MAUSAM, **77**, 4 (October 2026), Part-II, 1467-1480

MAUSAM

DOI: https://doi.org/10.54302/mausam.v77i4.7043
Homepage: https://mausamjournal.imd.gov.in/index.php/MAUSAM

UDC No.631.5:004.8

Data-driven approach to multiple crop yield prediction using general regression neural network model

DIPIKA SARKAR[1,2]*, ARINDAM GIRI[3] and SUBRATA DUTTA[1]

[1*]*Department of Computer Science and Engineering, National Institute of Technology Jamshedpur, Jharkhand, India*

[2]*Department of Computer Science and Engineering (Artificial Intelligence & Machine Learning), Haldia Institute of Technology, Haldia, West Bengal, India*

[3]*Department of Computer Science and Engineering, Haldia Institute of Technology, Haldia, West Bengal, India*

(*Received, 13 April 2025 Accepted 19 September 2025*)

***Corresponding author's email: sarkardipikacse@gmail.com**

सार – बढ़ती जनसंख्या की आवश्यकताओं को पूरा करने के लिए खाद्य उत्पादन पर दबाव बढ़ रहा है। ऐसी स्थिति में फसल उपज का सटीक पूर्वानुमान कृषि उत्पादन की योजना बनाने में महत्वपूर्ण भूमिका निभाता है। पारंपरिक रैखिक मॉडलों की अपर्याप्त पूर्वानुमान क्षमता को देखते हुए फसल उपज के लिए एक प्रभावी पूर्वानुमान मॉडल का विकास आवश्यक है। पूर्वानुमान मॉडल समझने में सरल, सटीक तथा प्रशिक्षण और सत्यापन के दौरान कम समय लेने वाला होना चाहिए। इसी उद्देश्य से वर्तमान अध्ययन में फसल उपज के पूर्वानुमान के लिए एक सटीक, लचीला और सरल अरैखिक मॉडल विकसित करने पर बल दिया गया है। एक नवीन सामान्य प्रतिगमन न्यूरल नेटवर्क (General Regression Neural Network–GRNN) मॉडल के माध्यम से विभिन्न स्वतंत्र विशेषताओं और फसल उपज के बीच प्रायोगिक अरैखिक संबंध निर्धारित किया गया। विभिन्न भौगोलिक क्षेत्रों में चावल और गेहूँ की उपज को प्रभावित करने वाले अनेक कृषि-जलवायवीय कारकों पर विचार किया गया। चावल की उपज के पूर्वानुमान के लिए जल स्तर, औसत तापमान, आर्द्रता स्तर और नाइट्रोजन की मात्रा जैसे प्रमुख पर्यावरणीय एवं मृदा संबंधी इनपुट मापदंडों का उपयोग किया गया, जबकि गेहूँ की उपज के लिए मृदा आर्द्रता, वर्षण स्तर, औसत तापमान और सूर्यप्रकाश की अवधि को प्रमुख मापदंड माना गया। परिणामों से पता चला कि संबंधित प्रतिक्रिया पृष्ठ पद्धति (Response Surface Methodology–RSM) मॉडल की तुलना में GRNN मॉडल ने पूर्वानुमान की सटीकता में सुधार किया। इसमें माध्य वर्ग त्रुटि (MSE) 0.011-0.017, मूल माध्य वर्ग त्रुटि (RMSE) 0.090-0.109 तथा R^2 का मान 0.91-0.94 के बीच रहा। GRNN मॉडल का उपयोग विश्व के विभिन्न भौगोलिक क्षेत्रों में समान या अन्य फसलों की उपज के पूर्वानुमान के लिए किया जा सकता है। इन परिणामों के विश्लेषण से किसान, कृषि विज्ञानी और शोधकर्ता विभिन्न कारकों के बीच अंतर्संबंध तथा समग्र उत्पादन पर उनके संभावित प्रभाव को बेहतर ढंग से समझ सकते हैं।

ABSTRACT. Food production is under pressure to fulfill the demands of growing populations. An accurate crop yield prediction helps in agricultural production planning in such a situation. Due to the inadequate prediction performance of the traditional linear models, the evolution of a crop yield prediction model is necessary. The prediction model should be easy to understand and precise and require less time during the training and validation stages. For these reasons, the current study emphasizes developing an accurate nonlinear model that is flexible and simple for crop yield prediction. An experimental nonlinear relationship between various independent attributes and crop yield has been derived from the novel General Regression Neural Network (GRNN) model. Numerous agro-climatic factors have been considered across diverse geographical regions influencing rice and wheat yield production. Key environmental and soil input parameters viz. water level, average temperature, humidity levels, and nitrogen content are considered as for rice yield prediction, whereas soil moisture, precipitation levels, average temperature, and sunlight hours are listed for wheat yield prediction. It has been observed that the model shows improvement in the prediction accuracy with mean square error (MSE) between 0.011-0.017, root mean square error (RMSE) between 0.090-0.109, and R^2 value between 0.91-0.94 compared to the corresponding response surface methodology (RSM) model. The GRNN model can predict different

crops from the same or other geographical regions worldwide. Farmers, agronomists, and researchers can better understand the interrelationship between different attributes and how they might impact overall production by analyzing these results.

Key words – Crop yield, prediction, nonlinear model, general regression neural network, response surface methodology.

1. Introduction

Agriculture is particularly a vulnerable sector due to varying climate patterns. The researchers predict yield to determine the proper policy response (Kumari *et al*., 2023). However, predicting agricultural production is challenging since it involves natural factors (like weather) and farmers' decisions or actions (such as acreage, crop rotation, and water irrigation). Crop yield exhibits nonlinear behavior in terms of both geographical and temporal variation (Drummond *et al*., 2003). The most popular techniques, such as crop simulation and data-driven models, are commonly used to estimate inputs affecting agricultural yield. However, statistical models consider the underlying eco-physiological conditions and forecast crop yields using historical climate, soil data, and observational yield data (Schlenker & Roberts, 2009). Process-based models generate more accurate results than statistical models, although statistical models are powerful in terms of the variety of data, volume, and velocity (Roberts *et al*., 2017). A better alternative to conventional regression models is statistical models implemented using machine learning models, which also come with tools for gaining deeper insights into the data. For crop yield prediction, a variety of machine learning models have been employed, from deep learning algorithms (Khaki *et al*., 2021; Rao &Manasa, 2019; Wang *et al*., 2018) to linear regression and decision trees (Jeong *et al*., 2016; Ranjan & Parida, 2019; Shah Hosseini *et al*., 2020; Vashisth *et al*., 2024; Bhatt *et al*., 2024).

RSM is considered a sophisticated mathematical tool for evaluating the correlation between multiple independent input parameters and the output responses. It can also optimize these input parameters for the best results (Taoufik *et al*., 2022). However, the nonlinear approach of RSM limits the range of input variables (Xue *et al*., 2023). Machine learning (ML) approaches address the nonlinear relationship between crop production and climate by mathematically expanding the objective function. According to Zeroual *et al*. (2009); Yonaba *et al*. (2010), and Zeroual *et al*. (2024); artificial neural network (ANN) is the most widely used machine learning technique that uses nonparametric structure. Safa and Samarasinghe (2011) used the ANN and, subsequently, the linear regression model to predict energy consumption in wheat production. Iran's paddy yield showed similar results (Abrougui *et al*. 2019; Taheri-Rad *et al*., 2017) using the ANN and multiple linear regression (MLR) techniques to evaluate the potato yield prediction behavior utilizing tillage schemes and soil parameters. As a result, the neural network model shows higher accuracy in figuring out how tillage, soil characteristics, and potato yield are related. In other research, viz. corn and soybean yield (Kaul *et al*., 2005), wheat yield (Guo *et al*., 2014), and dry matter yield(Majkovič *et al*.,2016) have all been predicted using ANN models using agricultural data. Linear models are now less compatible due to increased exposure to uneven climate conditions. A hybrid MLR-ANN model has already been used in some earlier research (Gopal and Bhargavi, 2019; Safa & Samarasinghe, 2011).

Because of the complex and unpredictable mechanism, predicting crop production using traditional methods is difficult. The outstanding ability of some advanced soft computing techniques, such as ANN, GRNN, and others, to depict the interaction and nonlinear relationship with input variables and learning from experimental data makes them noteworthy for their ability to predict these responses with great success (Yadav *et al*., 2015; Özener *et al*., 2013; Ouz *et al*., 2010). The probabilistic GRNN, a learning approach for function approximation, was first presented by Donald Specht in 1991 (Specht, 1991). The ability of GRNN to rapidly assess continuous variables during the training phase has made it preferred among several soft computing approaches. Additionally, it uses training data to evaluate the function immediately, eliminating the need to repeat the training set (Bendu *et al*., 2017). Any regression issue in which linearity is not justified can be handled with GRNN, a memory-based one-pass learning algorithm. Probabilistic neural networks, such as GRNN, are more valuable than other neural networks because when there are few training trials available, they can converge to the underlying function of the data (Jiang *et al*., 2024). The amount of additional information required for GRNN is minimal, and it can be executed without any additional information by the user (Liu *et al*. 2023; Zhang *et al*. 2024).

