The Pharma Innovation Journal 2021; 10(5): 1191-1200

www.ThePharmaJournal.com

The Pharma Innovation

ISSN (E): 2277- 7695
ISSN (P): 2349-8242
NAAS Rating: 5.23
TPI 2021; 10(5): 1191-1200

www.thepharmajournal.com
Received: 04-01-2021
Accepted: 18-02-2021

ANN based crop yield prediction from remotely sensed retrieved crop parameters using machine learning

Krupavathi K, M Raghu Babu, A Mani, PRK Parasad and L Edukondalu

Krupavathi K
Assistant Professor, Dr. NTR College of Agricultural Engineering, Bapatla, Andhra Pradesh, India

M Raghu Babu
Professor and University Head (IDE), College of Agricultural Engineering, Madakasira, Andhra Pradesh, India

A Mani
Associate Dean, College of Agricultural Engineering, Bapatla, Andhra Pradesh, India

PRK Parasad
Professor and Head, Department of Soil Science & Agricultural Chem, Agricultural College, Bapatla, Andhra Pradesh, India

L Edukondalu
Associate Professor, Dr. NTR College of Agricultural Engineering, Bapatla, Andhra Pradesh, India

Corresponding Author:
Krupavathi K
Assistant Professor, Dr. NTR College of Agricultural Engineering, Bapatla, Andhra Pradesh, India

Abstract
The application of remote sensing in crop studies at regional level is becoming very famous. Crop yield prediction and mapping at different scales other than field scale is a challenging task for the researchers using remote sensing. This article described about the successful development of a scientific model using artificial neural network to predict the crop yield on regional scale using well known feed forward and back-propagation algorithms with the help of remotely sensed retrieved crop parameters. The Feed Forward Back Propagating Neural Network (FFBPNN) model developed and was calibrated using the remote sensing retrieved parameters and ground truth data in Mat lab environment. The model gave accurate and stable results. The highest mean relative error was 6.166% and the lowest relative error was 0.133%. To test the performance of the developed model statistically, Coefficient of determination, root mean squared error, mean absolute error and the average ratio of predicted yield to target crop yield (R_{ratio}) and relative error were used. This study also tested the number of hidden neurons on the performance of the model. The statistical analysis confirmed the reliability of the developed ANN model for its applicability on remote sensing-based parameters for paddy yield estimation (The range of R^2 values are from 0.933 to 0.992 for training and same for testing it ranged from 0.928 to 0.989). Based on the results, it was concluded that the FFBPNN models performed better and could be applied successfully to estimate and map the crop yield of paddy.

Keywords: Crop yield, neural networks, feed forward, back propagation, NDVI, APAR, water stress index

1. Introduction
In almost all developing economies, agriculture is substantial portion of Gross Domestic Product (GDP). Crop yield estimation at regional level plays crucial role in planning for food security of the population. This is of greater important task for some wide applications including management of land and water management, crop planning, water use efficiency, crop losses and economy calculation etc. Traditional ground observation based methods of yield estimation, such as visual examination and sampling survey require continuous monitoring, regular recording of crop parameters. Owing to synoptic and repetitive coverage, the remotely sensed images offer great potential in estimating crop extent and yield over large areas (Xin *et al*., 2013) [30]. Spectral information from remote sensing images gives very accurate crop attributes.
The crop yield at regional level can be estimated by agronomic models and are based on mechanistic or empirical approaches (Poluektov and Topaj, 2001) [19]. Mechanistic models are complex mathematical functions and uses many input parameters (Basso *et al*., 2001; Wang *et al*., 2002) [1, 28]. On the other hand, empirical models require less data and are relatively simple but usage outside the data range for which they were created is not possible(Kaul *et al*. 2005). Bolton and Friedl, 2013 [3] determined an empirical relationship between a vegetation index and in-situ crop harvest. This developed relationship is often only valid for the particular crop type and RS data acquired in that season. Achieving accurate crop yield is difficult in traditional agronomic and statistical modelling (especially regression models) of nonlinear functions with multiple factors of crop (Jiang *et al*., 2004) [9].Some traditional nonlinear models give more realistic and accurate results. However, problems exist with these models is difficult to handle multiple factors in the cropping system.
Compared to traditional linear and nonlinear statistical modelling, machine learning algorithms have proved a more powerful empirical model and self-adaptive method of crop yield estimation (Jiang 2000, Jiang *et al*., 2004 and Kaul *et al*., 2005) [8-10] and relatively simple compared to mechanistic models. Machine learning algorithms, especially, Artificial Neural

Networks (ANN) are useful for estimating crop yield from remote sensing images.

In the present study, paddy yield prediction models were developed by adopting Feed-forward back-propagating neural network (FFBPNN) structure as illustrated in Fig. 1. A feed-forward network is an Artificial Neural Network (ANN) model that is generally fast requires less memory and executes fast (Lawrence, 1994; Kaul *et al*., 2005) [13, 10]. The main advantage of neural network is that they are able to use some a prior unknown information hidden in data. Process of ‘capturing’ the unknown information is called ‘learning of neural network’ or ‘training of neural network (Svozil *et al*., 1997) [25]. To train feed-forward network, backpropagation (BP) algorithm is used, which is form of supervised training and is based on minimizing the error of the network using the derivatives of the error function. The network knows the desired output (fixed as target), processes the inputs and compare resulted output against the targeted outputs simultaneously generates the weight coefficients. Errors are calculated as the difference between desired and predicted outputs. The error is then back propagated into the model, causes the system to readjust the weight coefficients. The system adjusts the network in such way, that the calculated and targeted outputs are as close as possible. Theerror reduces in each epoch (Iteration of the complete training set) by adjusting the weights in the network (Svozil *et al*., 1997) [25]. This process occurs repeatedly as the weights are continually tweaked.

The training and prediction are the two modes of operation in feed forward neural network. For training the data is divided in to “training set” and “test set”. The "training set" enables the training of a network (Kaul *et al*., 2005; Sirisha *et al*., 2014) [10, 24]. After it has undergone its training, a set of data used to test model i.e test data. ANN parameters like rate of learning, number of hidden nodes also affects the accuracy of yield predictions. Ji *et al*., 2007 [7] stated that smaller data sets required fewer hidden nodes and lower learning rates in model optimization. Also stated that although ANN models are superior than regression models, but time consuming in development (Ji *et al*. 2007) [7].

The paddy is the major crop growing in the study area. Butthere is no specific defined model developed to predict the paddy yield and following the traditional methods. Keeping the above discussion in mind, the present study aimed at developing simplified paddy yield prediction models with remote sensing parameters and historic yield data. The specific objectives include: (a) To investigate the effective yield prediction of artificial neural network (ANN) models from remote sensing retrieved crop parameters and historic yield data at regional level. (b) To observe the changes of model performance with variations of model parameters. (c) To test the model performance statistically.

2. Material and Methods

2.1 Study area

The present study was carried out in Krishna Central Delta (KCD) a part of Krishna Eastern Delta in Krishna district in the state of Andhra Pradesh, which is named after the holy river Krishna, bounded by the latitudes 16º 37' 15'' N and 15º 42' 15''N and longitudes 80º 34' 0''E and 81º 16' 0''E. It constitutes the command area of Bandar canal and Krishna Eastern bank canal which has an irrigated ayacut of 111223.83 ha in Krishna district. It irrigates about 18mandals in Krishna district (Fig. 1). In the *Kharif* season (July/August-November/December) the major crops grown are Paddy, Sugarcane, Turmeric and Vegetables and the predominant crops grown during the *Rabi* season (December- March) are Paddy, Sugarcane, Maize, Pulses and Cauliflower. At the upper reaches farmers are cultivating in summer also. Orchards are also grown specifically in south part of the study area. The main crop in *Kharif* if paddy in 95% of the study area is considered in the present study. The various land use pattern of the study area is presented in Table 1. The major area is under agriculture (1539.342 km^2) followed by aqua (402.971 km^2) and horticulture (57.648 km^2).

