

Statistical Performance Analysis of Hybrid Neural Network system to Predict Rice Production in Tamil Nadu, India

Arun Balaji.S^{#1} and Manimegalai Vairavan.P^{*2}

[#] PhD Scholar, Dept. Of Computer Science & Engineering, Karpagam University, Coimbatore, India

¹arunbalaji1983@yahoo.co.in

^{*} Professor, Dept. of Electronics and Communication Engineering, Karpagam University, Coimbatore, India

²manimegalai.vairavan@gmail.com

Abstract—This paper deals with the development of Hybrid Feed Forward Back Propagation Neural Network (FFBPNN) system to predict seasonal rice production in all the districts of Tamilnadu by incorporating the best fitting models using Curve Expert software. The predicted data was statistically compared with the observed data and analysed using Absolute Relative Error (ARE) and paired t test. It was found that 92,100 and 84.6 percent districts perform less than 50% ARE for the Kuruvai, Samba and Kodai seasons respectively for the area of rice cultivation data. It was also found that 96, 96 and 73.1 percent districts perform less than 50% ARE for the Kuruvai, Samba and Kodai seasons respectively for the rice production data. It was found that totally there are 42 paired t-test conducted between the FFBPNN system and Hybrid FFBPNN system at the rate of 21 paired t test for area of rice cultivation and another 21 paired t test for the rice production data. It was found that the entire 42 paired t test statistically convey the meaning that there is no significant difference between the observed data and predicted data.

Keyword - FFBPNN, Prediction, rice production, Tamilnadu

I. INTRODUCTION

Artificial Neural Network (ANN) can be defined as a model of reasoning based on the human brain. The brain consists of a densely interconnected set of nerve cells, or basic information-processing units, called neurons. The human brain incorporates nearly 10 billion neurons and 60 trillion connections, synapses, between them. Shepherd, G.M. and Koch, C. (1990) stated that by using multiple neurons simultaneously, the brain can perform its functions much faster than the fastest computers in existence today. Makinde, F.A. et al. (2012) stated that Feed Forward back propagation Neural Network (FFBPNN) resemble the human brain in acquiring knowledge through learning process and in storing knowledge in interneuron connection strength.

An ANN is configured for a specific application, such as pattern recognition or data classification, through a learning process. Learning in biological systems involves adjustments to the synaptic connections that exist between the neurons. This is true of ANNs as well. Neural networks are clusters of neurons that are interconnected to process information. Predicting the availability of rice from the three seasons of all the districts is a major input to take right decision by the Government planners to import rice or export rice. The planners normally used statistical and mathematical models, which leads to varied levels of success. The statistical and mathematical modelling needs input data from weather stations, field level data collected by agriculture departments from the farming community by conducting surveys.

Data collection for prediction needs more man power, time and financial resources. Proper data codification, data entry, data processing and information generation from the prediction models are also a cumbersome and time taking process. This leads the prediction system gives its output to the planners at a very late stage some time even after the harvesting period. Untimely prediction has no use to the planners to take right decisions. Hence, there is a need of a prediction system based on the previous year's data, which are readily available with planners. There is also a need to use modern artificial intelligence systems to be developed to quicken the supply of accurate and timely information to the decision makers. This issue made the researcher to take up the research on the Development of Hybrid Feed Forward Back Propagation Neural Network (FFBPNN) with best fitting models to predict seasonal rice production in Tamilnadu. The specific objectives of this paper are:

- To develop a hybrid FFBPNN system with best fitting models to predict the rice production during three seasons in different districts of Tamilnadu
- To carry out the statistical testing of finding Absolute Relative Error (ARE) between the predicted and observed values of area of rice cultivation and rice productions
- To analyse the statistical significance between the predicted and observed values of area of rice cultivation and rice productions using paired t test.

II. REVIEW OF LITERATURE

Laxmi et al. (2011) worked on Neural Networks for crop yields forecasting using multilayered neural network system with different learning algorithms at Uttar Pradesh. They considered crop productivity, maximum and minimum temperature, relative humidity and rainfall as input variables. They used stepwise regression techniques significant variables for selecting significant variables. They concluded that ANN models produced better results than statistical model.

Arun Balaji et al. (2013) stated that prediction of annual rice production in all the 31 districts of Tamilnadu using FFBPNN is an important decision for the Government of Tamilnadu. The predicted results were found to be exactly matching with the observed data.

Arun Balaji et al. (2014) stated the development of Multiple Linear Regression (MLR) equations between the years of rice cultivation and observed area of rice cultivation / rice production for different districts pertaining to Kuruvai, Samba and Kodai seasons in Tamilnadu.

Arun Balaji et al. (2015) stated that to get high accuracy of prediction, the curve expert software was integrated into the FFBPNN software. The curve fitting software developed the best fitting models among the 30 different linear and non linear models for Kuruvai, Samba and Kodai seasons of different districts of Tamilnadu. The curve expert produced the best fitting model to different districts during the three seasons. These developed models were used to simulate the best predicted area of rice cultivation and rice production. The type of fits are found to be 1) Quadratic Fit, 2) Linear Fit 3) User-Defined Model 4) Saturation Growth-Rate Model 5) Logarithm Fit 6) Hyperbolic Fit 7) Exponential Fit and 8) Gaussian Model.

III. METHODS ADOPTED

3.1 Data Collection

Tamilnadu is cultivating rice in three seasons namely 1) Kuruvai season (Jun.- Oct.), 2) Samba season (Aug.- Jan.) and 3) Kodai season (Dec.- Apr.). Rice is grown in all 31 districts of the state excepting Chennai. Data pertaining to the district wise area of cultivation in hectare and rice production in tonnes were collected for all the three seasons. Data were collected from the Seasons and Crop reports of the Dept. of Economics and Statistics, Government of Tamilnadu for seven years from 2005-06 to 2012-13 which are posted in the internet for different users to use it freely. The data for the year 2010-11 is not made available for the user in the net. Hence, the year 2010-11 is not included in the study. The data for 2013-14 and 2014-15 were yet posted in the net and hence it is not used in the study.

3.2 Steps adopted in modelling process

The primary objective of this study is to create a hybrid FFBPNN system and implement it using a program developed using Turbo C Programming language to predict the area of rice cultivation and rice production in different districts of Tamilnadu. The methodology adopted is based on the work of Makinde, F.A et al. (2012). There are generally five steps in the modelling process: 1) Assembling the training data and convert the data to a range of 0 to 1 using activation function, 2) Creating the FFBPNN network without best fitting models and implementation of training and testing of the system with the data collected 3) Creating the hybrid FFBPNN network with integration of Curve Expert Software 4) Testing the Hybrid FFBPNN system with best fitting models for various input, and 5) Statistical testing of the network response through Absolute Relative Error (ARE) and paired t test.

