Section A-Research paper
ISSN 2063-5346

Comparative analysis of Prediction on Under five mortality rates in India using Autoregressive Integrated Moving Average and Feed Forward Neural Network.

K. Bhaskar[1*], Dr. M. Uday Shankar[2], Dr. M. Raghavender Sharma[3]

1. *K. Bhaskar , Research Scholar, Department of statistics, University College of Science, Osmania University, Hyderabad, Telangana. Email: bhaskarstatistician@gmail.com*
2. *Dr. M. Udayshankar, University College of Science, Osmania University, Hyderabad, Telangana. Email:udayshankar.malyala@gmail.com*
3. *Dr. M. Raghavender Sharma, Assistant Professor, Department of Statistics, Osmania University, Hyderabad, Telangana. Email:drmrsstatou@gmail.com.*
4. *Corresponding author: K. Bhaskar , Research Scholar, Department of statistics, University College of Science, Osmania University, Hyderabad, Telangana. Email: bhaskarstatistician@gmail.com.*

DOI: 10.48047/ecb/2023.12.si4.1714

Abstract

The under-five mortality rate (U5MR) is an important measure of child well-being, which includes health status, and, more generally, of social and economic growth. This paper presents an empirical model to forecast the future under-five mortality rates in India using Auto Regressive Integrated Moving Average and Feed-forward Neural Network. The predictive models in the time series components are employed to forecast the U5MR in India. In this study, the training dataset was utilized to develop a model, and the test dataset was used to evaluate its performance. The performance was determined by computing error measures such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). The results of both models demonstrate that the FFNN has superior performance compared to the ARIMA model.

Keywords: Autoregressive Integrated Moving Average, Feed Forward Neural Network. Under-Five Mortality Rate. Mean Absolute Error, Mean Absolute Percentage Error,

1.Introduction:

According to UNICEF the under-five mortality rate(U5mr) refers to the probability a newborn would die before reaching exactly 5 years of age, expressed per 1,000 live births. Globally, infectious diseases, including pneumonia, diarrhoea, remain a leading cause of under -five deaths, along with preterm birth and intrapartum -related complications. The global under-five mortality rate declined by 59 percent, from 93 deaths per 1,000 live births in 1990 to 38 in 2021.

ISSN 2063-5346

Despite this considerable progress, improving child survival remains a matter of urgent concern. In 2021 alone, roughly 13,800 under-five deaths occurred every day, an intolerably high number of largely preventable child deaths [1].

Under five mortality rates for the country has shown significant decline of 1.4 points (Annual decline rate: 4.4%) from 2020 (30.6 per 1000 live births in 2021 against 32 per 1000 live births in 2020). The global goal is to lower the U5MR to at least 25 per 1,000 live births by 2030 in accordance with Sustainable Development Goal(SDG) number 3.Under five mortality is a crucial public health threat in INDIA and other Low-Middle Income Countries (LMICS).It has traditionally been used as an important health indicator for assessing population well-being and consistently gained visibility in the Millennium Development Goals (MDGs) and Sustainable Development Goals (SDGs). Sustainable Development Goal taken up by the United Nations in 2015 were developed to assist healthy lives and well-being for all children.

By 2030, the SDG 3.2.1 aims to eradicate deaths of infants and young children that could have been prevented. Accurate forecasts of child mortalities will show effective use of the limited health resources. Sound modeling path to improve child mortality estimates is essential. Such accurate forecasts will assist Indian and LMICs to plan and execute which can further aid in achieving to reach SDG.