2.2 Data acquisition and retrieval of crop parameters from remote sensing

Cropping system consists of nonlinear behavior with inherent sources of heterogeneity (Prasad *et al*., 2006). Crop yields are affected by typical factors like sunlight, temperature, rainfall, water supply, soil etc. With the advancement in Remote sensing technologies, the factors can be measuredby using appropriate indices extracted from remote sensing images (Jiang *et al*., 2004) [9]. Five remote sensing indices namely Normalized Difference Vegetation Index (NDVI), surface temperature (T_s), water stress index (WSI), Absorbed photo synthetically active radiation (APAR), and average crop yield over the last 5 yearswere selected. The first four parameters can be retrieved from Landsat 8remote sensing images.The average yield is calculated from statistical and ground.

The free available high resolution optical Landsat 8 satelliteis used in the present study. Spectral information from remote sensing images gives very accurate crop attributes. Several researchers reported that estimates from Landsat were considerably more accurate in yield estimates (Sibley *et al*., 2014, Hooda *et al*., 2006) [21, 6]. Landsat derived indices have good potential to use in prediction of yield and its variability during growth stages (Kumhalova *et al*., 2014). The Landsat 8 Level 1 images was downloaded from USGS Earth explorer. Using the radiance rescaling factor, Digital Numbers (DN)are converted to TOA spectral reflectance data. The selected indices were generated in ARCGIS 10.3 from remotesensing images. The retrieved indices were extracted to ground truth points surveyed to develop the model. The developed indices are as follows.

Normalized Difference Vegetation Index (NDVI)is one of the most efficient index of growing conditions for crops (Kaul *et al*., 2005, Jiang *et al*., 2004) [10, 9].The NDVI is the response index to greenness and vegetative cover. It is the normalized difference between the near infrared and visible RED reflectance bands.

$$NDVI = \frac{\rho_{NIR}-\rho_{RED}}{\rho_{NIR}+\rho_{RED}} \text{…………………………(1)}$$

High NDVI values reflect greater or greenness vegetation, whereas low NDVI values reflect to stress or senescence and low vegetation. The next important parameter is solar radiation. The amount of light available for photosynthesis is known as Photosynthetically Active Radiation (PAR) and ranges between 400 and 700 nanometers. Absorbed Photo Synthetically Active Radiation (APAR) is the portion absorbed for photosynthesis by crop leaves.

$$APAR=PAR*FAPAR \text{ ……………………….(2)}$$

Fraction of absorbed photosynthetically active radiation (FAPAR) is a function of absorbed photosynthesis active radiation can be used in estimation of light use efficiency to estimate crop yields at the pixel level (Patel *et al*., 2006) [16]. Sims *et al*., 2006 [22], Peng *et al*., 2010 [17]; Singh *et al*., 2012 [23] proposed a linear, scale-invariant relationship between FAPAR and the NDVI. In the present study, computed FAPAR using NDVI as suggested by Sims *et al*., 2006 [22] for Landsat images is adopted in this study as:

$FAPAR = 1.24 * NDVI - 0.168$………………………..(3)

Canopy surface temperature represents sunlight radiated onto leaves and also it ia an indication of evaporation intensity. Surface temperature is calculated as

$T_s = (BT / 1) + W * (BT / 14380) * \ln(\varepsilon)$ ………………(4)

Where, BT = Top of atmosphere brightness temperature (°C)
W = Wavelength of emitted radiance
ε = Land Surface Emissivity
Spectral radiance data can be converted to top of atmosphere (TOA) brightness temperature using the thermal constant (K_1 and K_2) values in Meta data file

$BT = K_2 / \ln (K_1 / L_\lambda + 1) - 272.15$ ……..………………….(5)

Where: BT = Top of atmosphere brightness temperature (°C)
L_λ = TOA spectral radiance (Watts/(m^2 * sr * µm))
K_1 = K_1 Constant Band (No.)
K_2 = K_2 Constant Band (No.)

NDVI/Ts (Crop Water Stress Index) is taken as one of the indicators of crop yield. The crop water stress index (CWSI) is a normalized index to quantify stress and overcome the effects of other parameters affecting the relationship between stress and plant temperature (Poblete *et al*, 2015) [18]. NDVI/T_s has been used for water stress monitoring in recent years with good results (Becker and Li 1990 and Jiang *et al*., 2014) [2]. There is a close relationship between NDVI/T_s and crop water content (Goward *et al*., 1985).
NDVI, T_s, CSWI and APAR maps were retrieved from remotely sensed images from transplanting to the harvesting stage of paddy crop. The values for the ground truth points of all parameters derived by remote sensing were extracted. Spatial distribution of sample points is shown in Fig. 3. The attributes of above derived thematic maps are extracted for all the sample points and exported to Excel as. csv file for preparation of input files to neural network structure.

2.3 Collection of crop cutting data
The crop yield data (Crop cutting data) is collected for five years. Wart *et al*., 2013 [29] reported 5 years of data is sufficient for estimates of yield potential for fully irrigated productionsystems and is adopted for present study. The crop yield per unit area of different crops for the years 2013 to 2017 is collected from Directorate of Economics and Statistics, Vijayawada. The location details of the data collected is in the form of Survey numbers of Revenue department. To know Latitude and Longitude and also to collect previous years data at each crop cutting point Ground truth is done by using EpiCollect (A mobile based App).

2.4 ANN Model design
The data extracted from the above maps is divided into two data sets for training and testingto analyze network results and testing the models. The typical architecture of three-layered MLFF perceptron used is shown in Fig 4. The derived five yield factors such as NDVI, APAR, Surface temperature, Water stress index, Average yield are taken as neurons for input layer. The output layer is one neuron i.e yield.
The hidden layer has different number of hidden neurons and is tested for optimum number of neurons. The optimum number of neurons or nodes in hidden layer and parameters of the model is determined by trial and error method. W_{ij} is the connecting weight between i^{th} input layer neuron to the j^{th} hidden layer neuron. The V_{jk} is weight between the j^{th} hidden layer neuron and the k^{th} output layer neuron (in this case k=1). Learning rate and the momentum are two main parameters for training which takes care of steepest-descent convergence and function preventing the solution from being trapped into local minima (Sirisha *et al*., 2014) [24]. The final weighting factors are used to simulate relationship between crop yield and corresponding crop growth factors. The final weighting factors generated by the trained network model were saved for prediction of new input data. The hidden layer neurons are varied from 1 to 30 in the developed models. Sigmoidal transfer function and linear activation functions are in hidden output layers.The code to develop the neural network is written in MATLAB programming language package.