3.3 Selection of Variables

The two most important variables identified are 1) the area of cultivation of rice and 2) rice production. There are three seasons like Kuruvai, Samba, and Kodai for rice cultivation. Each season contain 2 variables viz: area of rice and rice production. Hence the total variables selected were 2 variables multiplied by 3 seasons gives 6 variables.

3.4 Data preprocessing

The training and test data collected had some missing data marked with 0 values. If the study uses these 0 values in computation, this 0 may divide any real number leading to infinity condition. Computer cannot compute such infinity conditions. There are many data cleaning techniques available for data pre-processing. The present study adopted the use of a global constant 0.01 in place of 0 in training and testing data before processing is done. This replacement avoids the computational problem of avoiding infinity during computations.

3.5 Assembling training data, testing data and initial weights

The training data consist of the area (ha) and rice production (tonnes) for three seasons of Kuruvai, Samba and Kodai for the year 2009-10 pertaining to 31 districts. The test data comprises of the 6 years data from 2005-06 to 2008-09, 2011-12 and 2012-13 containing the area (ha) and rice production (tonnes) for three seasons of Kuruvai, Samba and Kodai seasons. The initial weights of 42 data within the range of 0 to 1 were assumed. The training data along with initial weights were stored in a input file. The training data sets were used to train the FFBPNN model. There are 31 districts; each district consists of 7 columns of data items like name of the district, three area of cultivation data and three rice production data for Kuruvai, Samba and Kodai seasons. It causes 217 data items. The initially assumed weights in the data file are 42. Hence, the total data item for training is 259. Six years testing data were stored in six different data files for processing by a computer program developed in the study.

3.6 Normalization of input data using sigmoid activation

The training data were normalized using sigmoid activation equation before being presented to the network for training. This step was taken to ensure that input data with different ranges were transformed into one similar range of 0 to 1 and allows for easier and faster model training. These training data are converted into sigmoid data between 0 to 1 using a sigmoid activation function and the corresponding converted data between 0 to 1 are represented as $x_1(i)$, $x_2(i)$ $x_6(i)$. The general format of the sigmoid action functional equation is given below:

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

Where

$S(x)$ is the sigmoid value. It varies from 0 to 1.

x is the independent input values like Area in Hectare and Rice Production in tonnes.

3.7 Creating the FFBPNN system

The multi layered FFBPNN system in this study begins with an input layer. The input layer is connected to a hidden layer; the hidden layer is then connected to output layer. In this study, the architecture used for the neural network consists of one input layer, one hidden layer and lastly, the output layer. The input layer converts the area of rice cultivation in hectare and the rice production in tonnes for three seasons into sigmoid values between 0 to 1 using the sigmoid activation function. The input layer is made with six neurons which are area of rice for Kuruvai, rice production for Kuruvai, area of rice for Samba, rice production for Samba, area of rice for Kodai and rice production for Kodai seasons. The output layer has only one neuron containing the predicted value from 0 to 1. The content of the output neuron will be multiplied with the six input observed values to get the targeted (predicted) values.

3.8 Development of the Hybrid FFBPNN system by incorporating the best fitting models

The hybrid FFBPNN system is created by incorporating the best fitted models developed using the Curve Expert Software into the FFBPNN system. Fig. 1 shows the hybrid FFBPNN system incorporated with Curve Expert system. The predicted output from the FFBPNN system is given as input into the Curve Expert software so as to develop the best fitting linear and nonlinear model among the 30 built in models based on the coefficient of correlation. Predictions from the Hybrid FFBPNN system were compared with the observed data also. This makes the system not only a FFBPNN system but also a Hybrid FFBPNN system incorporated with best fitting models developed from Curve Expert software for more realistic predictions. ARE was worked out between predicted data from the hybrid FFBPNN system and the observed data and then analysed.

3.9 Prediction of data from the Hybrid FFBPNN system

The predicted area of rice cultivation is done by inserting the years $x = (2005, 2006, 2007, 2008, 2009, 2011$ and $2012)$ into the best fitting equation for area of cultivation for a particular district and particular season. Similar procedure is adopted for finding the rice production for a particular district and particular season. These predicted data are called the best predicted data from the hybrid FFBPNN system as it is built with best fitting models. Further predictions from the Hybrid FFBPNN system were compared with the observed data also. The predicted area of cultivation from the Hybrid FFBPNN system with best fitted models will be examined. Some statistics like minimum, maximum mean and standard deviations of the predicted area of rice cultivation for different years during three seasons will be tabulated and discussed. Similarly, the predicted rice production from the hybrid FFBPNN system with best fitted models will be examined. Some statistics like minimum, maximum mean and standard deviations of the predicted rice production for different years during three seasons will be tabulated and discussed.

3.10 ARE between Observed and Hybrid FFBPNN systems of predicted data

The Absolute Relative Error is computed between the observed data and the Hybrid FFBPNN predicted data using the formula given below:

$$ARE \% = \frac{(Observed\ Data - Hybrid\ FFBPNN\ Predicted\ Data)}{Observed\ Data} \times 100$$

ARE percent will be computed for the area of cultivation of observed data and the Hybrid FFBPNN system for the Kuruvai, Samba and Kodai season for the dataset from 2005 to 2012 for all the districts. Average ARE % for all the districts in three seasons will be sorted in ascending order for both area of rice cultivation and rice production in separate tables. Some statistics like which districts has the lowest mean ARE% worked out from the years 2005 to 2012 in a particular seasons and which district has the highest ARE% in the same season will be pointed out and discussed. Information will be provided about how many percent of the districts are having < 50% ARE and the name of such districts will be arranged in ascending order of ARE. Similar information will also be made available for the districts having > 50% ARE.

3.11 Testing the statistical significance using paired t-test

The paired t-test will be used to test the significance between two sets of paired data. The first data field is the observed data and the second data field is the predicted data from Hybrid FFBPNN system. It will be done for both area of rice cultivation and rice production for every year for all the three seasons. The two data items for a district in a year form a pair. Such pairs of data items for all the districts cultivating rice form a set to be used for t test. The same procedure is used to compute the t value between the pairs of observed data and the Hybrid FFBPNN predicted data for the three seasons for both the area of rice cultivation and rice production. The number of observations (N) is the number of districts cultivating rice in a particular year in a particular season. The degree of freedom is N-1. The table t value for the degrees of freedom of (N-1) at 5% level of significance was taken up from the statistical t table. The calculation of t value between the pair of data consisting of the observed area of rice cultivation and Hybrid FFBPNN predicted data was done using MS-Excel for the three seasons. An example of how to carryout t test in MS Excel is given below:

EXCEL→INSERT→FUNCTION→STATISTIC→=TTEST(Array 1, Array 2, tails, type)

Where

TTEST - is the testing the paired t test

Array1 - is the first data set

Array2 - is the second data set

Tails - specifies the number of distribution tails to return: One tailed distribution =1. Two tailed distribution =2

Type-It is the kind of t test: Paired t test=1, Two sample equal variance = 2. Two sample unequal variance = 3

If the calculated t value is less than the table t value then there is no significant difference (NS) between the two samples. If the calculated t value is greater than the table t value then there is significant difference (S) between the two samples. This procedure is repeated for different season of area of rice cultivation and rice production for all the years of data collection. It is estimated that 42 t tests will be carried out.

