International Journal of Agriculture and Food Science 2026; 8(6): 94-103


**ISSN Print:** 2664-844X
**ISSN Online:** 2664-8458
NAAS Rating (2026): 4.97
IJAFS 2026; 8(6): 94-103
www.agriculturaljournals.com
Received: 16-03-2026
Accepted: 18-04-2026

**B Venkataviswateja**
Ph.D. Scholar, Department of Statistics and Computer Applications, College of Agriculture, Acharya N.G. Ranga Agricultural University (ANGRAU), Bapatla, Andhra Pradesh, India

**Sanketh Raj H**
M.Sc. Scholar, Department of Statistics and Computer Applications, College of Agriculture, Acharya N.G. Ranga Agricultural University (ANGRAU), Bapatla, Andhra Pradesh, India

**Corresponding Author:**
**B Venkataviswateja**
Ph.D. Scholar, Department of Statistics and Computer Applications, College of Agriculture, Acharya N.G. Ranga Agricultural University (ANGRAU), Bapatla, Andhra Pradesh, India

# Modelling of Jowar production in Karnataka state - A hybrid approach

**B Venkataviswateja and Sanketh Raj H**

**DOI:** https://www.doi.org/10.33545/2664844X.2026.v8.i6b.1582

**Abstract**
The current study used both linear and nonlinear time series techniques to forecast Jowar production from 1962-1963 to 2024-2025. The Autoregressive Integrated Moving Average (ARIMA) model was first used, and the best fit was chosen using diagnostic measures. Advanced machine learning techniques like Time Delay Neural Network (TDNN), Nonlinear Support Vector Regression (NLSVR), and their hybrid combinations of ARIMA-TDNN and ARIMA-NLSVR were used to look for any nonlinear patterns in the data. The adoption of hybrid models was validated by the BDS test on ARIMA residuals, which verified the existence of nonlinearity. RMSE, MAE and MAPE were used to assess the model performance for the nonlinear and hybrid techniques. In terms of forecast accuracy, the ARIMA (3,1,1)-TDNN (4-7-1) hybrid model outperformed the other models. Forecasts for up to 2030-2031 were also generated using the model, which predicted that Karnataka will produce 648.72 thousand tons of Jowar in 2030-31.

**Keywords:** Jowar, ARIMA, TDNN, NLSVR, RMSE, MAE, MAPE, $R^2$ and BDS

## 1. Introduction
Sorghum, (Sorghum bicolor), cereal grain plant of the grass family (Poaceae) and its edible starchy seeds. The plant likely originated in Africa, where it is a major food crop and has numerous varieties, including grain Sorghums used for food and grass Sorghums grown for hay, fodder and broomcorn, used in making brooms and brushes. In India Sorghum is known as Jowar, Cholam or Jola, in West Africa as Guinea corn and in China as Kaoliang. Sorghum is especially valued in hot and arid regions for its resistance to drought and heat. Globally, Sorghum produced about 626.9 lakh tonnes during 2024-25. United States stands top with 87.30 lakh tons (14%) followed by Nigeria with 64.21 lakh tons (10%), Brazil with 61 lakh tons (10%) and India with 50 lakh tons (8%) (fas.usda.gov, 2024-25)

India ranks fourth in total Sorghum production with 50 lakh tonnes grown in an area of 39.95 lakh hectares in 2024-25, where majority of Sorghum was produced during rabi season. The kharif Sorghum with 17.37 lakh tonnes was grown predominantly in Rajasthan, Uttar Pradesh, Haryana and Madhya Pradesh, while rabi Sorghum with 31.87 lakh tonnes grown predominantly in Maharashtra, Karnataka, Tamil Nadu and Andhra Pradesh (indiastat, 2024-25).

Karnataka is the eighth largest state in India, with a total size of 192 lakh square kilometres, or 6.3 per cent of the nation's total land area. It is estimated that 83.63 lakh ha of area under total food crops will be cultivated with the production of 146.31 lakh tonnes. Jowar is mainly grown in Kalaburagi, Belagavi, Vijayapura, Bagalkot, Raichur, Ballari, Yadgir, Bidar and Gadag districts and it is grown in an area of 6.48 lakh ha, with an annual production of 8.15 lakh tonnes and productivity of 1258 kg per hectare in Karnataka state (indiastat,2024-25).

Jowar production in Karnataka has changed significantly and declining during the last couple of years, indicating shrinking cultivation area, diversion to more remunerative crops, market-driven changes and regulatory initiatives. Nonetheless, nonlinear trends impacted by exogenous shocks and environmental conditions are also evident in the year-to-year fluctuation. The necessity for precise forecasting models that go beyond conventional linear techniques is highlighted by this complexity. In order to anticipate its production, Jowar has been chosen as one of the main crops for this study, which investigates sophisticated hybrid forecasting models such as ARIMA, TDNN, NLSVR, ARIMA-TDNN, and

ARIMA-NLSVR. For those concerned in agricultural planning, policy creation, and market regulation, these models seek to offer trustworthy insights.

## 2. Materials and methodologies

### 2.1 Data Source

The study has been illustrated with the time series data. The secondary data on the Jowar production in Karnataka state for 63 years (from 1962-63 to 2024-25) was obtained from the Directorate of Economics and Statistics, Dept. of Agriculture, Cooperation and Farmers Welfare, Ministry of Agriculture, Govt. of India.

### 2.2 Statistical Methodologies

In this study, Auto Regressive Integrated Moving Average (ARIMA), Time Delay Neural Network (TDNN), Nonlinear Support Vector Regression (NLSVR) and a proposed hybrid methodology combining ARIMA-TDNN or NLSVR are employed to forecast the Production of Jowar in Karnataka state, India. The analysis is carried out using R software. To examine the stationarity of the time series data, the Augmented Dickey-Fuller (ADF) test is performed. The Brock-Dechert-Scheinkman (BDS) test is used to detect nonlinear dependencies in the residuals, ensuring an appropriate model selection process. Additionally, the Box Pierce/Ljung-Box test is applied to check for autocorrelation in the residuals, validating the adequacy of the fitted models. The forecasting performance of the models is assessed using Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error which provide measures of predictive accuracy and reliability.