The data from the input layer sends to the hidden layer. The hidden layer multipliesthe inputs by suitable weights and sums. Thesigmoidal activation function was applied before the result sentto the output layer. Mathematical expression linear activation functionis

$x = \sum_{i=1}^{n} w_i x_i + b = w_1 x_1 + w_2 x_2 + w_3 x_3 + \ldots\ldots\ldots + w_n n + b$ …….(5)

The output *y* is expressed as:
$y = f(x) = f(\sum_{i=1}^{n} (w_i x_i + b)$………………………….(6)

where *f* is neuron activation or transfer function.The transfer function of each neuron is a sigmoid function given by:

$y = f(x) = \frac{1}{1+e^{-x}}$ ……………………(7)

The neuron activation function was shown in Fig. 4. The final form of the FFBPNN model with the substitution of weights is given as

$Y = \frac{1}{1+\exp[-(\sum_{j=1}^{q} V_{jk} H_{ij} - \gamma)]}$…………………….……(8)

Where, Y= Yield per unit area exported from the neural network model, q=number of nodes hidden, Vj=weight between j^{th} hidden node and k^{th}output node, c=threshold of the output node,

2.5 Model Training
The trained ANN recognizes the functional relationship between input parameters and desired outputs. The network training starts with random initialization of weights, proceeds by applying 'LM' algorithm and optimizes an error function (RMSE) (Marti and Gasque 2010; Tabari *et al*., 2010, Sirisha *et al*., 2014) [24]. The generated weights by neural network saves andalso remembers this functional relationship for further calculations. Yield prediction models were developed at the regional level for the paddy crops in kharif season. The

right time stopping of the training of neural network is called as early stopping is an important step to avoid over fitting. To achieve this the training, validation and test set was used to adjust the weights and biases, to stop the training process and for external prediction respectively. Initially 75% of samples are randomly selected for training, and the remaining 25% are used for testing to evaluate the model performance.

The data of different parameters have wide range of values. For uniformity and also to avoid the confusion of learning algorithm, all of the input data are normalized before training to represent 0 to minimum and 1 to maximum values. The output results (yields) are converted back into the same unit by a denormalization procedure. Learning rate, number of hidden nodes and training tolerance were adjusted. The initial selected number of hidden nodes was equal to inputs +1.

2.6 Model validation

Four statistical parameters were used for performance analysis of the developed FFBPNN models, namely coefficient of determination (R^2), root mean squared error (RMSE), mean absolute error (MAE), and the ratio of average output to target yield values (R_{ratio}). These parameters were calculated using the test data for finding out to optimize neural network. The criteria for optimum neural network is minimum RMSE, minimum MAE (Should be optimally 0), and the value of R^2 near to one. The R_{ratio} is used only to explain whether the model under- or over-predicted the simulated yield values. R_{ratio} less than 1 indicates under estimation, R_{ratio} is more than 1 indicates over estimation. The relative error for each data point is also calculated. In addition, the performance of each model was evaluated by plotting the simulated values against the measured values and by testing the statistical significance of regression parameters.

3. Results and Discussion

3.1 Collection of yield data of paddy crop for the last 5 years

The field data is selected such that at least two to a maximum of 15 samples are selected for field collection of yield data in each mandal out of 18 mandals by simple random sampling method. A separate project entitled “Old crop details” was created for crop yield data in EpiCollect website (Fig. 2) for collection of data from the sample points for the years 2013 to 2016. The data points for the year 2017 were taken at crop cutting spot exactly by creating another file as “New crop details”. Five year (2013–2017) crop yield per unit area collected at each ground point of the statistical data from the Department of Economics and Statistics (DES), Vijayawada. The five-year seasonal data was averaged and is used as another parameter input to the model.

3.2 Modeling and Validation crop yield using FFBPNN

Yield prediction FFBPNN models were developed at regional level. Developed FFBPNN model were trained with the scenario mentioned above and after each training run using 75% of the sample data. Data from the other 25% are used as testing dataset to validate the FFBP-NN model. The FFBPNN model is trained several thousand times until the average relative error was smaller than a predefined threshold and RMSE is minimum. The perceptron was trained with 75 out of the 100 possible inputs over 10000 epochs with 6 (I+1, I is input neurons) to 100 hidden neurons.The sample values of the weight parameters obtained from model training of paddy crop for best statistical parameters in the year 2015 are given in Table 2.The test data (25%) preserved in each year is used for validation of the FFBPNN models in the respective year for five consecutive years. Sample training data and results are shown in Table 3 representing minimum and maximum values of normalized data. The output results (yields) were converted into actual values in the end.

Relative error between the targeted and neural network model predicted yield values for the five years of study period is shown in Fig. 5. All relative errors of the model are smaller than 10% except for three readings. 70% of the relative errors between predicted and observed values are evensmaller than 5%.The data points varied in each year based on number of ground truth points surveyed. The highest mean relative error was-12.64% observed in 2014 in *Kharif* season and the lowest relative error was 0.133%. Relative error in the year 2017 is very less compared to all other years due to more consistency of the data.This shows that the developed model gives accurate result compared to regression models. It was also observed that the FFBPNN model gave more stable result. The observed results are on par with Jiang *et al* 2004 [9]. Hooda *et al*., 2006 [6], Li *et al*., 2 007 reported that remote sensing indices gave best results through regression models are 85%. This study shows the application of neural network further increased the prediction accuracy to 95% in 70% of cases.

The statistical parameters of the training and testing are given Table 4. The range of R^2 values are from 0.933 to 0.992 for training and same for testing it ranged from 0.928 to 0.989; and R_{ratio} values from 0.998 to1.063, the RMSE values ranged from 0.04 to 0.117; MAE ranged from 0.013 to 0.095 for training and 0.024 to 0.085 for testing; The lower values of normalized MAE, and RMSE, and values are close to 1 for R_{ratio}, and R^2 for all years indicates that the FFBPNN models performed better during testing. But the accuracy of predicted values is slightly less during simulation. Because of the simulation is highly dependent on the selection of testing dataset. Though the performance of the model is slightly less, the simulations produced highly satisfactory outcome in all five years. The results are in accordance with Kaul *et al*., 2005 [10]. This indicates that a well-trained FFBPNN model can be successfully used for crop yield prediction. The FFBPNN models overestimated yield values in the year 2014 and 2015 (R_{ratio} @ 1.040687 and 1.063936) as the ANNs are empirical models, and their performance mainly depends on data pertaining to training the networks which might cause under or over estimation of yield (Uno *et al*., 2005, Prasad *et al*., 2006, Sirisha *et al*., 2014) [27, 24].

During the study, it was also observed that for the model, the RMSE decreased with increasing number of hidden nodes from 1 to i+ 1, where i is the number of input nodes (i=5). Further, R^2 increased and MAE decreased with the increase in the number of hidden nodes from 1 to i+ 1, for all years. The results are in accordance with Kumar *et al*. 2002, Uno *et al*., 2005 [27], Sirisha *et al*., 2014 [24], Prasad *et al*., 2006. For 2013, 2016 and 2017 the best performance observed at i+1 hidden nodes. Kaul *et al*., 2005 [10] also reported that the Smaller data sets required fewer hidden nodes and lower learning rates in model optimization. After 20 hidden nodes the trails are conducted at a step of 10. After a number of trials with 100 to 10,000 epochs with a step of 100 upto 2000 and a step of 1000 upto 10000, the best results were found at 1,000 epochs for most of the cases. Though the performance increased with after i+1 epochs, the computation time increased with increase in no of nodes and epochs.

3.2 Comparison of observed and predicted yield of paddy
The scatter plots between observed and FFBPNN estimated crop yield for the five years is shown in Fig. 6(a-e) indicates of consistency between the training and testing data. The scatter plots confirm the statistics given in Table 4 (for all years). It indicates that there is a close agreement between reported and estimated yields of paddy using FFBPNN model and are distributed evenly on both sides of the 1:1 line (y=x). It is also observed that the model exhibits a highly satisfactory performance in training but the simulation is highly dependent on the selection of testing dataset (Patel *et al*., 2006, Wart *et al*., 2013, Guo and Xue 2014) [16 29, 5]; hence, the range of forecasting error is large. This indicates that a well-trained NN model produces consistently accurate results when data excludes "abnormal" datasets in both training and testing the same is expressed by Guo and Xue, 2014 [5].