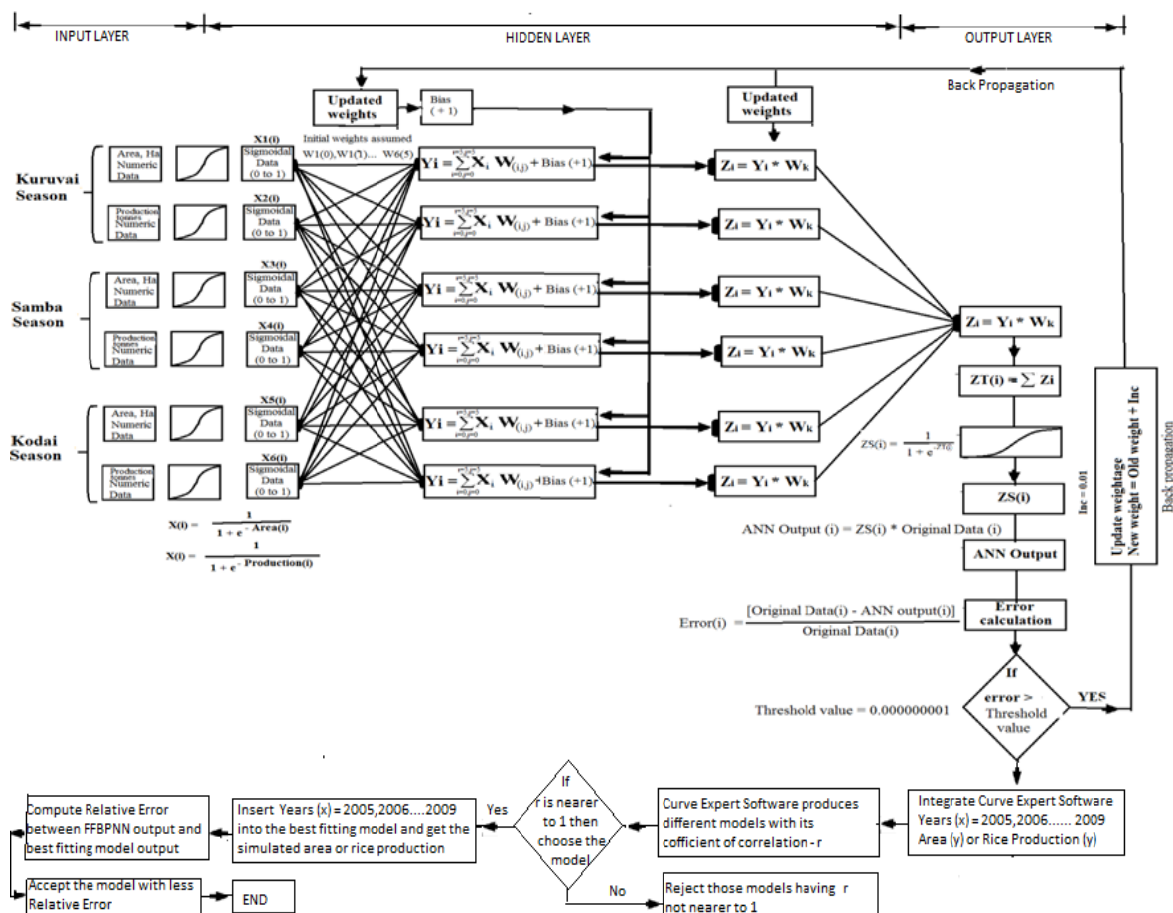


Fig. 1. Hybrid FFBPNN system with incorporation of best fitted models from Curve Expert Software

IV. RESULTS AND DISCUSSIONS

The hybrid FFBPNN is developed by connecting the output from the FFBPNN system with the best fitted models of area of rice cultivation and rice production for all the districts in three seasons. The output from the FFBPNN system was found out exactly same as observed data because FFBPNN system is 100% error free while comparing the observed and predicted data. The predicted output from the FFBPNN system is treated as input into the Hybrid FFBPNN system. In the best fitted models developed both for area of rice cultivation and rice production for different districts during the three seasons, the best predicted area of rice cultivation in hectare and rice production in tonnes were obtained by inserting $x = (2005, 2006, 2007, 2008, 2009, 2011 \text{ and } 2012)$ in the concerned models. The predicted data from the Hybrid FFBPNN system was analysed. Absolute Relative Error (ARE) was worked out between observed data and the predicted data from the Hybrid FFBPNN system and analysed. In addition, the testing of the statistical significance between the observed data and the predicted data from the Hybrid FFBPNN system was carried out by using paired t-test.

4.1 Prediction of the area of rice cultivation from the hybrid FFBPNN system

The predicted area of cultivation from hybrid FFBPNN prediction system is analysed. Some statistics of the predicted area of cultivation using the hybrid FFBPNN system is carried out for three seasons and are given in Table I. Table I brings out the fact that for the Kuruvai season, the minimum range of predicted area of rice cultivation varies from 154 ha to 1237 ha. It was also found out that the maximum predicted area varies from 30453 ha to 40730 ha. The average predicted area varies from 9946.04 ha to 13025.76 ha. Similarly, the standard deviation varies from 9581.74 ha to 10784.87 ha. It was found that the predicted area of rice cultivation gradually reduces from 2005 to 2012 due to non-availability of water in time. With regard to the Samba season, the minimum range of predicted area of rice cultivation varies from 1317 ha to 4889 ha. The maximum predicted area varies from 142734 ha to 142497 ha. The average predicted area varies from 44553.14 ha to 53194.61 ha. Similarly, the standard deviation varies from 43950.74 ha to 44477.83 ha.

