

## FORECASTING OF GINGER AREA AND PRODUCTION OF INDIA USING ARIMA AND ARTIFICIAL NEURAL NETWORK APPROACH

Divyashree<sup>1</sup>, Pavan Kumar S. T<sup>2</sup>, Ullas, M. Y<sup>4</sup>

<sup>1</sup>Master Scholar, Dept. of Mathematics, Davanagere University, Karnataka

<sup>2</sup>Assistant Professor, College of Community Science, Central Agricultural University,  
Tura, Meghalaya-794005

<sup>3</sup>SMS (Agronomy), U.A.H.S, Shivamogga, Karnataka, India.

DOI: <https://www.doi.org/10.58257/IJPREMS31675>

### ABSTRACT

In the present study, an attempt was made to forecast the Ginger Area and Production using Artificial Neural Networks and Box-Jenkins ARIMA models, and compared both the models on the basis of accuracy criteria's. The ANN models utilize back propagation and conjugate gradient methods to train the data in the hidden layers, and sigmoid function as a transformation function. The production of ginger was predicted closely by Neural Network 1:2-20:1 with less training (0.168) and testing error (0.051) with R-Square value 0.82 and ARIMA (2, 0, 3) forecasted the production with R-Square value 0.77. For predicting the area of ginger, ANN 1:2-7:1 outperformed the ARIMA (3, 0, 4) models with low RMSE value, lesser training (0.027) and testing errors (0.452). The ANN models were better in learning the complexity of the data series and to predict the out sample forecasts. Therefore, ANN model could be used as forecasting technique to get the time series projection of production and area of ginger for India.

**KEYWORDS:** Ginger, Area, Production, ARIMA, ANN and Forecasting, BP,CG, Statistica

### 1. INTRODUCTION

Ginger (*Zingiber officinale* Rose.) is an important commercial spice crop grown for its aromatic rhizomes, which are used both as a spice and a medicine. At present, India is the largest producer of ginger (1.12 m t) in the world accounting for about one-third of the total world output followed by China and Nigeria. The leading Ginger growing Indian states are Assam, Maharashtra, west Bengal, Gujarat, Kerala, Meghalaya, Mizoram, Karnataka and Arunachal Pradesh. (Annon.2020). Ginger is a tropical species native to Southeast Asia, belonging to the family Zingiberaceae. The English term 'ginger' originated from the Sanskrit word Sringavera. Botanically known as *Zingiber Officinale*, it is the most popular hot spice in the world. It is marketed in different forms such as raw ginger, dry ginger, bleached dry ginger, ginger powder, ginger oil, ginger oleoresin, gingerale, ginger candy, ginger beer, brined ginger, ginger wine, ginger squash, ginger flakes etc. Ginger has a long and well documented history of both culinary and medicinal use throughout world history, especially in Chinese, Indian and Japanese medicinal care. In Indian Ayurvedic medicine, ginger is used as an anti inflammatory (Bode and Dong, 2011). The Hindu epic Mahabharata describes a meal where meat stewed with ginger and other spices was served. In the Manasollasa literature written in the 11th century, ginger was mentioned as a flavouring agent for butter milk. ANN were first developed in the 1940s (Mc Culloch and Pitts, 1943), and the development has experienced a revival with Hopfield's effort (Hopfield, 1982) in iterative auto-associable neural networks. ANN has had wide application in many spheres of life. According to Maier and Dandy, in recent years, Artificial Neural Networks (ANNs) have become extremely popular for prediction and forecasting technique in a numeral of areas, including water resources, environmental science, power generation and as well as medicine and finance. The present study was taken to forecast the area and production of ginger well in advance using ARIMA and ANN approaches. The prediction of area and production of ginger on yearly time scales is scientifically challenging. Keeping this view in mind an attempt was made to exploit this approach as major tool important in forecasting for planning and devising agricultural strategies. Climate also play an important role in the production of ginger in major producing states. The trend (Pavan Kumar et al., 2021) of different weather parameters also affects the yield and quality of ginger. The prediction models should also consider the weather parameters as covariates, so that prediction will be better. Being a high value crop, Ginger needs intensive management practices starting from rhizome selection to harvest. Important agronomic practices included in ginger cultivation are, selection of good quality disease free rhizomes, proper rhizome treatment, proper land preparation and raised beds to provide adequate drainage, planting, weed management along with mulching, irrigation, and nutrient management. The cost of cultivation of ginger is very high. Hence, farmers think twice before going for cultivation of ginger. The price of ginger highly volatile in nature it fluctuates based on the quantity of produce available in market and export demand. In this situation farmer needs forecast of Area and production of Ginger in order to decide whether to opt for ginger or