Because of all these factors, GRNN is a helpful tool in real-world applications for predicting and evaluating the performance of any system. When there are more redundant inputs, it is not recommended to use GRNN since it cannot overlook those irrelevant inputs without significant modifications to the fundamental algorithm (Srinivasarao *et al*. 2025). In the fields of non-

conventional machining (Pradhan and Das, 2015; Das and Pradhan, 2014), electrical (Chen *et al*., 2009), medical (Land *et al*., 2012), image processing (Polat and Yıldırım, 2008), and various other fields (Chelgani and Jorjani, 2011; Rooki, 2016; Kumar and Malik, 2016; Cheng and Xiong, 2017; Yao *et al*., 2024; Dai *et al*., 2024; Izonin *et al*., 2024), only few researchers effectively used GRNN. Even though artificial data augmentation is used daily for these neural networks, most of them are utilized in image processing, machining, medicine, and electrical applications. However, one of the most important scientific and technical issues is developing numerical data augmentation techniques for resolving several agricultural issues.

Selecting the optimal parameters for accurate prediction, desired responses, and ongoing process monitoring are all crucial to meeting the demands of the global agricultural sector. However, selecting the optimal climate and agricultural factors is difficult due to the wide range of choices and the conflicting nature of multiple responses. It is clear from previous research that minimal work has been done to achieve the aforementioned goals. However, little consideration has been given to choosing the optimal conditions for jointly predicting rice and wheat crop yields using the innovative GRNN-RSM technique. Thus, there is a significant need to build a suitable model to improve agricultural performance and predict the best crop yield with the least error. This study is exclusive since it fills up these gaps in the literature. Thus, a GRNN model is built to generate rice and wheat crop yields as the output, climate variables, and agricultural inputs. The model's prediction accuracy has been compared with statistical models like Response Surface Methodology.

2. Data and methodology

RSM and GRNN models have been adopted for crop yield prediction and both models' prediction accuracy has been evaluated in the next section. Both the models were deployed and validated using MSE, RMSE, and coefficient of determination (R^2) values. The sub-sections below have showcased detailed discussions on the agricultural features considered. Fig.1 depicts the detailed methodology for this proposed research work.

2.1. *Data acquisition*

Datasets for model development have been collected from the IEEE database (Karthick Raghunath K.M., 2023). The datasets include records for two of the most widely grown crops: rice and wheat across diverse region. The dataset covers several crop growth seasons up to 2023, which ensure to help readers understand the temporal coverage. The datasets offer thorough insights into the range of environmental and soil-related factors that are conventionally important in determining crop yield.

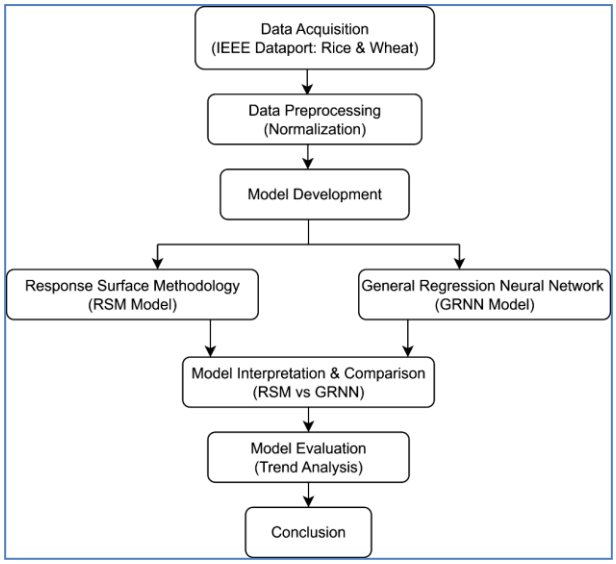


Fig. 1. Methodology of proposed work

2.2. *Data preprocessing*

The research aimed to predict the impact of various agro-climatic factors on yield of two widely produced crops, *i.e*., rice and wheat. Water level (WL), average temperature (AT), humidity levels (HL), nitrogen content (NC) has been considered for rice yield prediction. Similarly, soil moisture (SM), precipitation levels (PL), and sunlight hours (SH) and average temperature (AT) for wheat yield prediction. The sample of size 100, from a population as large as 5000 observations, was selected through the division of the entire range into equal intervals. Equal intervals were chosen to ensure that each part of the range was represented uniformly. This method minimized the potential for sampling bias, covering every section of the range equally. Data points were then collected at these evenly spaced intervals, comprehensively representing the overall range. Significant variations in data observations lead to issues during calibration. During the model's development, data normalization aids in resolving this issue and improves convergence.

2.3. *RSM*

RSM is a statistical and mathematical approach used to produce, improve, and optimize issues where several influencing variables impact a response variable. Analysis was set at a significance level of 5% and a confidence level of 95%. Fig. 2 shows workflow diagram for RSM model.

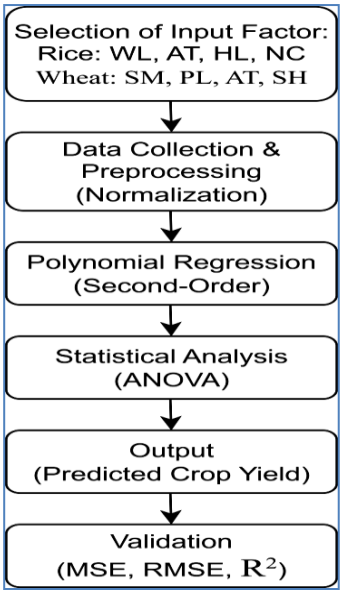


Fig. 2. RSM Workflow

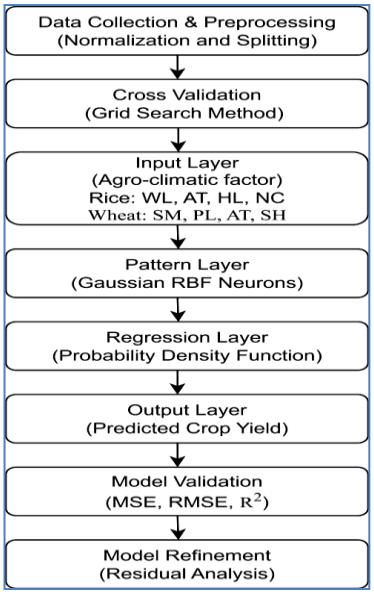


Fig. 3. GRNN workflow

Eq. (1) measured the optimum response *Y*, following the quadratic model or second-order polynomial.

$$Y = \beta_0 + \sum_{i=1}^{n} \beta_i x_i + \sum_{i=1}^{n} \beta_{ii} x_i^2 + \sum_{i=1}^{n} \sum_{j=2}^{n} \beta_{ij} x_i x_j \; ; (i \neq j) \qquad (1)$$

where β_0 is a constant; x_i and x_j are the input variables. The linear, quadratic, and interactive coefficients are denoted by β_i, β_{ii}, and β_{ij} respectively.

2.4. *GRNN method*

A neural network is originally inspired by biological systems composed of neurons. The model uses a supervised learning method to determine the expected outcome. The architecture is parallel-conveyed and has many neurons and connections between them. There is a weight associated with each connection that connects one hub to the next (Esme *et al*., 2009). Initially, data was collected and preprocessed, and the normalized data was segregated into training, testing and validation phases, respectively. The prediction model has been developed by employing the training set with an error correction mechanism, with the remaining being used for model testing and validation. Fig. 3 shows workflow diagram for GRNN model.

Input, pattern, regression, and output are the four layers that make up GRNN. The structure-provided data is shown as input in the input layer. These data are then processed collectively at the pattern layer. Thus, all of the experiments associated with training are the same as the total number of neurons (Izonin *et al*., 2024). The output layer estimates the yield, whereas the regression layer's extra neuron calculates the probability density function. According to Rooki, 2016; the GRNN model requires fewer training sets than traditional ANN models. Fig. 4 shows the architecture of the GRNN model representing four layers, respectively. Y_i

$$\hat{Y}(X) = \frac{\sum_{i=1}^{p} Y_i \;\; \exp[-D_i^2/2\sigma^2]}{\sum_{i=1}^{p} \exp[-D_i^2/2\sigma^2]} \qquad (2)$$

where, (X_i, Y_i)= sample of (X,Y); D_i^2= Scalar function; σ^2= Square of Gaussian RBF kernel bandwidth (Zhang *et al*. 2024); m = no. of testing patterns;

$$D_i^2 = (X - X_i)^T (X - X_i) \qquad (3)$$

The following is an outline of the GRNN model implementation (Zhang *et al*. 2024; Gao *et al*. 2025):

1st Step: At first, a total of 100 experimental datasets have been initialized. The model was trained on 60 datasets, validated on 20 datasets, and the predictive model was tested on the remaining 20 datasets. All the datasets have been chosen arbitrarily to maintain consistency.