Table I. Statistics of predicted area of cultivation based on hybrid FFBPNN system

| Parameter | Statistics of predicted area of cultivation based on the hybrid FFBPNN with best fitted models, Hectare | | | | | | |
|----------------|---|----------|----------|----------|----------|----------|----------|
| Year | 2005 | 2006 | 2007 | 2008 | 2009 | 2011 | 2012 |
| | Kuruvai season | | | | | | |
| Minimum | 1237 | 1101 | 800 | 627 | 560 | 308 | 154 |
| Maximum | 40730 | 41302 | 40936 | 39657 | 37550 | 36234 | 30453 |
| Mean | 13025.76 | 13739.48 | 12707.24 | 12626.52 | 12352.56 | 11329.24 | 9946.04 |
| Std deviation | 10784.87 | 11548.05 | 10539.62 | 10689.26 | 10531.85 | 10613.11 | 9581.74 |
| | Samba season | | | | | | |
| Minimum | 4889 | 3521 | 2752 | 2259 | 1919 | 1470 | 1317 |
| Maximum | 142497 | 141304 | 141750 | 143511 | 145304 | 144693 | 142734 |
| Mean | 53194.61 | 52804.14 | 51062.14 | 50220.21 | 49689.36 | 47438.86 | 44553.14 |
| Std. deviation | 43950.74 | 44763.45 | 44575.47 | 45146.00 | 45866.20 | 46082.05 | 44477.83 |
| | Kodai season | | | | | | |
| Minimum | 0 | 0 | -4 | 1 | -101 | -113 | 1 |
| Maximum | 34764 | 33971 | 33215 | 32492 | 31801 | 30505 | 29896 |
| Mean | 7120.39 | 5937.31 | 5244.54 | 4921.31 | 4696.23 | 4099.08 | 3719.00 |
| Std. deviation | 9328.28 | 8171.73 | 7490.50 | 6990.00 | 6658.67 | 6356.18 | 6412.70 |

It was found that for the Kodai season that the minimum range of predicted area of rice cultivation varies from -113 ha for Coimbatore district in 2011 to 1 ha for Karur district in 2008 and Thoothukudi district in 2012. The negative value of -113 ha cannot be acceptable prediction. The reason for the negative prediction of area is mainly due to wide variations of observed area of cultivation, which causes the regression equation not true representative to provide the correct prediction. This type of error can be avoided only when more input data are available to build the best fitted equations. It was also found that the negative and zero prediction of area was found only for Coimbatore and Nagapattinam districts. The prediction worked well for other districts. The maximum predicted area varies from 29896 ha to 34764 ha. The average predicted area varies from 3719 ha to 7120.39ha and the standard deviation varies from 6412.701ha to 9328.28 ha.

It was found that there is some difference of predicted area of rice cultivation between the observed data and the hybrid FFBPNN system. The predicted area of rice cultivation from the hybrid FFBPNN system with best fit model is having differences from the observed area of rice cultivation. This difference is a desired one because area of rice cultivation in a district varies every year according to availability of water resources. But, it is to note that for Kodai season of Coimbatore, Karur and Nagapattinam districts, the prediction of area of rice cultivation is a failure. It can be well studied if more number of years of consistent input data collection are done. For the Kuruvai and Samba season, the prediction worked well for all districts in three seasons.

4.2 Prediction of rice production from the hybrid FFBPNN system

The predicted rice production from hybrid FFBPNN prediction system was statistically analysed. Some statistics of the predicted rice production is obtained for three seasons and are given in Table II.

Table II. Statistics of predicted rice production based on hybrid FFBPNN system

| Parameter | Statistics of predicted rice production based on the hybrid FFBPNN system with best fitted models, tonnes | | | | | | |
|----------------|---|----------|----------|----------|----------|----------|----------|
| Year | 2005 | 2006 | 2007 | 2008 | 2009 | 2011 | 2012 |
| | Kuruvai season | | | | | | |
| Minimum | 4041 | 3328 | 2731 | 2465 | 2531 | 209 | 30 |
| Maximum | 136980 | 145052 | 148706 | 147597 | 141830 | 118843 | 117462 |
| Mean | 43498.7 | 45079.9 | 47418.4 | 47919.1 | 48187.0 | 43706.2 | 39898.6 |
| Std. deviation | 34587.3 | 35978.2 | 38005.7 | 38359.1 | 38385.8 | 36932.8 | 36737.6 |
| | Samba season | | | | | | |
| Minimum | 12620 | 15149 | 15426 | 13325 | 9764 | 3201 | 1432 |
| Maximum | 336477 | 434262 | 360200 | 324654 | 503468 | 548573 | 330015 |
| Mean | 121463.1 | 158201.7 | 128004.5 | 128168.7 | 169082.9 | 175887.8 | 120036.0 |
| Std. deviation | 90526.5 | 131039.5 | 95155.7 | 89403.9 | 141271.3 | 156125.9 | 100109.7 |
| | Kodai season | | | | | | |
| Minimum | 0 | 0 | 0 | 2 | 505 | 122 | -606 |
| Maximum | 108389 | 107391 | 106392 | 105393 | 104395 | 102397 | 101399 |
| Mean | 20297.2 | 18156.7 | 16333.7 | 16015.9 | 16195.5 | 15455.6 | 14324.5 |
| Std. deviation | 28597.6 | 26457.7 | 24761.6 | 23451.1 | 22643.0 | 22870.2 | 24276.7 |

Table II brings out the fact for the Kuruvai season the minimum range of predicted rice production varies from 30 tonnes in Perambalur district for 2012 to 4041 tonnes in Coimbatore district for 2005. It was also found out that the maximum predicted rice production varies from 148706 tonnes to 117462 tonnes. The average predicted rice production varies from 39898.6 tonnes to 48187.0 tonnes. Similarly, the standard deviation varies from 34587.3tonnes to 38385.8 tonnes.

With regard to the Samba season, the minimum predicted rice production varies from 1432 tonnes to 15426 tonnes. The maximum predicted rice production varies from 324654 tonnes to 548573 tonnes. The average predicted rice production varies from 120036.0 tonnes to 175887.8 tonnes. Similarly, the standard deviation varies from 89403.9 tonnes to 156125.9 tonnes.

With regard to the Kodai season, the minimum predicted rice production varies from -606 tonnes in 2012 for Coimbatore district to 505 tonnes in Namakkal district during 2009. The negative value of -606 tonnes in Coimbatore district for the year 2012 cannot be acceptable prediction. The reason for the negative prediction of rice production is mainly due to wide variations of observed rice production data during Kodai season, which causes the regression equation not true representative to provide the correct prediction. This type of error can be avoided only when more input data are available to build the best fitted equations. It was also found that the negative and zero prediction of rice production were found only for Coimbatore and Karur districts during Kodai seasons only. The prediction worked well for other districts. The maximum predicted rice production varies from 101399 tonnes to 108389 tonnes. The average predicted rice production varies from 14324.5 tonnes to 20297.2 tonnes. Similarly, the standard deviation varies from 22643.0 tonnes to 28597.6 tonnes.

4.3 Absolute Relative Error (ARE) Percent

ARE percent was calculated between the observed data and the hybrid FFBPNN system for predicted area of rice cultivation and rice production for all the districts in three seasons. Both ARE for area of rice cultivation data and rice production data are discussed. The following sections show the districts in ascending order of ARE percent. The districts having low ARE % has better prediction and the districts having higher ARE having poor matching of observed and predicted data.