not. Hence, in this paper attempt was made to predict the area and production of ginger using ARIMA and Artificial Neural Network Approach (ANN).

## 2. MATERIALS AND METHOD

Present study was based on ginger production and area data for 41 years (1970-71 to 2010-11) collected from indiastat.com to predict future projection of area and production. The data was tested for its stationarity using augmented dickey fuller (ADF) test (Dickey and Fuller, 1979) and analyzed using autoregressive integrated moving average (ARIMA) and Multilayer Perceptron (MLP) Artificial Neural Networks. The data set was divided into training (70%), selection (15%) and testing sets (15%) for feeding to neural network. For training neural network back propagation and conjugate gradient algorithm were utilized. The models i.e, ANN and ARIMA were compared based on the decision criteria's like R-Square, RMSE and accuracy of future prediction. The model with low error and highest coefficient of determination were proposed for prediction purpose.

### Methodology for ARIMA

Box – Jenkins time – series models written as ARIMA (p,d,q) was first popularized by Box, G.E.P. and Jenkins, G.M. (1976). This model amalgamate three types of processes, namely auto regressive of order p ; differencing to make a series stationary of degree d and moving average of order q. This method applies only to a stationary time series data. When the data is nonstationary which has to be brought into stationary by the method of differencing i.e.  $W_t = Y_t - Y_{t-1}$ . The series  $W_t$  is called the first differences of  $Y_t$ , and the second difference of the series is  $V_t = W_t - W_{t-1}$ . In many cases first differencing is sufficient to bring about a stationary mean and second differencing is done in few cases only.

In regression model it is usually assumed that the error terms are assumed to be uncorrelated. This implies that the various observations within a series are statistically independent. However, this assumption is rarely met in practice. Usually serial correlations in the observations often exist if the data are collected sequentially over time. That is, each observations of the observed data series  $\{Y_t\}$ , which being a family of random variables  $\{Y_t, t \in T\}$ , where T is the index set,  $T = \{0, \pm 1, \pm 2, \dots\}$  and apply standard time-series analysis technique to develop a model which will adequately represent the set of realizations and also their statistical relationship in a better way.

The statistical concept of correlation is to measure the relationships existing among the observations within the series. In these models, the values of correlations between the value of Y at time t (i.e.,  $Y_t$ ) and Y at earlier time periods (i.e.,  $Y_{t-1}, Y_{t-2}, \dots$ ) were examined. The algebraic forms of Autoregressive and Moving average processes are :

Autoregressive process

$$Z_t = C + \phi_1 Y_{t-1} + \varepsilon_t \quad \dots\dots\dots (A)$$

Moving average process

$$Z_t = C - \theta_1 a_{t-1} + \varepsilon_t \quad \dots\dots\dots (B)$$

A process involving past (time – lagged) Y terms is called an autoregressive (Abbreviated as AR) process. The longest time lag associated with a Y term on the right hand side is called the AR order of the process. The equation (A) is thus an AR process of order one, abbreviated as AR(1). On the left hand side,  $Y_t$  represents the set of possible observations on a time sequenced random variables  $Y_1$ . The co-efficient  $\phi_1$  has a fixed numerical value which tells how  $Y_t$  is related to  $Y_{t-1}$ , C is a constant term related to the mean  $\mu$  of the process. The constant term of an AR process is equal to the mean times the quantity one minus the sum of the AR co-efficients., i.e. for an AR(1) process  $C = \mu(1 - \phi_1)$ . Some of the past works also considered ARIMA models for forecasting agricultural commodities and found as significant models for the prediction purpose (Mastny, 2001, Li et al,2010).