2nd Step: The smoothening parameter was obtained using the cross-validation technique. The grid search method has been used to determine the optimum adaptive parameter (σ^2) for achieving the nominal cross-validation error.

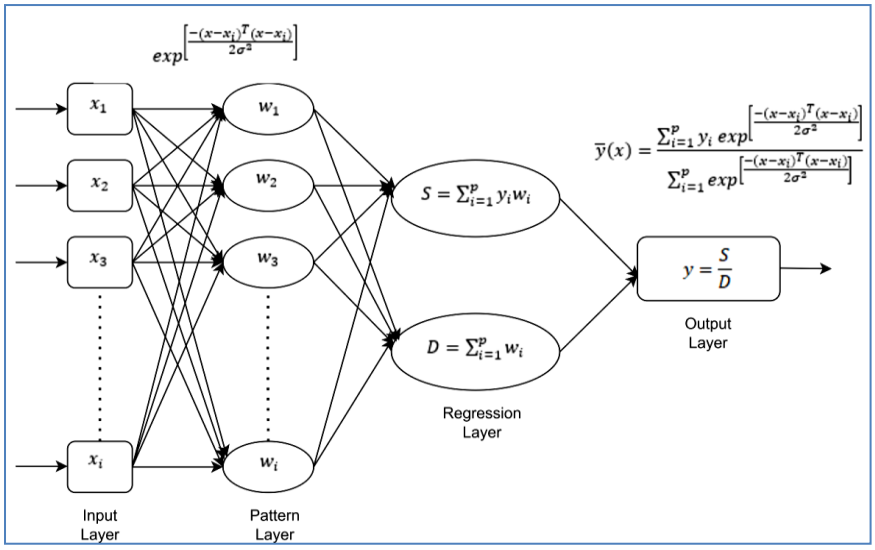


Fig. 4. Architecture of GRNN model

3rd Step: After that, a Scalar function (D_i^2) has been generated for each experimental setting (i = 1–100) using Eq. (3)

4th Step: By substituting σ and (D_i^2) in Eq. (2), the coefficient was obtained.

5th Step: In the pattern layer, the actual output value, Y_i was multiplied by the exponential term.

Step 6: An extra neuron in the regression layer estimated the output probability density function.

Step 7: Within the experience space, a weighted average of the GRNN-processed output was determined.

Step 8: Using the remaining 40% datasets, the effectiveness of the proposed model was investigated using a validation and testing methodology.

In GRNN, the only adaptive criterion is kernel width (σ). Therefore, overfitting of the data is offered by a smaller value of σ, while over-smoothing is produced by a larger value of σ (Dwivedi and Tajer, 2024). Therefore, for the generalization ability and prediction accuracy of GRNN (Bendu *et al*., 2017), the value of the smoothening parameter is significant. The grid search approach was employed to get the optimal kernel width value. For the cross-validation execution, the kernel width needed to obtain "N" values using Eq. 4. The optimal adaptive parameter (σ^2) value for this research is 3.64 (see Fig. 5).

$$\sigma = \sigma_{low} + (\rho - 1)\frac{\sigma_{high} - \sigma_{low}}{N} \quad (4)$$

where, $\rho = 1,2,3,\ldots\ldots n$; $\sigma_{low} = 1$; $\sigma_{high} = 10$; $N = 10$

2.5. *Model interpretation and comparison*

MSE and RMSE have been calculated to compare the available output value with all the generated outputs using the model (Eq. 5-6). Using Eq. (4), the R^2 or Coefficient of determination value is calculated (Eq. 7). The predicted model performs better when MSE, and RMSE, values are lower.

$$\text{MSE} = \left(\frac{\sum_{i=1}^{n}(Out_{obs} - Out_{est})^2}{n}\right) \quad (5)$$

$$\text{RMSE} = sqrt\left(\frac{\sum_{i=1}^{n}(Out_{obs} - Out_{est})^2}{n}\right) \quad (6)$$

$$R^2 = \frac{\left(\sum_{i=1}^{n}(Out_{obs} - \overline{Out_{obs}})(Out_{est} - \overline{Out_{est}})\right)^2}{\sum_{i=1}^{n}(Out_{obs} - \overline{Out_{obs}})^2 \sum_{i=1}^{n}(Out_{est} - \overline{Out_{est}})^2} \quad (7)$$

where Out_{obs} and Out_{est} represent the expected yield and estimated yield, respectively. There is a 'n' number of patterns for testing. Any value between 0 and 1 can be taken in R^2. This indicates how well the data is fitted by the model. The higher the value of R^2, the better the model fits the data.

3. Results and discussion

The correctness of prediction by GRNN has been compared with the RSM.

3.1. *Yield prediction using RSM*

RSM is a statistical technique that is helpful in modeling and analyzing real-life challenges. Several agricultural input features and performance metrics are correlated using RSM. This model is developed at a 95%

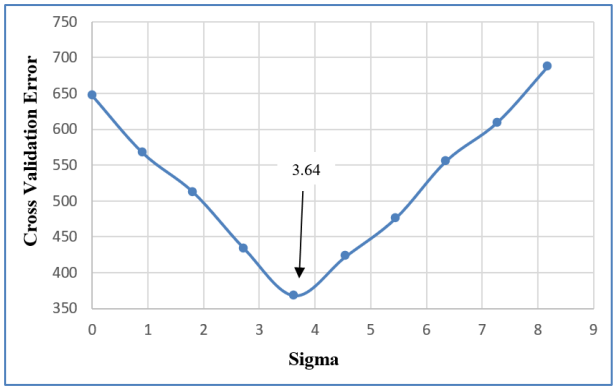


Fig. 5. Cross-validation curve for GRNN model

confidence level. RSM has been used to establish the former equations in MINITAB 17 software. The quadratic model was employed to predict the response values, as shown in Eqs. 8 and 9.

Yield (Rice) = 14887-171.7WL-947.2AT+51.8HL-831NC -0.254WL*WL+9.69AT*AT-0.8008HL*HL+545.2NC* NC+4.830WL*AT+0.152WL*HL+35.31WL*NC+3.430 AT*HL+11.6AT*NC-26.22HL*NC (8)

Yield(Wheat) = -6362+ 190.87SM+304.4AT+62.93PL-509.2SH-0.3987 SM*SM-2.230AT*AT-0.13624PL*PL+ 39.93SH*SH-1.014SM*AT-0.6547SM*PL-5.627SM*SH-0.5659AT*PL-9.783AT*SH+0.237PL*SH (9)

where WL, AT, HL, NC, SM, PL, and SH are water level, average temperature, humidity levels, nitrogen content, soil moisture, precipitation levels, and sunlight hours, respectively.

Probability plots are employed to evaluate whether a dataset adheres to a particular distribution, typically the normal distribution. If the data elements are approximately aligned with a straight line, the data follows the specified distribution. The actual yield closely resembles the predicted yield, exhibiting an approximately normal distribution. With minimal deviation between predicted and actual values, the prediction model appears to be robust. A few rare or outlier cases are suggested by only minor deviations at the tails, which is common in numerous real-world datasets. Fig. 6(a)–(b) depicts the Probability plots of actual and estimated values for both crop yields.

The best-fitted data are examined for the model, and S and R^2 values have been studied. Higher R^2 values are obtained using a generalized regression model, where R stands for correlation coefficient, and lower S values, where S stands for the sample's standard errors. The previously developed model shows that calculated S values for rice are 4.486 and wheat yield is 1.85,

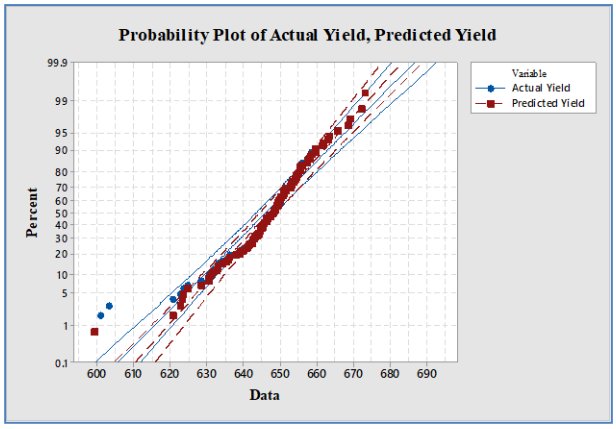


Fig. 6(a). Actual vs Estimated crop yield (rice) values in RSM model

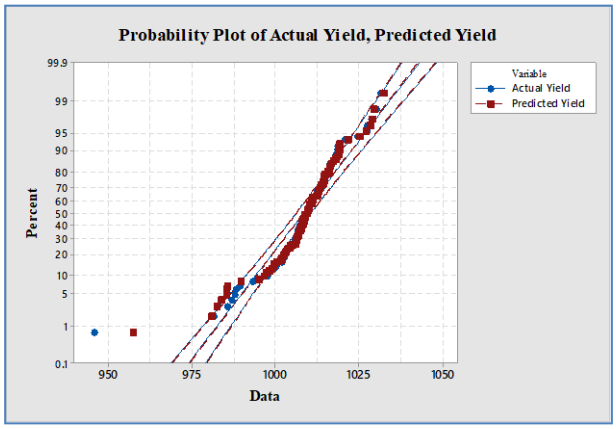


Fig. 6(b). Actual vs Estimated crop yield (wheat) values in RSM model

respectively. The corresponding R^2 values for both responses are 80.56% and 85.67%, respectively. A brief analysis will reveal that the established model best fits the collected data.