4.3.1 Average ARE percent for the Kuruvai season

ARE percent is the ratio of absolute difference between the observed data and the predicted data from Hybrid FFBPNN system to the observed data expressed in %. It was worked out for the three seasons separately and an extract and sorted order of relevant data for Kuruvai season alone is shown in Table III for better discussion.

Tables III. Average ARE percent for Kuruvai season from 2005 to 2012

| Kuruvai season | | | | | |
|-----------------------------|-------------------------------------|-----------------------------|-----------------|-------------------------------------|--------------------------------|
| Area of cultivation of rice | | | Rice production | | |
| Rank | Districts in ascending order of ARE | Mean ARE % for area of rice | Rank | Districts in ascending order of ARE | Mean ARE % for rice production |
| 1 | Thiruvallur | 2.7 | 1 | Thiruvannamalai | 4.7 |
| 2 | Coimbatore | 4.7 | 2 | Thiruvallur | 6.3 |
| 3 | Kanyakumari | 5.0 | 3 | Krishnagiri | 7.4 |
| 4 | Kancheepuram | 5.2 | 4 | Kancheepuram | 7.6 |
| 5 | Villupuram | 5.4 | 5 | Vellore | 8.6 |
| 6 | Theni | 6.5 | 6 | Theni | 8.8 |
| 7 | Cuddalore | 7.6 | 7 | Kanyakumari | 8.9 |
| 8 | Vellore | 7.7 | 8 | Erode | 9.3 |
| 9 | Krishnagiri | 7.9 | 9 | Coimbatore | 10.0 |
| 10 | Thiruvannamalai | 8.5 | 10 | Tiruchirapalli | 10.0 |
| 11 | The-Nilgiris | 9.0 | 11 | Cuddalore | 10.7 |
| 12 | Tiruchirapalli | 9.6 | 12 | Villupuram | 11.8 |
| 13 | Tirunelveli | 10.5 | 13 | The-Nilgiris | 12.8 |
| 14 | Erode | 10.7 | 14 | Dharmapuri | 15.7 |
| 15 | Pudukottai | 15.5 | 15 | Tirunelveli | 15.9 |
| 16 | Thoothukudi | 15.7 | 16 | Thoothukudi | 18.4 |
| 17 | Salem | 17.7 | 17 | Nagapattinam | 21.4 |
| 18 | Dharmapuri | 17.7 | 18 | Salem | 22.4 |
| 19 | Nagapattinam | 17.9 | 19 | Pudukottai | 22.6 |
| 20 | Dindigul | 20.3 | 20 | Dindigul | 25.8 |
| 21 | Namakkal | 22.8 | 21 | Namakkal | 28.9 |
| 22 | Thanjavur | 25.2 | 22 | Thanjavur | 30.0 |
| 23 | Madurai | 27.5 | 23 | Perambalur | 32.4 |
| 24 | Perambalur | 57.5 | 24 | Madurai | 36.9 |
| 25 | Thiruvarur | 66.3 | 25 | Thiruvarur | 73.1 |

With regard to the area of cultivation of rice in Kuruvai season, Table III reveals the fact that the lowest mean ARE of 2.7 % is noted for Thiruvallur district followed by 4.7 for Coimbatore district. It is also found that the highest ARE of 66.3% is noted for Thiruvarur district. The interesting feature of the study shows that 23 districts out of 25 districts, whose area of cultivation of rice in Kuruvai season has less than 50% ARE. It works out to be 92% of the districts perform better prediction for area of rice cultivation with less than 50% ARE.

With regard to the rice production in Kuruvai season,, Table III reveals the fact that the lowest ARE of 4.7 % is noted for Thiruvannamalai district followed by 6.3 for Thiruvallur district. It is also found that the highest ARE of 73.1% is the Thiruvarur district. The striking feature of the study shows 24 districts out of 25 districts, whose rice production in Kuruvai season has less than 50% ARE. It works out to be 96% of the districts perform better prediction for rice production with less than 50% ARE.

4.3.2 Average ARE percent for the Samba season

ARE percent is computed for the area of cultivation of observed data and the Hybrid FFBPNN system for the Samba season for the dataset from 2005 to 2012. It was worked out for the three seasons separately and an extract and sorted order of relevant data for Samba season alone is shown in Table IV for clear understanding.

Table IV. Average ARE percent for Samba season from 2005 to 2012

| Samba season | | | | | |
|-----------------------------|-------------------------------------|-----------------------------|-----------------|-------------------------------------|--------------------------------|
| Area of cultivation of rice | | | Rice production | | |
| Rank | Districts in ascending order of ARE | Mean ARE % for area of rice | Rank | Districts in ascending order of ARE | Mean ARE % for rice production |
| 1 | Thiruvarur | 0.4 | 1 | Villupuram | 2.6 |
| 2 | Cuddalore | 1.1 | 2 | Vellore | 5.7 |
| 3 | Nagapattinam | 1.2 | 3 | Krishnagiri | 6.0 |
| 4 | Ramanathapuram | 2.7 | 4 | Karur | 6.0 |
| 5 | Thanjavur | 2.9 | 5 | Kanyakumari | 7.3 |
| 6 | Kancheepuram | 3.1 | 6 | Tirunelveli | 7.6 |
| 7 | Villupuram | 3.1 | 7 | Kancheepuram | 7.8 |
| 8 | Pudukottai | 3.2 | 8 | Tiruchirapalli | 8.3 |
| 9 | Kanyakumari | 3.2 | 9 | Thiruvallur | 9.2 |
| 10 | Sivagangai | 4.6 | 10 | Thiruvannamalai | 9.3 |
| 11 | Theni | 4.9 | 11 | Theni | 9.4 |
| 12 | Thiruvallur | 5.2 | 12 | Pudukottai | 17.9 |
| 13 | Vellore | 6.0 | 13 | Sivagangai | 18.9 |
| 14 | Tirunelveli | 6.3 | 14 | Dindigul | 19.7 |
| 15 | Tiruchirapalli | 6.9 | 15 | Dharmapuri | 20.0 |
| 16 | Krishnagiri | 7.6 | 16 | Namakkal | 20.4 |
| 17 | Virudhunagar | 8.0 | 17 | Thanjavur | 23.9 |
| 18 | Thiruvannamalai | 10.3 | 18 | Thoothukudi | 25.3 |
| 19 | Karur | 12.5 | 19 | Salem | 26.0 |
| 20 | Dindigul | 13.8 | 20 | Perambalur | 29.4 |
| 21 | Thoothukudi | 18.3 | 21 | Coimbatore | 30.2 |
| 22 | Namakkal | 18.4 | 22 | Virudhunagar | 32.7 |
| 23 | Salem | 20.7 | 23 | Nagapattinam | 33.3 |
| 24 | Dharmapuri | 20.8 | 24 | Madurai | 33.4 |
| 25 | Madurai | 22.7 | 25 | Cuddalore | 40.8 |
| 26 | Perambalur | 32.3 | 26 | Thiruvarur | 47.2 |
| 27 | Coimbatore | 37.0 | 27 | Erode | 48.6 |
| 28 | Erode | 46.0 | 28 | Ramanathapuram | 68.8 |

