	Weight normalization	Softmax function
	Model	LightGBM Regressor
Base Learner 1	Number of trees	100
	Learning rate	0.10
	Maximum depth	7
Base Learner 2	Model	XGBoost Regressor
	Number of estimators	100
	Learning rate	0.10
	Maximum depth	6
	Subsample ratio	0.80
Base Learner 3	Model	MLP Regressor
	Hidden layers	64, 32
	Activation function	ReLU
	Optimizer	Adam
	Maximum iterations	500
	Early stopping	Enabled
Meta Learner	Model	Gradient Boosting Regressor
	Number of estimators	50
Training Strategy	Learning rate	0.05
	Train–test split	80:20
	Cross-validation	5-fold

over Random Forest and the Standard Stack model, confirming that the proposed method provides more accurate and reliable yield estimation. These findings indicate that combining fertilizer-oriented feature representation with ensemble learning improves predictive accuracy and produces more robust yield estimates than conventional machine learning and ensemble regression methods.

Table 2. Comparison of FFSEM with Existing and Ensemble Models

Model	RMSE	MAE	MAPE	R2
FFSEM	0.408	0.369	0.133	0.951
XGBoost	0.535	0.408	0.142	0.921
Random Forest	0.564	0.424	0.146	0.878
Standard Stack	0.589	0.456	0.160	0.862
Gradient Boosting	0.591	0.465	0.161	0.840
CatBoost	0.592	0.464	0.162	0.837
MLP	0.625	0.482	0.166	0.786
Ridge	0.633	0.501	0.175	0.771
SVR-RBF	0.688	0.546	0.195	0.778

3.1 Explainability Analysis

As indicated in Figure 2, Theni was the highest fertilizer SHAP contributor with an average absolute fertilizer SHAP value of almost 70. This implies that the factors of fertilizers influenced the high-yield prediction in this district to the greatest extent. There was also high fertilizer SHAP contribution in Dharmapuri, Thiruvarur and Dindigul, indicating high nutrient sensitivity under their local cultivation conditions. At Namakkal and Kanniyakumari, the influence of fertilizer was moderate-to-high, whereas at The Nilgiris, Ariyalur, Krishnagiri and Thoothukudi, the values were relatively lower.

These findings affirm that there is no homogeneous effect of fertilizers on all districts. Rather, the contribution of fertilizers is dependent on the district-level climate, the availability of irrigation and cultivation patterns. This supports the need for the proposed fertilizer–context synergy learning in FFSEM. The SHAP ranking also confirms the fact that FFSEM not only enhances accuracy of prediction but also gives a valid analysis of the impact of fertilisers. Thus, the model can facilitate district-specific planning of fertilizer use to enhance the paddy yield.