3.3 Predicted crop yield maps of Krishna Central Delta
After calibration and validation of the FFBPNN model, parameter weights were used to prepare yield maps of crops. The predicted paddy yields per unit area in *kharif* season were calculated and shown in Fig. 7 (a-e) for the five years. The spatial yields in KCD showed paddy yields increased from 2013 to 2017. The predicted yields in the year 2013 was lower as compared with remaining years. Yields were highest in the year 2017 in all the parts of study area. In 2017, less number of cloudy days, good water supplies due to interlinking Godavari river with Krishna river and favorable crop conditions improved the crop yields.

In all the years, the paddy yields were good in central parts of the study area and are matched with canal irrigation network. There was a constant yield range in upper part of the area. At lower parts of the KCD, the yields were low compared to other areas. The yields are less in Nagayalanka, Koduru, Machilipatnam and Pedana. The probable reasons for the less yields in these mandals are less canal water supply in tail reaches, salinity problem, salt water upcoming in the aquaculture area. In these areas, planting dates are generally in the month of September. The area with red tone is showing less or crop failure up to 2015 years. In 2016 and 2017, the area is left fallow due to non-availability of water supply and insufficient rainfall. The predicted average yields ranged from 3000 to 9020 kg/ha during different years. The results are on par with statistical data collected from Directorate of statistics, AP.

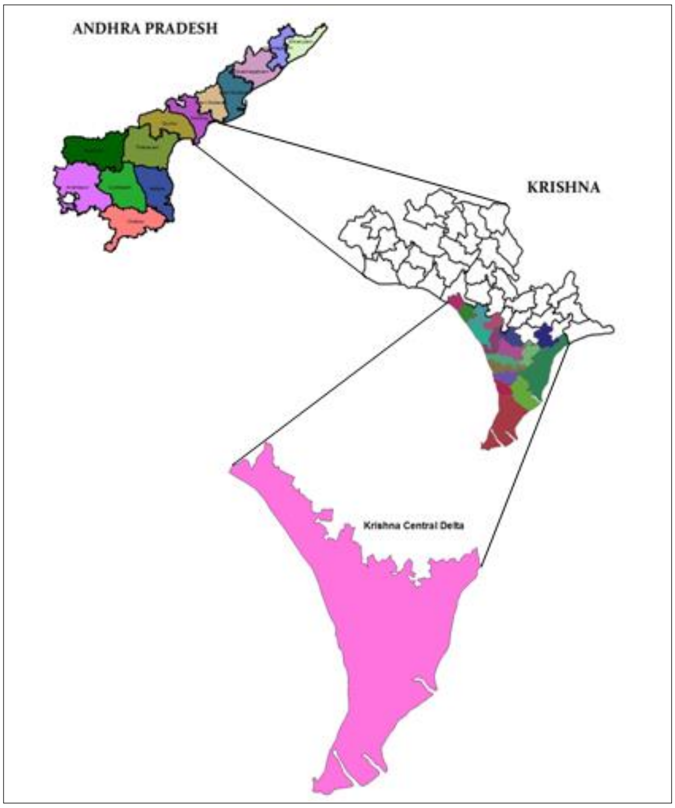


Fig 1: Location of Krishna Central Delta

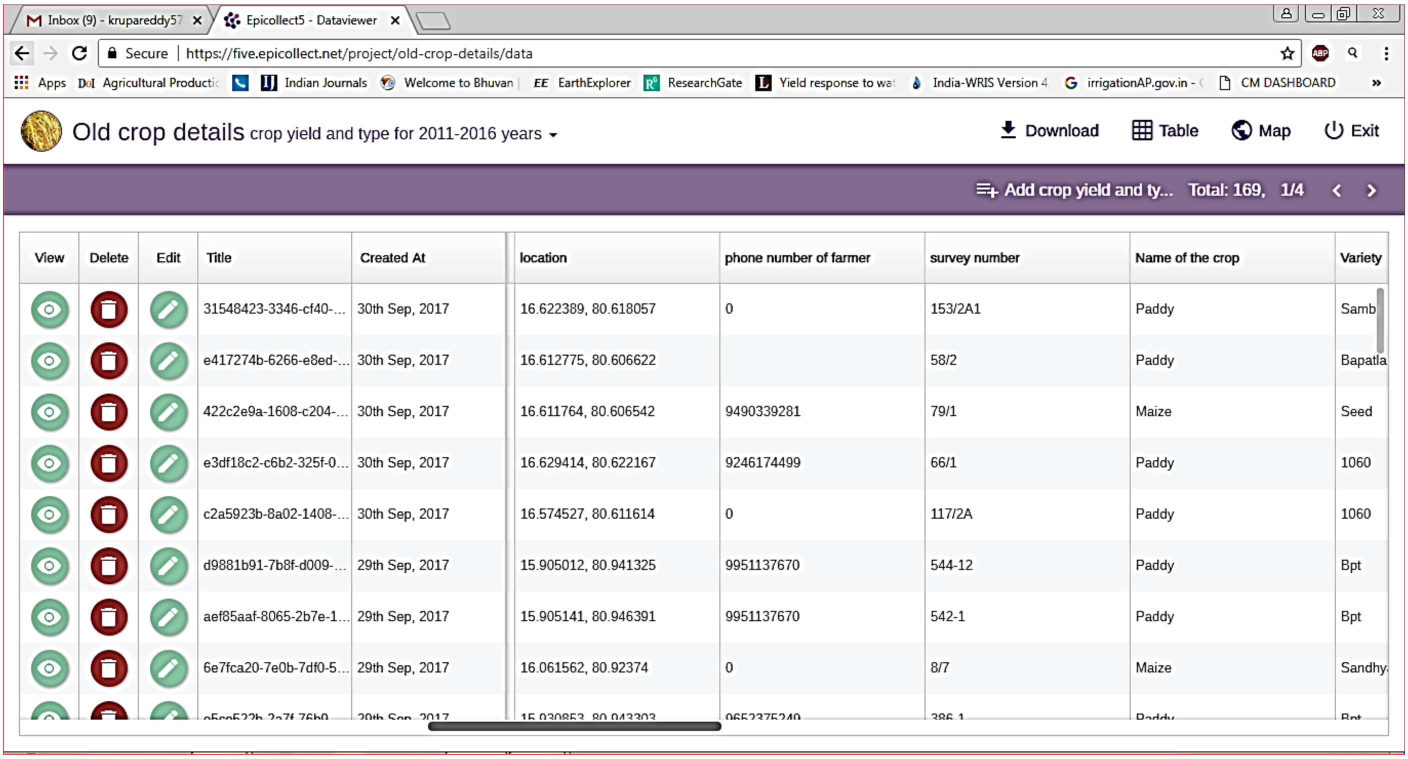


Fig 2: Collection of ground sample points and their attributes using Epi Collect app