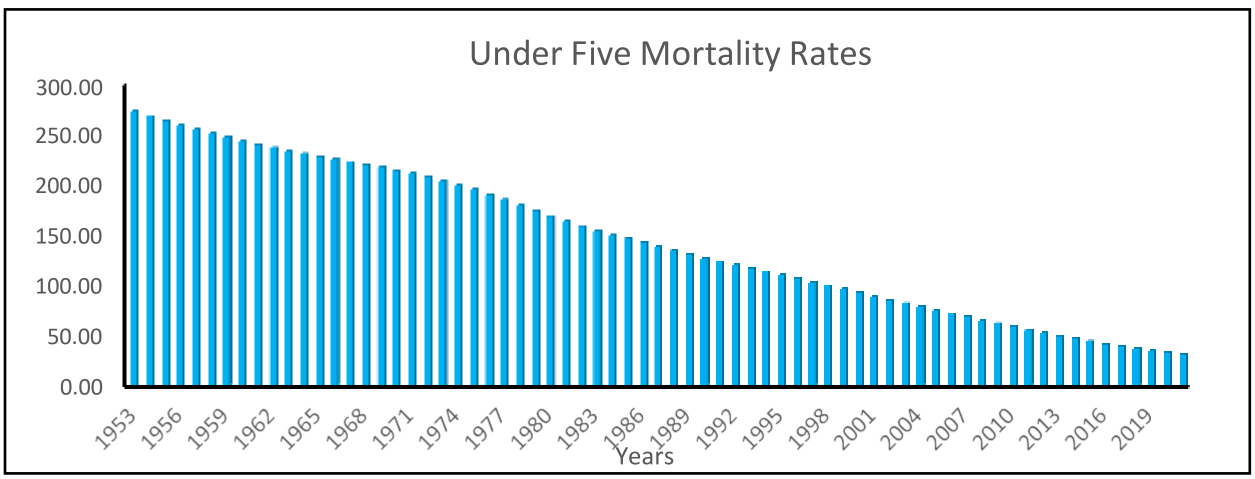

Figure 1: Time series plot for Under-Five Mortality Rates in India.

2. Material and Methods:

A historical data of under-five mortality rates in India is collected form the website of unicef.org from 1953 to 2021 in India. The Box-Jenkins procedure and feed forward neural network models are used for estimate or forecast the future mortality rates and to suggest the accurate model. R software is used for development a model and validation and charts and tables for Micro Soft Excel.

2.1 ARIMA Model:

The Box-Jenkins procedure is the best for determining the appropriate model by developing on real life data sets, and the methodology consists of multiple advantages for calculating a smaller number of parameters and checking the seasonality of the data. This method has mainly four steps for building the model: identification, Estimation, Diagnostic checking, and forecasting[12]. ARIMA (p, d, q) where "p" is the autoregressive coefficient, "d" is the order of difference made when the time series becomes stationary, and "q" is the number of moving average terms. Perform a stationarity test on the time series first, and if it is not stationary, utilize methods such as difference and logarithm to make it stationary. The ACF dies out for multiple lags and q spikes in the plot then q parameters will occur, and PACF dies out for multiple lags and p spikes in the plot, then this is the p parameter. Furthermore, we can use the unit root test to determine whether the data is from a stationary process: Stationarity is determined by the Augmented Dickey-Fuller (ADF) test. H_0: There is a unit root in the series; H_1: The series lacks a unit root. The sequence is stationary. If two or more models are developed, the Akaike Information Criterion (AIC) will be used to choose the optimal model (with the lowest AIC) among them. Next to check for diagnostics checking, test is used to verify the adequacy of the model with respect to errors that are random for using LJung – Box Q Statistics. The Q statistic checks the randomness of the error and the significance of the parameters. If the parameters are not significant or the error that are non-random, then proceed to test the other possible p and q orders until you get the significant p and q values. In this way, test several models for the train data set and find the best one based on the multiple criteria’s for forecasting the yearly under five mortality rates in India. The non-seasonal ARIMA model is

$$y_t = \mu + \text{Ø}_1 y_{t-1} + \text{Ø}_2 y_{t-2} + .. + \text{Ø}_p y_{t-p} - \theta_1 s_{t-1} - \theta_2 s_{t-2} - .. - \theta_q s_{t-q} + c_t \text{ ; } c_t \sim iid$$

Where Ø, θ are unknown parameters

2.2. Artificial Neural Network Model:

The ANN model is useful for the biological neural networks and those are linked with multiple groups of nodes. the following discern is a structure of neural network.

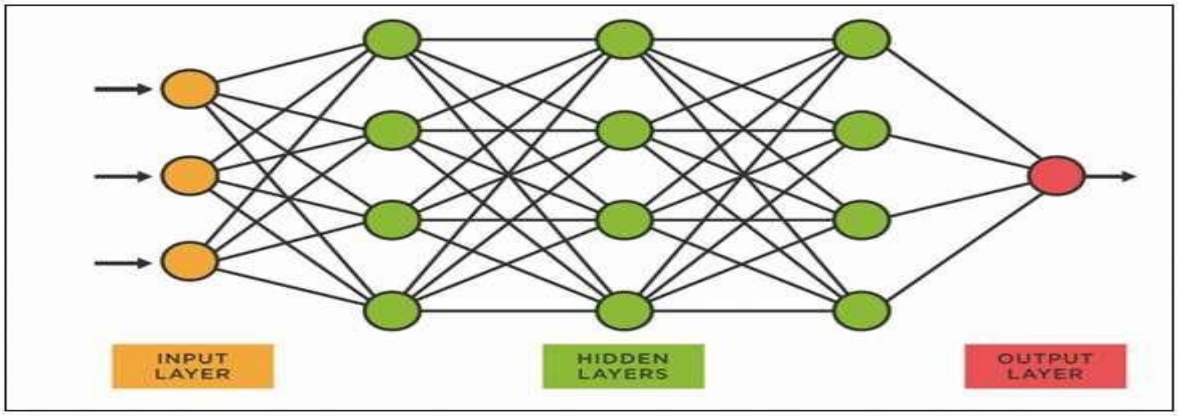

Figure 2: Three layered architecture of feed forward neural network.

