International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2024; SP-9(1): 387-394

https://www.mathsjournal.com
Received: 07-11-2023
Accepted: 09-12-2023

Obaid Zaffar
Ph.D. Scholar, Division of Farm Machinery and Power Engineering, Sher-e-Kashmir University of Agricultural Sciences and Technology of Jammu, Jammu and Kashmir, India

Sanjay Khar
Professor and Head, Division of Farm Machinery and Power Engineering, Sher-e-Kashmir University of Agricultural Sciences and Technology of Jammu, Jammu and Kashmir, India

Sushil Sharma
Dean, Faculty of Agricultural Engineering, Sher-e-Kashmir University of Agricultural Sciences and Technology of Jammu, Jammu and Kashmir, India

JP Singh
Professor, Division of Farm Machinery and Power Engineering, Sher-e-Kashmir University of Agricultural Sciences and Technology of Jammu, Jammu and Kashmir, India

Corresponding Author:
Obaid Zaffar
Ph.D. Scholar, Division of Farm Machinery and Power Engineering, Sher-e-Kashmir University of Agricultural Sciences and Technology of Jammu, Jammu and Kashmir, India

Forecasting of wheat production, productivity and cultivated area in India using artificial neural networks

Obaid Zaffar, Sanjay Khar, Sushil Sharma and JP Singh

Abstract
The population of the world is increasing at a rapid rate and more specifically in India resulting in increase of the food per capita requirement. Wheat plays an important role in ensuring global food with the demand of over 40 percent for the year 2030. In order to ascertain this requirement a reliable forecasting is essential for decision-makers to plan adequate policies and to establish the necessary logistical resources. In this sense, the Artificial Neural Networks was used to predict wheat production, productivity and cultivated area of wheat in India and compare the same with the classical methods of time series data. The time series data of 60 years from 1961 to 2021 of India was collected from the Food and Agriculture Organization of United Nations. The data was analysed using classical methods namely linear, exponential, logarithm, polynomial and power function. During the study the respective trendlines and equations were drawn using MS Excel for classical methods and also evaluation matrices parameter namely coefficient of determination (R^2), Mean Squared Error (MSE) and Mean Absolute Error (MAE) were estimated to determine the suitability of the model with respect to the time series data. The same data was then used in prediction analysis using Artificial Neural Networks (ANN) on R studio software and similar evaluation metrices were estimated. From the study it has been observed that the ANN proved to be the best model in predicting the production, productivity and cultivated area of the wheat with the highest R values of 0.9 and least error values of both MSE and MAE. For validation of data with the actual, the activation functions like linear and tangent function predicted the values nearer to the actual values. Furthermore, artificial neural networks (ANNs) are considered a dependable model for forecasting time series data, enabling stakeholders to plan in advance for the rising demands.

Keywords: Wheat, production, productivity, cultivated area, artificial neural networks, classical methods.

1. Introduction

The world population has tripled since the middle of the twentieth century, reaching almost 8 billion people in the year 2022 [16]. It has been also reported that the global population is growing by about 75 million every year, has increased nearly fourfold over the past 100 years and is projected to reach 9.2 billion by the year 2050 [15]. The largest population increases are projected to occur in Asia particularly in China, India and Southeast Asia, which are expected to collectively account for approximately 60% of the world's population by the year 2050. As population have increased there has been a corresponding rise in food consumption per capita, leading to an increase in global calorie intake per individual. In the year 2020, the United Nations reported that an alarming 2.4 billion people, which accounted for over 29 percent of the global population, experienced moderate to severe food insecurity, lacking consistent access to sufficient nourishment [10]. The figure increased by nearly 320 million people in just one year and according to the International Food Policy Research Institute's (IFPRI) Global Hunger Index, 33 nations are currently subject to "alarming" or "extremely alarming" levels of hunger. As of the Global Hunger Index (GHI) in 2022, India was positioned at 107 out of 121 countries, with only 14 nations performing worse. India's GHI score of 29.1 categorizes it as "serious," placing it below countries like Sri Lanka (64), Nepal (81), Bangladesh (84), and Pakistan (99) [11].

Wheat is one of the world's ancient and most extensively used crop and together with rice and maize forms a major component of the human diet, accounting for nearly half of the world's food calorie and two-fifths of protein intake.

Wheat plays a crucial role in ensuring global food/nutrition security supplying a fifth of global food calories and protein [8]. A significant portion of the world's population depends on wheat as their primary source of sustenance. Beyond addressing hunger, wheat serves as a fundamental source of both protein and carbohydrates. The projections indicate that the global demand for wheat is expected to surpass 40% by the year 2030. In order to meet this increasing worldwide demand for wheat, it becomes essential to increase the cultivation rates by approximately 2% annually [7].

Hence, our research objectives revolve around forecasting wheat production, productivity and cultivated area in India through the application of Artificial Neural Networks (ANN) and subsequently comparing the results with traditional methods of time series analysis. Kaur, 2023 [12] employed Artificial Neural Networks (ANN) to model and predict the input energy consumption associated with wheat production in India. The study also involved a comparison of the ANN's predictive accuracy with that of linear models. Notably, the study found that employing the sigmoid activation function in both the hidden and output layers provided the most favorable outcomes. The ANN technique is not only used with the time series data but can be used to draw comparison between different operational methods used in the agriculture. Basir *et al*., 2021 [4] conducted an experiment to investigate the relation of rice yield with respect to the mechanical rice transplanting parameters (i.e., seedling density in the tray, missing hill percentage, floating hill percentage and seedling number per hill) using Artificial Neural Networks (ANN). The results reported that the ANN model provided an accurate match in predicting yield from transplanting parameters with an R value of 0.994. Abraham *et al*. (2020) [1] employed the Artificial Neural Network (ANN) technique to estimate and predict the production of soybeans. Their study incorporated various input variables, which included monthly climatic factors such as air temperature, precipitation, global radiation, crop evapotranspiration, soil water storage, actual evapotranspiration, as well as water deficiencies and surpluses throughout the crop's growth cycle.