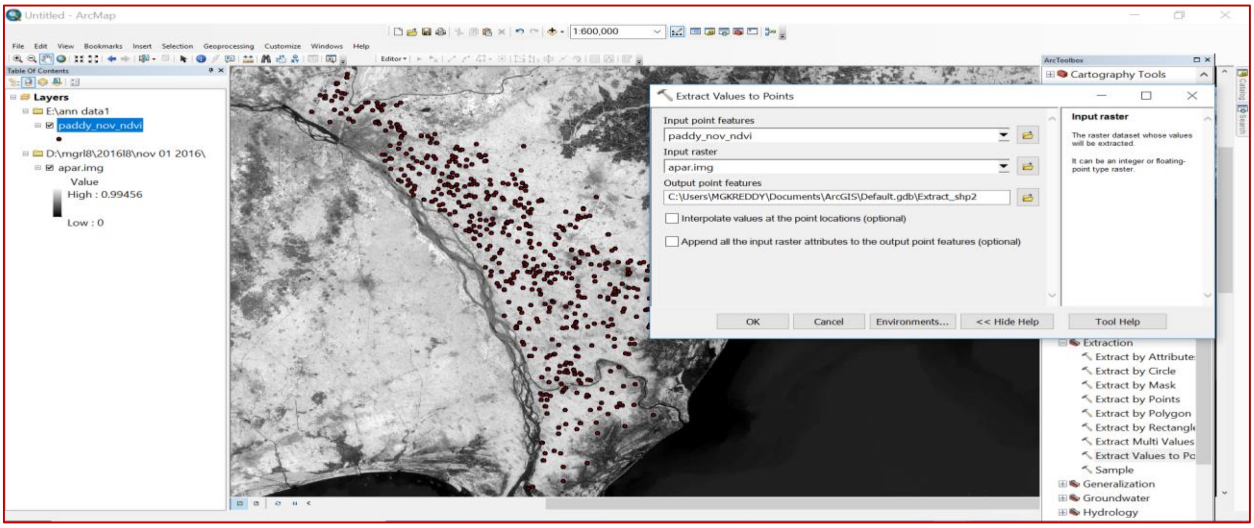


Fig 3: Synoptic view of spatial distribution of sample points of crop collected

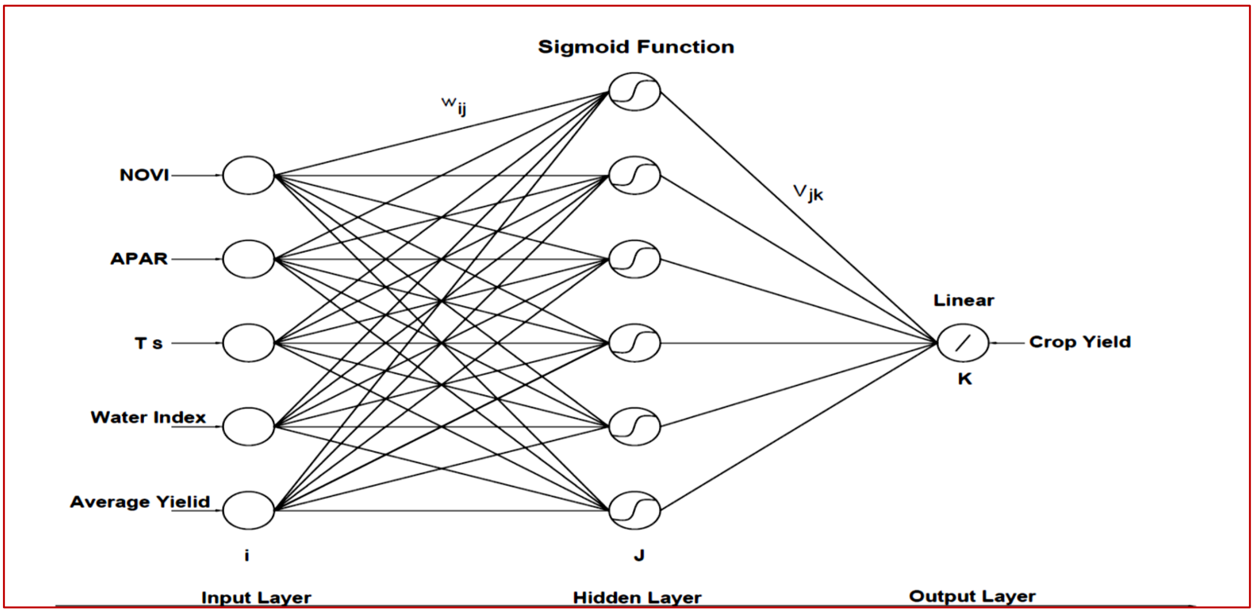


Fig 4: Architecture of the proposed FFBPNN model

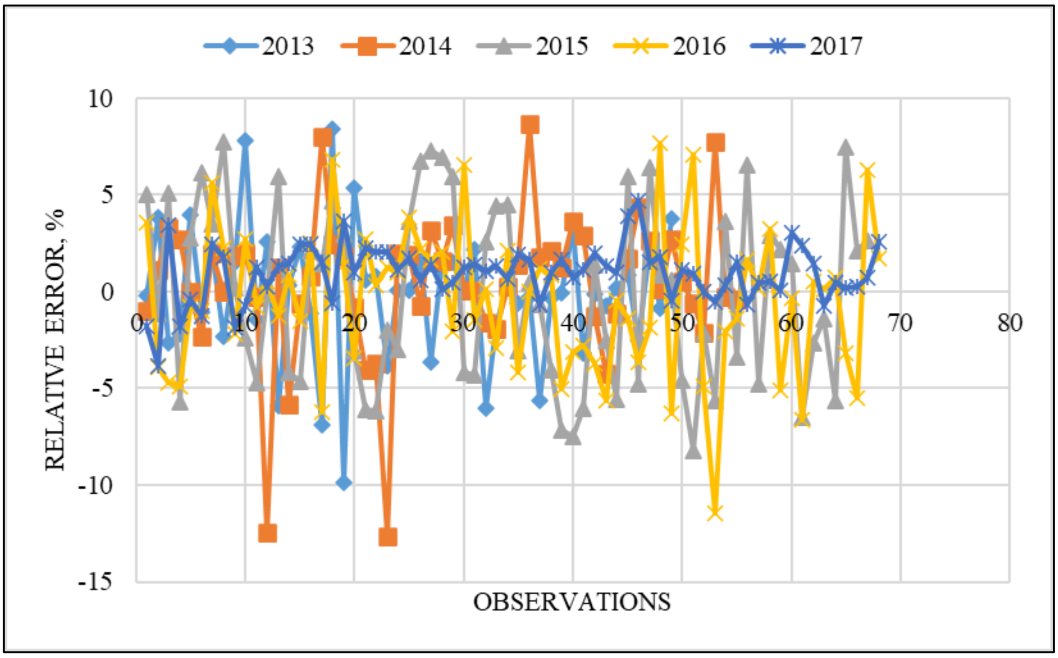
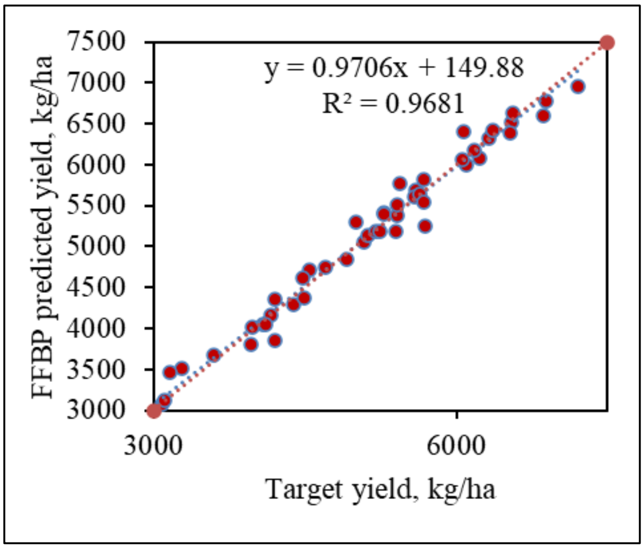
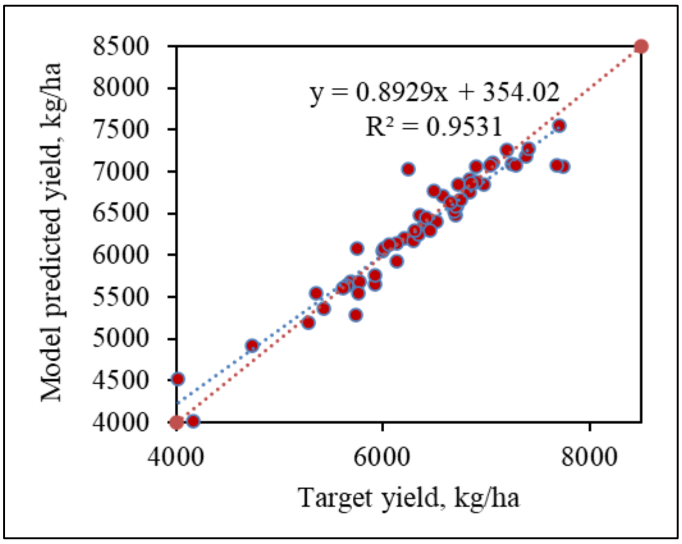