An ANOVA (Analysis of variance) result for multiple crop yields has been presented in Tables 1 and 2. ANOVA's significance test assessed the impact of the independent variable on the dependent variable. The model is highly significant, as indicated by its F-value of 53.98 and 254.54, and the **p**-value is less than 0.01, as shown in Tables 1 and 2.

3.2. *Yield prediction using GRNN*

The GRNN model has been implemented using experimental data where input parameters WL, AT, HL, NC for rice yield and SM, PL, AT, SH for wheat yield serve in the input layer and the corresponding crop yield component in the output layer. MATLAB R2015b software codes were used to simulate on a LENEVO computer (configuration: Intel Core i7-4th generation, 3.9

TABLE 1

ANOVA for rice yield

Source	Degrees of Freedom (DF)	The Sum of square (SS)	F-Value	P-Value
Model	14	15211.1	53.98	0.000
Linear	4	3764.8	46.76	0.000
WL	1	683.2	33.94	0.000
AT	1	1177.9	58.52	0.000
HL	1	1368.0	67.96	0.000
NC	1	716.5	35.60	0.000
Square	4	4805.1	59.68	0.000
WL*WL	1	64.6	3.21	0.077
AT*AT	1	1740.1	86.45	0.000
HL*HL	1	2399.1	119.20	0.000
NC*NC	1	905.2	44.97	0.000
2-Way Interaction	6	6386.3	52.88	0.000
WL*AT	1	2172.7	107.95	0.000
WL*HL	1	21.3	1.06	0.306
WL*NC	1	1235.7	61.39	0.000
AT*HL	1	1744.4	86.67	0.000
AT*NC	1	17.9	0.89	0.348
HL*NC	1	1024.8	50.92	0.000
Error	85	1710.8		

TABLE 2

ANOVA for wheat yield

Source	Degrees of Freedom (DF)	The Sum of square (SS)	F-Value	P-Value
Model	14	12089.7	254.54	0.000
Linear	4	2217.6	163.42	0.000
SM	1	411.7	121.35	0.000
AT	1	453.2	133.59	0.000
PL	1	147.6	43.50	0.000
SH	1	1369.7	403.73	0.000
Square	4	3976.0	292.99	0.000
SM*SM	1	101.2	29.83	0.000
AT*AT	1	448.1	132.08	0.000
PL*PL	1	2415.1	711.89	0.000
SH*SH	1	786.5	231.84	0.000
2-Way Interaction	6	4588.9	225.44	0.000
SM*AT	1	138.5	40.83	0.000
SM*PL	1	2667.8	786.38	0.000
SM*SH	1	311.5	91.82	0.000
AT*PL	1	606.6	178.81	0.000
AT*SH	1	388.3	114.46	0.000
PL*SH	1	8.6	2.54	0.115
Error	85	288.4		

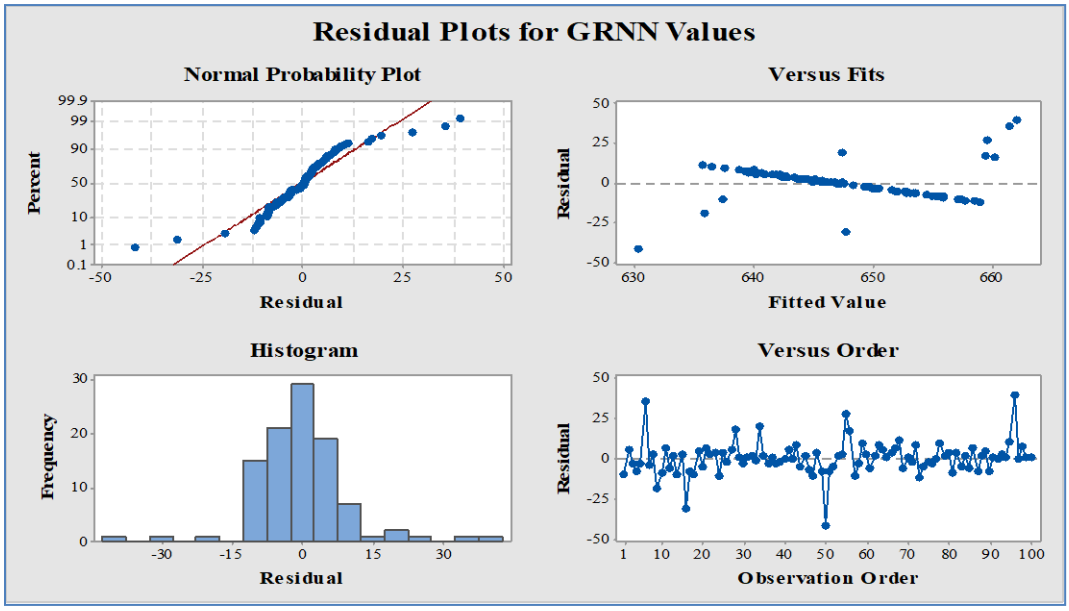


Fig.7a. Plot of residuals for the overall GRNN values (Rice Yield)

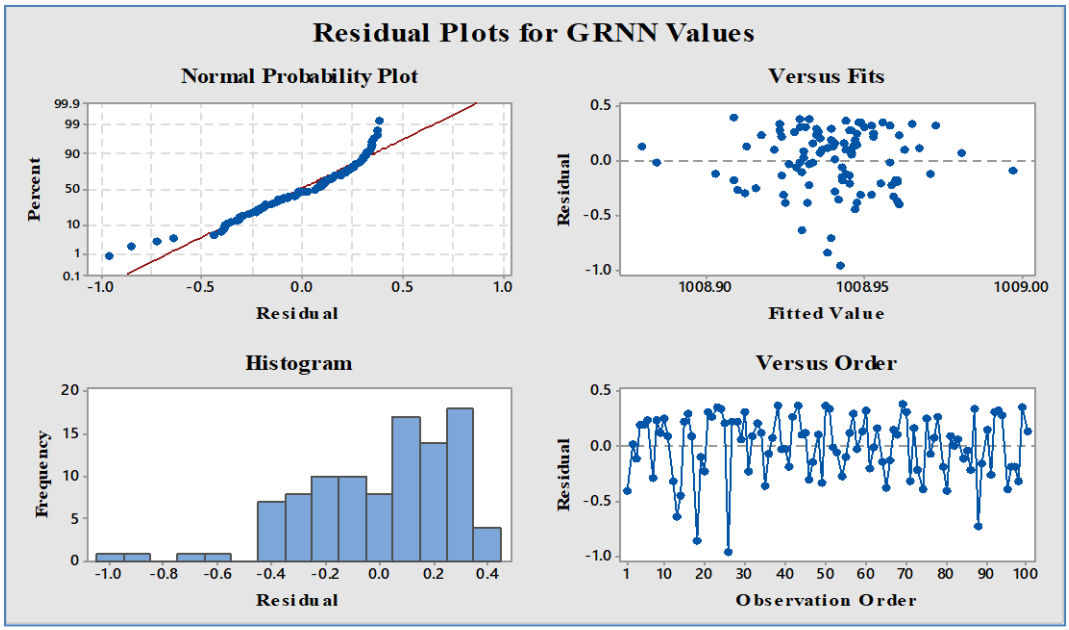


Fig.7b. Plot of residuals for the overall GRNN values (Wheat Yield)

GHz CPU, and 8 GB RAM). The models' results were compared against experimental data to evaluate their adequacy. The GRNN model used 100 processed data sets for training, validation, and testing. Of these, the network was trained using 60% of the datasets, testing was conducted using next 20%, and the model was validated using the remaining 20%.