The feed forward neural network incorporates three layers so it is called 3- layer architectural diagram. the first layer is data layer which encompass the raw information and second one is hidden layers and final one is output layer. The output layer must be one and the number of hidden layers according to the trial and error method. The hidden layers will get more than one layer along with weights. The below table 1 gives the information about the network.

Table1: Network Information

Input Layer	Covariates	Lag1
	Number of Units	Normalization
	Rescaling method of covariates	
Hidden Layer	Number of hidden layers	3
	Number of units in the hidden layer	3
	Activation function	Hyperbolic Tangent
Output Layer	Dependent variable	
	Number of units	1
	Rescaling method of scale dependent	Normalized
	Activation function	Identity
	Error function	Sum of Squares

3.Results and discussion:

3.1. ARIMA Model:

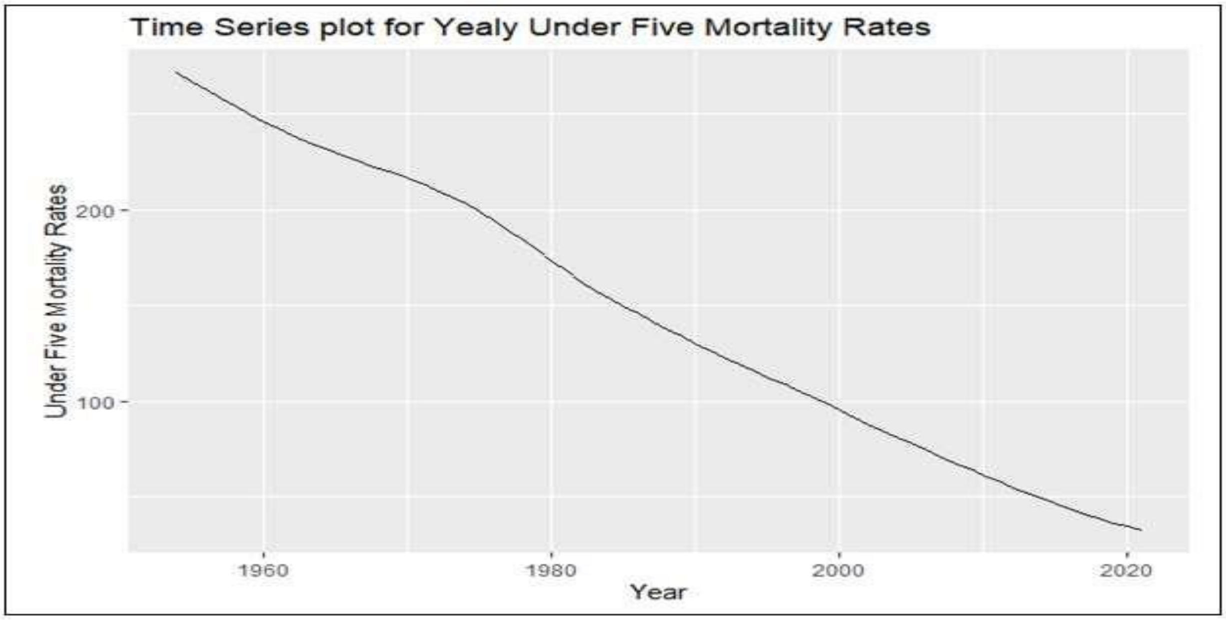

Figure 3: Time series plot on under five mortality rates in India from 1953 to 2021.

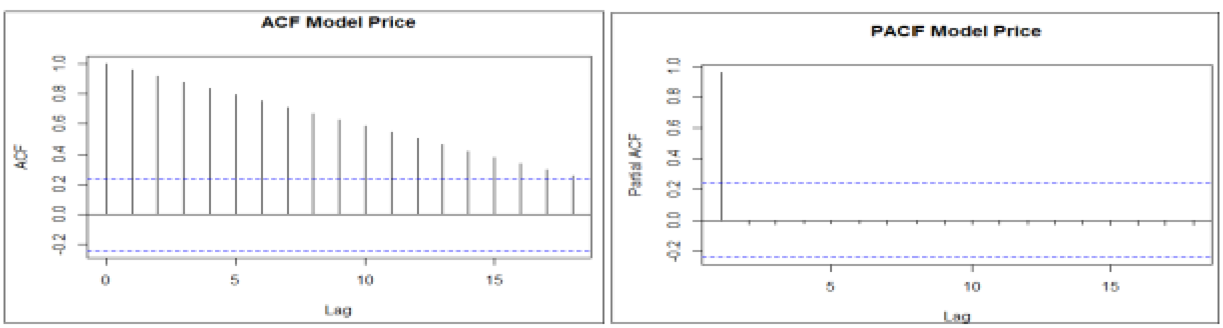

Figure 4: ACF and PACF plots on under-five mortality rates in India

Figure 3&4 depicts the data flow of U5mr in India from 1953 to 2021.The mortality rates are completely decreasing year by year due to implementing more health programmes for Antenatal care (ANC) and children by the Government of India. The U5mr is 252 in the year 1953 and 30.6 in 2021. The mortality rate has declined to 87.9% (yearly 1.3%) from 1953 to 2021. It shows gradually constant decline.