With regard to the area of cultivation of rice in Samba season, Table IV reveals the fact that the lowest mean ARE of 0.4% is noted for Thiruvarur district followed by 1.7% for Cuddalore district. It is also found that the highest ARE of 46.0% is noted for Erode district. The interesting feature of the study shows that all the 28 districts cultivating rice in Samba season has their area of cultivation of rice has less than 50% ARE. It works out to be 100% of the districts perform better prediction for area of rice cultivation with less than 50% ARE.

With regard to the rice production in Samba season,, Table IV reveals the fact that the lowest ARE of 2.6 % is noted for Villupuram district followed by 5.7 for Vellore district . It is also found that the highest ARE of 68.8% is reported for the Ramanathapuram district. The striking feature of the study shows 27 districts out of 28 districts producing rice in Samba season has less than 50% ARE. It works out to be 96% of the districts perform better prediction for area of rice cultivation with less than 50% ARE.

4.3.3 Average ARE percent for the Kodai season

ARE percent is computed for the area of cultivation of observed data and the Hybrid FFBPNN system for the Kodai season for the dataset from 2005 to 2012. It was worked out for the three seasons separately and an extract and sorted order of relevant data for Kodai season alone is shown in Table V for clear understanding.

With regard to the area of cultivation of rice in Kodai season, Table V reveals the fact that the lowest mean ARE of 6.2% is noted for Pudukottai district followed by 7.4% for Thiruvallur district. It is also found that the highest ARE of 48.7% is noted for Tiruchirapalli district at the 22nd rank of performance. The striking feature of the study shows 22 districts out of 26 districts producing rice in Kodai season has less than 50% ARE. It works out to be 84.6% of the districts perform better prediction for area of rice cultivation with less than 50% ARE.

Another noted issue in this study is four districts namely Karur, Namakkal, Coimbatore and Krishnagiri have failed to predict the area of rice cultivation because of their data distribution is erratic without any type of distributions the curve expert tried. If more years of input data are collected then their performance will be amenable for a scientific study.

Table V. Average ARE percent for Kodai season from 2005 to 2012

| Kodai season | | | | | |
|-----------------------------|-------------------------------------|-----------------------------|-----------------|-------------------------------------|--------------------------------|
| Area of cultivation of rice | | | Rice production | | |
| Rank | Districts in ascending order of ARE | Mean ARE % for area of rice | Rank | Districts in ascending order of ARE | Mean ARE % for rice production |
| 1 | Pudukottai | 6.2 | 1 | Thiruvallur | 8.2 |
| 2 | Thiruvallur | 7.4 | 2 | Kancheepuram | 9.7 |
| 3 | Cuddalore | 9.1 | 3 | Thanjavur | 10.5 |
| 4 | Thanjavur | 12.6 | 4 | Villupuram | 11.1 |
| 5 | Villupuram | 12.9 | 5 | Thiruvannamalai | 11.6 |
| 6 | Thiruvannamalai | 14.8 | 6 | Cuddalore | 13.0 |
| 7 | Virudhunagar | 15.5 | 7 | Krishnagiri | 13.1 |
| 8 | Vellore | 16.3 | 8 | Pudukottai | 13.3 |
| 9 | Kancheepuram | 19.0 | 9 | Vellore | 16.1 |
| 10 | Dindigul | 20.7 | 10 | Nagapattinam | 17.6 |
| 11 | Ramanathapuram | 23.1 | 11 | Virudhunagar | 19.3 |
| 12 | Thiruvarur | 30.3 | 12 | Ramanathapuram | 22.7 |
| 13 | Salem | 33.2 | 13 | Thiruvarur | 24.2 |
| 14 | Nagapattinam | 34.6 | 14 | Salem | 27.1 |
| 15 | Madurai | 39.4 | 15 | Dindigul | 30.3 |
| 16 | Theni | 40.6 | 16 | Madurai | 34.5 |
| 17 | Perambalur | 42.7 | 17 | Theni | 38.3 |
| 18 | Tirunelveli | 43.5 | 18 | Perambalur | 43.4 |
| 19 | Thoothukudi | 44.8 | 19 | Dharmapuri | 47.3 |
| 20 | Erode | 46.2 | 20 | Tiruchirapalli | 51.2 |
| 21 | Dharmapuri | 48.3 | 21 | Thoothukudi | 52.9 |
| 22 | Tiruchirapalli | 48.7 | 22 | Tirunelveli | 62.4 |
| 23 | Karur | 100.0 | 23 | Erode | 68.1 |
| 24 | Namakkal | 101.5 | 24 | Karur | 88.7 |
| 25 | Coimbatore | 259.9 | 25 | Namakkal | 112.1 |
| 26 | Krishnagiri | 422.8 | 26 | Coimbatore | 236.8 |

With regard to the rice production in Kodai season is concerned, Table V reveals the fact that the lowest ARE of 8.2% is noted for Thiruvallur district followed by 9.7% for Kancheepuram district. It is also found that the highest ARE of 47.3% is reported for the Dharmapuri district. The striking feature of the study shows 19 districts out of 26 districts producing rice in Kodai season has less than 50% ARE. It works out to be 73.1% of the districts perform better prediction for area of rice cultivation with less than 50% ARE. Another noted issue in this study is six districts namely Tiruchirapalli, Thoothukud, Tirunelveli, Erode, Karur, Namakkal and Coimbatore have failed to predict the rice production properly because of their data distribution is erratic without any type of distributions the curve expert tried. If more years of input data are collected then their performance will be amenable for a scientific study.

4.4 Testing the statistical significance using paired t-test

The paired t-test is used to test the significance between two sets of paired data. The first data field is the observed data and the second data field is the predicted data from Hybrid FFBPNN system. The two data items for a district in a crop year form a pair. Such pairs of data items for all the districts cultivating rice form a set to be used for t test. The same procedure is used to compute the t value between the pairs of observed data and the Hybrid FFBPNN system for the three seasons.