A maximum inaccuracy of 9.30% has been found when comparing the experimentally observed and GRNN-predicted crop yield values, as shown graphically in Fig. 7. Residual analysis is a primary diagnostic tool to evaluate the model efficiency. Figs 7a and 7b show a histogram of the standardized residual as a normal probability plot against the fitted value and observation order. Additionally, it is clear that there is no outlier, and the process follows normal distribution. Fig. 7a and 7b show that the plot between the observed levels and the standardized residuals does not follow a noticed pattern *i.e.,* structureless. It is clear from Figs. 7a and 7b that a

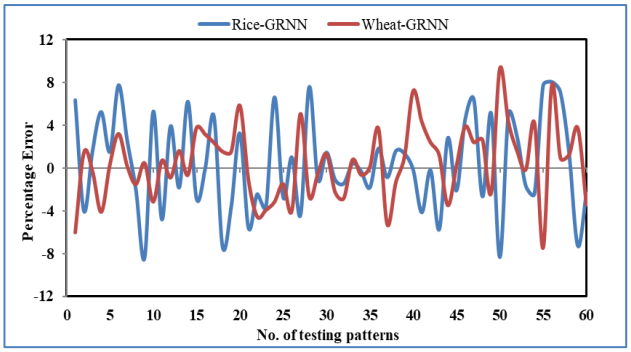


Fig. 8. Visual depiction of error percentage in GRNN trainee data

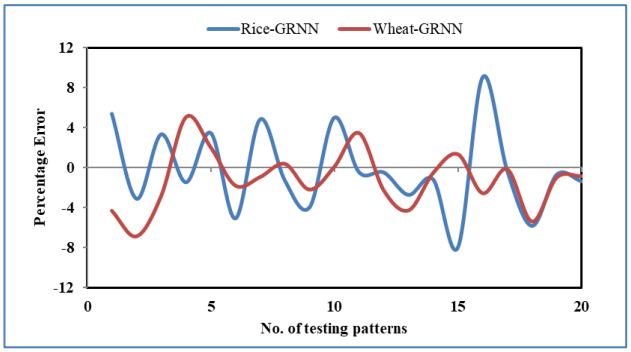


Fig. 9. Visual depiction of error percentage in GRNN testing data

linear approximation between values predicted by GRNN and the experimental data has been revealed. This implies that GRNN can accurately predict crop responses with minimal error. Therefore, it can be said that the model performs satisfactorily.

During the training phase of GRNN, a maximum 9.30% error and a minimum -7.49% error were observed. Fig. 8 illustrates the error percentages within the training datasets. After training, the developed GRNN model underwent testing and validation before being compared with experimental results. Fig. 9 displays the error percentages observed during the testing phase, with an utmost error of 9.05% and a lowest of -7.95%. During the validation phase, the error percentage depicted in Fig. 10 ranged from a minimum of -5.40% to a maximum of 7.29%. These findings indicate that the GRNN model's predictions closely align with the experimental data. Based on the results obtained from the training, validation, and testing phases, it can be concluded that the GRNN model is suitable for predicting or analyzing crop responses with an acceptable error margin of ±10%.

3.3. *Comparison of RSM and GRNN*

Error analysis plays a crucial role in the research, offering valuable insights into how precise and dependable the predictive models are. The graphical representations of MSE, RMSE, and R^2 values are shown

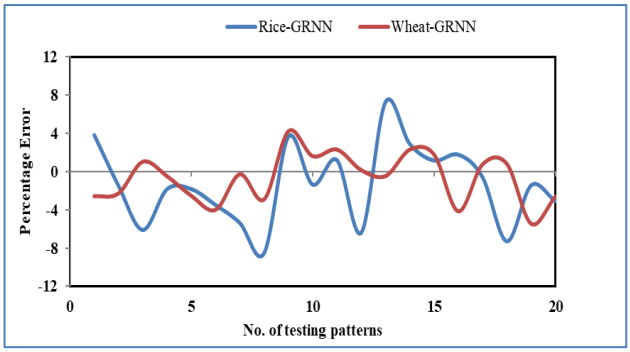


Fig. 10. Visual depiction of error percentage in GRNN validation data

in Figs. 11-14, respectively. Figs. 11 and 13 show the RMSE values for rice and wheat crop yields. MSE and R^2 values are shown in Figs. 12 and 14, respectively. During the training and testing phases, the GRNN model predicted crop yields with notably superior accuracy over the traditional RSM model for both crops. Using the GRNN model, the training phase RMSE value for rice was found to be 0.090. However, the RMSE value using RSM was 0.285. This is in good agreement with previous studies (Thurkkaivel *et al*. 2021; Son *et al*. 2022; Satpathi *et al*. 2023; and Yu *et al*. 2023). It is noted that there is a slight discrepancy between the RMSE values calculated in this study and earlier similar studies; this discrepancy may result from modifications to the variable specification and the caliber of the testing pattern used.

For wheat, the training phase RMSE value was 0.070 using the GRNN model, whereas the same for the RSM model was 0.307. These results are in good agreement with the previous study (Basso and Liu, 2019; Srivastava *et al*., 2022; Kaur 2023). A minor disagreement has been seen between the RMSE values computed in other pertinent studies and the current study; this discrepancy may result from differences in the definition of the variable and the caliber of the testing pattern used. Figures 12 & 14 show the R^2 values, which indicate the estimated models' goodness of fit. Using the GRNN model, the R^2 values for rice and wheat yields fall between 0.91 and 0.94, whereas the RSM model's R^2 values fall between 0.80 and 0.85. A comparison of R^2 values for both crops also shows that the GRNN model outperforms the regression model regarding goodness of fit.

The aforementioned observation indicates that, compared to the conventional RSM approach, the GRNN model predicts crop yield with noticeably more accuracy. Thus, compared to the RSM, GRNN shows significantly higher capacity in terms of generalization. While RSM can only approximate a second-order polynomial, the GRNN can approximate the system's nonlinearity universally, which accounts for its superior predictive accuracy.

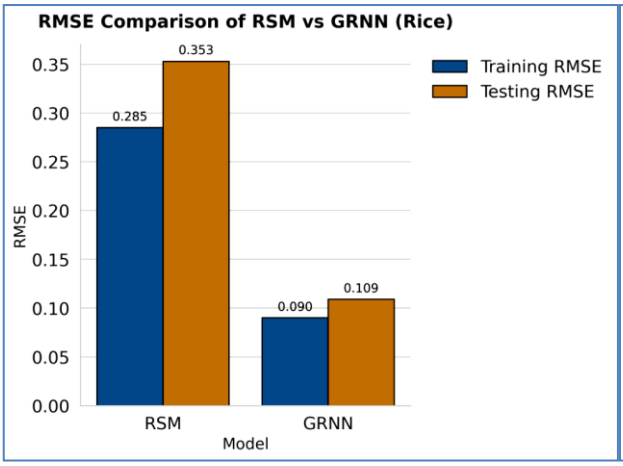


Fig. 11. RMSE comparison of GRNN and RSM models (rice yield)

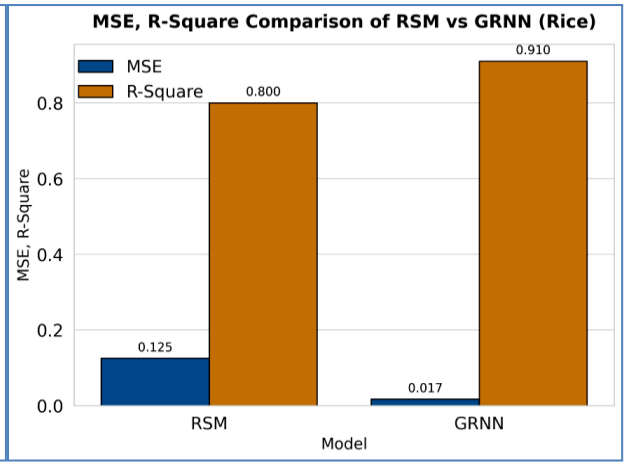


Fig.12. MSE &R-Square comparison of GRNN & RSM models (rice yield)

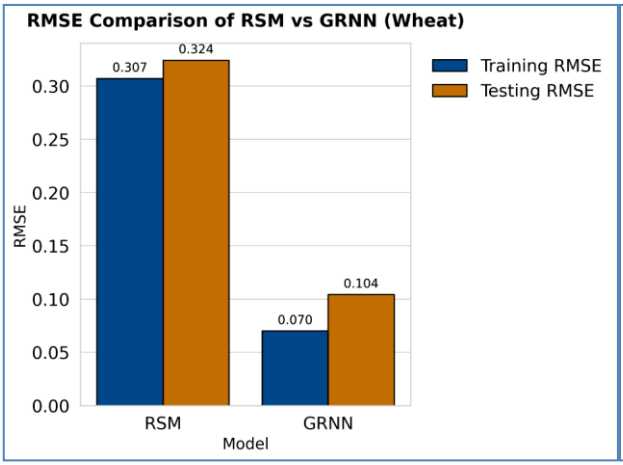


Fig. 13. RMSE comparison of GRNN and RSM models (wheat yield)

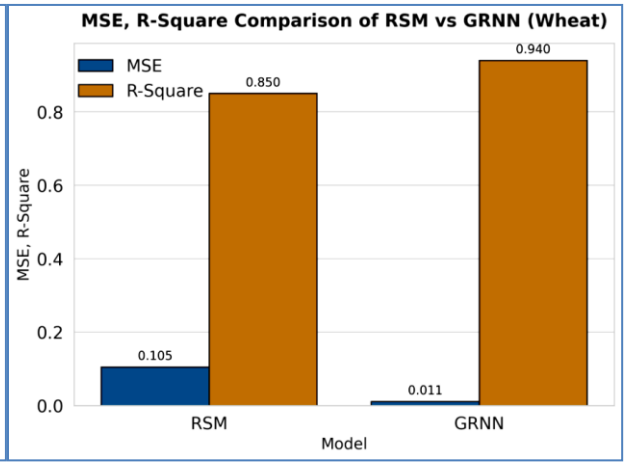


Fig. 14. MSE and R-Square comparison of GRNN and RSM models (wheat yield)

3.3.1. *Trend analysis*

The predictive performance comparison between RSM and GRNN has been illustrated via trend analysis in Figs. 15 and 16. Because RSM uses a fixed quadratic surface, that can underfit if the yield relationship is more complex than a parabola. Whereas, GRNN is non-parametric and adjusts to local variances, identifying complex patterns that RSM misses.