### 2.2.1 Box-Jenkins ARIMA model

Before ARIMA, the concept of validation of data and Stationery was discussed briefly.

**Validation of data**

For model development and evaluation, the dataset comprising 63 annual observations was divided into training and testing sets. Out of these, 57 data points (1962-63 to 2018-19) were utilized for training the models, while the remaining 6 data points (2019-20 to 2024-25) were reserved as a testing set to assess the predictive accuracy and generalization performance of the models. This split ensures that the models are trained on a sufficiently large sample while also being validated on unseen data to avoid overfitting and to obtain realistic forecasting performance.

**Stationary**

A stochastic process is said to be stationary if its mean and variance are constant over time and the value of the covariance between the two time periods depends only on the disturbance or lag between the two time periods and not on the actual time at which the covariance is computed.
It can be identified through Autocorrelation Function (ACF) of actual data, if the ACF does not die out rapidly, it indicates that the data is non-stationary. Under this situation, the auto correlation corresponding to most of lags are statistically significant.
For reducing the data to stationary, selected data has to transform by taking first order differences (d=1). If the autocorrelation functions of first differenced data indicate a rapid decrease, then it can be concluded that the transformed data is stationary. If not, again the data has to be transformed by taking second order differences (d=2). Continuing in a similar way as that of d=1, thus the order of differencing i.e.,'d' can be determined.

**Test of Stationary:** The order of differencing can be assessed by performing unit root test. The most widely used test for detecting the unit root (non-stationary) of time series is Augmented Dickey Fuller (ADF) test. In conducting Dickey Fuller (DF) test, it has assumed that the error term is uncorrelated. But in most of cases correlated, hence Dickey-Fuller (1979) has developed another test, known popularly known as the ADF test. This test is conducted by augmenting the regression $(\Delta Y_t = \beta_1 + \delta Y_{t-1} + e_t)$, by adding the lagged values of dependent variable$(\Delta Y_t)$. The ADF test consist of estimating following regression,

$$\Delta Y_t = \beta_1 + \delta Y_{t-1} + \sum_{i=1}^{m} \alpha_i \, \Delta Y_{t-i} + e_t \ldots \qquad (1)$$

**Where;**

- m is number of lagged difference terms required so that the error term
- $e_t$ is serially independent.

The test can be carried out by performing a t (= τ) -statistic of $'\delta'$ with dickey fuller table values at 5% and 10% LOS. If absolute t (= τ) statistic of $'\delta'$ is greater than its table value, it indicates that the series is Stationary.

**ARIMA (Auto Regressive Integrated Moving Average)**

It's denoted by ARIMA (p,d,q) which is a combination of Auto Regressive (AR) and Moving Average (MA) with an order of integration or differencing (d), where p and q are the order of autocorrelation and moving average respectively (Box and Jenkins, 2015) [2].
The Auto-regressive model of order p denoted by AR (p) is as follows:

$$Y_t = c + \varnothing_1 Y_{t-1} + \varnothing_2 Y_{t-2} + \ldots + \varnothing_p Y_{t-p} + e_t \ldots \qquad (2)$$

The Moving Average (MA) model of order q or MA (q) can be written as:

$$Y_t = c - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \ldots \theta_q e_{t-q} + e_t \ldots \qquad (3)$$

**Where;**

- c is constant term.
- $\varnothing_p$ is the $p^{th}$ autoregressive parameter.
- $\theta_q$ is the $q^{th}$ moving average parameter and
- $e_t$ is the error term at time 't'.

ARIMA in general form is as follows:

$$\Delta^d Y_t = c + (\varnothing_1 \Delta^d Y_{t-1} + \ldots + \varnothing_p \Delta^d Y_{t-p}) - (\theta_1 e_{t-1} + \ldots + \theta_q e_{t-q}) + e_t \ldots \qquad (4)$$

Where Δ denotes difference operator like

- $\Delta Y_t = Y_t - Y_{t-1}$(data form of first order differentiation)... (5)
- $\Delta^2 Y_{t-1} = \Delta Y_t - \Delta Y_{t-1}$ (data form of second order differentiation)... (6)

Here, $Y_{t-1}, \ldots, Y_{t-p}$ are values of past series with lag 1, …, p, respectively.