4.4.1 Paired t test for the area of rice cultivation

The t-test was conducted between the observed data and the Hybrid FFBPNN systems for the area of cultivation of rice. The number of observations (N) is the number of districts cultivating rice in a season. The degree of freedom is N-1. The table t value for the degrees of freedom of (N-1) at 5% level of significance was taken up from the statistical t table. The calculation of t value between the pair of data consisting of the observed data and Hybrid FFBPNN systems was done using MS-Excel for the three seasons. If the calculated t value is less than the table t value then there is no significant difference (NS) between the two samples. This procedure is repeated for different season of area of rice cultivation like Kuruvai, Samba and Kodai seasons. The summary of the t test for area of rice cultivation is given in Table VI.

Table VI. Result of t test for the area of rice cultivation for three seasons

| Area of rice cultivation in Kuruvai Season | | | | | | | | |
|--|---|----------|----------|----------|----------|----------|----------|--------|
| No. of District cultivating rice (N) = 25. Degrees of Freedom (N-1) = 24 | | | | | | | | |
| Table t at 5% level | Calculated t value for area of rice cultivation for different years | | | | | | | Remark |
| | 2005 | 2006 | 2007 | 2008 | 2009 | 2011 | 2012 | |
| 2.064 | 0.00002 | 0.00003 | 0.00009 | 0.00010 | 0.00001 | 0.00062 | 0.00029 | NS |
| Area of rice cultivation in Samba season | | | | | | | | |
| No. of District cultivating rice (N) = 28. Degrees of Freedom (N-1) = 27 | | | | | | | | |
| Table t at 5% level | Calculated t value for area of rice cultivation for different years | | | | | | | Remark |
| | 2005 | 2006 | 2007 | 2008 | 2009 | 2011 | 2012 | |
| 2.052 | 0.000003 | 0.000016 | 0.000085 | 0.000025 | 0.000002 | 0.000005 | 0.000638 | NS |
| Area of rice cultivation in Kodai season | | | | | | | | |
| No. of District cultivating rice (N) = 26. Degrees of Freedom (N-1) = 25 | | | | | | | | |
| Table t at 5% level | Calculated t value for area of rice cultivation for different years | | | | | | | Remark |
| | 2005 | 2006 | 2007 | 2008 | 2009 | 2011 | 2012 | |
| 2.060 | 0.0003 | 0.0003 | 0.0022 | 0.0011 | 0.0002 | 0.0004 | 0.0034 | NS |
| Note: NS – No significant difference between the observed and predicted data | | | | | | | | |

Table VI shows the fact that the calculated t value is less than the table t value at 5% level of significance for all the years and for all the three seasons for the periods from 2005 to 2012, it means that there is 95% confidence level between the observed data and the Hybrid FFBPNN system of prediction for the area of rice cultivation. The researcher is of the opinion that the input data collection should be for more than 10 years for better modelling and simulations though the hybrid system of FFBPNN worked well in this research. It is found that there are 21 t-test conducted because of three seasons and seven years. It was found that all 21 tests are showing no significant difference between the observed and predicted area of rice cultivation. It straightly conveys the meaning that 100% of predictions are statistically showing no significant difference between the observed and predicted area of rice cultivation.

4.4.2 Paired t test for the rice production

The t-test was conducted between the observed data and the Hybrid FFBPNN systems for the rice production. The number of observations (N) is the number of districts involved in rice production in a season.

Table VII. Result of t test for the rice production for three seasons

| Rice production in Kuruvai Season | | | | | | | | |
|--|--|-------|-------|-------|-------|--------|-------|--------|
| No. of District cultivating rice (N) = 25. Degrees of Freedom (N-1) = 24 | | | | | | | | |
| Table t at 5% level | Calculated t value for rice production for different years | | | | | | | Remark |
| | 2005 | 2006 | 2007 | 2008 | 2009 | 2011 | 2012 | |
| 2.064 | 0.166 | 0.185 | 0.097 | 0.012 | 0.005 | 0.007 | 0.012 | NS |
| Rice production in Samba season | | | | | | | | |
| No. of District cultivating rice (N) = 28. Degrees of Freedom (N-1) = 27 | | | | | | | | |
| Table t at 5% level | Calculated t value for rice production for different years | | | | | | | Remark |
| | 2005 | 2006 | 2007 | 2008 | 2009 | 2011 | 2012 | |
| 2.052 | 0.043 | 0.010 | 0.244 | 0.199 | 0.002 | 0.0003 | 0.009 | NS |
| Rice production in Kodai season | | | | | | | | |
| No. of District cultivating rice (N) = 26. Degrees of Freedom (N-1) = 25 | | | | | | | | |
| Table t at 5% level | Calculated t value for rice production for different years | | | | | | | Remark |
| | 2005 | 2006 | 2007 | 2008 | 2009 | 2011 | 2012 | |
| 2.060 | 0.007 | 0.189 | 0.079 | 0.060 | 0.093 | 0.075 | 0.008 | NS |
| Note: NS – No significant difference between the observed and predicted data | | | | | | | | |

The degree of freedom is N-1. The table t value for the degrees of freedom of (N-1) at 5% level of significance was taken up from the statistical t table. The calculation of t value between the pair of data consisting of the predicted rice production (observed data) from the observed data and Hybrid FFBPNN systems was done using MS-Excel for the three seasons. If the calculated t value is less than the table t value then there is no significant difference (NS) between the two samples. This procedure is repeated for different season of rice production like Kuruvai, Samba and Kodai seasons. The summary of the t test for rice production is given in Table VII.

Table VII shows the fact that the calculated t value is less than the table t value at 5% level of significance for all the years and for all the three seasons for the periods from 2005 to 2012, it means that there is 95% confidence level between the observed data and the Hybrid FFBPNN system of prediction for the rice production. The researcher is of the opinion that the input data collection should be for more than 10 years for better modelling and simulations though the hybrid system of FFBPNN worked well in this research. It is found that there are 21 t-test conducted because of three seasons and seven years of data. It was found that all 21 tests are showing no significant difference between the observed and predicted rice production. It straightly conveys the meaning that 100% of predictions are statistically showing no significant difference between the observed and predicted rice production.

V. SUMMARY AND CONCLUSIONS

The development of the hybrid FFBPNN system was done by incorporating the best fitting models developed for different districts and different seasons. The statistics of the predicted values from the hybrid system was analysed and also the statistical testing between the observed and predicted data was carried out. The following are some of the unique conclusions drawn from the research.