In both graphs (Figs. 15 and 16) the GRNN curve more closely resembles the actual yield line than the RSM curve, indicating that GRNN predictions are more in line with reality. It is visually evident from the plots that GRNN would be the better option in practical applications. Better fit in yield prediction results in more reliable recommendation for climate-resilient farming.

4. Conclusions

The agricultural system requires the application of nonlinear methods of yield prediction. The application is necessary due to the nonlinear feature of numerous environmental parameters that influence the efficiency of crop yield. Hence, more accurate prediction techniques must be used than traditional statistical modeling methods. In this work, the RSM and GRNN models were developed, and their performances in terms of modeling and predictive abilities were compared. Even with fewer testing patterns, the GRNN outperformed the RSM with higher accuracy and predictive ability. Because of the well-organized structure, RSM is still helpful for immediately gaining insight into the system (*e.g.*, relationships between features). However, GRNN has consistently done better than RSM in both category of yield. With an MSE of 0.017, 0.011, RMSE of 0.109,

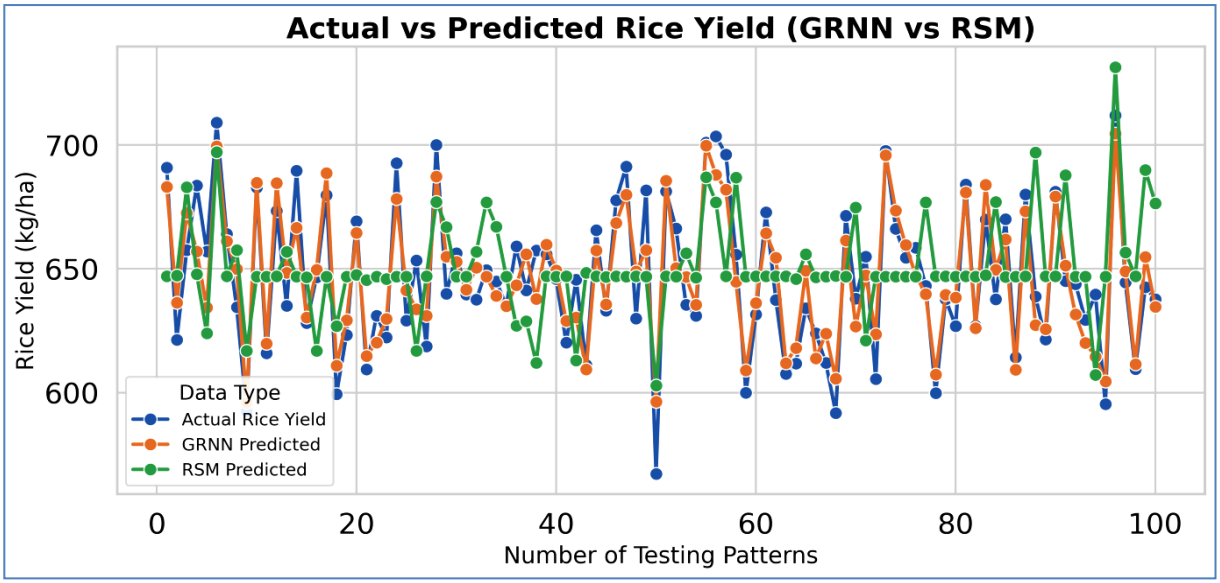


Fig. 15. Trends of GRNN - RSM along testing patterns (Rice yield)

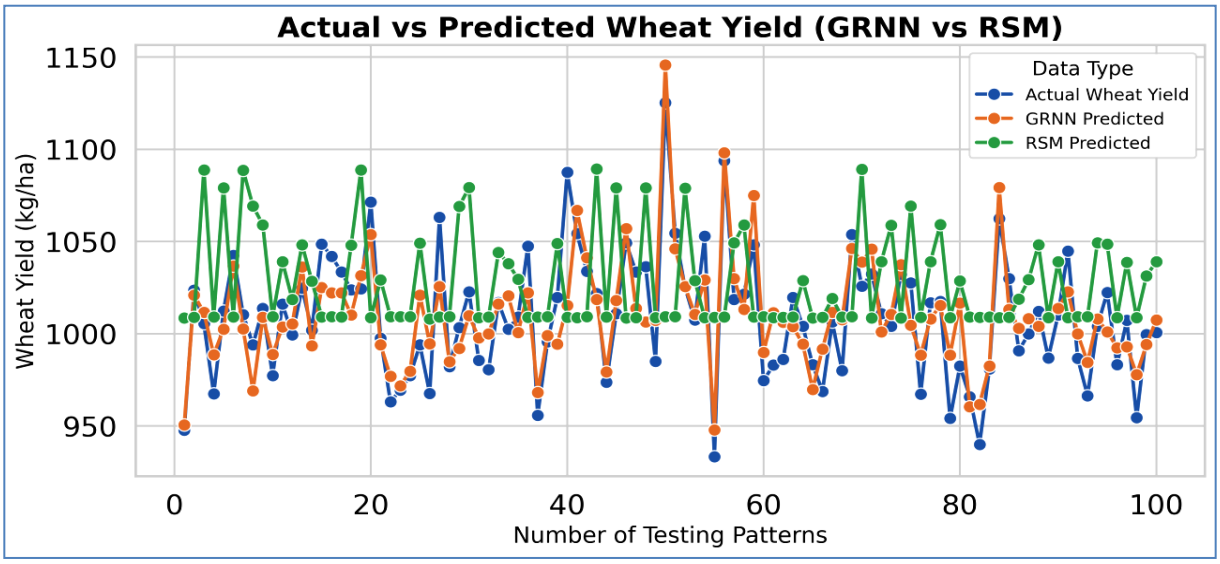


Fig. 16. Trends of GRNN - RSM along testing patterns (Wheat yield)

0.104 and an R^2 value of 0.91 and 0.94, respectively, the GRNN model was able to forecast the production of rice and wheat simultaneously. The model can be extended to other regions of the country by adding more input features, which will be helpful for many researchers in future studies.

4.1. *Future scope*

Although the suggested GRNN-based method shows potential for predicting multi-crop output, there are a few things to remember in future research. The method excludes domain-specific agronomic knowledge due to confidentiality agreements, which could improve the robustness and dependability of the model. Last but not least, although being selected for its capacity to represent nonlinear interactions, GRNN's relative performance could be evaluated more broadly by comparing with other machine learning based methods. Future research should concentrate on resolving these issues by enhanced data preprocessing, hybrid modeling, and thorough comparison with other approaches.

Despite all the limitations, there are some practical strategies that might strengthen the real-world applicability of the model. The GRNN based model has

been developed by employing datasets from diverse regions so that it can predict different crops from the same or other geographical regions worldwide. For instance, if predicted yield is not as expected in heavy rainfall or drought affected area, farmers can adopt intensive crop cultivation by using modern machinary, presticide and fertilizer optimization, irrigation scheduling to produce estimated yield in such vulnerable areas. Decision makers may even follow crop rotation to avoid risk arising of insufficient production and promote food safety. Moreover, the selection of key agro-climatic input parameters has been used to identify high-risk areas for declining yield production under various environmental conditions. The proper selection of input factor plays a vital role to build up the proposed GRNN model robust and accurate. Because of its high accuracy and low processing cost, it can give policymakers information on the gaps in the future food supply as Climate-resilient crop planning is supported by policy, which is relevant to SDG 2 (Zero Hunger). In order to increase model robustness and dependability, future work will combine GRNN modeling with expert-driven feature engineering and region-specific agronomic knowledge as per India Meteorological Department (IMD) guidelines.