2. Materials and Methods

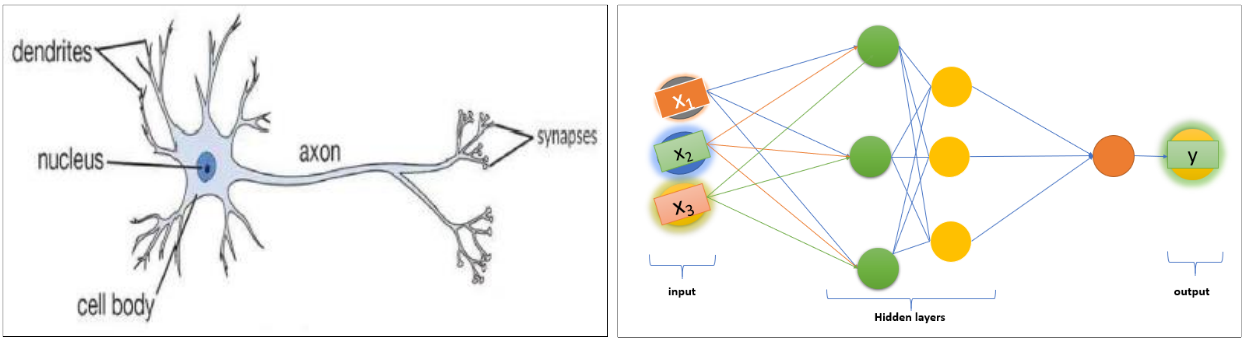
2.1 Artificial neural networks

The Artificial Neural Networks (ANN) is a non-parametric method that performs same operations as that of the human brain as shown in Figure 1 [5]. ANNs consists of flexible mathematical structures for adjustments to identify the complex non-linear relationships between input and output data using the same data. The ANN neurons are grouped in layers namely input layer, hidden layers and the output layer [2]. The hidden layers are between the input and output layers and the signals are transferred from input layer through hidden layers to output layer. Each neuron in the hidden layers and output layer sums the corresponding weighted inputs and computes its output according to a transfer function. In the case of a hidden layer, this output is passed on to the next layer; whereas, in case of the output layer, neuron(s) output is the network output. The functional form of an artificial neural network can be given as

$$ne = \sum_{i=1}^{n} XiWi + bi$$

$$u = f(ne)$$

Where Xi…Xn are the input values of the data set, Wi…Wn are the weights, and b is the activation threshold (bias) in the neuron potential ne [6, 14].



Nuerons of a Human brain   Artificial Nueral Netwoks

Fig 1: Pictorial analogy of Human brain neurons and Artificial neural networks

2.2 Time series and classical methods

Time series analysis was utilized to examine the patterns in historical data by employing various classical methods (as detailed in Table 1). This analysis aims to uncover underlying trends within the dataset and assess how effectively particular methods or equations replicate these historical data patterns. To determine the most appropriate model for a given dataset, evaluation metrics including the coefficient of determination (R), mean absolute error (MAE), and mean squared error (MSE) were employed. These metrics help gauge the accuracy and goodness of fit of the models to the historical data.

Table 1: List of the classical methods with their characteristic

Classical method	Trend formula	Features
Linear function	$y = ax \pm c$	A linear curve is defined as a curve of the first degree, often represented by a straight line.
Logarithm function	$\ln(y) = ax \pm c$	Exponential refers to a transcendental curve characterized by rapid growth or decay. In this context, 'e' denotes the base of natural logarithms, with a constant value approximately equal to 2.7813.
Exponential function	$y = ae^{bx}$	Exponential function is the inverse of the logarithm function
Polynomial function	$y = ax^2 \pm bx \pm c$	A polynomial function comprises of variables raised to non-negative integer powers. It represents a wide range of curves and shapes in mathematics. The second-degree polynomial functions give rise to parabolic curves, which have a characteristic U-shape.
Power function	$y = ax^b$	A power curve is typically associated with hyperbolic functions. These curves are characterized by their asymptotic behavior, where the graph approaches but never reaches certain lines

2.3 Evaluation matrices

There are various matrices to evaluate the performance of the model in regression analysis namely Mean Square Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Coefficient of Determination (R). During the study the MSE, MAE and R were used to determine the suitable model or function and also to compare the same with each other.

Mean Squared Error (MSE) serves as a statistical metric used to assess the quality of a predictive model, particularly in the context of regression analysis. It quantifies the average of the squared differences between the actual (original) values and the corresponding predicted values within a dataset [1]. A lower MSE indicates that the model's predictions are closer to the actual values.

$$Mean\ Squared\ Error\ (MSE) = 1/N \sum_{i=1}^{N} (y_i - \hat{y})^2$$

Mean Absolute Error (MAE) is used to measures how close the predictions are to the actual values, without considering the direction of the errors (overestimation or underestimation). A lower MAE indicates a better fit of the model to the data.

$$Mean\ Absolute\ Error\ (MAE) = 1/N \sum_{i=1}^{N} |y_i - \hat{y}|$$

Coefficient of Determination (R^2) quantifies the extent to which a linear regression model can elucidate the relationship between variables. It signifies the goodness of fit. The closer this number is to one, the more fitted is the model [1].

$$Coeffiecent\ of\ Determination = 1 - \frac{\sum (y_i - \hat{y})^2}{\sum (y - \bar{y})^2}$$

Where y represents the real value of the series, ŷ is the expected value (value of the regression line approaching the actual value), ȳ and is the average value of the series

2.4 Data set

In order to predict the wheat production in India, a huge data was required which was acquired from the Food and Agriculture Organization of the United Nations [9] related to the cultivated area (million hectares), production (million tonnes) and the productivity (tonnes/ hectare) of the wheat crop between the year of 1961-2021. The data was firstly subjected to time series analysis in R software and the various graphs and trendlines were derived. The details of various time series analysis equation derived from the data are presented in the table 2 to 4 for production, productivity and harvested area respectively.