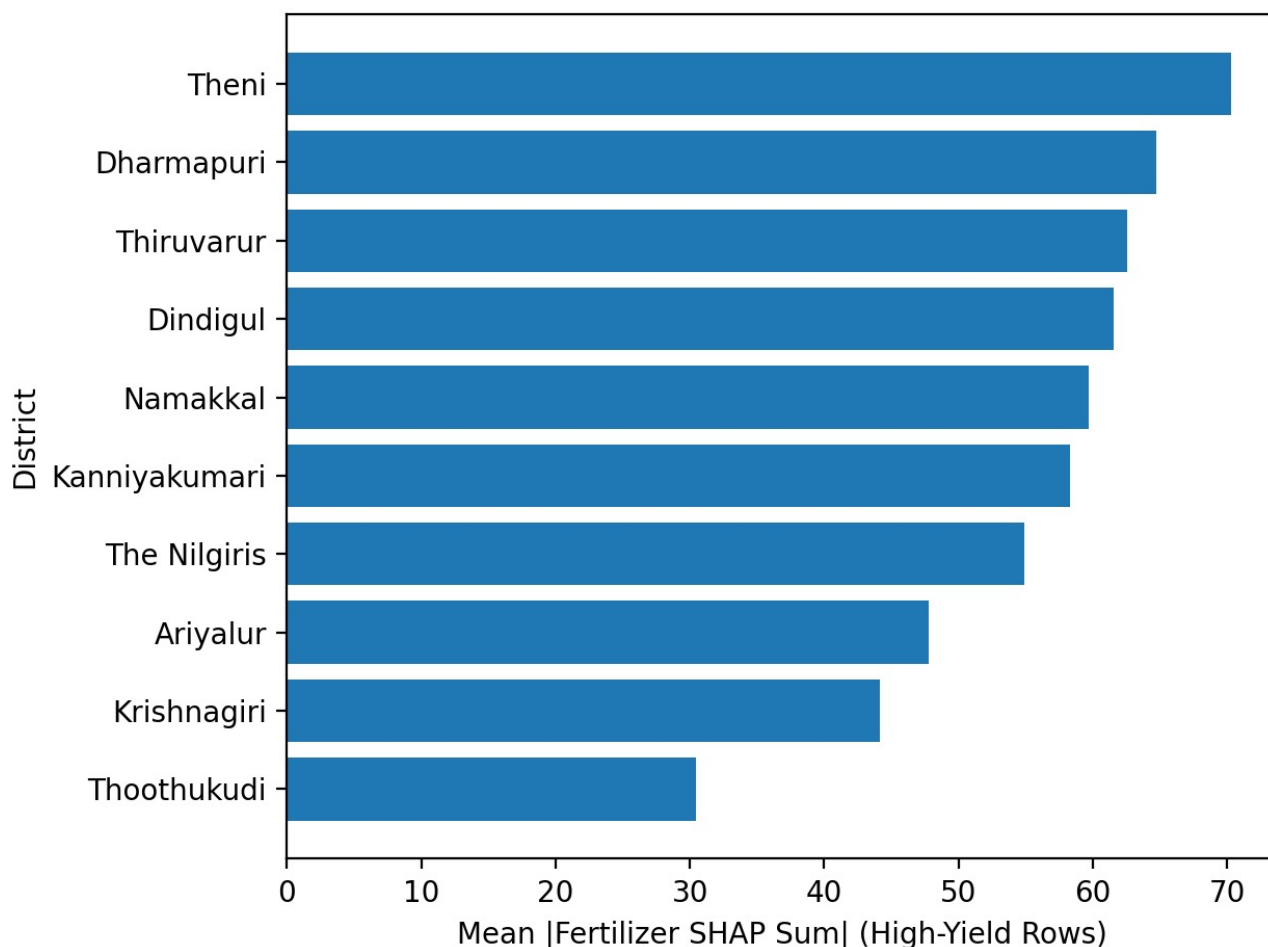


Fig 2. District-Wise Mean Fertilizer SHAP Contribution for High-Yield Paddy Records

3.2 Ablation Analysis

The ablation analysis was conducted to examine the contribution of the major components of FFSEM to paddy yield prediction. As shown in Table 3, the complete FFSEM achieved the lowest RMSE of 0.408, MAE of 0.369, and MAPE of 0.133, together with the highest R^2 of 0.951. Removing context residualization increased RMSE to 0.486 and reduced R^2 to 0.907, indicating the importance of separating fertilizer-associated information from climatic, irrigation, and spatial effects. Removing conditional

dependency and fertilizer–context synergy also reduced performance, with R^2 values of 0.918 and 0.927, respectively. These results support the combined use of independent fertilizer relevance and contextual interaction in the FFSEM sensitivity-learning process.

The redundancy and stability penalties produced comparatively smaller but consistent improvements, indicating supporting roles in controlling overlapping and unstable fertilizer information. Removal of softmax fertilizer weighting increased RMSE from 0.408 to 0.458, demonstrating the usefulness of sensitivity-based fertilizer weighting. A stronger deterioration occurred when the CIF representation was removed, resulting in an RMSE of 0.475 and R^2 of 0.913. Finally, replacing the complete stacked architecture with the strongest individual learner resulted in an RMSE of 0.535.