Figure 5: Transformed series plot on yearly under five mortality rates in India.

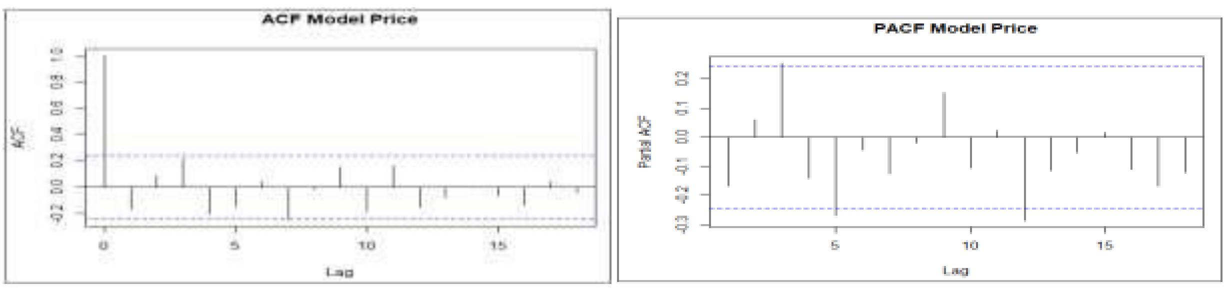

Figure 6: Transformed ACF and PACF plot on yearly under five mortality rates in India.

According to fig 5&6 the data shows stationary with constant mean and variance after applied the transformation such as second difference (d=2) and auto correlation and partial auto correlation functions are dies out with the second difference. The one more test is to check the stationary of the data that is Augmented Dicky-Fuller (ADF) test. The ADF test on second difference the p-value is 0.01 (<0.05). Now reject the null hypothesis and the data is stationary. The following table is second order difference ADF test results are in the table 2.

Table 2: ADF test results

ADF Test	Lag Order	P Value
-4.11	**4**	**0.01 < 0.05**

From the above table 2, the data was found to be stationary with a second difference, and, the different orders of p and q were evaluated, as well as multiple models and to find the optimum one from the several models. Table 3 depicts the few possible models of ARIMA.

Table 3: Possible model of ARIMA

ARIMA(p,d,q)			AIC	BIC	Significance of the parameter	LJung Box Q statistics	P-Value	Adequacy
1	2	0	-73.84	-69.9	Significant	25.916	0.002	Inadequacy
0	2	1	-50.82	-46.88	Significant	37.804	0.000	Inadequacy
1	2	1	-72.03	-66.12	Insignificant	24.993	0.002	Inadequacy
2	2	0	-72.11	-66.2	Insignificant	24.721	0.002	Inadequacy
0	2	2	-57.19	-51.28	Significant	36.223	0.000	Inadequacy
2	2	1	-70.21	-62.33	Insignificant	25.031	0.001	Inadequacy
1	2	2	-74.45	-66.57	Insignificant	15.794	0.027	Inadequacy
3	2	0	-71.25	-63.37	Insignificant	29.274	0.000	Inadequacy
0	**2**	**3**	**-74.62**	**-66.73**	**Significant**	**9.7967**	**0.200**	**Adequacy**
1	2	3	-83.6	-73.74	Insignificant	10.576	0.102	Adequacy
2	2	3	-75.74	-63.92	Insignificant	12.487	0.029	Inadequacy
3	2	1	-77.16	-67.31	Insignificant	24.652	0.000	Inadequacy
3	2	2	-72.57	-60.75	Insignificant	11.015	0.051	Inadequacy
0	2	4	-81.09	-71.24	Significant	16.158	0.013	Inadequacy
1	2	4	-83.1	-71.28	Insignificant	14.183	0.014	Inadequacy
2	2	4	-80.52	-66.73	Insignificant	11.654	0.020	Inadequacy

After observing several models from the above table 3, concluded that the ARIMA (0,2,3) is identified as the best model as compared with other similar models shown in the above table 3. The chosen model ARIMA (0,2,3) has the lowest Akaike information criterion (AIC) and Bayesian information criterion (BIC) with parameters that are significant, and it was also examined for model adequacy. The parameters of the ARIMA (0,2,3) model are listed in the below table. **Table 4: ARIMA (0,2,3) model parameters.**