Table 2: Effective functions for forecasting wheat production

Model	Trend formula
Linear function	Y= 1.64x – 3201.25
Logarithm function	ln(Y) = 0.037x -71.12
Exponential function	Y = 1.290285e-31 × $e^{0.0376x}$
Polynomial function	Y= 7.408737x^2 +224.8405x + 54.11644
Power function	Y = 1.738899e-246 $x^{74.99}$

Table 3: Effective functions for forecasting wheat productivity

Model	Trend formula
Linear function	Y= 0.45x - 875.86
Logarithm function	ln(Y) = 0.023x - 44.69
Exponential function	Y = 3.867329e-20 * $e^{0.02394279x}$
Polynomial function	Y = -3.667054x^2+ 61.96297x + 21.27143 +
Power function	Y= 7.692526e-157 $x^{47.71192}$

Table 4: Effective function for forecasting wheat cultivated area

Model	Trend formula
Linear function	Y= 0.29x -563.59
Logarithm function	ln(Y) = 0.01367371x - 24.09902
Exponential function	Y = 3.419238e-11 * $e^{0.01367371x}$
Polynomial function	Y = -7.256599x^2 + 40.54765x + 23.47892
Power function	Y = 2.7598e-89 $x^{27.25522}$

2.5 Model classification

The discrepancies between the actual (original) values and the predicted values were calculated using three common metrics namely R, MAE and MSE. A comparison was made between the classical models and neural networks by organising evaluation matrices in which error metrices i.e., MSE and MAE were sorted in ascending order i.e, from lowest to highest. However, regression measures R were sorted in descending order from highest to lowest [1]. The equation for the same is given as;

$$\text{Rank of model} = \frac{(\text{MAE} \times 0.5)+(\text{MSE} \times 0.5)+(\text{R}+1)}{2} \qquad \text{(A)}$$

Where MAE is mean absolute error, MSE is mean squared error and R is the coefficient of determination

3. Results and Discussion

3.1 Time series analysis using classical methods

The time series analysis was carried out using classical methods on production, productivity and cultivated area of wheat to determine their accuracy in predicting the actual variation of the dataset of the period from 1961-2021. The evaluation matrices like MSE, MAE and R^2 were used to determine it and details are given in subsequent heading below.

Production of wheat crop

From the time series dataset, in early 1960's the production ranged from 9.85 to 12.26 units and significant increase in production was observed in the late 1960s. Subsequently, the production continued to rise, reaching its highest point at 109.59 million tonnes in the year 2021. The data shows periods of growth and decline, with some years experiencing significant jumps in production, such as the early 1970s, early 1980s, and late 2000s. Overall, upward trajectory with periodic variations was observed (Figure 2). In regard to find suitable function for predicting the production of wheat the polynomial and linear function were found most suitable on the basis of R, MAE, and MSE as shown in Table 4.

Table 5: Effective functions for forecasting of production in wheat.

Rank	Model	R	MAE	MSE
1	Polynomial model	0.992	2.8796	13.392
2	Linear model	0.991	2.9651	14.293
3	Power model	0.933	8.3326	108.864
4	Logarithm model	0.931	8.4230	111.754
5	Exponential model	0.931	8.4230	111.754