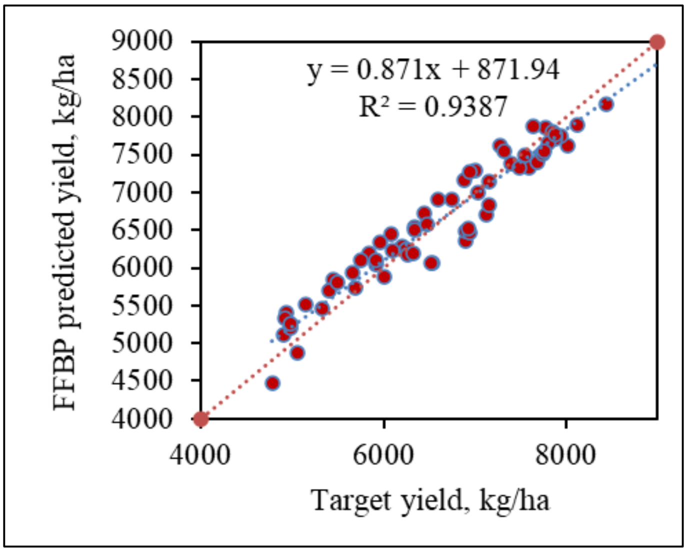
Fig 5: Variation of relative error of crop yield in different years of cropping season