Acknowledgment

This work is partially supported by the INDIA ASEAN project funded by SERB (File No: CRD/2022/000504), Government of India.

Authors' Contributions

Dipika Sarkar: Manuscript writing, model formation and data analysis.
Arindam Giri: Co-Supervision and advisory support. (*email: ari_giri111@rediffmail.com*).
Subrata Dutta: Supervision, technical guidance, methodology support and manuscript review. (*email-sdutta.cse@nitjsr.ac.in*).

Disclaimer: The contents and views expressed in this research article are the views of the authors and do not necessarily reflect the views of the organizations they belong to.

References

Abrougui, K., Gabsi, K., Mercatoris, B., Khemis, C., Amami, R. and Chehaibi, S., 2019, "Prediction of organic potato yield using tillage systems and soil properties by artificial neural network (ANN) and multiple linear regressions (MLR)", *Soil and Tillage Research*, **190**, 202-208, doi: https://doi.org/10.1016/j.still.2019.01.011.

Bendu, H., Deepak, B. and Murugan, S., 2016, "Application of GRNN for the prediction of performance and exhaust emissions in HCCI engine using ethanol", *Energy Conversion and Management*, **122**, 165-173, doi: https://doi.org/10.1016/j.enconman.2016.05.061.

Bendu, H., Deepak, B. and Murugan, S., 2017, "Multi-objective optimization of ethanol fuelled HCCI engine performance using hybrid GRNN-PSO", *Applied Energy*, **187**, 601-611, doi: https://doi.org/10.1016/j.apenergy.2016.11.072.

Bhatt, C.K. and Nain, A.S., 2024, "Integration of Sentinel-1A SAR data with crop simulation model for rice yield prediction in Udham Singh Nagar, Uttarakhand", *MAUSAM*, **75**, 3. 649-658, doi: https://doi.org/10.54302/mausam.v75i3.5905.

Chelgani, S.C. and Jorjani, E., 2011, "Microwave irradiation pretreatment and peroxyacetic acid desulfurization of coal and application of GRNN simultaneous predictor", *Fuel*, **90**, 3156-3163, doi: https://doi.org/10.1016/j.fuel.2011.06.045

Chen, T.-C. and Yu, C.-H., 2009, "Motion control with deadzone estimation and compensation using GRNN for TWUSM drive system", *Expert Systems with Applications*, **36**, 10931-10941, doi: https://doi.org/10.1016/j.eswa.2009.02.025

Cheng, J. and Xiong, Y., 2017, "The quality evaluation of classroom teaching based on FOA-GRNN", *Procedia Computer Science*, **107**, 355-360, doi: https://doi.org/10.1016/j.procs.2017.03.117.

Dai, Y., Liu, C., Min, C. et al., 2024, "Research on the thermal error of electric spindle based on ISCA-GRNN", *International Journal of Advanced Manufacturing Technology*, **134**, 5805-5819, doi: https://doi.org/10.1007/s00170-024-14496-6.

Das, R. and Pradhan, M., 2014, "General regression neural network and back propagation neural network modeling for predicting radial overcut in EDM: A comparative study", *World Academy of Science, Engineering and Technology, International Journal of Mechanical, Aerospace, Industrial, Mechatronics and Manufacturing Engineering*, **8**, 799-805, doi: https://doi.org/10.5281/zenodo.2661480

Drummond, S.T., Sudduth, K.A., Joshi, A., Birrell, S.J. and Kitchen, N.R., 2003, "Statistical and neural methods for site-specific yield prediction", *Transactions of the ASAE*, **46**, 1. 5-14, doi: https://doi.org/10.13031/2013.12541.

Dwivedi, A. and Tajer, A., 2024, "GRNN-Based Real-Time Fault Chain Prediction", *IEEE Transactions on Power Systems*, **39**, 1. 934-946, doi: https://doi.org/10.1109/TPWRS.2023.3258740.

Esme, U., Sagbas, A. and Kahraman, F., 2009, "Prediction of surface roughness in wire electrical discharge machining using design of experiments and neural networks", *Iranian Journal of Science and Technology, Transaction B: Engineering*, **33**, B3. 231-240, doi: https://sid.ir/paper/557501/en

Gao, X., Tang, Z., Zhang, H., Xie, Y. and Gui, W., 2025, "Upstream process condition monitoring for froth flotation based on feature performance evaluation and parameter-mapped GRNN", *Expert Systems with Applications*, **259**, 125261, doi: https://doi.org/10.1016/j.eswa.2024.125261.

Gopal, P.M. and Bhargavi, R., 2019, "A novel approach for efficient crop yield prediction", *Computers and Electronics in Agriculture*, **165**, 104968, doi: https://doi.org/10.1016/j.compag.2019.104968.

Guo, W.W. and Xue, H., 2014, "Crop yield forecasting using artificial neural networks: A comparison between spatial and temporal models", *Mathematical Problems in Engineering*, **2014**, 1-7, doi: https://doi.org/10.1155/2014/857865

Izonin, I., Kazantzi, A.K., Tkachenko, R. et al., 2025, "GRNN-based cascade ensemble model for non-destructive damage state

identification: Small data approach", *Engineering with Computers,* **41**, 1. 723-738, doi: https://doi.org/10.1007/s00366-024-02048-1.

Jeong, J.H., Resop, J.P., Mueller, N.D., Fleisher, D.H., Yun, K., Butler, E.E., Timlin, D.J., Shim, K.M., Gerber, J.S., Reddy, V.R. and Kim, S.H., 2016, "Random forests for global and regional crop yield predictions", *PLOS ONE,* **11**, 6. 1-15, doi: https://doi.org/10.1371/journal.pone.0156571.

Jiang, W., Wang, J., Hsiung, K.-L. and Chen, H.-Y., 2024, "GRNN-Based detection of eavesdropping attacks in SWIPT-enabled smart grid wireless sensor networks", *IEEE Internet of Things Journal,* **11**, 22. 37381-37393, doi: https://doi.org/10.1109/JIOT.2024.3443277.

Kaul, M., Hill, R.L. and Walthall, C., 2005, "Artificial neural networks for corn and soybean yield prediction", *Agricultural Systems,* **85**, 1. 1-18, doi: https://doi.org/10.1016/j.agsy.2004.07.009

Kaur, K., 2023, "Artificial Neural Network Model to Forecast Energy Consumption in Wheat Production in India", *Journal of Statistical Theory and Applications*, **22**, 1. 19-37, doi: https://doi.org/10.1007/s44199-023-00052-w

Khaki, S., Pham, H. and Wang, L., 2021, "Simultaneous corn and soybean yield prediction from remote sensing data using deep transfer learning", *Scientific Reports,* **11**, 1. 1-14, doi: https://doi.org/10.1038/s41598-021-89779-z.

Kumar, G. and Malik, H., 2016, "Generalized regression neural network based wind speed prediction model for Western region of India", *Procedia Computer Science,* **93**, 26-32, doi: https://doi.org/10.1016/j.procs.2016.07.177.

Kumari, S., Venkatesh, V.G., Tan, F.T.C. et al., 2023, "Application of machine learning and artificial intelligence on agriculture supply chain: A comprehensive review and future research directions", *Annals of Operations Research,* **348,** 1573–1617, doi: https://doi.org/10.1007/s10479-023-05556-3.

Land, W.H., Ma, X., Barnes, E., Qiao, X., Heine, J., Masters, T. and Park, J.W., 2012, "PNN/GRNN ensemble processor design for early screening of breast cancer", *Procedia Computer Science,* **12**, 438-443, doi: https://doi.org/10.1016/j.procs.2012.09.101

Liu, K., Lin, T., Zhong, T., Ge, X., Jiang, F. and Zhang, X., 2023, "New methods based on a genetic algorithm back propagation (GABP) neural network and general regression neural network (GRNN) for predicting the occurrence of trihalomethanes in tap water", *Science of the Total Environment,* **870**, 161976, doi: https://doi.org/10.1016/j.scitotenv.2023.161976

Majkovič, D., O'Kiely, P., Kramberger, B., Vračko, M., Turk, J., Pažek, K. and Rozman, Č., 2016, "Comparison of using regression modeling and an artificial neural network for herbage dry matter yield forecasting", *Journal of Chemometrics,* **30**, 4. 203-209, doi: https://doi.org/10.1002/cem.2770

Oğuz, H., Saritas, I. and Baydan, H.E., 2010, "Prediction of diesel engine performance using biofuels with artificial neural network", *Expert Systems with Applications,* **37**, 9. 6579-6586, doi: https://doi.org/10.1016/j.eswa.2010.02.128.