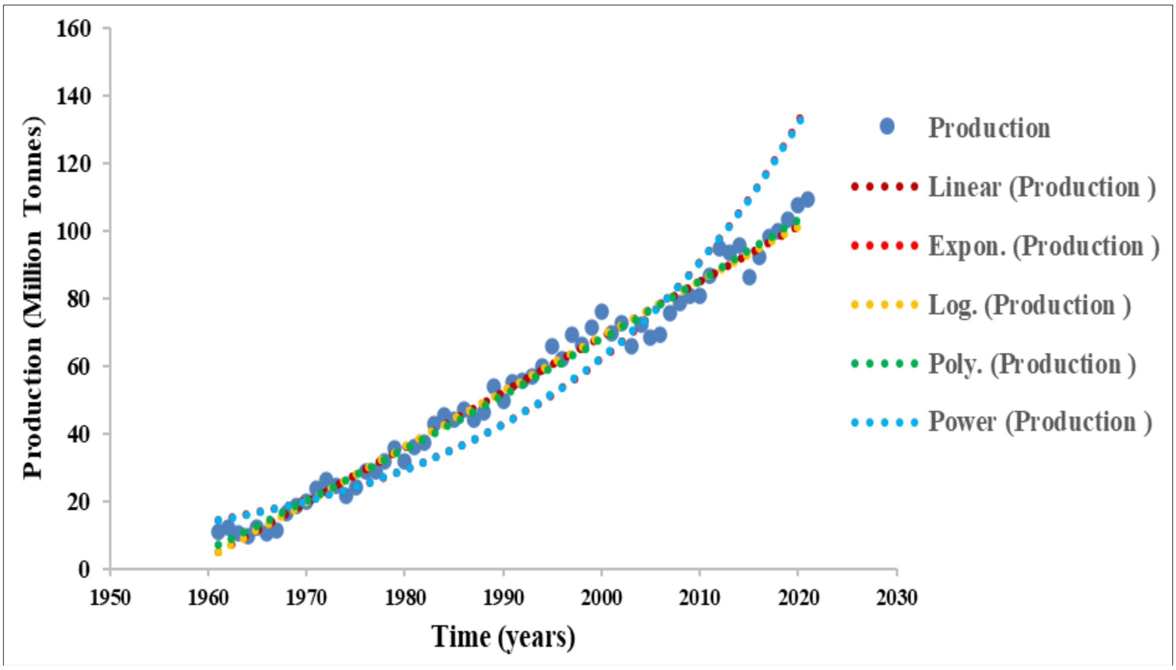


Fig 2: Graphical trend line of various classical methods w.r.t to the actual time series production data of wheat

Productivity of wheat crop

From the dataset it was observed the productivity was at its lowest in the year 1964 with a value of 7.299 q/ha and peaked in the year 2019 with productivity of 35.334 q/ha. In general, both the production and productivity values showed an upward trend and the significant increase in productivity was during the early 1970s. The subsequent years witnessed consistent growth, although there were fluctuations along the way as illustrated in the Figure 3. In terms of the suitability of the model among the classicals models the polynomial and linear function was found most suitable on the basis of R, MAE, and MSE as shown in Table 6 below.

Table 6: Effective functions for forecasting productivity of wheat.

Rank	Model	R	MAE	MSE
1	Polynomial model	0.988	1.0055	1.5531
2	Linear model	0.986	1.0303	1.7736
3	Power model	0.996	2.0356	6.3380
4	Logarithm model	0.949	2.0594	6.4902
5	Exponential model	0.949	2.0594	6.4902

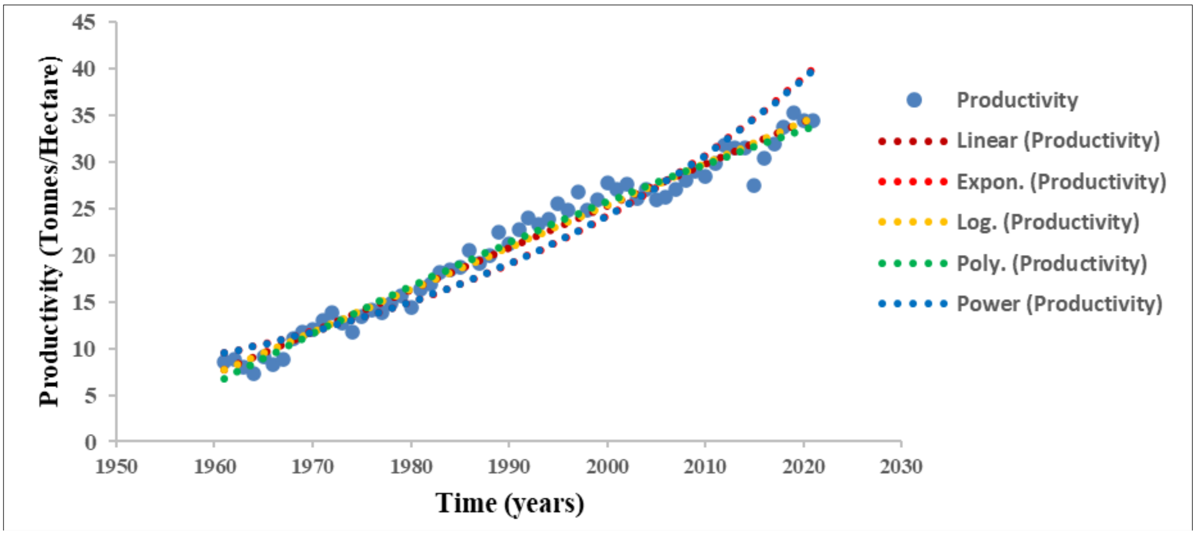


Fig 3: Graphical trend line of various classical methods w.r.t to the actual time series productivity data of wheat

Cultivated area of wheat crop

From the dataset for the period of 1961 to 2021, the distinctive trends in cultivated area were observed. In the earlier years a relative stable cultivable area was observed and a sharp rise was recorded in cultivated area from the late 1960s to the late 1970s reaching its zenith around 1979 at 22.64 million hectares. From approximately 2010 onwards a steady climb was observed culminating in 31.61 million hectares by 2021.

In terms of the suitability of the model among the classicals models the polynomial and linear function was found most suitable on the basis of R, MAE, and MSE as shown in Table 7 below.

Table 7: Effective functions for forecasting harvested area of wheat.