1. It was found that there is some difference of predicted rice production between the observed data and the hybrid FFBPNN system. This difference is a desired one because rice production in a district varies every year according to weather, soil, manure, and other management practices. It was found that for the Kuruvai, Samba and Kodai seasons, the minimum range of predicted rice production varies from 30 tonnes to 4041 tonnes, 1432 tonnes to 15426 tonnes and -606 tonnes in 2012 for Coimbatore district to 505 tonnes in Namakkal district during 2009 respectively. It was also found that the negative and zero prediction of rice production were found only for Coimbatore and Karur districts during Kodai seasons only. The reason for the negative and zero prediction of rice production is mainly due to wide variations of observed rice production data during Kodai season, which causes the regression equation not true representative to provide the correct prediction. This type of error can be avoided only when more input data are available to build the best fitted equations. It was found that for the Kuruvai, Samba and Kodai seasons, the maximum predicted rice production varies from 148706 tonnes to 117462tonnes, 324654 tonnes to 548573 tonnes, and 101399 tonnes to 108389 tonnes respectively. It was found that the average predicted rice production varies from 39898.6

- tonnes to 48187.0 tonnes, 120036.0 tonnes to 175887.8 tonnes and 14324.5 tonnes to 20297.2 tonnes for Kuruvai, Samba and Kodai seasons respectively.
2. With regard to the area of cultivation of rice in Kuruvai season, the lowest ARE for area of cultivation of rice is 2.7 % for Thiruvallur district followed by 4.7% for Coimbatore district and the highest ARE of 66.3% is noted for Thiruvarur district. It was also found that 92% of the districts perform better prediction for area of rice cultivation with less than 50% ARE. With regard to the rice production in Kuruvai season, the lowest ARE of 4.7 % is noted for Thiruvannamalai district followed by 6.3% for Thiruvallur district. It is also found that the highest ARE of 73.1% is noted for the Thiruvarur district. It is found that 96% of the districts perform better prediction for rice production with less than 50% ARE in Kuruvai season.
 3. With regard to the area of cultivation of rice in Samba season, the lowest mean ARE of 0.4% is noted for Thiruvarur district followed by 1.7% for Cuddalore district. It is also found that the highest ARE of 46.0% is noted for Erode district. It was found that 100% of the districts perform better prediction for area of rice cultivation with less than 50% ARE during Samba season. With regard to the rice production in Samba season, the lowest ARE of 2.6 % is noted for Villupuram district followed by 5.7 for Vellore district. It is also found that the highest ARE of 68.8% is reported for the Ramanathapuram district. It was found that 96% of the districts perform better prediction for rice production with less than 50% ARE during Samba season.
 4. It was found that with regard to the area of cultivation of rice in Kodai season, the lowest mean ARE of 6.2% is noted for Pudukottai district followed by 7.4% for Thiruvallur district. It is also found that the highest ARE of 48.7% is noted for Tiruchirapalli district. It was found that the 84.6% of the districts perform better prediction for area of rice cultivation with less than 50% ARE during Kodai season. With regard to the rice production in Kodai season, the lowest ARE of 8.2% is noted for Thiruvallur district followed by 9.7% for Kancheepuram district. It is also found that the highest ARE of 47.3% is reported for the Dharmapuri district. It was found that 73.1% of the districts perform better prediction for rice production with less than 50% ARE during Kodai season for rice production.
 5. It was found that totally there are 42 paired t-test conducted between the observed data and Hybrid FFBPNN system at the rate of 21 paired t test for area of rice cultivation and another 21 paired t test for the rice production data. It was found that all the 42 paired t test statistically convey the meaning that there is no significant difference between the observed data and predicted data. The study also reported that there are wide absolute relative error between the observed data and predicted data in some districts but statistically such differences are not significantly affecting the overall prediction system for all the districts in a season. The researcher discussed some districts are having more than 100% ARE in Kuruvai, Samba and Kodai seasons and hence they are failed in proper predictions but the statistical test of significance considers the whole system of prediction of all districts put together as one entity without isolated thinking of individual failed districts and convey the meaning that there is no significant difference between the observed and predicted rice production data. However, the researcher is of the opinion that the input data collection should be for more than 10 years for better modeling and simulations though the hybrid system of FFBPNN worked well in this research.

ACKNOWLEDGMENT

The PhD scholar first and foremost thank the almighty God for his abundant blessings and protection during my studies and research work. I am highly indebted to Dr. P. Manimegalai, my research guide and Professor, Dept. of Electronics and Communication Engineering, Karpagam University, Coimbatore for her involvement, encouragement, suggestions and constructive criticism to make the research a grand success. I also thank the members of my research committee Dr. M. Mohanapriya, Head of the Department of Computer Science & Engineering, Karpagam University Coimbatore for the suggestions and guidance to make my work more applied to the society. I also thank the teaching staff and the administration of Karpagam University for guidance and encouragement. Finally, I thank Dr. K.Baskaran, Professor of Computer Science and Engineering, Government College of Technology, Coimbatore for his advises and suggestions to carry out the task more efficiently.

REFERENCES

- [1] F.A.Makinde and C.T.Ako et al., Prediction of crude oil viscosity using Feed-Forward Back-Propagation Neural Network (FFBPNN), *Journal of Petroleum & Coal*, ISSN 1337-7027, *Petroleum & Coal*, 54 (2) 120-131, 2012.
- [2] G.M.Shepherd and C.Koch, Introduction to synaptic circuits, *The Synaptic Organisation of the Brain*, Oxford University Press, New York, 1990, pp. 3 – 31.
- [3] Laxmi Ratna Raj and Amrender Kumar. Weather based forecasting model for crops yield using neural network approach, *Statistics and Applications*, Volume 9, Nos. 1&2, (New Series), pp. 55-69, 2011.
- [4] S.Arun Balaji and K.Baskaran, Feed Forward Back Propagation Neural Network coupled with Rice Data Simulator for prediction of rice production in Tamilnadu, *International Journal of Computer Science, Engineering and Information Technology (IJCSSEIT)*, Vol. 4, No.5, October 2014.
- [5] S.Arun Balaji and K.Baskaran, Integration of Feed Forward Back Propagation Neural Network (FFBPNN) and Curve Expert software to predict rice production in Tamilnadu, India, *International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCCE)*, Vol. 3, Issue 10, 2015.

- [6] S.Arun Balaji and K.Baskaran, 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, No.1, February 2013.
- [7] S.Arun Balaji and P.Manimegalai Vairavan, Effect of different statistical measures in error reduction in Feed Forward Back Propagation Neural Network (FFBPNN) to predict rice production, International Journal of Applied Engineering Research ISSN 0973-4562 Volume 10, Number 24, pp 44570-44578 © Research India Publications,2015

AUTHOR PROFILE

Mr. S.Arun Balaji is a PhD Scholar in the Department of Computer Science & Engineering, Karpagam University. He holds B.E (Information Technology), M.Tech (Information Technology) and MBA (Information Systems). He has a total of 5½ years experience in teaching, research and development works in Computer Science and Engineering both in India and abroad.

Prof. Dr. P. Manimegalai Vairavan is a Professor of Electronics and Communication Engineering, Karpagam University, Coimbatore, India. Her areas of Interest are Biosignal processing and image processing. She has more than 20 papers published in national and international journals. She is a potential PhD guide.