

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452

NAAS Rating (2025): 4.49

Maths 2025; 10(9): 49-55

© 2025 Stats & Maths

<https://www.mathsjournal.com>

Received: 16-07-2025

Accepted: 18-08-2025

Purnima Sahu

M.Sc. Scholar, Department of
Agricultural Statistics and Social
Science (L.), College of
Agriculture, IGKV, Raipur,
Chhattisgarh, India

Sweta Ramole

Professor and HOD, Department
of Agricultural Statistics and
Social Science (L.), College of
Agriculture, IGKV, Raipur,
Chhattisgarh, India

Tripti Swarnkar

Professor and HOD, Department
of Computer Application, NIT,
Raipur, Chhattisgarh, India

Corresponding Author:

Purnima Sahu

M.Sc. Scholar, Department of
Agricultural Statistics and Social
Science (L.), College of
Agriculture, IGKV, Raipur,
Chhattisgarh, India

Application of artificial neural networks for rice and wheat crop yield prediction in Chhattisgarh

Purnima Sahu, Sweta Ramole and Tripti Swarnkar

DOI: <https://www.doi.org/10.22271/math.2025.v10.i9a.2162>

Abstract

Climate variability and increasing demand for food necessitate accurate crop yield prediction models to enhance agricultural planning and ensure food security. This study evaluates the predictive performance of three modeling techniques: Artificial Neural Network (ANN), Multiple Linear Regression (MLR), and Ridge Regression for rice and wheat yield forecasting in the Chhattisgarh Plains. Using 27 years (1997-2023) of climatic and yield data, including rainfall, temperature, humidity, and wind parameters, the models were trained and validated on standardized datasets. Results indicate that ANN outperformed other methods, achieving the highest accuracy ($R^2 = 0.87$ for rice and 0.94 for wheat) with the lowest error rates. Ridge Regression provided moderate accuracy, while MLR recorded the weakest performance with low explanatory power and higher errors. The findings underscore ANN's effectiveness in capturing complex non-linear relationships between climatic variables and crop yields, making it a promising tool for precision agriculture and climate-adaptive farming strategies.

Keywords: Artificial neural network, ridge regression, multiple linear regression, crop yield prediction, climate variability

1. Introduction

Climate variability poses significant challenges to agricultural productivity, necessitating accurate and robust crop yield prediction models to support sustainable farming practices [Pooja *et al.* (2025)]^[11]. As climate anomalies continue to increase in frequency and intensity, the need for accurate crop yield predictions becomes imperative for the assurance of food security and the optimization of resource management [Ray *et al.* (2015)]^[12]. Crop yield prediction is one of the most difficult tasks in precision agriculture. The ability to forecast crop yields enables the relevant authorities to make appropriate decisions to ensure food security. In addition to soil, genotype, and management techniques, weather conditions have a significant impact on crop yield [Singh *et al.* (2014)]^[14]. Traditional forecasting methods frequently face significant challenges in accurately representing the intricate relationships between climate variables and agricultural productivity [Chlingaryan *et al.* (2018)]^[6]. The comparative analysis revealed that, among the evaluated models, Multiple Linear Regression, Ridge Regression, and Artificial Neural Network, the ANN consistently outperformed the traditional approaches, achieving the highest predictive accuracy and demonstrating superior capability in capturing complex, non-linear relationships between climatic variables and crop yields [Satpathi *et al.* (2023)]^[13]. However, further research is needed in crop yield prediction. The primary objective of this study is to evaluate the effects on crop yield prediction through the implementation of ANN, MLR, and Ridge Regression. This study analyzes historical climate and yield data to determine which model can provide the most accurate crop yield predictions.

2. Materials and Methods

2.1 Study Area

This study examines crop yield prediction in the Chhattisgarh Central Plains, a major agricultural region of India known for rice and wheat production. Covering 1, 35, 194 sq. km, the area's farming is shaped by diverse soils, fluctuating temperatures, variable rainfall, and irrigation from rivers (Mahanadi, Shivanth, Hasdeo, Arpa) and canals. Geographically located between 17°46'-24°06' N latitude and 80°15'-84°20' E longitude, the region has three main

agricultural zones: Raipur, Bilaspur, and Durg, which were later subdivided into smaller districts. For analysis, historical weather and crop yield data were consolidated at the regional level to maintain continuity. These regions were chosen for their agricultural importance, data availability, and climatic

variability. The research utilized a dataset collected from various websites between 1997 and 2023. These sources included the Directorate of Economics and Statistics (DES), ICRISAT-District level data, and the meteorological website <https://power.larc.nasa.gov>.