Rank	Model	R	MAE	MSE
1	Polynomial model	0.978	0.905453	1.2411
2	Linear model	0.708	1.194718	2.1044
3	Power model	0.936	1.641869	3.6188
4	Logarithm model	0.935	1.654918	3.6772
5	Exponential model	0.935	1.654918	3.6772

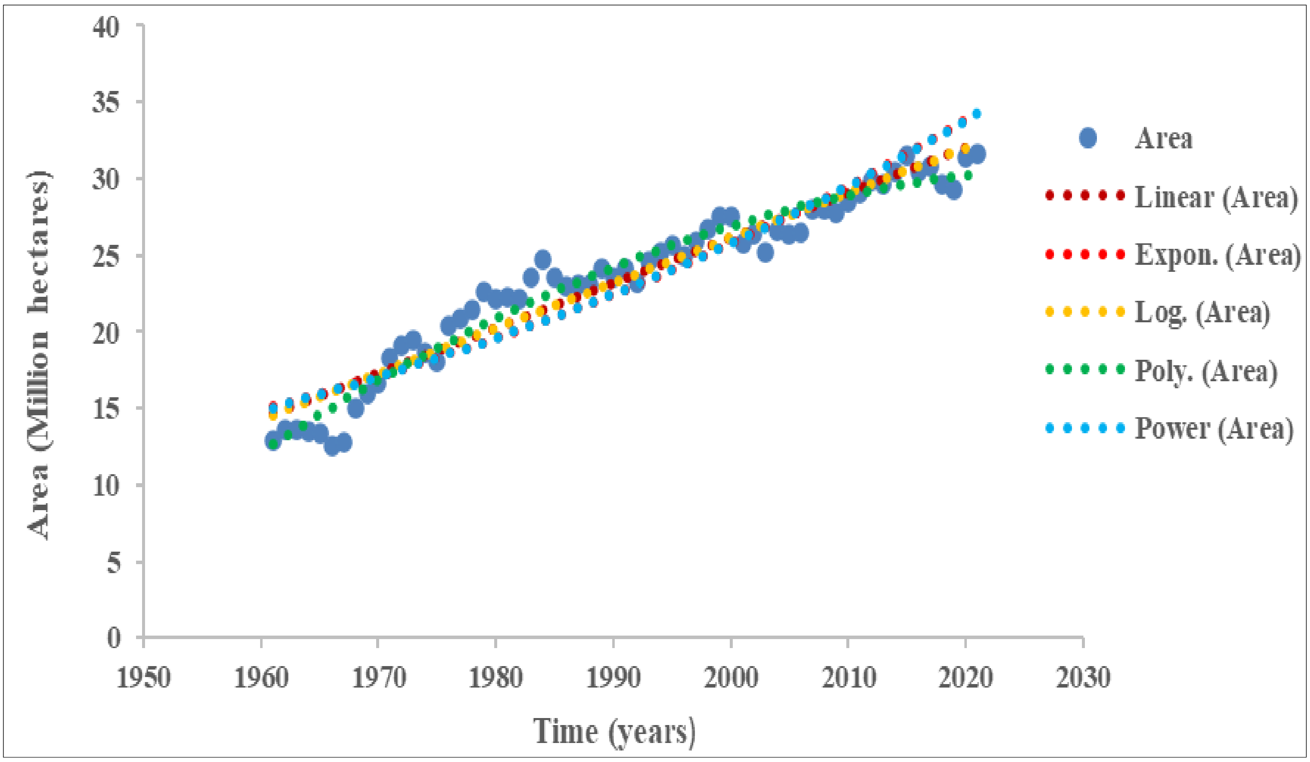


Fig 4: Graphical trend line of various classical methods w.r.t to the actual time series cultivated area data of wheat

3.2 Time series analysis using Artificial Neural Network

The prediction of time series data of production, productivity and cultivated area using artificial neural network (ANN) showed better adjustments. The classical methods specifically polynomial and linear function provided fare values of the coefficient of determination R i.e., nearer to 1 however all classical function produced large values of errors. But ANN provided desirable R value with minimum errors production, productivity and cultivated area. The trend line of an ANN is shown in the Figure 4 to 6.