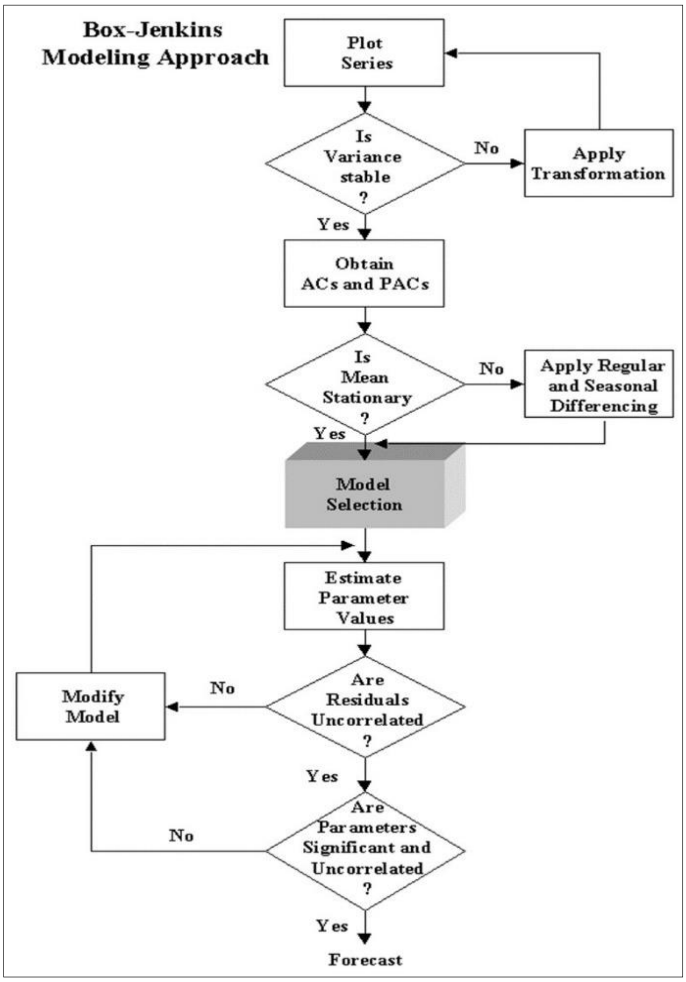


**Fig 2.1:** Flow chart of Box-Jenkins Methodology

**2.2.2 Time Delay Neural Network (TDNN)**

The ANN for time series analysis is termed as Time Delay Neural Network (TDNN). The time series phenomenon can be mathematically modelled using neural network with implicit functional representation of time, whereas static neural network like multilayer perceptron is presented with dynamic properties (Haykin, 1999). One simple way of building artificial neural network for time series is the use of time delay also called as time lags. These time lags can be considered in the input layer of the ANN. The TDNN is the class of such architecture. In this study, the R package "nnfor" was customized and implemented for developing Time Delay Neural Network (TDNN) models.

This package provides a flexible and powerful framework for fitting artificial neural networks specifically designed for time series forecasting problems. In view of the residual nonlinearity detected through the BDS test, Time Delay Neural Network (TDNN) models were employed to model the residual series obtained from the ARIMA models for Jowar production in Karnataka. The TDNN models are well suited for capturing complex, nonlinear and dynamic dependencies in the time series data through their feedback and delay structure.

For each data series, multiple TDNN networks were trained by varying the number of neurons in the hidden layer from 1 to 10, with the sigmoid function employed as the activation function in the hidden layer, while keeping the output layers fixed and input layer varying from 2 to 5. To identify the optimal network configuration based on performance metrics such as RMSE, MAE, MAPE and model adequacy through the Ljung Box test p values. The structure of each network is denoted in the format Input-Hidden-Output. Following is the general expression for the final output $Y_t$ of a multi- layer feed forward time delay neural network.

The TDNN can be mathematically expressed as:

$$\mathrm{Y}_t = f(Y_{t-1}, \mathrm{Y}_{t-2}, \ldots\ldots, Y_{t-p}; W) \ldots \qquad (12)$$

**Where;**

- $Y_t$ is the predicted value at time t
- $f\,(\cdot)$ represents the non-linear activation function
- $Y_{t-1}, Y_{t-2}, \ldots, Y_{t-p}$ are the lagged inputs
- $W$ represents the weights of the network

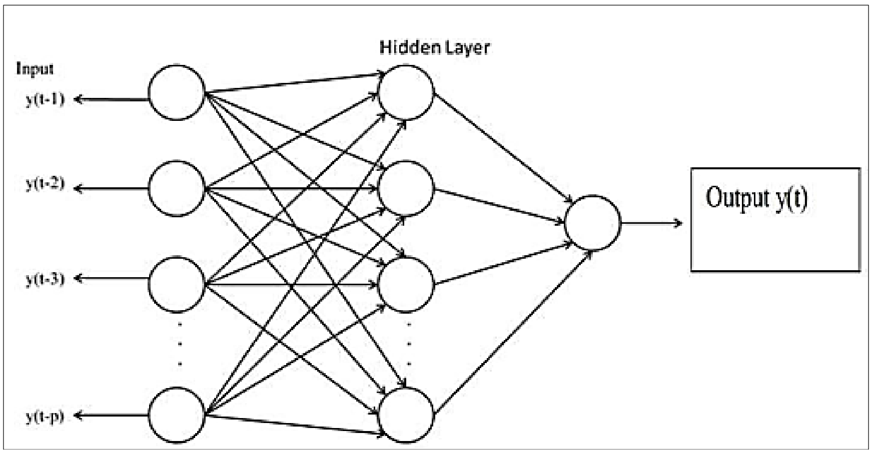


**Fig 2.2:** Architecture of Neural Network

**2.2.3 Non-Linear Support Vector Regression (NLSVR) model:** Support Vector Machine (SVM) is a supervised machine learning technique which was originally developed for linear classification problems. Later in the year 1997, the support vector machine for regression problems were developed by Vapnik by introducing $\varepsilon$-insensitive loss function (Vapnik *et al.*, 1997) [11] and it has been extended to the nonlinear regression estimation problems. Modeling of such problems is called as Nonlinear Support Vector Regression (NLSVR) model. The basic principle involved in Non-Linear Support Vector Regression (NLSVR) is to transform the original input time series into a higher-dimensional feature space using a nonlinear kernel function, where a linear regression function is then fitted, enabling the model to capture complex, nonlinear relationships present in the data. The general form of NLSVR is:

$$f(x) = \sum_{i=1}^{N} \alpha_i K(x_i, x) + b \; .... \qquad (13)$$

**Where;**

- $\alpha_i$ are the coefficients,
- $K(x_i, x)$ is the kernel function and
- $b$ is the bias term.

In this study R package, "e1071" was customized and implemented. This package provides several options for SVR model fitting using different types of kernels like linear, radial basis, sigmoid and polynomial. Although the most often used and recommended function is radial basis.

After detecting potential nonlinear patterns in the ARIMA residuals using the BDS test, a nonlinear model such as Nonlinear Support Vector Regression was applied to capture the nonlinear structures not modelled by ARIMA. The NLSVR models were fitted to the residuals of the optimal ARIMA models for Jowar production dataset. The radial basis function (RBF) kernel was chosen due to its flexibility and ability to model complex nonlinear relationships. The optimal hyperparameters for the NLSVR models, include Cost, Gamma, Epsilon were determined through iterative tuning to minimize the error while ensuring good generalization ability. A 10-fold cross validation procedure was adopted to ensure the robustness and generalization capability of the selected hyperparameters. This method involves dividing the datasets into ten subsets, iteratively training the model on nine subsets and validating it on the remaining one thereby minimizing overfitting and improving model reliability. The number of support vectors selected in each model reflects the complexity of the fitted regression function. The best performance value represents the minimal error achieved during the cross-validation process for each model.