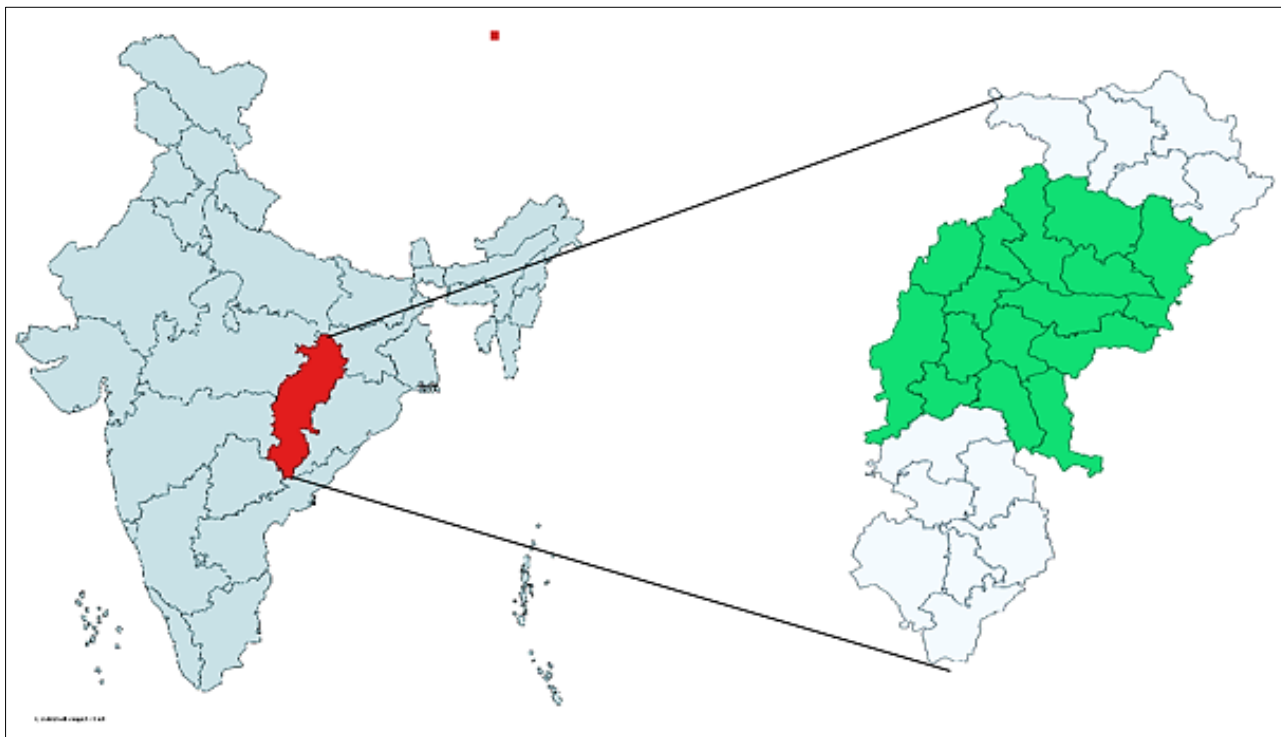


Fig 1: Location map of the study area consisting of different central plains in Chhattisgarh.

2.2 Methodology

In this study, crop yield prediction was evaluated using machine learning (ML) model as well as traditional models. We collected historical climate data and crop yield records and performed data preprocessing to handle missing values, normalize continuous variables, and identify key climatic indicators affecting crop growth. Three models, ANN, MLR, and Ridge Regression, were implemented using R software. The dataset was divided into training (80%) and testing (20%) subsets to ensure an unbiased evaluation. Model performance was assessed using statistical metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2).