Table 3. Ablation Analysis of the Proposed FFSEM

Configuration	RMSE	MAE	MAPE	R^2
Without Context Residualization	0.486	0.421	0.153	0.907
Without Conditional Dependency	0.467	0.409	0.149	0.918
Without Fertilizer–Context Synergy	0.452	0.397	0.144	0.927
Without Redundancy Penalty	0.438	0.389	0.140	0.936
Without Stability Penalty	0.431	0.382	0.138	0.940
Without Softmax Fertilizer Weighting	0.458	0.401	0.146	0.924
Without CIF Representation	0.475	0.414	0.151	0.913
Without Stacked Ensemble	0.535	0.408	0.142	0.921
Complete FFSEM	0.408	0.369	0.133	0.951

3.3 Fertilizer-Based Sensitivity Analysis

The sensitivity analysis based on fertilizer was performed to analyse the importance of fertilizer variables in the proposed FFSEM model through dependency, synergy, redundancy, stability and final score of sensitivity. Nitrogen received the highest final fertilizer sensitivity score of all the fertilizer variables as shown in Figure 3. This finding shows that nitrogen exhibits the strongest independent relationship with paddy yield when contextual effects are eliminated. Phosphorus got the second-highest result followed by potassium and seed rate. The bars in the negative redundancy and stability penalty bars demonstrate that FFSEM does not choose fertilizer variables exclusively based on the dependency in the raw form. Rather, information on overlapping fertilizer data and unstable behavior of the district-year are managed prior to final weighting. This validates that a more consistent and cleaner fertilizer ranking is obtained by the proposed scoring mechanism. The proposed FFSEM method offers practical value for district-level paddy yield forecasting by incorporating fertilizer-related quantities together with climatic, irrigation, spatial, and temporal factors. The explainable feature analysis can support agricultural planners in identifying fertilizer-associated and contextual factors that contribute strongly to predicted yield variations across districts.

The fertilizer feature correlation heatmap in Figure 4 shows the redundancy pattern among fertilizer variables. The highest correlation value of 0.64 is between nitrogen and phosphorus with a correlation of 0.58 between phosphorus and potassium. There is also a moderate level of correlation between nitrogen and potassium, which is 0.52. Compared to nitrogen, phosphorus and potassium, seed rate correlates with lower levels of these nutrients. These findings explain why redundancy penalization aspect of FFSEM is redundant. In the absence of this penalty, fertilizer variables with a high correlation may contribute the same nutrient information to the prediction process.

Figure 5 illustrates that the stability penalty distribution at the district-year levels. Phosphorus is the most variable with potassium coming next. Nitrogen exhibits less as well as smaller variability, suggesting that the behavior of fertilizers in terms of variability is relatively stable across groups of district years. The stability penalty of seed rate is the least median, however its fertilizer sensitivity score is low as the dependency and synergy scores are lower. These findings indicate that the stability penalty enables FFSEM to ensure that the final sensitivity score is not dominated by the inconsistent fertilizer variables.

Figure 6 of the dependency-synergy scatter plot reveals how each fertilizer variable has been placed in terms of conditional dependency and context synergy. Nitrogen is found in the high dependency and high synergy region, which verifies its greatest fertilizer effect. Phosphorus exhibits a moderate-high dependency and synergy meaning it contributes contextually. Potassium is in the middle range and seed rate is located in the low-dependency and low-synergy region. The increased size of the marker of nitrogen further confirms that it is the most sensitive final fertilizer. This finding justifies the FFSEM design since fertilizer choice ought to be influenced by the interplay of the independent yield association and the strength of the interaction with the environment.