1. a high dimensional feature space and then build the regression model in a new feature space.
2. Let us consider a vector of data set $Z = \{x_i\, y_i\}_{i=1}^{N}$ where $x_i \in R^n$ is the input vector, $y_i$ is the scalar output and $N$ is the size of data set. The general equation of Nonlinear Support Vector Regression estimation function is given as follows:

$$f(x) = W^T \phi(x) + b... \qquad (14)$$

where, $\phi(.)$: $R^n \rightarrow R^{nh}$ is a nonlinear mapping function which maps the nonlinear mapping function which maps the original input space into a higher dimensional feature space vector. $W \in R^{nh}$ is weight vector, b is bias term and superscript T denotes the transpose.

**Hyperparameter Tuning**

The performance of NLSVR depends significantly on the choice of three key hyperparameters:

- **Cost (C)**: Controls the trade-off between model complexity and the degree to which deviations larger than ε are tolerated.
- **Gamma (γ)**: Defines the influence of each data point on the regression function.
- **Epsilon (ε)**: Specifies the margin of tolerance within which no penalty is given for prediction errors.

**2.2.4 The Proposed Hybrid Methodology**

The hybrid method considers the time series $y_t$ as a combination of both linear and non-linear components. This approach follows the Zhang's (2003) [12] hybrid approach. Accordingly, the relationship between linear and nonlinear components can be expressed as follows

$$y_t = \hat{L}_t + \hat{N}_t \; ... \qquad (15)$$

Where, $\hat{L}_t$ and $\hat{N}_t$ represent the linear and nonlinear component, respectively. In this work, the linear part is modeled using ARIMA model and non-liner part by TDNN or NLSVR. The methodology consists of three steps. Firstly, an ARIMA model is employed to fit the linear component.

Let the prediction series provided by ARIMA model be denoted as $\widehat{L}_t$. In the second step, the residuals ($e_t = y_t - \widehat{L}_t$) obtained from ARIMA model are tested for non-linearity by using BDS test (Brock *et al.*, 1996) Once the residuals confirm the non-linearity, then, they are modelled and predicted using TDNN and NLSVR. Finally, the forecasted linear and nonlinear components are combined to generate aggregate forecast.

$$\hat{y}_t = \hat{L}_t + \widehat{N}_t \ldots \quad (16)$$

Where $\hat{L}_t$ and $\widehat{N}_t$ represent the predicted linear and nonlinear component, respectively.

**2.2.4(a) Hybrid ARIMA-TDNN Model:** The hybrid ARIMA-TDNN model is constructed by first fitting the ARIMA model to capture the linear patterns in the data and then applying the TDNN model to the residuals from the ARIMA model to account for non-linear patterns.
The hybrid model can be formulated as:

$$\hat{Y}_t = \hat{Y}_{ARIMA_t} + \hat{Y}_{TDNN_t} \ldots \quad (17)$$

**Where;**
- $\hat{Y}_t$ is the final forecasted value
- $\hat{Y}_{ARIMA_t}$ is the ARIMA forecasted value at time $t$
- $\hat{Y}_{TDNN_t}$ is the TDNN forecasted residual value at time $t$

**2.2.4(b) Hybrid ARIMA + NLSVR Model**
A hybrid ARIMA + NLSVR (Non-Linear Support Vector Regression) model combines the strengths of both ARIMA and NLSVR to forecast time series data.
The hybrid model can be formulated as:

$$\hat{Y}_t = \hat{Y}_{ARIMA}(t) + \hat{R}\,(t)\ldots. \quad (18)$$

**Where**
- $\hat{Y}_t$ is the final forecasted value
- $\hat{Y}_{ARIMA_t}$ is the ARIMA forecasted value at time $t$
- $\hat{R}$ (t) is the NLSVR forecast applied to residuals R(t)

The graphical representation of hybrid methodology has shown below.

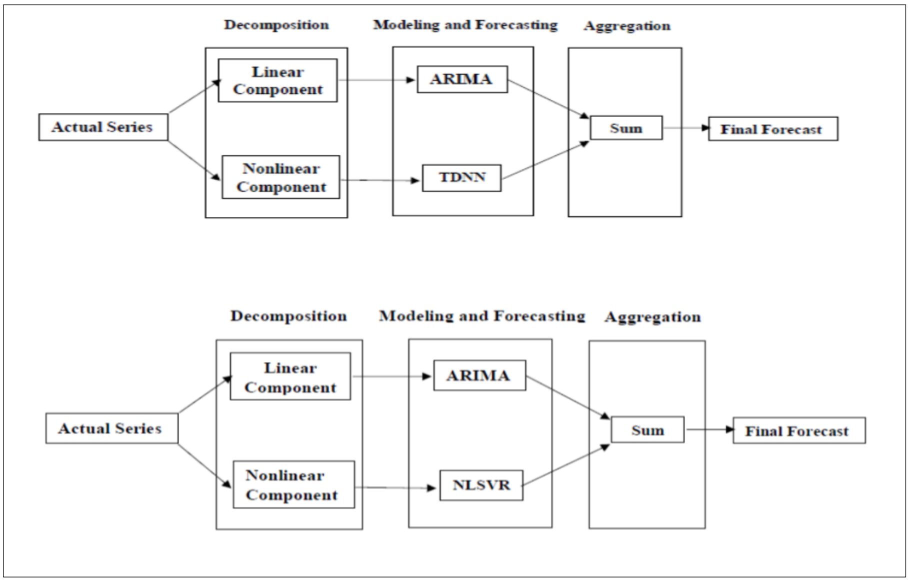


**Fig 2.3:** Graphical representation of Hybrid Methodology

**2.2.5 Performance Criteria**
The model's performance was evaluated using four key metrics: Coefficient of Determination $R^2$, which measures goodness-of-fit; Root Mean Square Error (RMSE), Mean Absolute Error (MAE), which quantify prediction errors in absolute terms and Mean Absolute Percentage Error (MAPE), which expresses errors as a percentage. Furthermore, the BDS test was used to check for non-linear structures in the data, while the Augmented Dickey Fuller (ADF) test confirmed if the time series data was statistically stationary or not.