2.3 Models

2.3.1 Artificial Neural Network

Artificial Neural Networks are computational models inspired by the biological neural networks of the human brain. The human brain consists of billions of neurons working together as a highly complex and efficient network. Similarly, ANNs are made up of interconnected artificial neurons (nodes) that process data collaboratively. These models are particularly useful for solving problems involving pattern recognition,

classification, and prediction. The architecture of ANNs allows them to capture complex and nonlinear relationships in data, distinguishing them from traditional statistical methods. An ANN typically consists of three types of layers: an input layer, one or more hidden layers, and an output layer. The input layer receives external data that represents the features or variables of interest. The hidden layers contain neurons that perform computations on the input data and learn underlying patterns. Networks with multiple hidden layers are called “deep” neural networks. Finally, the output layer provides the network’s decision or prediction. Each connection between neurons has an associated weight that signifies the strength or influence of the input on the network’s output. ANN has its ability to learn from data. The learning process generally involves two phases: forward propagation and back-propagation. In forward propagation, the input data passes through the network, layer by layer, until an output is generated. The predicted output is then compared with the actual result, and the error is calculated. During back-propagation, this error is sent backward through the network, and the connection weights are adjusted to minimize the error. This cycle repeats multiple times, allowing the network to gradually improve its accuracy and make better predictions.

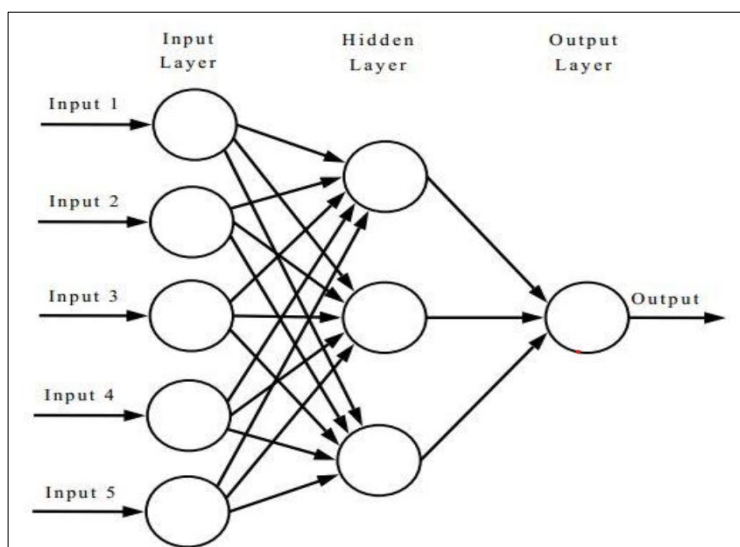


Fig 2: Basic Artificial Neural Network model

2.3.2 Multiple Linear Regression

Multiple regression analysis involves incorporating several independent variables simultaneously to estimate the impact on a dependent variable (Snedecor and Cochran, 1967). In this study, the multiple linear regression equation was developed with crop yield as the dependent variable and various weather parameters as the independent variables. The generalized representation of the multiple regression model is provided below.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \epsilon$$

In the regression model, Y represents the dependent variable, while β_0 denotes the intercept. The term β_i corresponds to the regression coefficient of the i th independent parameter ($i = 1, 2, \dots, n$), and X_i represents the i th weather parameter. The model also includes an error term, ϵ , which accounts for unexplained variability in the dependent variable.

2.3.3 Ridge Regression

Ridge Regression, also known as Tikhonov regularization, is a traditional statistical model and a modified form of Multiple Linear Regression. It improves model stability and prediction accuracy by shrinking the coefficients toward zero. This method is especially useful when explanatory variables suffer from multicollinearity, a condition where predictors are highly correlated with each other. In such cases, the coefficients estimated by Ordinary Least Squares become unstable and sensitive to small data changes. Ridge Regression addresses this by adding a penalty term to the regression objective function. Similar to OLS, Ridge Regression minimizes the Residual Sum of Squares (RSS), but it adds a tuning parameter λ (lambda) as a penalty to control coefficient size. The objective function is:

$$\text{Min} \sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij})^2 + \lambda \sum_{j=1}^p |\beta_j|^2$$

In the model, y_i represents the dependent variable or response variable, while x_{ij} denotes the j -th predictor variable for the i -th observation. β_0 is the intercept term, and β_j is the coefficient corresponding to predictor j . The total number of predictors is represented by p , and the number of observations is n . Additionally, λ is the tuning parameter that controls the size of the penalty applied in the model.