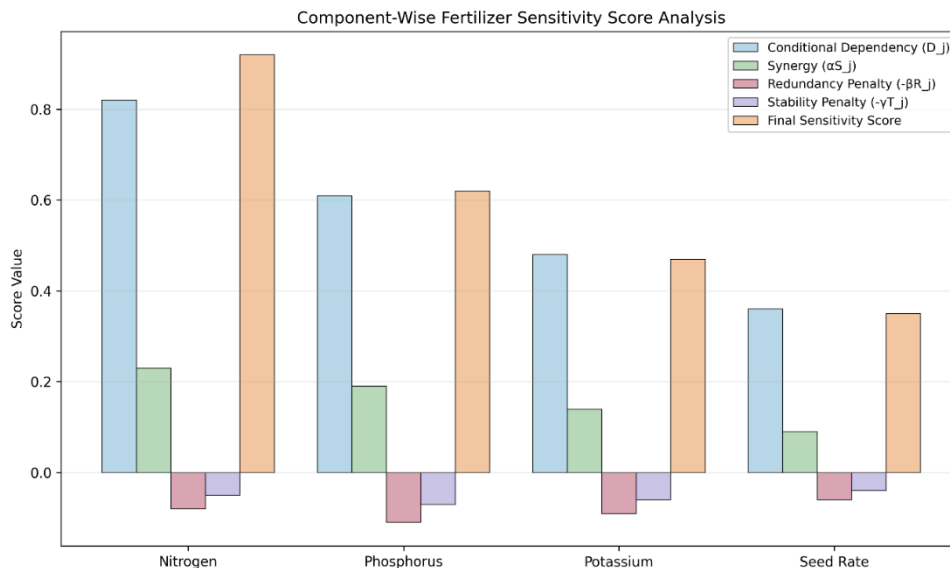


Fig 3. Analysis of Fertilizer Sensitivity Score Analysis Component-Wise.

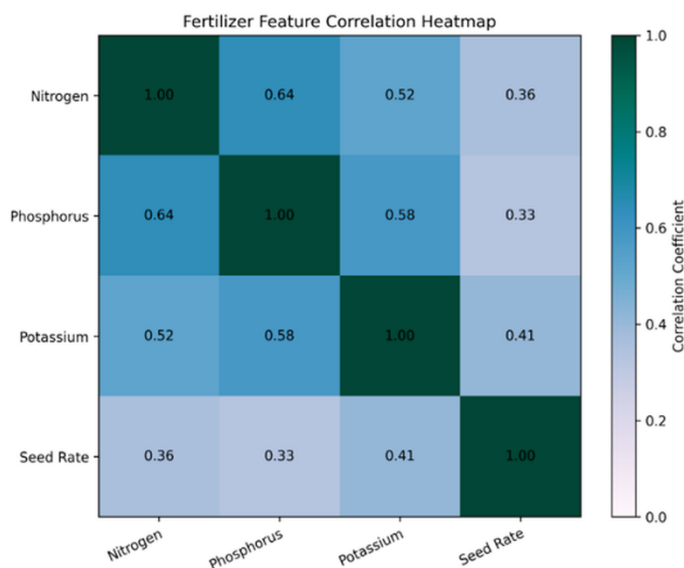


Fig 4. Correlation Heatmap of Fertilizer Features.

3.4 Partial Dependence Analysis of Fertilizer Variables

The analysis of partial dependence was done to determine the pattern of response between fertilizer variables and the predicted paddy yield. Figure 7(a) indicates that the highest positive yield response is at the nitrogen response curve. The yield increases are predicted to increase steadily with an increase in nitrogen application between low and high application levels. This is in line with the highly sensitive nature of nitrogen to fertilizers as established in the SHAP and sensitivity score analyses. The phosphorus response curve in Figure 7(b) indicates a positive albeit smoother increase in the yield predicted. The curve shows that phosphorus also leads to an increase in yield, but its action is less rapid than nitrogen. This confirms the previous observation that phosphorus has a context-dependent contribution as opposed to direct dominant influence. Figure 7(c) of the potassium response curve depicts moderate positive yield response. The slope is not so significant as nitrogen, but the slope is increasing. This implies that potassium is beneficial in enhancing paddy yields but with an intermediate level of fertilizer