Parameter	Estimate	Std.error	z value	Pr(>|z|)
ma1	0.64	0.10	6.20	<0.001
ma2	0.89	0.23	3.81	<0.001
ma3	0.77	0.21	3.62	<0.001

The parameters of the ARIMA (0,2,3) is significant and this is the best model as compared to the other model for forecast the future mortality rates in India. Then the ARIMA (0,2,3) model is

$$Y_t = 2Y_{t-1} - Y_{t-2} - 0.64\, e_{t-1} - 0.89 e_{t-2} - 0.77\, e_{t-3}$$

Now, check the model adequacy using the LJung – Box Q statistics. This test uses the error of the series after developing the model. Then the hypothesis of the model will be H_0: Model is adequate and H_1: Model is Inadequate

Table 5: LJung- Box Q Statistics

LJung Box Q Statistics		
Statistics	df	Significance
9.80	7	0.20

From the above table 5, the hypothesis value is more than the p-value as 0.05, and then accept the null hypothesis and concluded that the ARIMA (0,2,3) model is a adequate model. The following table 6 gives the forecasting of under-five mortality rates in India from the year of 2009 to 2021.

Table 6: Forecasted of Under-Five mortality rates in India using ARIMA (0,2,3) from 2009 to 2021

Year	Actual U5MR	Forecasted U5MR
2009	61.36	61.43
2010	58.15	58.30
2011	55.04	55.16
2012	51.97	52.03
2013	49.03	48.90
2014	46.22	45.77
2015	43.53	42.63
2016	40.97	39.50
2017	38.56	36.37
2018	36.33	33.24
2019	34.37	30.11
2020	32.63	26.97
2021	30.60	23.84

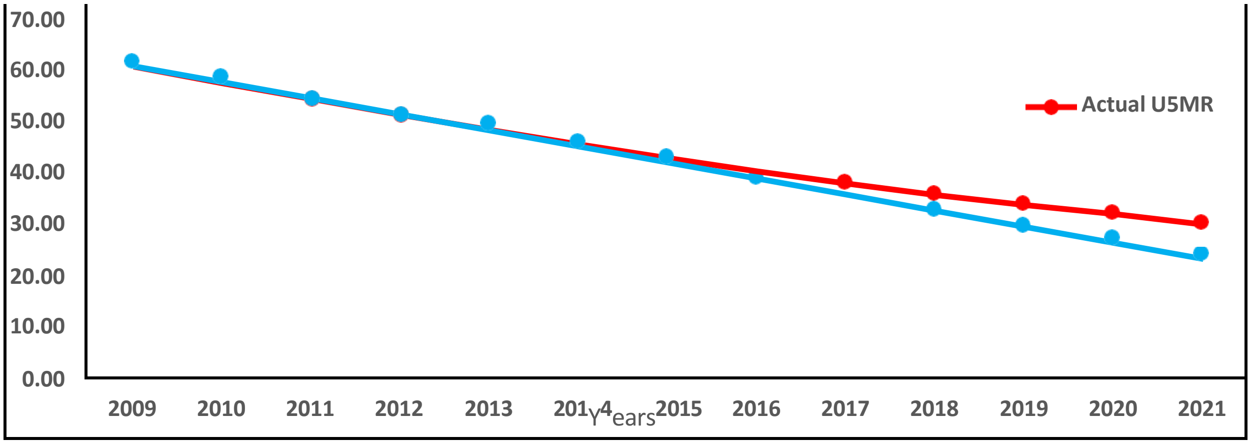

Figure 7: Forecasted of Under-Five mortality rates in India using ARIMA(0,2,3)

To test the model performance using the validation data set apply the ARIMA (0,2,3) model and determined the error measures. The following table 7 shows the model performance based on the error measures.

Table 7: Performance of the ARIMA (0,2,3) model

Data	RMSE	MAE	MAPE
Training Set	0.12	0.09	0.05
Test Set	2.28	1.43	3.97

The performance of the model from the above table 7, the identified model ARIMA(0,2,3) gives the better error results. The error results are minimum and very near to training and testing data sets for the selected model. The mean square error values are 0.09 to 1.43 on training and test. The mean absolute percentage error values are 0.05 to 3.97 respectively on training and test data sets.

3.2 Feed Forward Neural Network Model

The feed forward neural network model has one input neuron, lag1, and there is no actual method to identify the number of hidden layers without preparing the forward or backward selection process to determine the number of hidden layers. Here, the tangent function is used for the activation function on the back-propagation methodology, and the forward selection process is used to find some possible models at different orders on hidden layers and identify the best model based on which hidden layer contains minimum error measures such as MAPE, MAE, and RMSE. The following table 8 has some possible models on multiple hidden layers.