Model accuracy evaluation: The coefficient of determination (R^2) employed in this study to evaluate the

extent to which the explanatory variables align with crop production outcomes. The precision of the crop production estimations was assessed through the use of RMSE and MSE.

The formulas are;

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2}$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}$$

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

Where, N stands for the number of data, y_i for the actual yield value and $\hat{y}_i$ for the predicted model value. For optimal model performance, it is crucial that R^2 is maximized and RMSE, MSE and MAE are minimized.

4. Results and Discussion

This section presents the results of three models ANN, MLR and Ridge Regression in predicting rice and wheat yield in the central plains of Chhattisgarh using historical weather and crop yield data. The dataset was split into 80% for training and 20% for testing to ensure unbiased evaluation. Key performance metrics such as RMSE, MSE and R^2 were used to assess model effectiveness. Before implementation, we will load and preprocess the climate and crop yield datasets.

Table 1: Performance Correlation between weather parameters and crop yield prediction

Weather Parameters	Rice	Wheat
Maximum temperature	0.1294	-0.1073
Minimum temperature	-0.4903	-0.2113
Rainfall(RF)	0.1051	-0.0108
Relative humidity(RH)	-0.2745	0.1778
Wind speed (WS)	-0.4293	-0.3231

Table: 1 shows the correlation between various weather parameters and the crop yields of rice and wheat. Maximum temperature has a slight positive effect on rice but a small negative effect on wheat. Minimum temperature negatively affects both crops, especially rice. Rainfall shows a mild

positive correlation with rice and almost no effect on wheat. Relative humidity has a negative correlation with rice but a positive one with wheat. Lastly, wind speed negatively impacts both crops, more so for rice. These correlations help in understanding how the weather influences crop yields.

Predicting crop yield using the artificial neural network

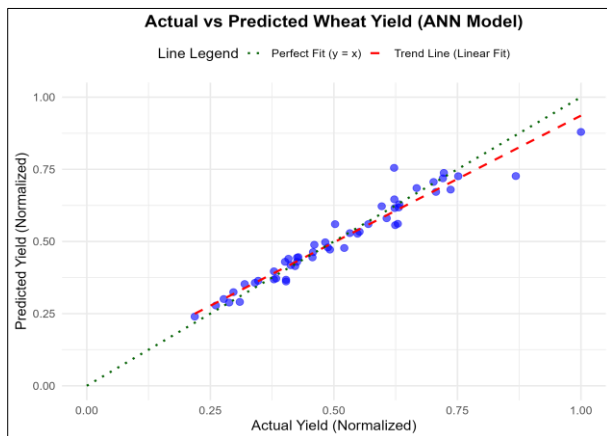
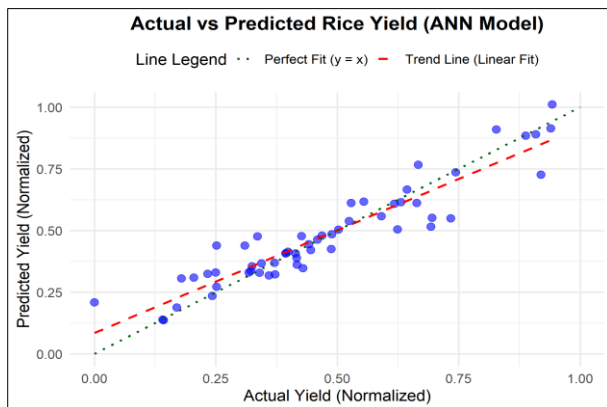


Fig 2: Actual vs Predicted rice and wheat yield using artificial neural network

Table 2: Performance of Artificial Neural Network model

Performance Metrics	Rice	Wheat
RMSE	0.0799	0.0413
MSE	0.0063	0.0017
R ²	0.8688	0.9357