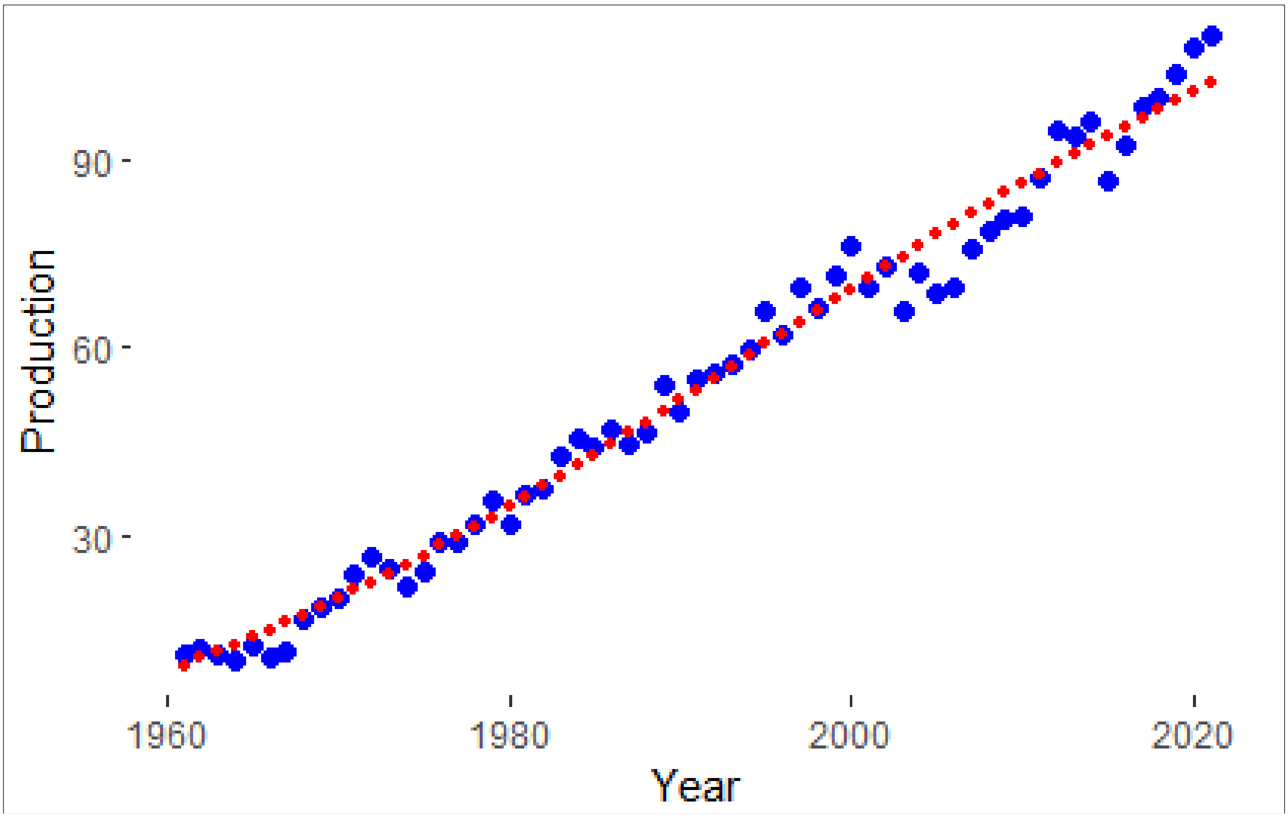


Fig 5: Graphical trend line of ANN w.r.t to the actual time series production of wheat

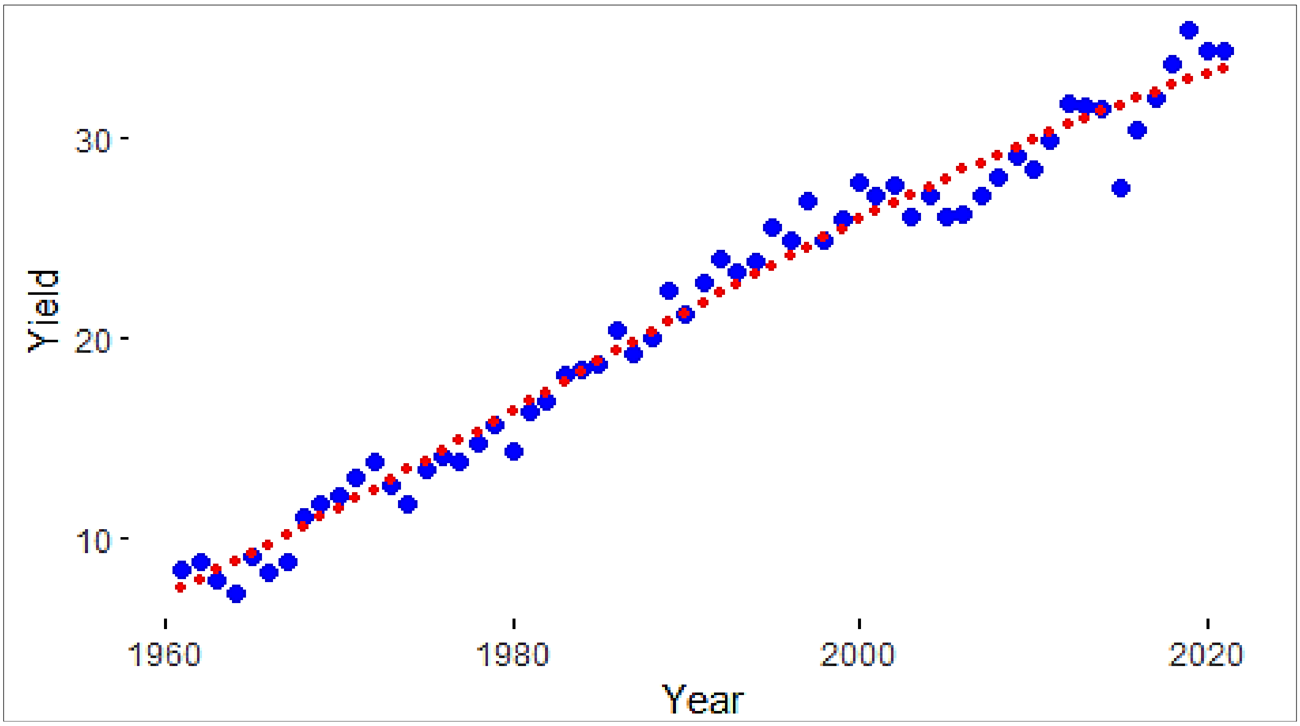


Fig 6: Graphical trend line of ANN w.r.t to the actual time series productivity of wheat