### ANN MLP

Multilayer feed forward neural network or multi layer perceptron (MLP),(Anifat Olawoyin and Yangjuin Chen, 2018) a very popular and powerful system is used more than other neural network type for solving complex phenomenon and to find a existing relationship. Multilayer feed forward neural network learned by back propagation algorithm is based on supervised procedure, i.e., the network constructs a model based on examples of data with known output.

MLP can use any number of inputs and one or more hidden layers with any number of nodes, which usually uses the sigmoid function as activation function in the hidden layer and MLP can uses any number of output with any activation function. Has connections between the input layer and the first hidden layer, between the hidden layers, and between the last hidden layer and the output layer. An MLP with just one hidden layer can learn to approximate virtually any function to any degree of accuracy. Due to all these reasons MLPs are known as universal

approximates and can be used when we have little prior knowledge of the relationship between input and targets. A sigmoid activation function uses the sigmoid function to determine its activation. A sigmoid function is a mathematical function having an "S" shape (sigmoid curve). Sigmoid function is given by the following formula

$$f(x) = \frac{1}{1 + e^{-x}}$$

Sigmoid functions are very similar to the input-output relationships of biological neurons, although not exactly the same. Sigmoid function exhibits smoothness and has the desired asymptotic properties. As 'x' goes to minus infinity, f(x) goes to 0. As 'x' goes to infinity, f(x) goes to 1. As x = 0, f(x) = 0.5 (Arun Balaji and Baskaran. 2013).

ANN models were applied to a wide variety of data sets and at variety of disciplines like Agriculture and allied sciences, Social Sciences (Cheng and Titterington. 1994), Weather related aspects (Vivekanadan. 2011, Luis et al., 2008, Cigizoglu. 2003), Banking sector, Public health and natural sciences. This model has potential to learn the complex data at a faster rate and converges quickly to local minima (Anderson. 2003), due to which ANN models (Zhang, et al. 1998) can able to produce better forecasting accuracy than any other linear models. Therefore, an attempt had been made to forecast the area and production of ginger using MLP ANN and Box-Jenkin's ARIMA models.

### ACCURACY MEASURES

Neural network and ARIMA model were compared based on the criteria's viz., Error standard deviation, Absolute Error Mean, SD Ratio (standard deviation), Correlation, R-squared, Root Mean Square Error (RMSE), Mean Absolute Percent Error (MAPE), Mean Absolute Error (MAE), MaxAPE, MaxAE, Normalized BIC (Salazar Fernández. 2017).

### 3. RESULTS AND DISCUSSION

The importance of different forecasting techniques leads to lot much confusion to use the appropriate techniques, so the comparison of techniques made to avail the better technique for use with less statistical error. During the present study results for two important forecasting techniques has been utilized for the forecasting of ginger production and its area in India for comparison. Results of the predicting ginger production using ARIMA and ANN revealed that, Artificial Neural Network Multilayer perceptron model 1:2-20-1:1 produced a best result with less training (0.16), Selection (0.009) and Testing (0.051) error. The Back propagation and conjugate gradient (7b) method as a training algorithm used for training the Network for 100 iterations (epochs). The sigmoid function used as a transformation function in the hidden layer and network has single hidden layer, and was sufficient for prediction purpose (Bandyopadhyay and Chattopadhyay. 2007). Neural network models have higher convergence and learning rate and reduce the possibility of over fitting or under fitting.