Predicting crop yield using the Multiple Linear Regression

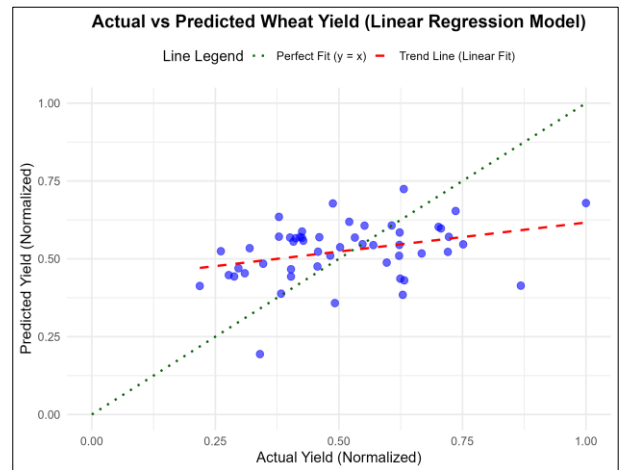
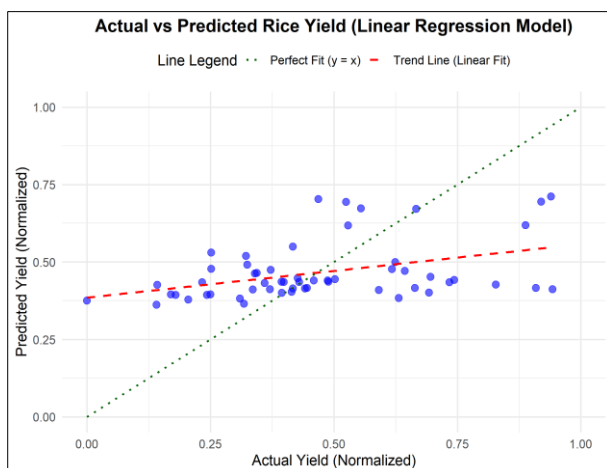


Fig 3: Actual vs Predicted rice and wheat yield using Multiple Linear Regression

Table 3: Performance of Multiple Linear Regression model

Performance Metrics	Rice	Wheat
RMSE	0.2017	0.1591
MSE	0.0407	0.0253
R ²	0.3431	0.1605

Predicting crop yield using the Ridge Regression

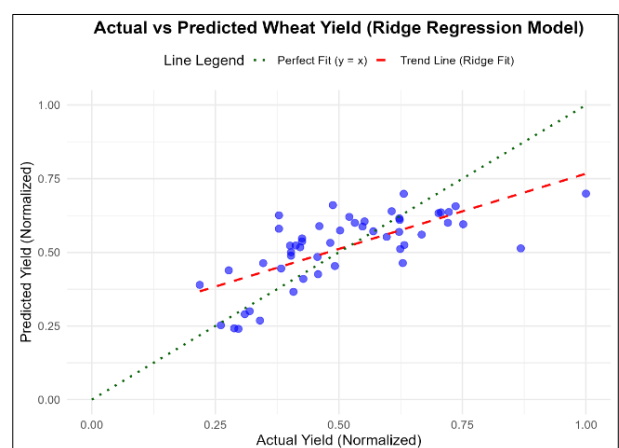
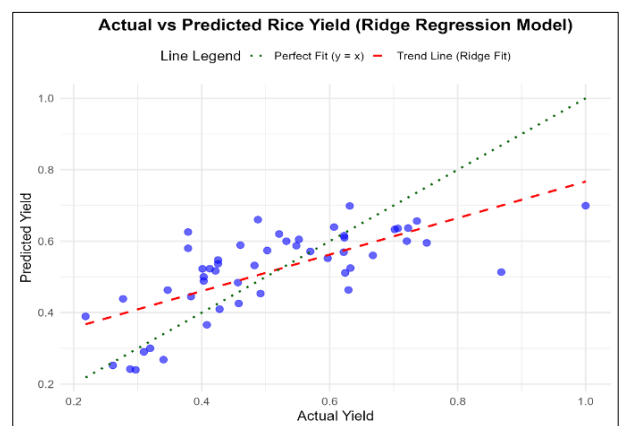


Fig 4: Actual vs Predicted rice and wheat yield using Ridge Regression

Table 4: Performance of Ridge Regression Model

Performance Metrics	Rice	Wheat
RMSE	0.1255	0.1168
MSE	0.0157	0.0136
R ²	0.6768	0.4874

Model Comparison Analysis

This study presents a comparative analysis of three models: Artificial Neural Network, Multiple Linear Regression, and Ridge Regression based on statistical performance metrics: Coefficient of Determination (R^2), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). R^2 measures the proportion of variance in the dependent variable (Yield) that is explained by the independent variables (Weather parameters); a higher R^2 value indicates better explanatory power. MSE is the average of the squared differences between actual and predicted values, whereas RMSE, being the square root of MSE, expresses the error in the original unit, making it more interpretable.