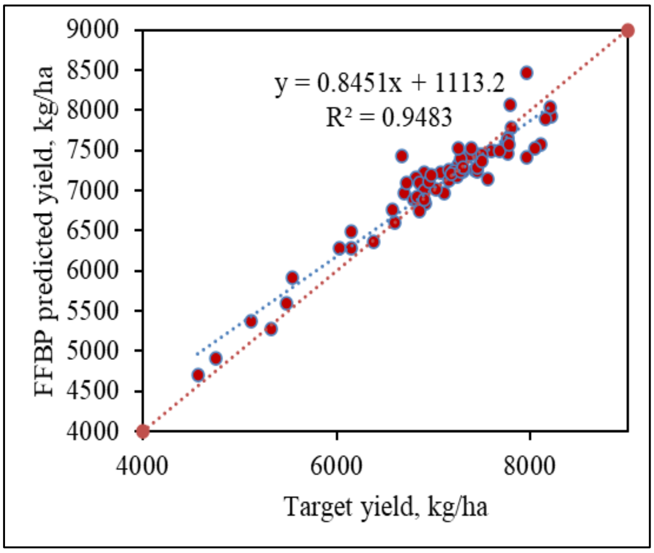
(a) 2013



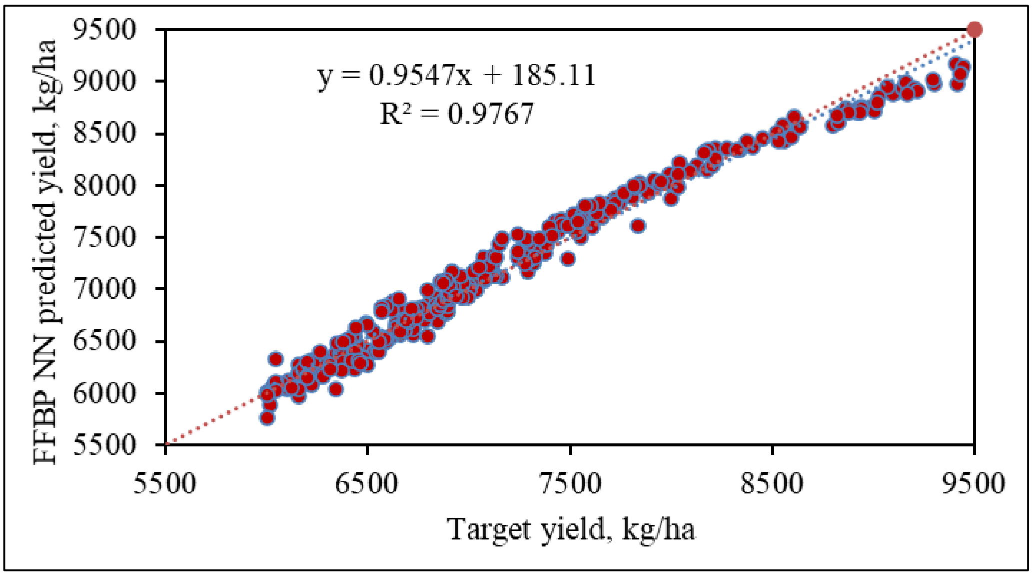
(b) 2014



(b) 2015

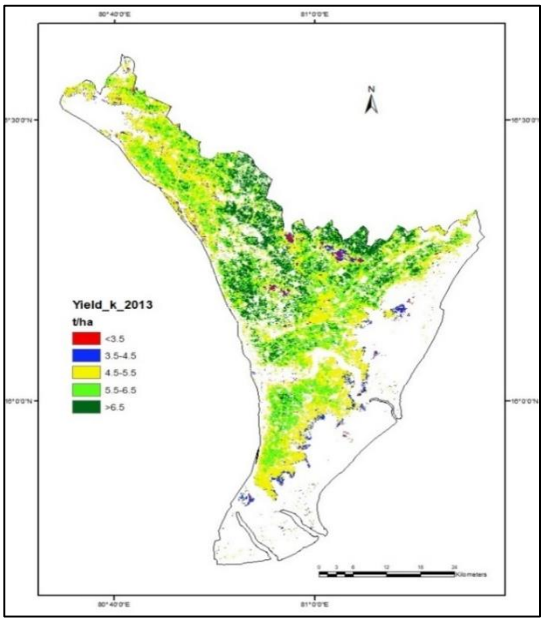


(d) 2016

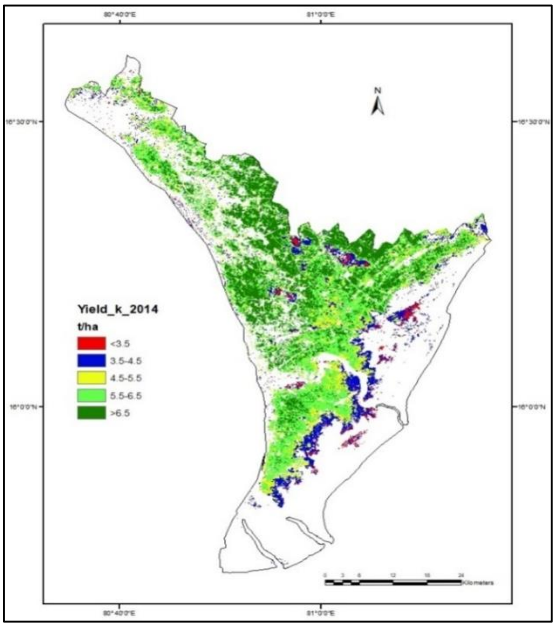


(e) 2017

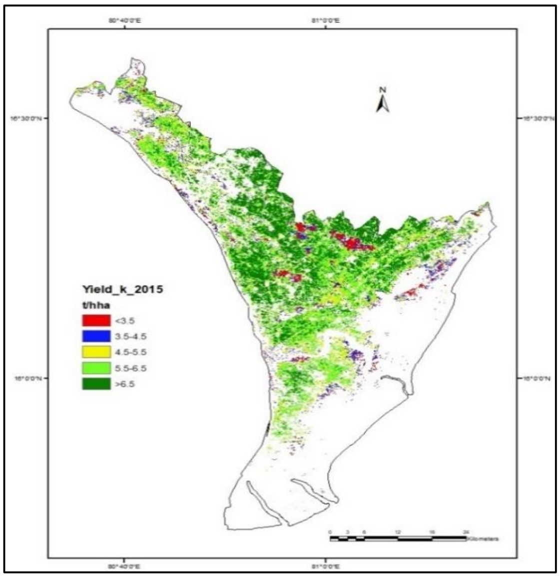
Fig 6 (a-e). Scatter plots of actual and FFBP NN model predicted yield of paddy crop in *kharif* season during different years.



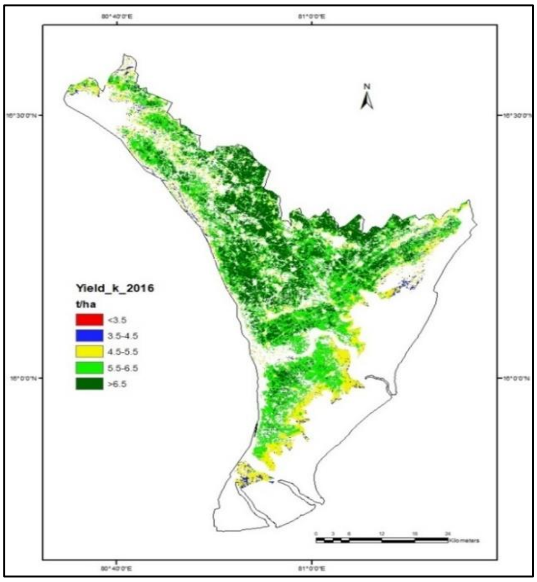
A.



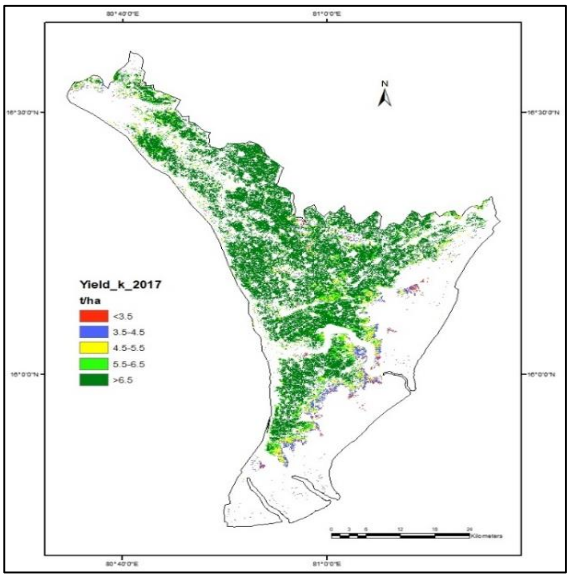
B.



C.



D.



E.

Fig 7 (a - 6). Yield per unit area of paddy crop in *kharif* season during 2013, 2014, 2015, 2016 and2017

Table 1: Land use of the study area in the year 2014-2015

S. No.	LULC class	Area (km^2)	Percentage (%)
1	Agriculture	1539.342	69.1924
2	Water bodies/Aqua	402.971	18.11328
3	Other waste land	18.91	0.849992
4	Buildup	117.369	5.275659
5	Decidious forest	88.487	3.977432
6	Plantation/orchard	57.648	2.591239

Table 2: Sample weight parameters of FFBPNN yield estimation model for paddy crop in *Kharif* 2015

j	1	2	3	4	5	6
w1j	-0.0403	0.057611	-0.19018	-0.13112	-0.02701	-0.13502
w2j	-0.14199	-0.16274	0.041413	-0.26324	0.031121	-0.04462
w3j	-0.02086	0.013857	-0.09399	-0.0513	0.16498	0.015007
w4j	-0.00745	0.063022	-0.05538	-0.18956	-0.03317	0.007654
w5j	-0.05268	-0.89495	-0.34146	-0.49691	-1.12633	-0.10711
Ѳj	0.158842	0.065105	-0.16914	0.013203	0.252625	-0.05552
Vj	-0.04343	-0.33426	-0.19198	-0.26668	-0.57053	-0.09973

Table 3: Sample data and training result of FFBPNN yield estimation model of kharif paddy in KCD in 2015

S. No	Mandal name	Normalized NDVI	Normalized Ts	Normalized APAR	Normalized Water Index	Normalized Average yield
1	Vijayawada rural	0.542	0.831	0.775	0.354	0.674
2	Kankipadu	0.548	0.148	0.778	0.448	0.855
3	Challapalle	0.539	0.792	0.773	0.180	0.544
4	Pamarru	0.528	0.577	0.768	0.224	0.533
5	Vuyyuru	0.531	0.330	0.770	0.325	0.744
6	Movva	0.687	0.363	0.844	0.723	0.725
7	Thotlavalluru	0.687	0.439	0.844	0.686	0.450
8	Avanigada	0.563	0.454	0.785	0.359	0.000
9	Pamidimukkala	0.575	0.687	0.791	0.302	0.900
10	Guduru	0.459	0.767	0.735	0.000	0.915
11	Penamaluru	0.478	0.706	0.744	0.063	0.525
12	Koduru	0.748	0.482	0.874	0.821	0.900
13	Pamarru	0.771	0.390	0.885	0.929	0.619
14	Machilipatnam	0.533	0.452	0.771	0.284	0.921
15	Pedana	0.491	0.203	0.750	0.266	0.957
16	Mopidevi	0.496	0.790	0.752	0.080	0.544
17	Nagayalanka	0.511	0.521	0.760	0.203	0.750

Table 4: Statistical results of neural network training and testing

	Training				Testing			
Year	RMSE	R_{ratio}	MAE	R^2	RMSE	R_{ratio}	MAE	R^2
2013	0.040	0.998	0.030	0.977	0.035	1.000	0.024	0.969
2014	0.058	1.040	0.040	0.933	0.065	0.958	0.045	0.928
2015	0.117	1.063	0.095	0.946	0.108	1.065	0.085	0.936
2016	0.068	1.002	0.054	0.948	0.084	1.020	0.063	0.947
2017	0.017	0.998	0.013	0.992	0.027	1.000	0.024	0.989