The ANN model with 2 inputs, 20 hidden and 1 output units (2-20-1) was fitted to data and provided better forecast for production with less error standard deviation, absolute error mean, and standard deviation ratio and with high correlation (0.91) as compared to model 2-15-1 (Table 1). In ARIMA methodology the auto-correlation up to fourteen lags were worked out and found the computed auto-correlations  $\gamma_k$  values tail off towards zero. The stationarity was also confirmed by examining the realization visually and it was found that, the mean and variance were constant over the time. The ARIMA (4,0,4) and ARIMA(2,0,3) models were fitted for the data and found that the values of RMSE, MAPE, MaxAPE, Normalized BIC and MaxAE were low for ARIMA (2,0,3). Also ACF and PACF (Figure-3, 4) figures of residuals confirmed that residuals were random.

After comparing ARIMA and ANN models it was found that ANN model 2-20-1 showed high R-square (0.82) than ARIMA (2, 0, 3) model (0.77). The production was forecasted using both the models and observed values were predicted accurately by ANN MLP model. Hence ANN was found suitable for out sample prediction of ginger production than ARIMA model (Figure 2) and the model with input; hidden and output units are also showed in the Figure1. For forecasting area of ginger, the ANN model with 2 input, 7 hidden and 1 output nodes trained using back propagation and conjugate gradient algorithm showed better accuracy criteria's after running the model for 100 epochs (iterations) and CG29b and CG40b (Table 4). The ANN model 2-7-1 with CG29b provided better model accuracy values compared to 2-7-1 with CG40b (Table 5). The ARIMA and ANN provided same R-square values but other accuracy criteria's were better with less error standard deviation (11416) and absolute error mean (3276.59) in case of ANN model compared to ARIMA (3,04) model with RMSE value 11810 and mean absolute error 5561 (Table 6). For the selected ARIMA models auto-correlation up to seventeen lags were worked out. Since the computed auto-correlations  $\gamma_k$  values tail off towards zero, the original series was found to be stationary. The stationarity was also confirmed by examining the realization visually. It was found that the mean and variance were constant over the time.

ARIMA (4,0,4) and ARIMA(3,0,4) models were fitted for the data and found that value of RMSE, MAPE, MaxAPE, Normalized BIC and MaxAE were lower for ARIMA (3,0,4). Also ACF and PACF (Figure-7, 8)

Figures of residuals confirmed that residuals were random. Observed and Forecasted values from the ANN and ARIMA were depicted in the line graphs and ANN was closely forecasted the area of ginger than the Box-Jenkin's ARIMA methodology. It was clear that ANN and ARIMA predicted the data series with similar trend but the estimates and model accuracy was better for neural network model as compared to the ARIMA models (Shabri et al., 2009) both in case of area and production of ginger.