Table 8: Possible model on FFNN

Number of layers			Train			Test		
Input	hidden	Output	RMSE	MAE	MAPE	RMSE	MAE	MAPE
1	2	1	1.76	1.40	0.90	0.95	0.79	2.03
1	3	1	1.15	0.98	0.74	1.00	0.85	2.18
1	4	1	1.73	1.33	0.83	3.15	2.88	7.13
1	5	1	1.74	1.42	0.88	4.59	3.85	9.92

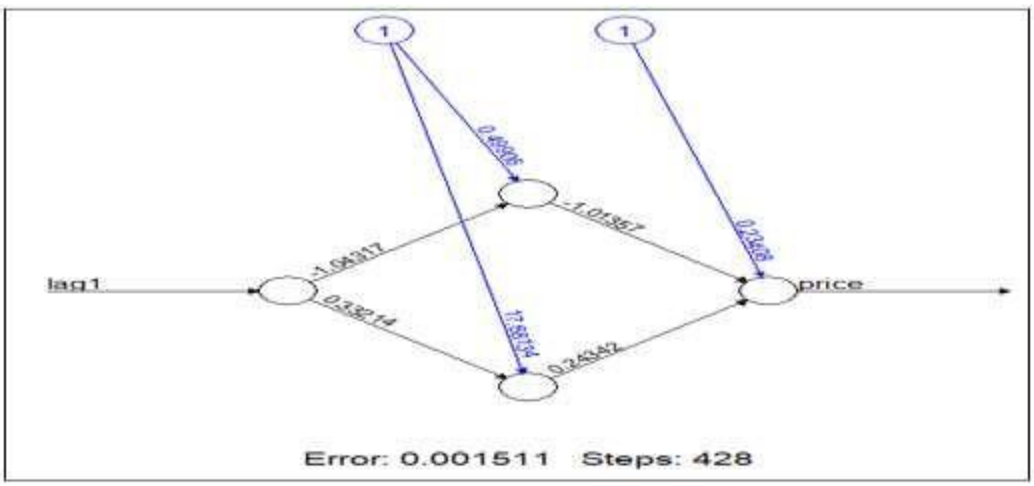

Figure 8: architecture for FFNN (1-2-1) model

The R program is utilized in the above build of models, and the (1-2-1) model parameters are presented in table 9.

Table 9: Model parameters on FFNN(1-2-1)

Parameter	
Error	0.002
reached threshold	0.009
Steps	428
Intercept.to.1layhid1	0.499
lag1.to.1layhid1	-1.043
Intercept.to.1layhid2	17.681
lag1.to.1layhid2	0.332
Intercept.to.price	0.234
1layhid1.to.price	-1.014
1layhid2.to.price	0.243

Now, the hidden neurons are,

$H_1 = \tanh[0.499+1.043\ z_{t-1}]$

$H_2 = \tanh[17.681-0.332\ z_{t-1}]$

Where z_{t-1} is rescaled information, Then the FFNN(1-2-1) model equation is

$O_t = -1.014\ H_1 + 0.243\ H_2 - 0.234$

The model performance of the FFNN (1-2-1) model in the development and validation stages is presented in the following table 10.

Table 10: Model performance FFNN (1-2-1)

Data	RMSE	MAE	MAPE
Training Set	1.76	1.40	0.90
Test Set	0.95	0.79	2.03

In comparison to other models, the performance of the model from the above table 10, the identified model FFNN (1-2-1) provides the best and minimum errors. For the specified model, the error results are minimal and extremely close to the training and testing data sets. The mean square error values are 01.40 and 0.79 on training and testing. The mean absolute percentage error values are 0.90 and 2.03 respectively on training and testing data sets.

Table 11: Forecasted of under-five mortality rates in India using FFNN (1-2-1) model

Year	Actual U5MR	Forecasted U5MR
2009	61.36	61.27
2010	58.15	58.31
2011	55.04	55.42
2012	51.97	52.64
2013	49.03	49.92
2014	46.22	47.36
2015	43.53	44.92
2016	40.97	42.62
2017	38.56	40.45