In this study, 27 years (1997-2023) of weather and yield data, including parameters such as temperature, rainfall, humidity,

and wind speed, were used to predict the yield of rice and wheat. ANN, a machine learning-based non-linear model, was found capable of capturing complex relationships, while MLR and Ridge Regression, as standard statistical approaches, offered the benefits of interpretability, stability, and lower computational complexity. Ridge Regression employs regularization to mitigate multicollinearity, whereas ANN achieved higher prediction accuracy but required longer training time and careful parameter tuning. The entire analysis and modeling were performed using R software, with 80% of the data used for training and 20% for testing. The results are based on the historical data and climatic conditions of the central plains of Chhattisgarh.

Table 5: Comparison of model performance for rice

Model	RMSE	MSE	R^2
Artificial Neural Network	0.0799	0.0063	0.8688
Multiple Linear Regression	0.2017	0.0407	0.3431
Ridge Regression	0.1255	0.0157	0.6768

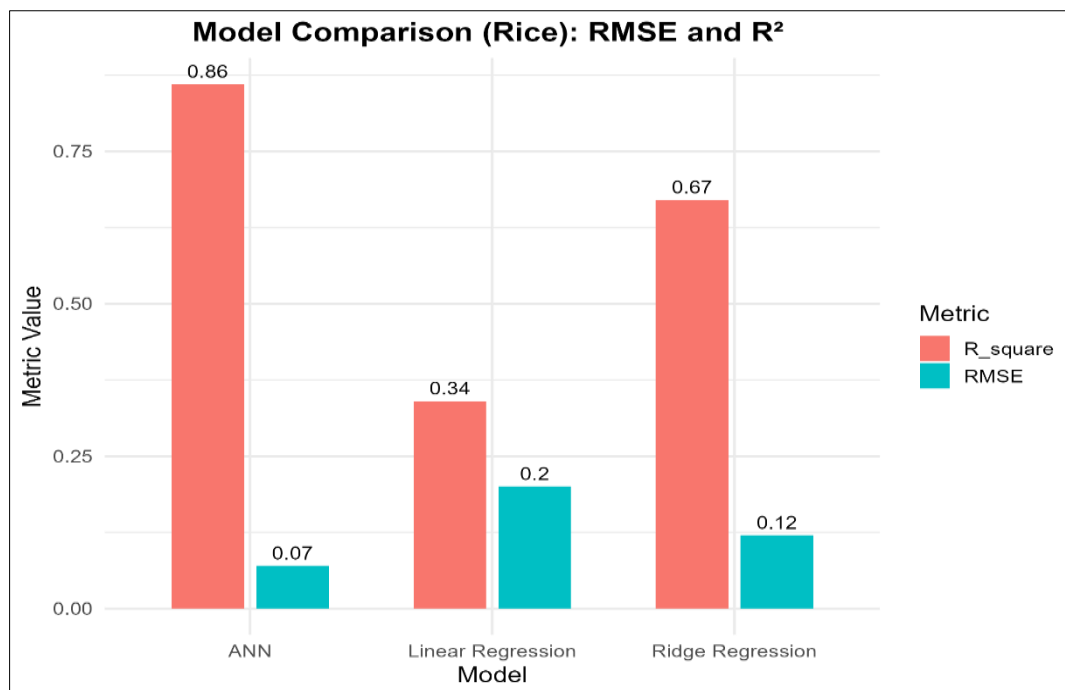


Fig 5: Comparison of R^2 and RMSE for rice yield prediction

The performance comparison for rice yield prediction shows that the Artificial Neural Network model outperforms Multiple Linear Regression and Ridge Regression across all evaluation metrics. ANN achieved the lowest RMSE (0.0799) and MSE (0.0063), indicating minimal prediction errors, and the highest R^2 value (0.8688), meaning it explains about 86.88% of the variability in rice yield. In contrast, MLR recorded the highest RMSE (0.2017) and MSE (0.0407), along with the lowest R^2 (0.3431), reflecting weaker accuracy and explanatory power. Ridge Regression performed moderately well, with an RMSE of 0.1255, MSE of 0.0157, and R^2 of 0.6768, showing clear improvement over MLR but still lagging behind ANN. Overall, ANN proved to be the most accurate and reliable model for rice yield prediction in this study.