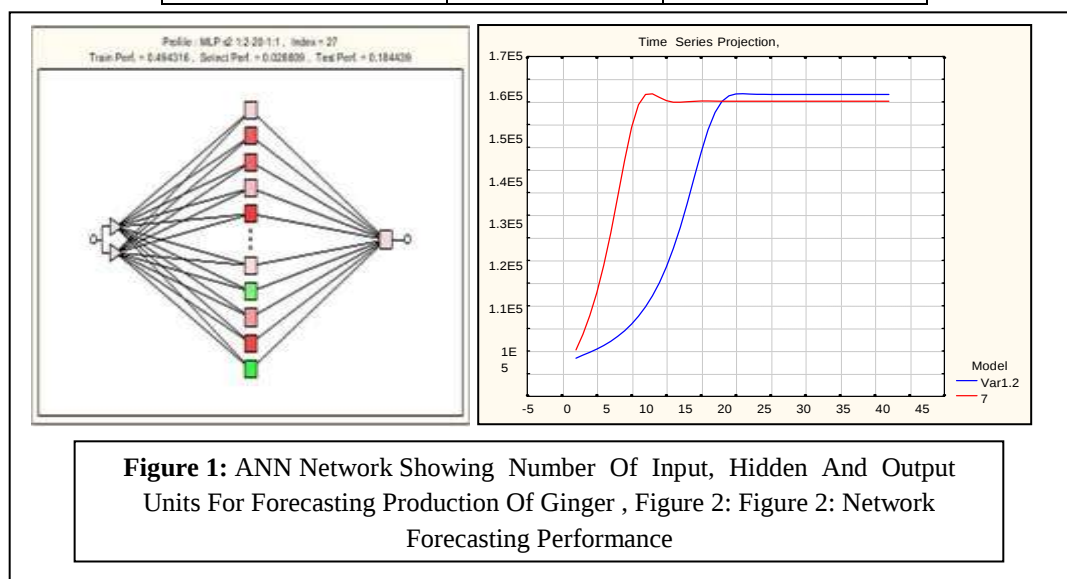
Superiority of neural network models was studied (Rama Krishna Singh and Prajneshu. 2008) for forecasting agriculture crop yield and it was concluded that ANN performed better as compare to MLR and also compared artificial neural network (p, d, q) model for time series forecasting (Mehdi Khashei and Mehdi Bijari. 2010), indicated that ANN can be used as a alternative time series forecasting tool for higher accuracy of prediction. Review of research articles revealed the importance of artificial neural networks in prediction with higher accuracy as compare to other techniques. Artificial neural networks are powerful tools and gaining more importance in agriculture sector for prediction of production, area, yield (Srinath Reddy et al., 2014), productivity incidence of disease pest and commodity prices (Sinha. 2013, Shahwan and Odening. 2007 and Li. 2010) across markets etc.

**Table1.** ANN Model Summary For Ginger Production Forecasting

| Model | Profile        | Train perf | Selec. Perf | Test Perf | Train.E roor | Sel. Error | Test Error | Training Members | Input | Hidden |
|-------|----------------|------------|-------------|-----------|--------------|------------|------------|------------------|-------|--------|
| 27    | MLP<br>s2 1:2- | 0.4943     | 0.028       | 0.1844    | 0.16892      | 0.009      | 0.051      | BP100,CG7b       | 1     | 20     |
| 28    | MLP<br>s2 1:2- | 0.5107     | 0.015       | 0.1821    | 0.17449      | 0.0049     | 0.049      | BP100,CG29b      | 1     | 15     |

**Table 2:** Descriptive Statistics for ANN Networks for Ginger Production Forecasting

| MODEL       | 1        | 2        |
|-------------|----------|----------|
| Data Mean   | 77855.74 | 77855.74 |
| Data S.D.   | 75229.76 | 75229.76 |
| Error S.D.  | 31187.06 | 32183.23 |
| Abs E. Mean | 13176.69 | 14405.55 |
| S.D. Ratio  | 0.41     | 0.43     |
| Correlation | 0.91     | 0.9      |