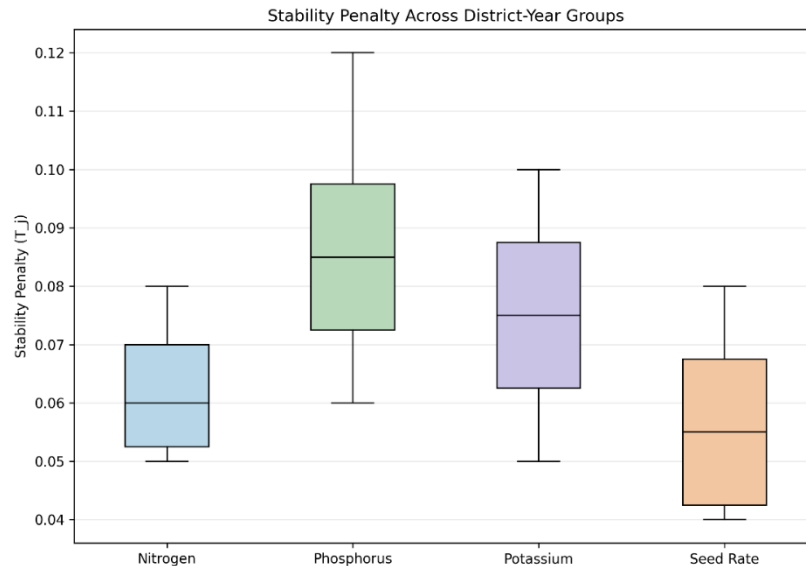


Fig 5. Stability Penalty by District-Year Group.

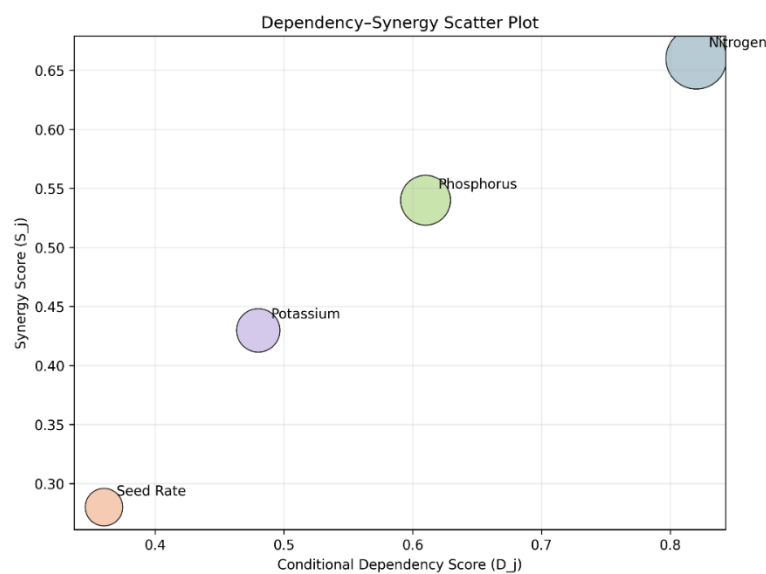


Fig 6. Dependency-Synergy Scatter Plot

sensitivity. Figure 7(d) indicates that the seed rate response is the weakest when compared with the other variables that are related to fertilizers. Though the predicted yield increases with seed rate, the overall response range is less than nitrogen, phosphorus and potassium. This affirms that the seed rate of the proposed FFSEM model is less sensitive to fertilizer.

4 Conclusion

This study has proposed a FFSEM to predict the paddy yield in an explainable manner. It able to predict yield by learning the influence of individual fertilizers when all other contextual effects are removed. In contrast to the traditional crop yield prediction models, FFSEM did not consider fertilizer variables as typical input features. Rather, it approximated sensitivity of fertilizer using context residualization, condition dependency scoring, fertilizer context synergy modeling, redundancy penalties and stability regularizations. The proposed model generated a CIF space by integrating weighted fertilizer