2018	36.33	38.42
2019	34.37	36.56
2020	32.63	34.95
2021	30.60	32.82

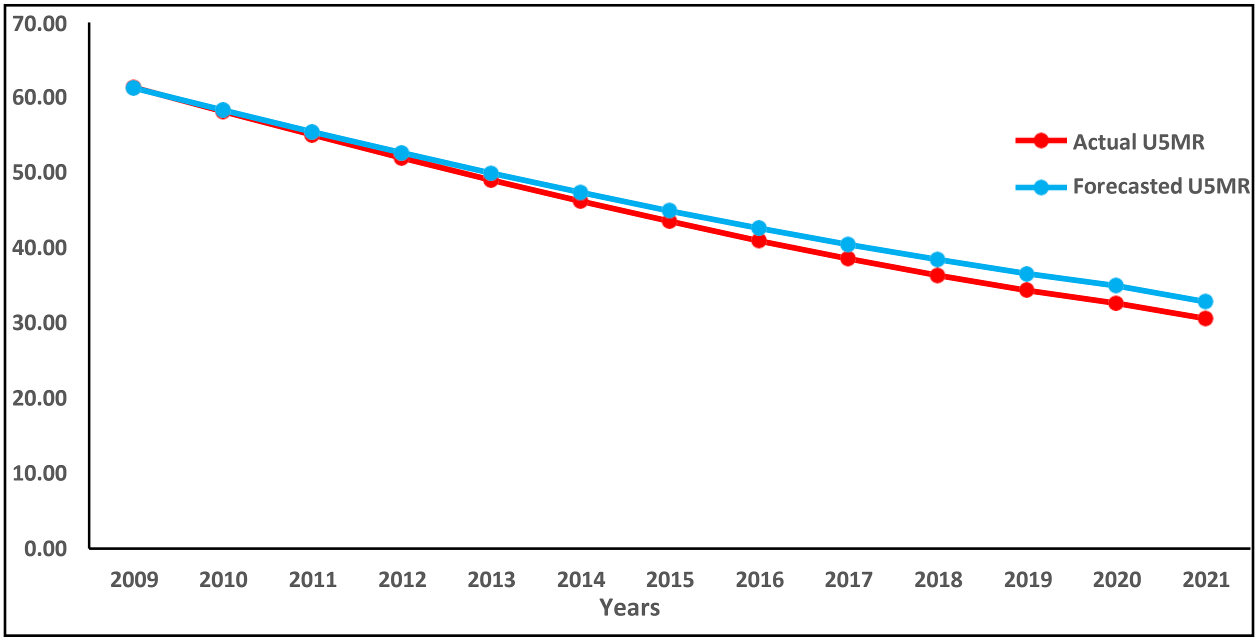

Figure 9: Forecasted of Under-Five mortality rates in India using FFNN (1-2-1) model

Figure 9 shows that the FFNN model performs better, and the predicted trend is quite close to the actual mortality rates in India. Tables 12 and 13 and Figure 10 provide a comparison of ARIMA and FFNN's projected U5MR with actual mortality rates.

Table 12: Comparison on Forecasts of ARIMA (0,2,3) and FFNN (1-2-1) models

Year	Actual U5MR	ARIMA Forecasted U5MR	FFNN Forecasted U5MR
2009	61.36	61.43	61.27
2010	58.15	58.30	58.31
2011	55.04	55.16	55.42
2012	51.97	52.03	52.64
2013	49.03	48.90	49.92
2014	46.22	45.77	47.36
2015	43.53	42.63	44.92
2016	40.97	39.50	42.62
2017	38.56	36.37	40.45

2018	36.33	33.24	38.42
2019	34.37	30.11	36.56
2020	32.63	26.97	34.95
2021	30.60	23.84	32.82

The model performance of the ARIMA (0, 2, 3) and FFNN (1-2-1) models in the development and validation stages is presented in the following table 13.

Table 13: Comparison of ARIMA and FFNN models

Model	ARIMA			FFNN		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE
Train set	0.12	0.09	0.05	1.76	1.40	0.90
Test set	2.28	1.43	3.97	0.95	0.79	2.03