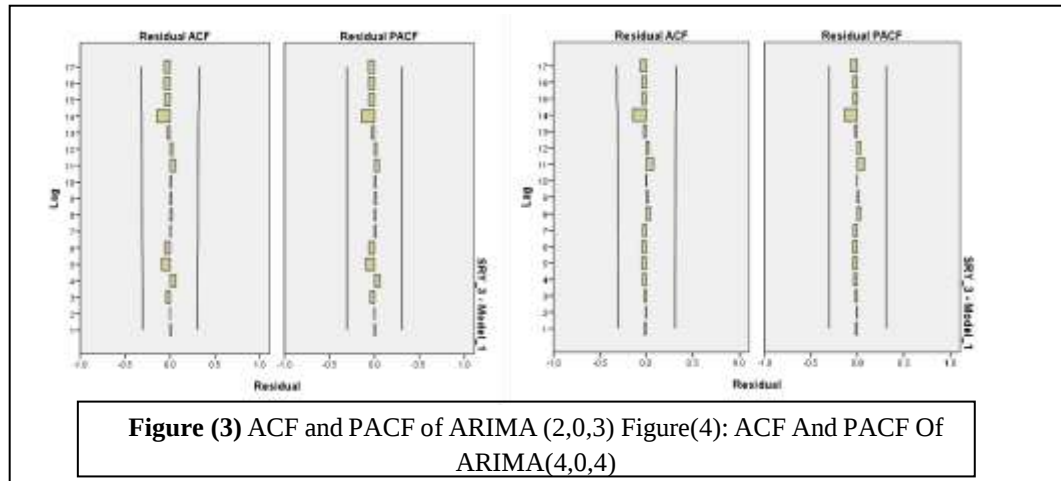


**Table 3:** ARIMA Models For Forecasting Production Of Ginger.

| Model Type   | R-squared | RMSE  | MAPE | MAE   | MaxAPE | MaxAE  | Normalized BIC |
|--------------|-----------|-------|------|-------|--------|--------|----------------|
| ARIMA(4,0,4) | 0.774     | 39820 | 2660 | 18700 | 72830  | 183600 | 21.999         |
| ARIMA(2,0,3) | 0.771     | 38380 | 2456 | 18610 | 73100  | 184300 | 21.654         |

**Table 4:** ANN Model Summary For Ginger Area Forecasting

| Model | Profile | Train Perf. | Select Perf. | Test Perf. | Train Error | Select Error | Test Error | Training/Members | Input | Hidden |
|-------|---------|-------------|--------------|------------|-------------|--------------|------------|------------------|-------|--------|
| 127   | MLP s2  | 0.0803      | 0.0636       | 1.5441     | 0.0274      | 0.0288       | 0.4528     | BP100,CG29       | 1     | 7      |
| 128   | MLP s2  | 0.081       | 0.0622       | 1.5606     | 0.0277      | 0.0277       | 0.4566     | BP100,CG40       | 1     | 7      |