## 3. Results and discussions
### 3.1 Descriptive Statistics of Jowar Production of Karnataka State
A detailed descriptive statistical analysis was carried out to examine the characteristics and distribution of the production (in thousand tonnes) of Jowar in Karnataka for the period from 1962-23 to 2024-25. These measures provide information on central tendency, dispersion, distribution pattern and behaviour of the data. The summary of descriptive statistics is presented in Table 3.1. The production data set been plotted vs time as seen in Figure 3.1

**Table 3.1:** Descriptive statistics of Jowar production

| Statistic | Value | Statistic | Value (in thousand tonnes) |
|---|---|---|---|
| Mean (in thousand tonnes) | 1454.95 | Range | 1426.4 |
| SD (in thousand tonnes) | 355.15 | Minimum | 682 |
| CV% | 24.41 | Maximum | 2108.40 |
| Kurtosis | -0.67 | Skewness | -0.42 |

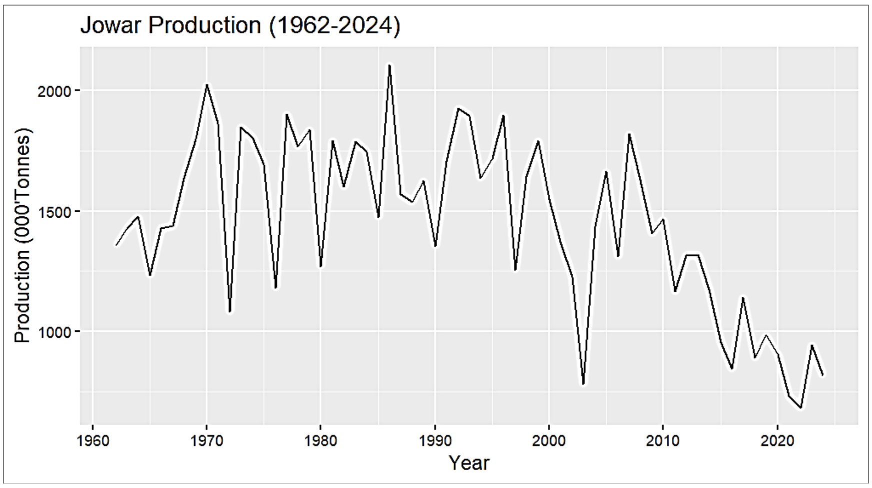


**Fig 3.1:** Time series plot of Jowar production

### 3.2 Fitting of Box Jenkins (ARIMA) model for Jowar Production in Karnataka State

Significant autocorrelations are present at many lags in the ACF plots of the trained data series of Jowar production shown in Figure 3.2, which have a slow and progressive decline trend. This pattern indicates that the data series needed to be differenced in order to become stationary, which is typical of a non-stationary time series. The production of Jowar in Karnataka State was subjected to the Augmented Dickey-Fuller test in order to statistically validate this. The data series was non-stationary and became stationary at first difference, according to Table 3.2, since the null hypothesis was rejected at the five percent level of significance because the p value was less than 0.05.

**Table 3.2:** Augmented Dickey-Fuller Stationarity Test for Jowar production

| Jowar crop | Data type | ADF statistic | critical value | Decision |
|---|---|---|---|---|
| Production | ADF at zero difference level | -2.58 | 0.34 | Non-Stationary |
| | ADF at 1st difference | -5.55 | 0.01 | Stationary |

Consequently, the ACF plots of the differenced series for production was examined in Figure 3.2. The differenced series exhibited a rapid decay of autocorrelations with most of the autocorrelation coefficients falling within the 95 per cent confidence limits, indicating that the data had attained stationarity.

After setting d=1 to achieve stationarity, several ARIMA models were developed. The models were compared using $R^2$, RMSE, MAE and MAPE criteria which was shown in Table 3.3. Among them, ARIMA (3,1,1) was identified as best model with high $R^2$, low RMSE, MAE and MAPE values.