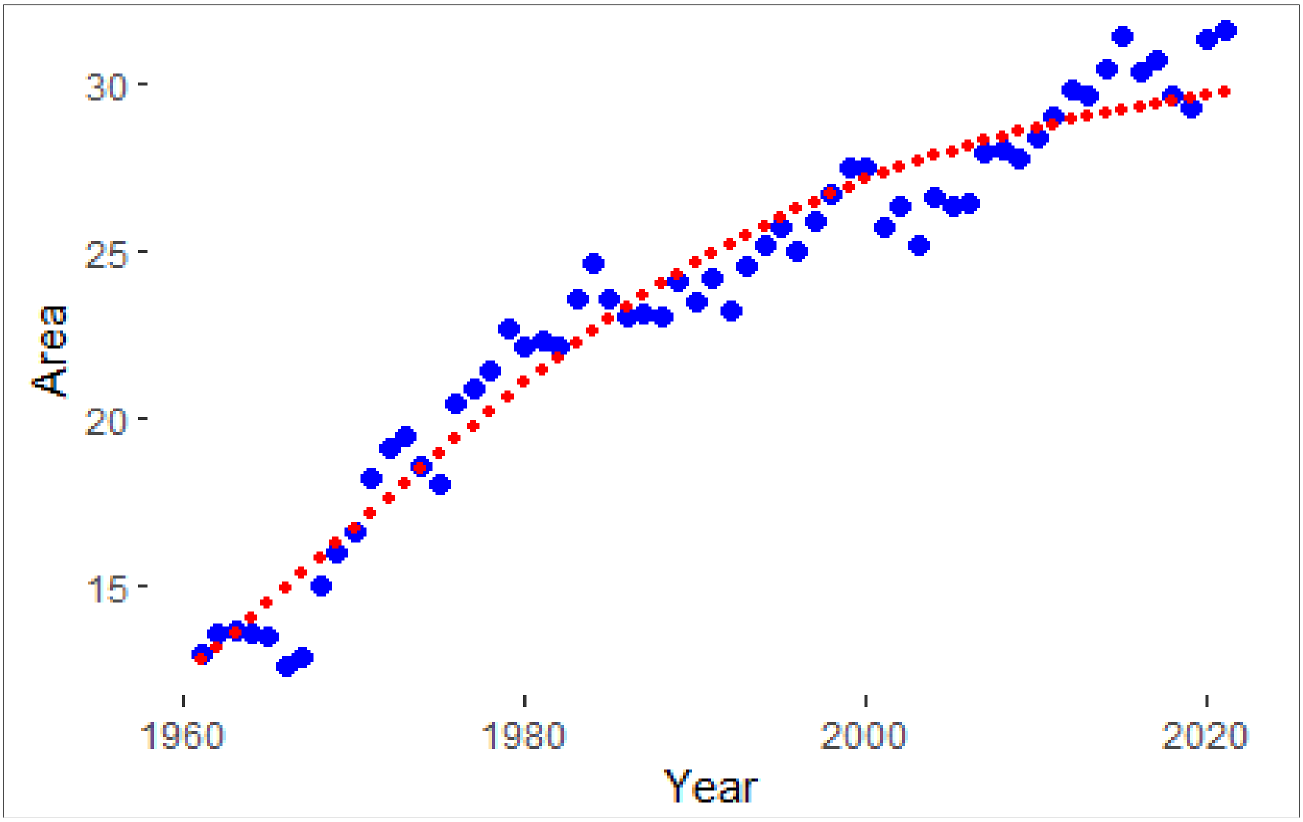


Fig 7: Graphical trend line of ANN w.r.t to the actual time series of cultivated area of wheat

3.3 Comparison between artificial neural networks and time series classical models

The comparison between Artificial Neural Networks (ANN) and classical time series models was conducted to predict wheat production, productivity, and cultivated area based on evaluation metrics namely coefficient of determination (R), Mean Squared Error (MSE), and Mean Absolute Error (MAE). Using equation (A), rankings were assigned to these models.

Upon comparison, the results indicated that Artificial Neural Networks achieved the highest ranking, demonstrating their superior performance in predicting wheat production, productivity, and cultivated area. Following ANN, polynomial and linear models were ranked next in line for their predictive abilities in these areas, as shown in Table 4 to 6. On the other hand, exponential and logarithmic models were found to be the least suitable for predicting these aspects of wheat crop in India.

Table 8: Effective functions versus ANN for predicting production of wheat for the period of 1961-2021

Rank	Model	R	MAE	MSE
1	ANN	0.991	0.0301	0.0015
2	Polynomial	0.992	2.8796	13.392
3	Linear	0.991	2.9651	14.293
4	Power	0.933	8.3326	108.864
5	Logarithm	0.931	8.4230	111.754
6	Exponential	0.931	8.4230	111.754

From the table it is observed the R value of ANN, polynomial and linear model are almost same i.e., 0.991 but in both polynomial and linear model MAE and MSE value are significantly higher in comparison to the ANN value. Thus, ANN model was ranked first in predicting the production of wheat for time series data of 1961 to 2021.

Table 9: Effective functions versus ANN for predicting productivity of wheat for the period of 1961-2021

Rank	Model	R	MAE	MSE
1	ANN	0.988	0.0353	0.0019
2	Polynomial	0.988	1.0055	1.5531
3	Linear	0.986	1.0303	1.7736
4	Power	0.996	2.0356	6.3380
5	Logarithm	0.949	2.0594	6.4902
6	Exponential	0.949	2.0594	6.4902

From the Table 9, similar results as that of production was observed for predicting productivity of wheat crop with highest rank to the ANN model followed by polynomial and linear with the corresponding values of R as 0.988, 0.988 and 0.986 respectively. The MAE and MSE was also found favorable in case of ANN. The logarithm and exponential model were least favorable for predicting the productivity of wheat.

In case of cultivated area, the R was found highest for the polynomial model followed by power and ANN model with R^2 values of 0.978, 0.9369 and 0.933 respectively. However, the MAE and MSE values of the ANN were found minimum with the values of 0.0363 and 0.002 respectively and was ranked first for prediction of the cultivated area of wheat. Thus, the ANN model has more predictive capabilities for predicting the production, productivity and cultivated area of wheat crop in contrast to the existing classical methods in respect to the time series dataset.