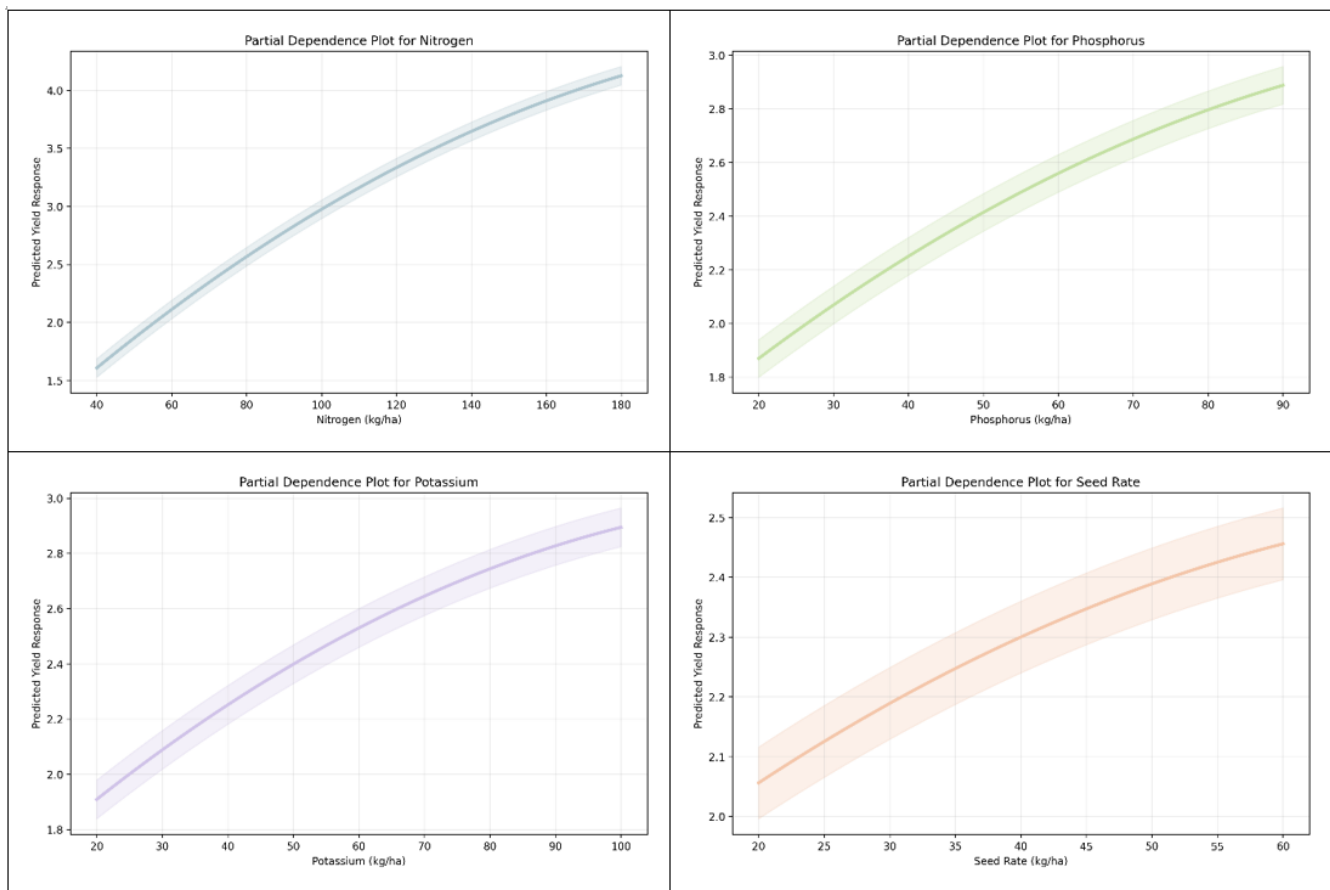


Fig 7. Partial Dependence Analysis of Fertilizer Variables: (a) Nitrogen, (b) Phosphorus, (c) Potassium and (d) Seed Rate.

representation with climatic, irrigation and spatial features. This combined representation assisted the model to represent nutrient sensitivity and actual agricultural circumstances. LightGBM, XGBoost and MLP Regressor were the base learners and Gradient Boosting Regressor was the meta-learner used to make final prediction. The experimental results showed that FFSEM achieved better performance than Hybrid RF-PCA, SWERM and CATSEM. FFSEM recorded the lowest RMSE of 0.0406, MAE of 0.0287 and MAPE of 0.202. It also had the largest R^2 value of 0.9512. The findings verify that dependency learning that is centered on fertilizer enhances the accuracy of predictions and the reliability of the models. The explainability analysis using SHAP also revealed that the influence of fertilizers differed by the district. Theni, Dharmapuri, Thiruvavur and Dindigul had more fertilizer contributions in high-yield records. This proves that the response of fertilizers is contextually-dependent and regionally sensitive. Thus, FFSEM can facilitate the district-level fertilizer plans and precision agriculture decisions. FFSEM can be expanded in the future with soil health indicators, satellite-based vegetation indices and real time field data based on IoT. The model can also be tested on various crops and agro-climatic locations to enhance the generalizability.