In the context of Chhattisgarh central plains, where rice is the dominant crop and yield is highly influenced by seasonal

rainfall, temperature, and soil conditions, the Artificial Neural Network model proved most effective for accurate prediction. With an R^2 of 0.8688, it captured local yield variability far better than Ridge Regression (0.6768) and Multiple Linear Regression (0.3431). Its low RMSE and MSE values indicate strong adaptability to the region's complex climatic patterns. Ridge Regression showed moderate success, while MLR struggled to reflect the area's diverse agricultural conditions. This makes ANN highly suitable for planning and decision-making in the central plains.

Table 6: Comparison of model performance for wheat

Model	RMSE	MSE	R^2
Artificial Neural Network	0.0413	0.0017	0.9357
Multiple Linear Regression	0.1591	0.0253	0.1605
Ridge Regression	0.1168	0.0136	0.4874

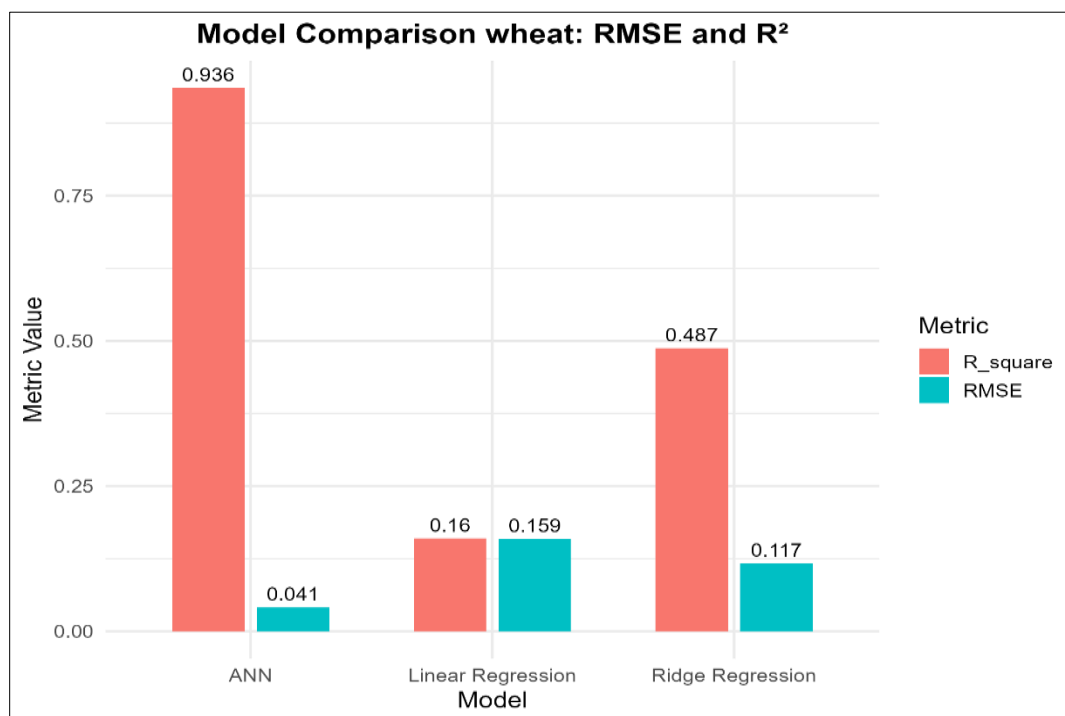


Fig 6: Comparison of R^2 and RMSE for wheat yield prediction

The performance comparison for wheat yield prediction shows that the Artificial Neural Network model showed the best performance among the three approaches. ANN recorded the lowest RMSE (0.0413) and MSE (0.0017), indicating very small prediction errors, and achieved the highest R^2 value (0.9357), meaning it explains about 93.57% of the variability in wheat yield. In comparison, Multiple Linear Regression (MLR) had the highest RMSE (0.1591) and MSE (0.0253), with a low R^2 (0.1605), reflecting weak predictive ability. Ridge Regression performed moderately well with an RMSE of 0.1168, MSE of 0.0136, and R^2 of 0.4874, showing improvement over MLR but still far behind ANN. Overall, ANN is the most accurate and reliable model for predicting wheat yield in this study.