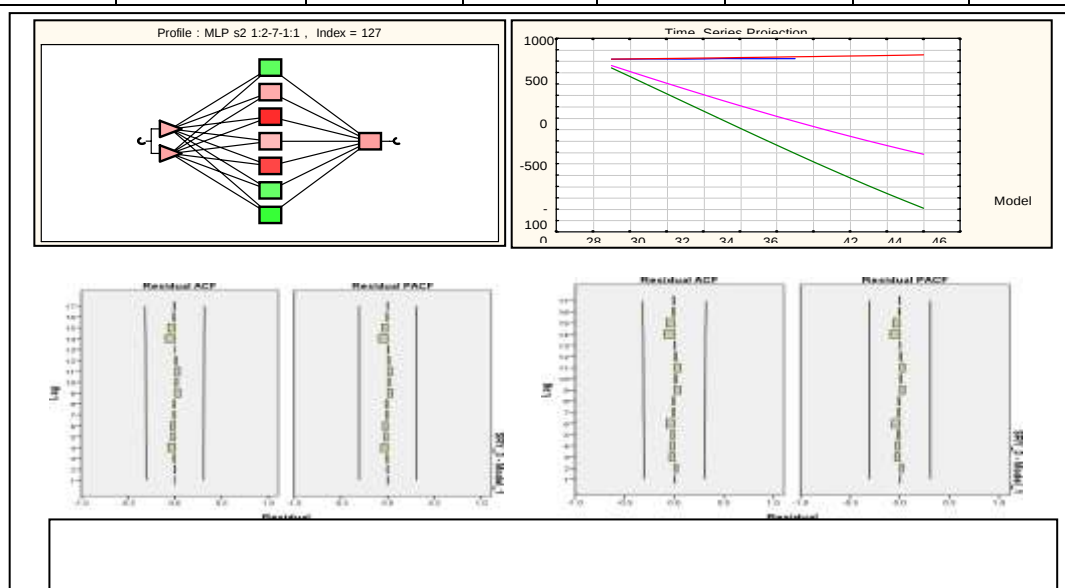


**Table 5:** Descriptive Statistics For ANN Networks For Ginger area Forecasting

| Model       | Var1.127 | Var1.128 |
|-------------|----------|----------|
| Data Mean   | 29961.44 | 29961.44 |
| Data S.D.   | 24865.03 | 24865.03 |
| Error S.D.  | 11416.14 | 11497.80 |
| Abs E. Mean | 3276.59  | 3461.66  |
| S.D. Ratio  | 0.46     | 0.46     |
| Correlation | 0.90     | 0.90     |

**Table 6:** ARIMA Models For Forecasting Area of Ginger

| Model type   | R-squared   | RMSE  | MAPE | MAE  | MaxAPE       | MaxAE        | Normalized BIC |
|--------------|-------------|-------|------|------|--------------|--------------|----------------|
| ARIMA(3,0,4) | <b>0.81</b> | 11810 | 2650 | 5561 | <b>76690</b> | <b>57980</b> | <b>19.478</b>  |
| ARIMA(4,0,4) | 0.809       | 12030 | 2629 | 5442 | 78070        | 59020        | 19.606         |



**Figure 5:** ANN Network Showing Number Of Input, Hidden And Output Units For Forecasting Area Of Ginger, Figure 6: Network Forecasting Performance, Figure 7:ACF and PACF of ARIMA(3,0,4), Graph 8:ACF and PACF of ARIMA (4,0,4)

#### 4. CONCLUSION

Most of the times underlying data may not be linear, in this case nonlinear models are best suited to obtain the better estimates compare to linear models. The Artificial Neural Networks are becoming more popular in forecasting studies. Results from both techniques revealed that production and area was predicted closely using Artificial Neural Network and outperformed Autoregressive Integrated Moving Average (ARIMA) model with respect to accuracy criteria's. Some authors were suggested using hybrid ANN-ARIMA models which can learn better compared to traditional time series models in forecasting (Luis A. Di'az-Robles et al, 2008). Therefore ANN model could be used as better alternative (Peter Zhang, 2003) model to obtain accurate time series projection of production and area of ginger for India.