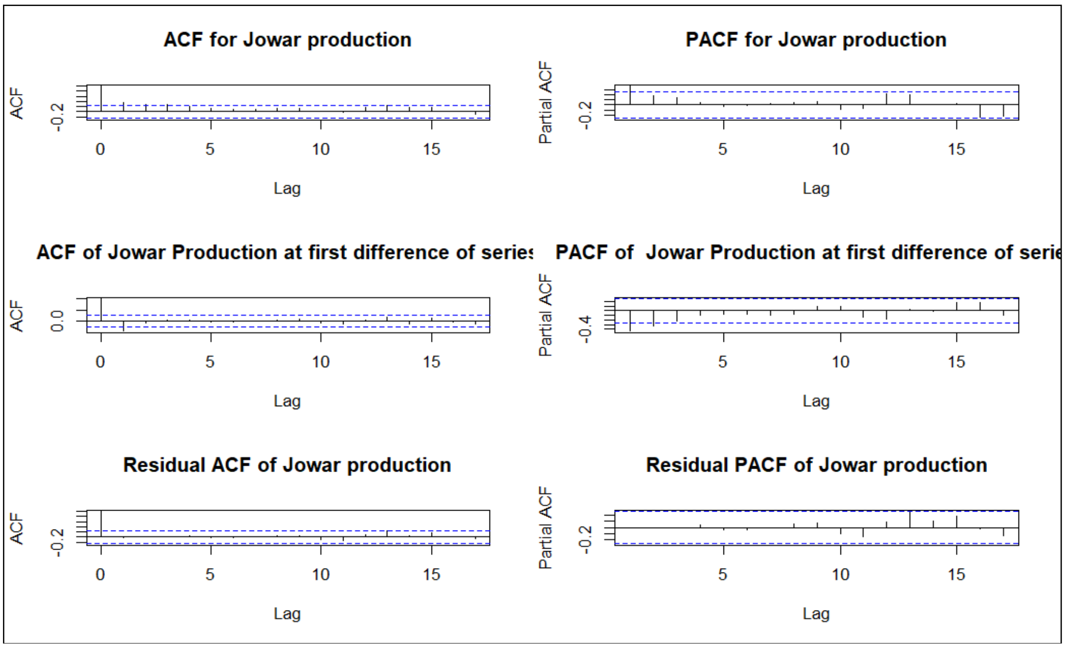


**Fig 3.2:** ACF and PACF plots for Jowar production

**Table 3.3:** Model fit statistics for Jowar Production

| ARIMA model | c | Parameters Estimate | | | | | Goodness of Fit | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Autoregressive Coefficient | | | Moving Average Coefficient | | $R^2$ | RMSE | MAE | MAPE |
| | | AR1 | AR2 | AR3 | MA1 | MA2 | | | | |
| ARIMA (0,1,2) | -7.48 | | | | -0.71 | -0.05 | 0.224 | 267.70 | 216.39 | 15.61 |
| ARIMA (1,1,0) | -7.41 | -0.43 | | | | | 0.046 | 296.94 | 240.29 | 17.05 |
| ARIMA (1,1,1) | -7.48 | 0.06 | | | -0.78 | | 0.224 | 267.72 | 216.44 | 15.62 |
| ARIMA (1,1,2) | -7.47 | 0.38 | | | -1.11 | -0.02 | 0.224 | 267.75 | 216.63 | 15.63 |
| ARIMA (2,1,1) | -7.5 | 0.05 | -0.02 | | -0.77 | | 0.224 | 267.69 | 215.93 | 15.59 |
| ARIMA (2,1,2) | -7.51 | -0.15 | -0.06 | | -0.54 | -0.54 | 0.225 | 267.63 | 215.71 | 15.57 |
| ARIMA (3,1,1) | -7.52 | 0.09** | -0.01 | 0.05* | -0.81 | | 0.226 | 267.37 | 215.72 | 15.56 |
| ARIMA (3,1,2) | -7.50 | -0.58 | 0.016 | -0.001 | -0.13 | -0.5 | 0.224 | 267.65 | 215.87 | 15.58 |

**Significant at 1%, *Significant at 5%

The ARIMA (3,1,1) fitted equation:

Jowar production in Karnataka ($\Delta$ $Y_t$) = $-7.52+0.09Y_{t-1}-0.01Y_{t-2}+0.05Y_{t-3}-0.81\varepsilon_{t-1}+e_t$

Further, a residual analysis was conducted to assess the adequacy of the selected model. The analysis revealed that none of the lag values in the residual ACF and PACF plots were significant, as shown in the Figure 3.2. Additionally, Ljung-Box test and Kolmogorov-Smirnov test were conducted to check whether the residuals were independently and normally distributed or not. The Ljung-Box test statistic was Q = 0.008 (p = 0.92) K-S test statistic was D = 0.06 (p value = 0.96). The p values were greater than 0.05 indicating that the residuals were independently and normally distributed. This confirms the validity and goodness of fit of the selected ARIMA models for forecasting Jowar production in the Karnataka State.

### 3.3 Fitting of Time Delay Neural Network (TDNN) for Jowar Production in Karnataka State

From Table 3.4, it is shown that the best performing network was 4-4-1, which achieved low RMSE, MAE and MAPE values. Further, residual analysis was also carried out to check the adequacy of the selected neural network model and it was discovered that none of the lags of residual ACF chart were found to be significant as per Figure 3.3. Furthermore, the p-value (0.92) for the Ljung-Box test was greater than 0.05, indicating that the residuals were independently distributed.

**Table 3.4:** Performance of TDNN Models with Varying Network Structures on Jowar production

| Particulars | Network structure | RMSE | MAE | MAPE |
|---|---|---|---|---|
| Production | 4-1-1 | 243.19 | 194.25 | 14.10 |
| | 4-2-1 | 201.91 | 161.83 | 11.63 |
| | 4-3-1 | 151.07 | 114.69 | 8.11 |
| | 4-4-1 | 130.82 | 94.15 | 6.58 |
| | 4-5-1 | 141.51 | 104.25 | 7.16 |

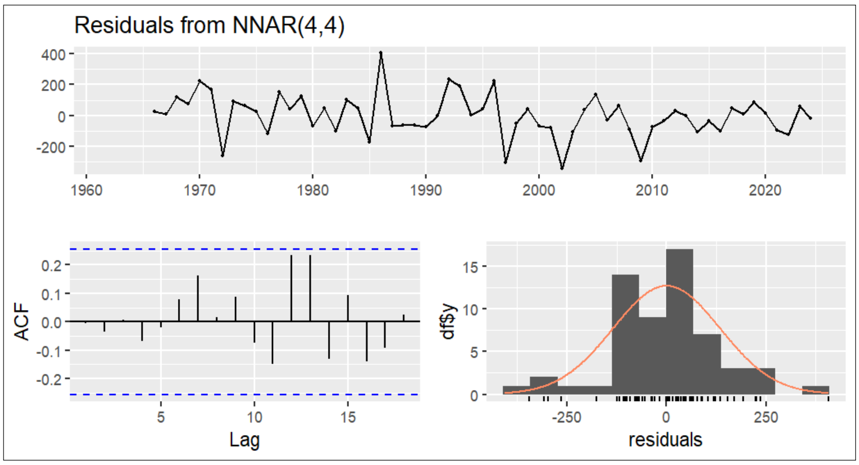


**Fig 3.3:** Residual plot of NNAR (4, 4)

**3.4 Fitting of Nonlinear Support Vector Regression (NLSVR) Model on Jowar production:** Table 3.5 displays the findings of the Non-Linear Support Vector Regression (NLSVR) model for Karnataka Jowar production. With an Epsilon ($\varepsilon$) value of 0.01, a Cost (C) value of 4, and a Gamma ($\gamma$) value of 0.25, the model was constructed using a Radial Basis Function (RBF) kernel under the EPS-Regression type. The number of data points that helped define the regression function was indicated by the final model's use of 43 support vectors. The goal of this setup was to avoid overfitting while improving forecast accuracy. The model's optimal performance value, 91228.49, was the lowest value of the objective function, which typically indicates the error measure, through tuning. The non-linear trends in the Jowar production data must be sufficiently captured. These parameter choices aimed to achieve a balance between forecast accuracy and model complexity.