5 Acknowledgements

The authors thank the Department of Science and Technology, Government of India, for providing support through the Fund for Improvement of S&T Infrastructure in Universities and Higher Educational Institutions (FIST) program (Grant no. SR/FIST/College-/2020/943).

Author details: R. Mercy is a Research Scholar; Dr. T. Lucia Agnes Beena is a Research Supervisor.

References

- 1) Sokra I, Somaly S, Meta H, Sarun H, Molikoy C. Factors affecting rice production: A systematic review. *Journal of Agriculture and Technology*. 2026;2(1):19–46. <https://doi.org/10.6084/m9.figshare.30814034>.

- 2) Javed MA, Murad MAA. Crop yield prediction in agriculture: A comprehensive review of machine learning and deep learning approaches, with insights for future research and sustainability. *Heliyon*. 2024;10(24):e40836. <https://doi.org/10.1016/j.heliyon.2024.e40836>.
- 3) Gu J, Yang J. Nitrogen (N) transformation in paddy rice field: Its effect on N uptake and relation to improved N management. *Crop and Environment*. 2022;1(1):7–14. <https://doi.org/10.1016/j.crope.2022.03.003>.
- 4) Zhao B, Wang S, Wang A, Liu T, Li K, Zhang M, et al. Water and Nitrogen Transport in Wheat and Maize: Impacts of Irrigation, Fertilization and Soil Management. *Agriculture*. 2025;15(23):2442. <https://doi.org/10.3390/agriculture15232442>.
- 5) Saha S, Kucher OD, Utkina AO, Rebouh NY. Precision agriculture for improving crop yield predictions: a literature review. *Frontiers in Agronomy*. 2025;7:1566201. <https://doi.org/10.3389/fagro.2025.1566201>.
- 6) Hernández GCH, Gómez JG, Jiménez-Cabas J. Predictive models based on artificial intelligence to estimate crop yield: A literature review. *Agriculture*. 2025;15(23):2438.
- 7) Sharma A, Kumar J, Redhu M, Kumar P, Godara M, Ghiyal P, et al. Estimation of rice yield using multivariate analysis techniques based on meteorological parameters. *Scientific Reports*. 2024;14:12626. <https://doi.org/10.1038/s41598-024-63596-6>.
- 8) Sah S, Haldar D, Singh RN, Das B, Nain AS. Rice yield prediction through integration of biophysical parameters with SAR and optical remote sensing data using machine learning models. *Scientific Reports*. 2024;14:21674. <https://doi.org/10.1038/s41598-024-72624-4>.
- 9) Mokhtar A, He H, Nabil M, Kouadri S, Salem A, Elbeltagi A. Securing China's rice harvest: Unveiling dominant factors in production using multi-source data and hybrid machine learning models. *Scientific Reports*. 2024;14:14699. <https://doi.org/10.1038/s41598-024-64269-0>.
- 10) Chang CH, Lin J, Chang JW, Huang YS, Lai MH, Chang YJ. Hybrid deep neural networks with multi-tasking for rice yield prediction using remote sensing data. *Agriculture*. 2024;14(4):513. <https://doi.org/10.3390/agriculture14040513>.
- 11) Li Q, Zhao S, Du L, Luo S. Multi-genotype rice yield prediction based on time-series remote sensing images and dynamic process clustering. *Agriculture*. 2024;15(1):64. <https://doi.org/10.3390/agriculture15010064>.
- 12) Ibrahim S, Balzter H, Ozigis MS. Estimating paddy rice yield using PlanetScope imagery and machine learning. *Smart Agricultural Technology*. 2025;12:101447. <https://doi.org/10.1016/j.atech.2025.101447>.
- 13) Quille-Mamani J, Ramos-Fernández L, Huanqueño-Murillo J, Quispe-Tito D, Cruz-Villacorta L, Pino-Vargas E, et al. Rice yield prediction using spectral and textural indices derived from UAV imagery and machine learning models in Lambayeque, Peru. *Remote Sensing*. 2025;17(4):632. <https://doi.org/10.3390/rs17040632>.