Table 10: Effective functions versus ANN for predicting cultivated area of wheat for the period of 1961-2021

Rank	Model	R	MAE	MSE
1	ANN	0.933	0.0363	0.002
2	Polynomial	0.978	0.905453	1.2411
3	Linear	0.708	1.194718	2.1044
4	Power	0.936	1.641869	3.6188
5	Logarithm	0.935	1.654918	3.6772
6	Exponential	0.935	1.654918	3.6772

3.4 Validation of the Artificial Neural Network Model

The validation of the ANN model was performed on forecast basis for production, productivity and cultivated area of wheat in India for the year of 2022. The predicted value from the model was compared with the actual data of the year 2022. The various functions were used in order to match the actual values of the data of wheat crop of year 2022. The details are provided in Table 11 below.

Table 11: Comparison of actual and prediction data of wheat using Artificial Neural Networks with different activation functions

Variable	Actual data (2022)	Linear activation function	Sigmoid activation function	Tangent activation function	Leaky ReLU activation function
Production (Mt)	107.6	102.66	92.66	104.88	104.79
Productivity (t/ha)	35.07	32.91	31.05	32.53	32.53
Cultivated area (Mha)	31.35	31.00	30.38	30.75	29.71

In general, the ANN provided better results in comparison to the classical methods however in ANN itself the production was predicted more accurately by using tangent activation function in it. In case of productivity (t/ha), the linear activation gave the results nearer to the actual values. However, in case of cultivated land all functions predicted near to the actual value expect that of Leaky ReLu activation functions thus for prediction of cultivated area of wheat any of the four functions in ANN can be utilized.

4. Conclusions

The study comprised of the time series prediction analysis of the production, productivity and cultivated area from the year 1961-2021 of wheat crop in India. The study can be concluded on a note that the Artificial Neural Networks proved to be a best approach to predict wheat production, productivity and cultivated area. However, the classical methods can also use for the purpose but in comparison to the ANN technique the classical methods of time series fall short in terms of error values i.e., Mean Squared Error (MSE) and Mean Absolute Error (MAE). The R value for ANN technique in all cases was greater than 0.9 with least error in order of decimals which was not found in case of the classical methods. Furthermore, we can also conclude that the Artificial Neural Network model was even effective for short time series i.e., 60 years using different activation function namely tangent, linear and sigmoid for production, productivity and cultivated area respectively which in turn reveals a robustness of the ANN model.

5. References

1. Abraham ER, Reis MDJG, Vendrametto O, Neto OCPLD, Toloi CR, Souza AED, *et al.* Time series prediction with artificial neural networks: An analysis using Brazilian soybean production. Agriculture (Basel). 2020;10(10):113-130.
2. Arifin F, Robbani H, Annisa T, Ma'Arof NNM. Variations in the number of layers and the number of neurons in artificial neural networks: Case study of pattern recognition. Journal of Physics: Conference Series. 2019 Nov;1413(1):1-6.
3. Santos BDV, Santos AMFD, Rolim GDS. Estimation and forecasting of soybean yield using artificial neural networks. Agronomy Journal. 2021;113(4):3193-3209.
4. Basir MS, Chowdhury M, Islam MN, Rabbani AEM. Artificial neural network model in predicting yield of mechanically transplanted rice from transplanting parameters in Bangladesh. Journal of Agriculture and Food Research. 2021;5:100186.
5. Dastres R, Soori M. Artificial neural network systems. International Journal of Imaging and Robotics (IJIR). 2021;21(2):13-25.
6. Demuth H, Beale M, Hagan M. Neural Network Toolbox User's Guide. Natick, MA, USA: The MathWorks, Inc.; c2017.
7. Ejaz N, Abbasi S. Wheat yield prediction using neural network and integrated SVM-NN with regression. Pakistan Journal of Engineering, Technology & Science. 2020;8(2):77-97.
8. Erenstein O, Jaleta M, Mottaleb KA, Sonder K, Donovan J, Braun HJ. Global trends in wheat production, consumption and trade. In: Braun HJ, ed. Wheat Improvement: Food Security in a Changing Climate. Cham: Springer International Publishing; c2022. p. 47-66.
9. FAO. Food and Agriculture Organization of United Nations, Crops and Livestock. Online Available: https://www.fao.org/faostat/en/#data/QCL. Accessed on 10 August, 2023.
10. FAO. Food and Agriculture Organization of United Nations, UN Report: Global hunger numbers rose to as many as 828 million in 2021. Online Available: https://www.fao.org/newsroom/detail/un-report-global-hunger-SOFI-2022-FAO/en. Accessed on 10 August, 2023.
11. GHI. Global Hunger Index Food Systems Transformation and Local Governance; c2022. Online Available: https://www.globalhungerindex.org/pdf/en/2022.pdf. Accessed on 8 August, 2023.
12. Kaur K. Artificial neural network model to forecast energy consumption in wheat production in India. Journal of Statistical Theory and Applications; c2023. p. 1-19.
13. Russell S, Norvig P. Artificial Intelligence: A Modern Approach. 3rd Ed. Bengaluru, India: Pearson Education India; c2015.
14. Silva G. Feeding the World in 2050 and Beyond-Part 1: Productivity Challenges. Michigan State University Extension-3; c2018 December. Online Available: https://www.canr.msu.edu/news/feeding-the-world-in-2050-and-beyod-part-1. Accessed on 06 August 2023.
15. United Nations, Department of Economic and Social Affairs. World Population Projected to Reach 9.8 Billion in 2050 and 11.2 Billion in 2100. Online Available: https://www.un.org/en/desa/world-population-projected-reach-98-billion-2050-and-112-billion-2100. Accessed on 10 August, 2023.
16. United Nations, Department of Economics and Social Affairs Economic Analysis. Online Available: https://www.un.org/development/desa/dpad/publication/un-desa-policy-brief-no-130-why-population-growth-matters-for-sustainable-development/. Accessed on 10 August, 2023.