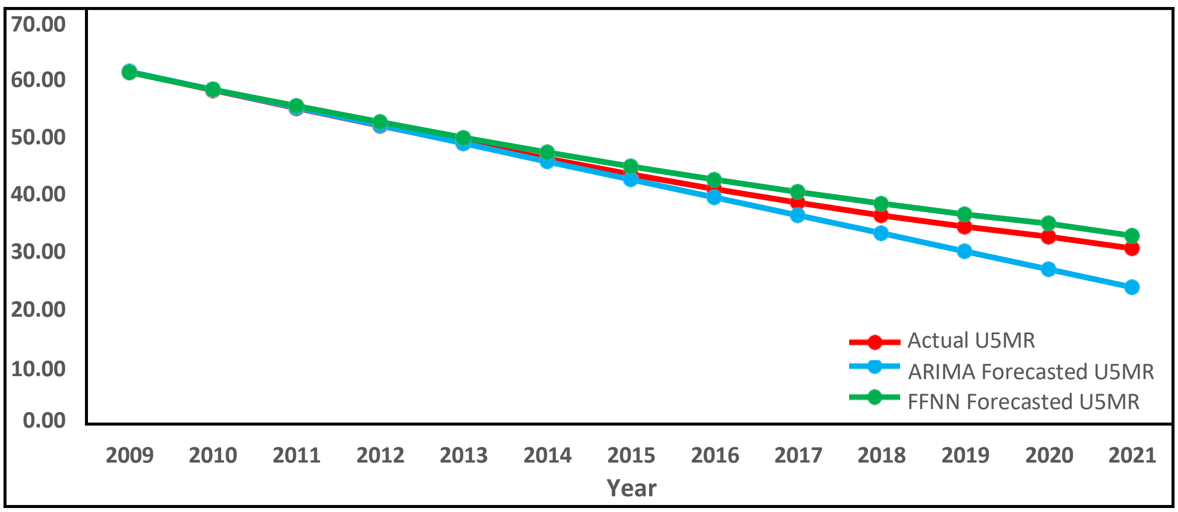

Figure 10: Comparison on Forecasts of ARIMA (0,2,3) and FFNN (1-2-1) models.

5. Conclusion:

The results of the autoregressive integrated moving average model and the feedforward neural network model were compared and it was found that the feedforward neural network model gave better results in predicting future mortality in India according to the performance of the model and forecasted pattern. The ARIMA regression might not be suitable for long-term forecasting of Under-five mortality rates for India. Therefore, FFNN might be more suitable for non-linear data, such as U5MR.FFNN increases forecasting accuracy of mortality rates, so that the Indian government can disperse future resources and prepare for children's services.

6. References

1. UNICEF(2023) https://data.unicef.org/topic/child-survival/under-five-mortality/#status Last accessed on 26.06.2023.
2. Wan Zakiyatussariroh Wan Husin, Ratna Zuarni Ramli, Ainaa Nabilah Muzaffar, Nur Fadhilah Abd Nasir And Siti Nura Effa Rahmat.TREND ANALYSIS AND FORECASTING MODELS FOR UNDER-FIVE MORTALITY RATE IN MALAYSIA, PJAEE, 17 (10) (2020).
3. W. Liang, F. Liang, J. Bai, Y. Hu, R. Zheng, L. Pan, & Z. Li (2020), Mortality Analysis And Model Prediction Of Children Under 5 Years Old In A City Of Northwest China, A Epidemiol Public Health, Vol. 3(1), Pp. 1019,
4. M. Uday Shankar (2020) et al, Forecasting Of Egg Prices In Telangana Using R-Software, Indian Journal Of Applied Research, Volume -10 | Issue - 3 | March - 2020 | PRINT ISSN No. 2249 - 555X | DOI: 10.36106/ijar.
5. D Srinivas (2022) et al, Forecasting of SBI Stock Prices in India using ARIMA models, Juni Khyat Journal, ISSN:2278-4632, Volume -12| | Issue - 08 | August – 2022.
6. Anderson, B. and Ledolder, J.: Statistical Methods for Forecasting, John Wiley & Sons, New York, 1983
7. Bollerslev, T(1986). Generalized autoregressive conditional heteroskedasticity, Journal of Econometrics 31(3)309-317.
8. Srinivas Dande, Dr. M. Udayshankar, Dr. M. Raghavender Sharma [2022].: Forecast of SBI stock prices in India Using ARIMA models, JUNI KHYAT JOURNAL ISSN: 2278-4632.
9. Banhi Guha and Gautam Bandyopadhyay, Gold Price Forecasting using Arima model, Journal of Advanced Management Science, Vol 4, No.2, March 2016.
10. Raghavender, M. (2009), Forecasting paddy yield in Andhra Pradesh using seasonal Time Series Model, Bulletin of Pure and applied sciences. Jan. 2019.
11. Hornik, 1993, Some New result on neural network approximation, Neural Network, 6(1993), pp. 1069-1072.
12. Box, G. E. P. and Pierce, D (1970).: Distribution of Residual Autocorrelations in Autoregressive Integrated Moving Average Time Series Models, Journal of the American Statistical Association 65 1509-1526.

13. M.A. Khan, “Forecasting of Gold Prices (Box Jenkins Approach) (2013)”. ‘International Journal of Emerging Technology and Advanced Engineering, Vol.3, PP.662-67.
14. Raghavender, M. (2010), Forecasting paddy production in Andhra Pradesh with ARIMA model, Int. J. Agricult. Stat. Sci., Vol 6 (1), pp 251-258.
15. Wan Zakiyatussariroh Wan Husin, Ratna Zuarni Ramli, Ainaa Nabilah Muzaffar, Nur Fadhilah AbdNasir And Siti Nura Effa Rahmat (2020). Trend Analysis and Forecasting Models For Under-Five Mortality Rate In Malaysia, PJAEE, 17 (10).