**Table 3.5:** Parameters Specification of NLSVR on Jowar production

| Particulars | SVM Type | SVM Kernel | Cost | Gamma | Epsilon | No. Of Support Vectors | Best Performance |
|---|---|---|---|---|---|---|---|
| Production | EPS-Regression | Radial | 4 | 0.25 | 0.01 | 55 | 91228.49 |

### 3.5 BDS Test for Nonlinearity in ARIMA Residuals of ARIMA (3,1,1) model for Jowar Production

| BDS Test | | | | | | | | | Decision |
|---|---|---|---|---|---|---|---|---|---|
| Embedding Dimension | 2 | | | | 3 | | | | Presence of nonlinearity |
| Epsilon | 134.82 | 269.64 | 404.46 | 539.28 | 134.82 | 269.64 | 404.46 | 539.28 | |
| p-value | 0.76 | 0.98 | 0.03* | 0.85 | 0.001** | 0.52 | 0.93 | 0.97 | |
| **Significant at 1%, *Significant at 5% | | | | | | | | | |

The residuals of the fitted ARIMA model for Jowar production were tested for nonlinearity using the Brock-Dechert-Scheinkman (BDS) test. As seen in Table 4.6, a significant p-value was found at Embedding Dimensions 2 and 3, suggesting the existence of a nonlinear structure not captured by the fitted ARIMA model. These findings showed that while certain nonlinear patterns persisted, ARIMA models successfully captured the linear structure in the data series. Therefore, using a nonlinear model to represent the nonlinear structures was justified.

### Table 3.6. BDS Test on ARIMA (3,1,1) Residuals

In the present study, to enhance the forecasting accuracy nonlinear models are integrated with the ARIMA structure. Further the ARIMA residuals are fitted by nonlinear models such as Time Delay Neural Network (TDNN) and Nonlinear Support Vector Regression (NLSVR). This hybrid ARIMA-TDNN and ARIMA-NLSVR approach combines the strengths of ARIMA for capturing linear trends and nonlinear models for addressing remaining nonlinear structures in the data.

### 3.6 Fitting of Time Delay Neural Network (TDNN) Models on ARIMA Residuals

Table 3.7 demonstrates that 4-7-1 was the best-performing network, achieving low RMSE, MAE, and MAPE values. In order to assess the suitability of the chosen neural network model, residual analysis was also performed. According to Figure 3.4, none of the delays in the residual ACF chart were found to be significant. Additionally, the Ljung-Box test's p-value (0.93) was higher than 0.05, suggesting that the residuals had an independent distribution. These results imply that the chosen model provides a good match to the ARIMA residuals.

**Table 3.7:** Performance of TDNN Models with Varying Network Structures on ARIMA Residuals

| Particulars | Network structure | RMSE | MAE | MAPE |
|---|---|---|---|---|
| Production | 3-6-1 | 98.42 | 74.91 | 73.77 |
| | 3-7-1 | 80.06 | 59.98 | 44.89 |
| | 4-6-1 | 63.63 | 50.82 | 53.10 |
| | 4-7-1 | 47.76 | 39.11 | 39.68 |
| | 4-8-1 | 49.88 | 44.19 | 40.90 |

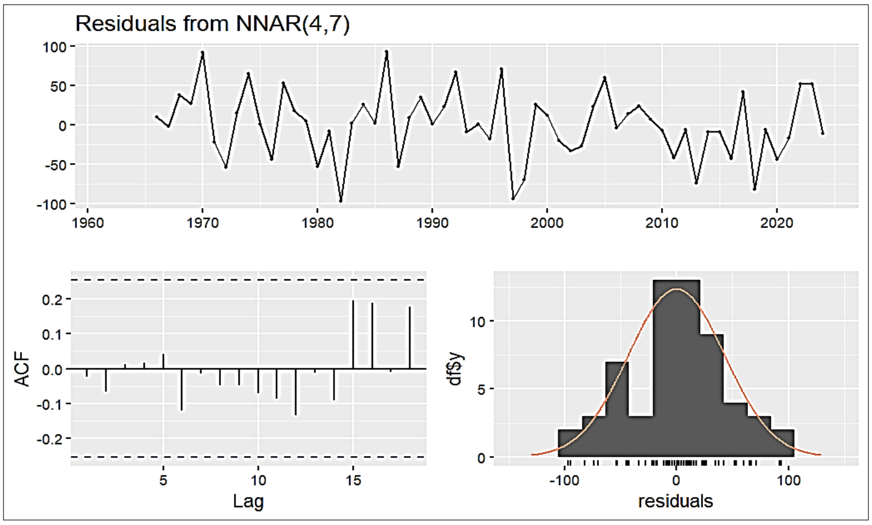


**Fig 3.4:** Residual plot of NNAR (4,7)

### 3.7 Nonlinear Support Vector Regression (NLSVR) Model Fitting on ARIMA Residuals

Table 3.8 displays the outcomes of the Non-Linear Support Vector Regression (NLSVR) model for Karnataka Jowar production. With an Epsilon (ε) value of 0.001, a Cost (C) value of 8, and a Gamma (γ) value of 0.03125, the model was constructed using a Radial Basis Function (RBF) kernel under the EPS-Regression type. The number of data points that helped define the regression function was indicated by the final model's use of 45 support vectors. The goal of this setup was to avoid overfitting while improving forecast accuracy. The model's optimal performance value, 66122.37, was the lowest value of the objective function, which typically indicates the error measure, throughout tuning. These parameter selections aimed to achieve a balance between forecast accuracy and model complexity in order to effectively represent the non-linear patterns present in the Jowar production data.

**Table 3.8:** Parameters Specification of NLSVR on ARIMA (3,1,1) Residuals

| Particulars | SVM Type | SVM Kernel | Cost | Gamma | Epsilon | No. Of Support Vectors | Best Performance |
|---|---|---|---|---|---|---|---|
| Production | EPS-Regression | Radial | 8 | 0.03125 | 0.001 | 45 | 66122.37 |

### 3.8 Performance Comparison of Fitted Models

Both training datasets (1962-63 to 2018-19) and testing datasets (2019-20 to 2024-25) were used to assess the fitted models using RMSE, MAE and MAPE. The accuracy of the hybrid techniques was higher than that of the individual models (ARIMA, TDNN, and NLSVR), which performed moderately well which is shown in the Table 3.9. On both datasets, ARIMA-TDNN consistently performed better than the others, demonstrating the benefit of combining linear and nonlinear approaches.