Özener, O., Yüksek, L. and Özkan, M., 2013, "Artificial neural network approach to predicting engine-out emissions and performance parameters of a turbocharged diesel engine", *Thermal Science,* **17**, 1. 153-166, doi: 10.2298/TSCI120321220O

Polat, Ö. and Yıldırım, T., 2008, "Genetic optimization of GRNN for pattern recognition without feature extraction", *Expert Systems with Applications,* **34**, 4. 2444-2448, doi: https://doi.org/10.1016/j.eswa.2007.04.006

Pradhan, M. and Das, R., 2015, "Application of a general regression neural network for predicting radial overcut in electrical discharge machining of AISI D2 tool steel", *International Journal of Machining and Machinability of Materials,* **17**, 3-4. 355-369, doi: https://doi.org/10.1504/IJMMM.2015.071998

Raghunath, K.M.K., 2023, "Rice and Wheat crop yield prophesy", *IEEE Dataport,* doi: https://doi.org/10.21227/bcpj-af28.

Ranjan, A.K. and Parida, B.R., 2019, "Paddy acreage mapping and yield prediction using Sentinel-based optical and SAR data in Sahibganj district, Jharkhand (India)", *Spatial Information Research,* **27**, 4. 399-410, doi: https://doi.org/10.1007/s41324-019-00246-4.

Rao, D.T.V.N. and Manasa, S., 2019, "Artificial neural networks for soil quality and crop yield prediction using machine learning", *International Journal on Future Revolution in Computer Science & Communication Engineering,* **5**, 1. 57-60, doi: https://www.ijfrcsce.org/index.php/ijfrcsce/article/view/1835

Ray, R. and Roy, L.B., 2024, "Probabilistic stability analysis of earthen slope using ANN, PSO-ANN, GPR and GA-ANFIS soft computing techniques", *Indian Geotechnical Journal,* doi: https://doi.org/10.1007/s40098-024-01039-9.

Roberts, M.J., Braun, N.O., Sinclair, T.R., Lobell, D.B. and Schlenker, W., 2017, "Comparing and combining process-based crop models and statistical models with some implications for climate change", *Environmental Research Letters,* **12**, 9, 095010, doi: https://doi.org/10.1088/1748-9326/aa7f33.

Rooki, R., 2016, "Application of general regression neural network (GRNN) for indirect measuring pressure loss of Herschel-Bulkley drilling fluids in oil drilling", *Measurement,* **85**, 184-191, doi: https://doi.org/10.1016/j.measurement.2016.02.037

Safa, M. and Samarasinghe, S., 2011, "Determination and modeling of energy consumption in wheat production using neural networks: A case study in Canterbury province, New Zealand", *Energy,* **36**, 8. 5140-5147, doi: https://doi.org/10.1016/j.energy.2011.06.016.

Satpathi, A., Setiya, P., Das, B., Nain, A.S., Jha, P.K., Singh, S. and Singh, S., 2023, "Comparative analysis of statistical and machine learning techniques for rice yield forecasting for Chhattisgarh, India", *Sustainability,* **15**, 3. 2786, doi: https://doi.org/10.3390/su15032786

Schlenker, W. and Roberts, M.J., 2009, "Nonlinear temperature effects indicate severe damage to U.S. crop yields under climate change", *Proceedings of the National Academy of Sciences,* **106**, 37, 15594-15598, doi: https://doi.org/10.1073/pnas.0906865106.

Shahhosseini, M., Hu, G. and Archontoulis, S., 2020, "Forecasting corn yield with machine learning ensembles", *Frontiers in Plant Science,* **11**, 01120, doi: https://doi.org/10.3389/fpls.2020.01120.

Son, N.T., Chen, C.F., Cheng, Y.S., Toscano, P., Chen, C.R., Chen, S.L., Tseng, K.H., Syu, C.H., Guo, H.Y. and Zhang, Y.T., 2022, "Field-scale rice yield prediction from Sentinel-2 monthly image composites using machine learning algorithms", *Ecological Informatics,* **69**, 101618, doi: https://doi.org/10.1016/j.ecoinf.2022.101618.

Specht, D.F., 1991, "A general regression neural network", *IEEE Transactions on Neural Networks,* **2**, 6. 568-576, doi: https://doi.org/10.1109/72.97934.

Srinivasarao, M., Srinivasarao, C., Kumari, A.S., Rao, B.J., Gandhi, P., PraveenKumar, S., Samuel, O.D., Mustafa, A., Enweremadu,

C.C. and Elboughdiri, N., 2025, "Combustion enhancement and emission reduction in an IC engine by adopting ZnO nanoparticles with calophyllum biodiesel/diesel/propanol blend: A case study of General Regression Neural Network (GRNN) modeling", *Industrial Crops and Products*, **227**, 120812, doi: https://doi.org/10.1016/j.indcrop.2025.120812

Srivastava, A.K., Safaei, N., Khaki, S. et al., 2022, "Winter wheat yield prediction using convolutional neural networks from environmental and phenological data", *Scientific Reports*, **12**, 3215, doi: https://doi.org/10.1038/s41598-022-06249-w.

Taheri-Rad, A., Khojastehpour, M., Rohani, A., Khoramdel, S. and Nikkhah, A., 2017, "Energy flow modeling and predicting the yield of Iranian paddy cultivars using artificial neural networks", *Energy*, **135**, 405-412, doi: https://doi.org/10.1016/j.energy.2017.06.089.

Taoufik, N., Boumya, W., Elmoubarki, R., Elhalil, A., Achak, M., Abdennouri, M. and Barka, N., 2022, "Experimental design, machine learning approaches for the optimization and modeling of caffeine adsorption", *Materials Today Chemistry*, **23**, 100732, doi: https://doi.org/10.1016/j.mtchem.2021.100732

Thurkkaivel, T., Dheebakaran, G.A., Geethalakshmi, V., Patil, G.S. and Bhuvaneshwari, K., 2021, "Comparison statistical rice yield prediction with multiple weather parameters", *International Journal of Plant & Soil Science*, **33**, 22, 31-38, doi: https://hal.science/hal-05350047/

Vashisth, A., Aravind, K.S., Das, B. and Krishnan, P., 2024, "Multi stage wheat yield estimation using multiple linear, neural network and penalised regression models: Wheat yield estimation", *MAUSAM*, **74**, 3. 833-846, doi: https://doi.org/10.54302/mausam.v74i3.1923.

Wang, A.X., Tran, C., Desai, N., Lobell, D. and Ermon, S., 2018, "Deep transfer learning for crop yield prediction with remote sensing data", *ACM SIGCAS Conference on Computing and Sustainable Societies (COMPASS)*, **50**, 1-5, doi: https://doi.org/10.1145/3209811.3212707

Xue, Z., Fu, J., Fu, Q., Li, X. and Chen, Z., 2023, "Modeling and optimizing the performance of green forage maize harvester header using a combined response surface methodology-artificial neural network approach", *Agriculture*, **13**, 1890, doi: https://doi.org/10.3390/agriculture13101890

Yadav, R.N. and Yadava, V., 2015, "Application of soft computing techniques for modeling and optimization of slotted-electrical discharge diamond face grinding process", *Transactions of the Indian Institute of Metals*, **68**, 981-990, doi: https://doi.org/10.1007/s12666-015-0536-6

Yao, D., Shen, J., Hou, J. et al., 2024, "Online vehicle velocity prediction based on an adaptive GRNN with various input signals", *International Journal of Automotive Technology*, doi: https://doi.org/10.1007/s12239-024-00172-x.

Yonaba, H., Anctil, F. and Fortin, V., 2010, "Comparing sigmoid transfer functions for neural network multistep ahead streamflow forecasting", *Journal of Hydrologic Engineering*, **15**, 4. 275-283, doi: https://doi.org/10.1061/(ASCE)HE.1943-5584.0000188

Yu, W., Yang, G., Li, D., Zheng, H., Yao, X., Zhu, Y., Cao, W., Qiu, L. and Cheng, T., 2023, "Improved prediction of rice yield at field and county levels by synergistic use of SAR, optical, and meteorological data", *Agricultural and Forest Meteorology*, **342**, 109729, doi: https://doi.org/10.1016/j.agrformet.2023.109729

Zeroual, A., Fourar, A. and Djeddou, M., 2019, "Predictive modeling of static and seismic stability of small homogeneous earth dams using artificial neural network", *Arabian Journal of Geosciences*, doi: https://doi.org/10.1007/s12517-018-4162-6.

Zhang, D., Liu, Z., Song, K., Zhai, Z., Zhang, Y. and Gao, Z., 2024, "Parameters optimization for laser-cladding CoCrFeNiMn high-entropy alloy layer based on GRNN and NSGA-II", *Materials Today Communications*, **39**, 108615, doi: https://doi.org/10.1016/j.mtcomm.2024.108615

Zhang, T., Zhang, J., Hou, H. and Xu, Y., 2024, "The prediction of sound absorption coefficient of film multi-cavity materials based on generalized regression neural network (GRNN)", *Applied Acoustics*, **224**, 110134, doi: https://doi.org/10.1016/j.apacoust.2024.110134