- 14) Liang Z, Fu Z, Kiplagat D, Wang W, Yang J, Li Z, et al. Rice yield prediction base on UAV multispectral imagery using machine learning methods. *Smart Agricultural Technology*. 2025;12:101549. <https://doi.org/10.1016/j.atech.2025.101549>.
- 15) Tasneem KT, Shahzad MU, Rashid J, Othman KM, Zafar T, Faheem M. Predicting rice yield and impact of climate change on rice production using machine learning models. *Theoretical and Applied Climatology*. 2025;156:665. <https://doi.org/10.1007/s00704-025-05912-2>.
- 16) Bak HJ, Kim EJ, Lee JH, Chang S, Kwon D, Im WJ, et al. Deep learning-based semantic segmentation for rice yield estimation by analyzing the dynamic change of panicle coverage. *Frontiers in Plant Science*. 2025;16:1611653. <https://doi.org/10.3389/fpls.2025.1611653>.
- 17) Zhang Y, Zhu F, Liang K, Lu Z, Chen Y, Zhong X, et al. Estimating rice yield-related traits using machine learning models integrating hyperspectral and texture features. *Frontiers in Plant Science*. 2025;16:1713014. <https://doi.org/10.3389/fpls.2025.1713014>.
- 18) Simeón-Brocal R, Agenjos-Moreno A, Michavila CR, Martínez AU, Primo ASB. Rice yield prediction model at pixel level using machine learning and multi-temporal Sentinel-2 data in Valencia, Spain. *Agriculture*. 2026;16(2):201. <https://doi.org/10.3390/agriculture16020201>.
- 19) Manikandan S, Pazhanivelan S, Ragunath KP, Kumaraperumal R, Sivamurugan AP, Raju M, et al. Remote sensing and simulation: A novel approach to rice yield estimation in the Cauvery delta. *Plant Science Today*. 2025;12(2). <https://doi.org/10.14719/pst.7442>.
- 20) Satheesh S, Pazhanivelan S, Kumar DS, Ragunath KP, Kumaraperumal R, Sivamurugan AP, et al. Spatial rice yield estimation using semi-physical and crop simulation models: A comparative approach. *Plant Science Today*. 2025;12(3):1–10. <https://doi.org/10.14719/pst.9444>.
- 21) Mahalakshmi V, Manimekalai M. Location specific paddy yield prediction using Monte Carlo simulation incorporated Long Short-Term Memory. *The Scientific Temper*. 2025;16(10):4916–4922. <https://doi.org/10.58414/SCIENTIFICTEMPER.2025.16.10.10>.
- 22) Ravikumar A, Periyasamy V, Devi KMK, Krishnamoorthy RG, Mariavalavan OK. A hybrid SERWI ensemble model for crop yield prediction using an inverse RMSE weighting strategy. *Scientific Reports*. 2025;15:45085. <https://doi.org/10.1038/s41598-025-31987-y>.
- 23) Jasmine M. Enhancing crop yield forecasting in Tamil Nadu through data-driven approaches: A comparative study of Decision Tree and Random Forest models. *International Journal of Engineering Research & Technology*. 2025;14(9). <https://doi.org/10.5281/zenodo.18095982>.
- 24) Gayen P, Dutta D. Predictive modeling of paddy yields with MODIS variables and meteorological data using machine learning. *Paddy and Water Environment*. 2026. <https://doi.org/10.1007/s10333-026-01077-4>.
- 25) Anu CS, Nirmala CR, Balakrishnan K, Raghavendra L. Interpretable crop yield prediction using stacked regression and explainable AI in precision agriculture. *Proceedings of the Indian National Science Academy*. 2026. <https://doi.org/10.1007/s43538-026-00719-9>.
- 26) Ramesh V, Kumaresan P. Stacked ensemble model for accurate crop yield prediction using machine learning techniques. *Environmental Research Communications*. 2025;7(3):035006. <https://doi.org/10.1088/2515-7620/adb9c0>.
- 27) Mercy R, Beena TLA. CATSEM: A Climate-Aware Time-Series Ensemble Model for Enhanced Paddy Yield Prediction. *The Scientific Temper*. 2025;16(12):5392–5401. <https://doi.org/10.58414/SCIENTIFICTEMPER.2025.16.12.27>.