**Table 3.9:** Performance Metrics Comparison of the Fitted Models of Jowar Production in Karnataka State

| Model/Criteria | Training Dataset | | | Testing Dataset | | |
|---|---|---|---|---|---|---|
| | RMSE | MAE | MAPE | RMSE | MAE | MAPE |
| ARIMA | 267.37 | 215.72 | 0.15 | 233.64 | 205.62 | 0.29 |
| TDNN | 130.82 | 94.15 | 0.65 | 220.96 | 198.78 | 0.37 |
| NLSVR | 281.55 | 226.52 | 0.16 | 599.44 | 564.96 | 0.39 |
| ARIMA-TDNN | 45.69 | 34.00 | 0.023 | 214.46 | 179.51 | 0.28 |
| ARIMA-NLSVR | 252.05 | 194.38 | 0.13 | 292.45 | 201.51 | 0.47 |

**Table 3.10:** Forecasted Jowar production in Karnataka State from 2025-26 to 2030-31 using the ARIMA-TDNN Hybrid Model

| Year | 2025-26 | 2026-27 | 2027-28 | 2028-29 | 2029-30 | 2030-31 |
|---|---|---|---|---|---|---|
| Forecasted values ('000 tonnes) | 856.56 | 770.77 | 594.47 | 545.95 | 594.56 | 684.72 |

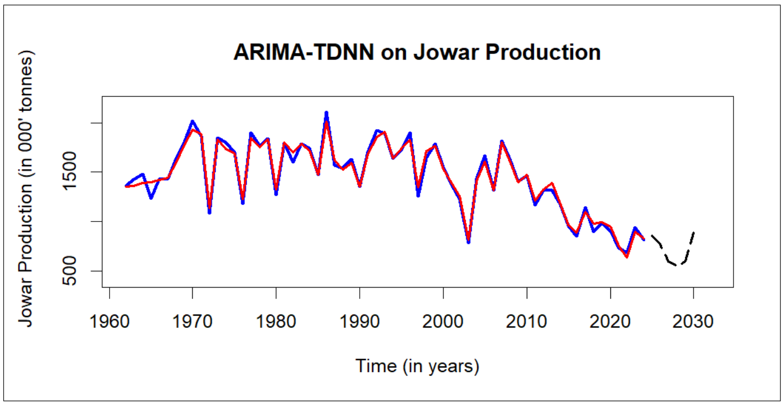


**Fig 3.5:** Actual vs ARIMA-TDNN fitted for Jowar production in Karnataka State

Finally, the actual and fitted values of Jowar production in Karnataka State were generated using the ARIMA (3,1,1) - TDNN (4-7-1) hybrid model as shown in the Figure 3.5. The forecasted values of Jowar production for the year 2030-31 were determined to be 684.72 thousand tonnes. The results obtained from this hybrid model, between 2025 and 2029, jowar output is expected to experience a period of severe shrinkage, as a result of severe market pressure and climate volatility affecting rain-fed regions. However, a strong recovery quickly follows this steep decrease, indicating a fundamental structural shift fuelled by increased market demand for this climate-resilient grain by 2030-2031 and renewed official support for millets.

## 4. Conclusion

A comparison of other forecasting models showed that the hybrid ARIMA-TDNN model generated the best accuracy of all fitted models, as evidenced by the lowest error metrics on both training and testing datasets. The enhanced performance of this hybrid model demonstrates how well linear and nonlinear methods used to generate precise forecasts. Decades of policy that favoured irrigated crops like wheat and rice drove Jowar to marginal, rain-fed fields, reducing cultivation area and causing yield instability. This led to the historical drop in Jowar production. However, estimates indicate a sharp rise by 2030 due to Jowar's low water requirements and natural climatic resilience, making it an essential option in the face of increasing water constraint. The government's reinvigorated marketing of millets as "Nutri Cereals," along with rising consumer demand for their industrial and nutritional applications, are accelerating this comeback and driving focused research to reduce the yield gap.

**References**


1. Ali AM, Mushi J. Investigation of a suitable hybrid time series model for predicting clove rice. J ICT Syst. 2023;1(2):1-16.
2. Box GE, Jenkins GM, Reinsel GC, Ljung GM. Time series analysis: Forecasting and control. 5th ed. Hoboken (NJ): John Wiley & Sons; 2015.
3. Devra SJ, Patel DV, Shitap MS, Raj SR. Time series forecasting of price for oilseed crops by combining ARIMA and ANN. Int J Stat Appl Math. 2023;8(4):40-54.
4. Kumar S, Arora K, Singh P, Gupta AK, Sharma I, Vatta K. Performance comparison of ARIMA and time delay neural network for forecasting of potato price in India. Agric Econ Res Rev. 2022;35(2):1-12.
5. Kumar TM, Prajneshu. Development of hybrid models for forecasting time-series data using nonlinear SVR enhanced by PSO. J Stat Theory Pract. 2015;9(4):699-711.
6. Jha GK, Sinha K. Time-delay neural networks for time series prediction: An application to the monthly wholesale price of oilseeds in India. Neural Comput Appl. 2014;24(3):563-571.
7. Rathod S, Mishra GC. Statistical models for forecasting mango and banana yield of Karnataka, India. J Agric Sci Technol. 2018;20(4):803-816.
8. Rathod S, Singh KN, Patil SG, Naik RH, Ray M, Meena VS. Modelling and forecasting of oilseed production of India through artificial intelligence techniques. Indian J Agric Sci. 2018;88(1):22-27.
9. Ray M, Rai A, Ramasubramanian V, Singh KN. ARIMA-WNN hybrid model for forecasting wheat yield time-series data. J Indian Soc Agric Stat. 2016;70(1):63-70.
10. Raghu KN. Statistical analysis of jowar production in Karnataka [doctoral dissertation]. Jabalpur: Jawaharlal Nehru Krishi Vishwa Vidyalaya; 2020.
11. Vapnik V, Golowich S, Smola A. Support vector method for function approximation, regression estimation and signal processing. In: Mozer M, Jordan M, Petsche T, editors. Advances in Neural Information Processing Systems. Cambridge (MA): MIT Press; 1997. p. 281-287.
12. Zhang GP. Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing. 2003;50:159-175.