In the central plains of Chhattisgarh, where wheat is cultivated mainly in the rabi season under irrigated conditions, the Artificial Neural Network model emerged as the most accurate for yield prediction. With an R^2 of 0.9357, it explained over 93% of the yield variability, effectively capturing the region's specific agro-climatic patterns. Ridge Regression showed moderate adaptability with an R^2 of 0.4874, while Multiple Linear Regression struggled with an R^2 of only 0.1605, indicating weak explanatory power. The ANN's very low RMSE (0.0413) and MSE (0.0017) make it highly reliable for local planning. This model can greatly assist policymakers and farmers in making informed decisions in the central plains.

5. Conclusion

The study confirms that the ANN offers the most reliable and precise predictions for rice and wheat yields in Chhattisgarh's central plains. Its ability to handle complex non-linear relationships and respond to varying weather conditions allows it to consistently outperform traditional models. MLR and Ridge Regression were useful in identifying general trends but were limited by their linear frameworks, restricting their capability to account for multidimensional climate yield interactions. ANN's adaptability, capacity to manage large datasets, and iterative learning process provide it with a clear edge, making it a powerful instrument for agricultural

planning, resource optimization, and climate risk mitigation. Incorporating ANN into precision agriculture can greatly enhance crop management and help build resilient, sustainable food systems.

References

1. Agrawal R, Singh KN, Lama A, Gurung B, Haque MA, Singh KK, *et al.* Weather based models for pre-harvest crop yield forecasting. *J Indian Soc Agric Stat.* 2023;77(1):59-69.
2. Aravind KS, Vashisth A, Krishanan P, Das B. Wheat yield prediction based on weather parameters using multiple linear, neural network and penalised regression models. *J Agrometeorol.* 2022;24(1):18-25.
3. Bejo SK, Mustaffha S, Ismail WI. Application of artificial neural network in predicting crop yield: A review. *J Food Sci Eng.* 2014;4:1-9.
4. Bhojani SH, Bhatt N. Wheat crop yield prediction using new activation functions in neural network. *Neural Comput Appl.* 2020;32(17):13941-13951.
5. Champaneri M, Chachpara D, Chandvidkar C, Rathod M. Crop yield prediction using machine learning. *Int J Sci Res.* 2020;9(4):645-648.
6. Chlingaryan A, Sukkariéh S, Whelan B. Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. *Comput Electron Agric.* 2018;151:61-69.
7. Das B, Nair B, Reddy VK, Venkatesh P. Evaluation of multiple linear, neural network and penalised regression models for prediction of rice yield based on weather parameters for west coast of India. *Int J Biometeorol.* 2018;62(10):1809-1822.
8. Demirel MH, Baran MF, Gokdogan O. Effect of input usage on wheat yield: An application of artificial neural networks. *J Agric Sci Technol.* 2024;26(2):122-134.
9. Dhekale BS. Statistical models for prediction of crop yield. *J Indian Soc Agric Stat.* 2014;68(3):331-338.
10. Kakati N, Deka RL, Das P, Goswami J, Khanikar PG, Saikia H. Forecasting yield of rapeseed and mustard using multiple linear regression and ANN techniques in

- the Brahmaputra valley of Assam, North East India. *Theor Appl Climatol.* 2022;150:1201-1215.
11. Pooja, Kumar PS, Ramole S, Narsimhaiah L, Saxena RR. Impact of climate variability on groundnut yield prediction using AI models in Chhattisgarh. *Int J Agric Stat Sci.* 2025;21(1):237-243.
 12. Ray DK, Gerber JS, MacDonald GK, West PC. Climate variation explains a third of global crop yield variability. *Nat Commun.* 2015;6:5989.
 13. Satpathi A, Setiya P, Das B, Nain AS, Jha PK, Singh S, *et al.* Comparative analysis of statistical and machine learning techniques for rice yield forecasting for Chhattisgarh, India. *Sustainability.* 2023;15(3):2786.
 14. Singh RS, Patel C, Yadav MK, Singh KK. Yield forecasting of rice and wheat crops for eastern Uttar Pradesh. *J Agrometeorol.* 2014;16:199-202.
 15. Verma OP, Pandey D, Singh P. Evaluation of multiple linear, neural network and penalised regression models for prediction of rice yield based on weather variables. *Int J Agric Stat Sci.* 2016;12(1):123-129.