#### 5. REFERENCES

- [1] Anderson, J. A. (2003). An Introduction to neural networks. Prentice Hall.
- [2] Anifat Olawoyin, and Yangjuin Chen. (2018). Predicting the Future with Artificial Neural Network, Complex Adaptive Systems Conference with Theme: Cyber Physical Systems and Deep Learning, CAS 2018, 5 November – 7 November 2018, Chicago, Illinois, USA. Procedia Computer Science., 140 (2018) 383– 392
- [3] Anonymous. (2020). <https://www.indiaagronet.com/indiaagronet/crop%20info/ginger.htm>. Last Accessed on 22.04.2020
- [4] Arun Balaji, S. and Baskaran. K. (2013). Design And Development Of Artificial Neural Networking (Ann) System Using Sigmoid Activation Function To Predict Annual Rice Production In Tamilnadu. International Journal of Computer Science, Engineering and Information Technology (IJCEIT)., Vol.3 (1):13-31 DOI :10.5121/ijcseit.2013.3102
- [5] Bandyopadhyay, G., Chattopadhyay, S. (2007). Single hidden layer artificial neural network models versus multiple linear regression model in forecasting the time series of total ozone. Int. J. Environ. Sci. Technol., 4: 141–149. <https://doi.org/10.1007/BF03325972>
- [6] Bode AM, Dong Z. (2011). The Amazing and Mighty Ginger. In: Benzie IFF, Wachtel-Galor S, editors. Herbal Medicine: Biomolecular and Clinical Aspects. 2nd edition. Boca Raton (FL): CRC Press/Taylor & Francis; 2011. Chapter 7. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK92775/>
- [7] Cheng, B. and Titterington, D. M. (1994). Neural networks: A review from a statistical perspective. Statistical Science., 9: 2-54.
- [8] Cigizoglu, H.K. (2003). Inter-comparison of ARMA models into flow forecasting by artificial neural networks. Environmetrics., 14: 417-427.
- [9] Dickey, D. A. and Fuller, W. A. (1979). "Distribution of the Estimators for Autoregressive Time Series with a Unit Root". Journal of the American Statistical Association., 74 (366):427:431. doi:10.1080/01621459.1979.10482531. JSTOR 2286348.
- [10] Li, G., Xu, S., and Li, Z. (2010). Short-Term Price Forecasting For Agro-products Using Artificial Neural Networks. Agriculture and Agricultural Science Procedia., 1: 278–287. doi:10.1016/j.aaspro.2010.09.035
- [11] Luis A. Di'az-Robles, Juan C. Ortega, Joshua S. Fu, Gregory D. Reed, Judith C. Chow, John G. Watson, Juan A. and Moncada-Herrera. (2008). A hybrid ARIMA and artificial neural networks model to forecast particulate matter in urban areas: The case of Temuco. Chile Atmospheric Environment., 42: 8331–8340
- [12] Mastny, V. (2001). The use of Box-Jenkins methodology to predict the price development of agricultural commodities. Acta Unierisatatis Agriculturae et Silviculturae- Mendelianae- Brunensis., 49(2): 165-172.
- [13] Mehdi Khashei, and Mehdi Bijari (2010). An artificial neural network (p, d,q) model for timeseries forecasting. Expert Systems with Applications., 37 :479–489
- [14] Rama Krishna Singh and Prajneshu. (2008). Artificial Neural Network Methodology for Modelling and Forecasting Maize Crop Yield. Agricultural Economics Research Review. 21:5-10
- [15] Salazar, Fernando , Rafael Morán, Miguel Ángel Toledo, and Eugenio Oñate. (2017). Data-Based Models for the Prediction of Dam Behaviour: A Review and Some Methodological Considerations. Archives of Computational Methods in Engineering., 24: 1-21.
- [16] Shabri, A., Samsudin, R. and Ismail, Z. (2009). Forecasting of the Rice Yields Time Series Forecasting using Artificial Neural Network and Statistical Model. Journal of Applied Sciences. 9: 4168-4173.
- [17] Shahwan T. and Odening, M. (2007). Forecasting Agricultural Commodity Prices using Hybrid Neural Networks. In: Chen SH., Wang P.P., Kuo TW. (eds) Computational Intelligence in Economics and Finance. Springer, Berlin, Heidelberg

- 
- [18] Sinha Kanchan. (2013). Agricultural Price Forecasting Using Neural Network Model: An innovative information delivery system. *Agricultural economics research.*, 26: 229-239.
- [19] Srinath Reddy, A. B., Havaladar, Y. N., Pavan Kumar, S.T., and Adam Kamei. (2014). Forecasting of Areca Nut (Areca catechu) Yield Using ARIMA Model for Uttara Kannada District of Karnataka. *International journal of science and research.*, 3(7):1002-1006
- [20] Vivekanadan, N. (2011). Prediction of Annual run off using artificial neural network and regression approaches. *MAUSAM.*, 62(1):11-20.
- [21] Zhang, G., Patuwo, B. E. and Hu, M. Y. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting.*, 14: 35-62.
- [22] Zhang, G.P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing.*, 50: 159–175
- [23] Pavankumar, S.T., Lahiri, B. and Alvarado, R. (2021) Estimation of trends in Covid-19 Infections and deaths across WHO regions and India: Multiple change point analysis approach. *Spat. Stat.* 49: 100538. <https://doi.org/10.1016/j.spasta.2021.100538>