4. Conclusions

The objective of this study was to develop simplified paddy yield prediction models with remote sensing parameters and historic yield data for the study area. The proposed methodology was able to estimate Paddy yield, the physical range yield estimates obtained by introducing the indices from Landsat 8 in FFBPNN model and were very close to observed yield variation within KCD. The arrived prediction errors were less than 10 percent, which indicates best suitabilityof crop yield prediction at regional level. It was observed that the ANN models were stable, quite efficient in capturing the complex relationship between crop yield and remote sensed retrieved parameters. The analysis confirms that the FFBPNN model discussed in the present paper reasonably minimized inconsistency and errors in yield prediction giving high R^2-values. To test the suitability, the study taken up in 5 years. In all years of study, the performance of model increased after i+1 epochs, but the computation time increased with increase in no of nodes and epochs. In all years, the FFBPNN models proved suitable in predicting the paddy yields with good accuracy from the remotely sensed parameters. The ANN models proved to be a superior model for predicting paddy yield with 95% accuracy. The crop yield prediction model discussed in the present paper will further improve in future with the use of long period dataset. Similar model can be developed for different crops of other locations.

5. References

1. Basso B, Ritchie JT, Pierce FJ, Braga RP, Jones JW. Spatial validation of crop models for precision

agriculture. Agricultural Systems 2001;68(2):97-112.
2. Becker F, Li ZL. Temperature-independent spectral indices in thermal infrared bands. Remote sensing of environment 1990;32(1):17-33.
3. Bolton DK, Friedl MA. Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics. Agricultural and Forest Meteorology 2013;173:74-84.
4. Goward SN, Cruickshanks GD, Hope AS. Observed relation between thermal emission and reflected spectral radiance of a complex vegetated landscape. Remote Sensing of Environment 1985;18(2):137-146.
5. Guo WW, Xue H. Crop yield forecasting using artificial neural networks: A comparison between spatial and temporal models. Mathematical Problems in Engineering 2014.
6. Hooda RS, Yadav M, Kalubarme MH. November. Wheat production estimation using remote sensing data: An Indian experience. In Workshop Proceedings: Remote Sensing Support to Crop Yield Forecast and Area Estimates, Stresa, Italy. 2006, 85-89.
7. Ji B, Sun Y, Yang S, Wan J. Artificial neural networks for rice yield prediction in mountainous regions. The Journal of Agricultural Science 2007;145(3):249.
8. Jiang D. Study on crop yield forecasting model using remote sensed information supported by artificial neural network. Doctoral dissertation, Institute of Geographic Science and Natural Resources Research, Chinese Academy of Sciences 2000.
9. Jiang D, Yang X, Clinton N, Wang N. An artificial neural network model for estimating crop yields using remotely sensed information. International Journal of Remote Sensing 2004;25(9):1723-1732.
10. Kaul M, Hill RL, Walthall C. Artificial neural networks for corn and soybean yield prediction. Agricultural Systems 2005;85(1):1-18.
11. Kumar M, Raghuwanshi NS, Singh R, Wallender WW, Pruitt WO. Estimating evapotranspiration using artificial neural network. Journal of Irrigation and Drainage Engineering. 2002;128(4):224-233.
12. Kumhálová J, Zemek F, Novák P, Brovkina O, Mayerová M. Use of Landsat images for yield evaluation within a small plot. Plant, Soil and Environment 2014;60(11):501-506.
13. Lawrence J. Introduction to Neural Networks. California Scientific Software Press, Nevada City, CA 1994.
14. Li A, Liang S, Wang A, Qin J. Estimating crop yield from multi-temporal satellite data using multivariate regression and neural network techniques. Photogrammetric Engineering & Remote Sensing 2007;73(10):1149-1157.
15. Martí P, Gasque M. Ancillary data supply strategies for improvement of temperature-based ETo ANN models. Agricultural water management 2010;97(7):939-955.
16. Patel NR, Bhattacharjee B, Mohammed AJ, Tanupriya B, Saha SK. Remote sensing of regional yield assessment of wheat in Haryana, India. International Journal of Remote Sensing 2006;27(19):4071-4090.
17. Peng J, Liu Y, Zhao X, Ye C, Song P, Zhao D. October. A novel approach to calculating water stress index. In 2010 3rd International Congress on Image and Signal Processing IEEE. 2010;5:2085-2088.
18. Poblete-Echeverría C, Espinace D, Sepúlveda-Reyes D, Zúñiga M, Sanchez M. June. Analysis of crop water stress index (CWSI) for estimating stem water potential in grapevines: Comparison between natural reference and baseline approaches. In VIII International Symposium on Irrigation of Horticultural Crops 2015;1150:189-194.
19. Poluektov RA, Topaj AG. Crop modeling: nostalgia about present or reminiscence about future. Agronomy Journal 2001;93:653-659.
20. Prasad AK, Lim Chai, Ramesh P Singh, Menas Kafatos. Crop yield estimation model for Iowa using remote sensing and surface parameters International Journal of Applied Earth Observation and Geoinformation 2006;8:26-33.
21. Sibley AM, Patricio G, Nancy ET, Cassman KG, Lobell DB. Testing Remote Sensing Approaches for Assessing Yield Variability among Maize Fields. Agronomy Journal 2014;106(1):24-32.
22. Sims DA, Luo H, Hastings S, Oechel WC, Rahman AF, Gamon JA. Parallel adjustments in vegetation greenness and ecosystem CO2 exchange in response to drought in a Southern California chaparral ecosystem. Remote Sensing of Environment 2006;103(3):289-303.
23. Singh RK, Liu S, Tieszen LL, Suyker AE, Verma SB. Novel approach for computing photosynthetically active radiation for productivity modeling using remotely sensed images in the Great Plains, United States. Journal of Applied Remote Sensing 2012;6(1):063522.
24. Sirisha A, Raghuwanshi NS, Ashok Mishra, Mukesh K Tiwari. Evapotranspiration Modeling Using Second-Order Neural Networks. Journal of Hydrologic Engineering. 2014;19(6):1131-1140.
25. Svozil D, Kvasnicka V, Pospichal J. Introduction to multi-layer feed-forward neural networks. Chemometrics and Intelligent Laboratory Systems. 1997;39(1):43-62.
26. Tabari H, Marofi S, Sabziparvar AA. “Estimation of daily pan evaporation using artificial neural network and multivariate nonlinear regression.” Irrigation Science. 2010;28(5):399-406.
27. Uno Y, Prasher SO, Lacroix R, Goel PK, Karimi Y, Viau A *et al*. Artificial neural networks to predict corn yield from Compact Airborne Spectrographic Imager data. Computers and electronics in agriculture 2005;47(2):149-161.
28. Wang F, Fraisse CW, Kitchen NR, Sudduth KA. Site-specific evaluation of the CROPGROW-soybean model on Missouri claypan soils. Agricultural Systems. 2002;76:985-1005.
29. Wart JV, Christian K, Peng K, Milner S, Cassman M, KG. Estimating crop yield potential at regional to national scales. Field Crops Research 2013;143:34-43.
30. Xin Q, Peng G, Chaoqing Y, Le Y, Mark B, Suyker AE, Myneni RB. A Production Efficiency Model-Based Method for Satellite Estimates of Corn and Soybean Yields in the Midwestern US. Remote Sensing 2013;5(11):5926-5943.