

An Artificial Neural Network-Based Model for Predicting Boro Rice Yield in Bangladesh Using AVHRR-Based Satellite Data

Kawsar Akhand*, Mohammad Nizamuddin, Leonid Roytman

The City College of New York, Department of Electrical Engineering, New York, USA

Abstract Rice is the staple food of Bangladesh and plays a crucial role in its agro-based economy. Due to the ever-increasing population and decreasing agricultural land area, rice production has become a key issue for the Bangladeshi economy because the sector is under pressure from increasing food demand to maintain food security for its large population. This study demonstrates the successful application of Artificial Neural Network (ANN) and remote sensing satellite data in developing a reliable prediction model using Advanced Very High Resolution Radiometer (AVHRR) sensor-based vegetation health indices (the Vegetation Condition Index (VCI) and Temperature Condition Index (TCI)) and statistical yield data to predict Boro rice yield, the main rice variety in Bangladesh. We compared the predicted yield data with the actual Boro rice yield statistical data. The results obtained from this model are highly promising given that the error of prediction is less than 10%. This model is fundamentally important to the government, agricultural stakeholders and policy makers for monitoring food security and for planning crop production and trade.

Keywords Prediction Model, Artificial Neural Network, Remote Sensing, Vegetation Health, Rice Yield

1. Introduction

An accurate and timely method of crop yield prediction can help to ensure an uninterrupted flow of goods and services among the producing, processing, and marketing sectors of agriculture while minimizing the uncertainties accompanied with production, prices, trade and reservation. Bangladesh is one of the highly populated countries globally; close to 160 million people live on a land mass of 147,570 km², where agriculture plays a pivotal role in growing its economy, improving food security, reducing poverty, generating employment and bringing prosperity to its citizens. Agriculture is the single largest contributor to its economy and makes a significant contribution to the nation's employment and Gross Domestic Product (GDP)-around 45% of the total labor force is employed in this sector and contributes about 17% of the country's GDP [1]. Bangladesh is essentially an agro-based developing country where people's livelihood and economic progress is predicated on agriculture, because around 76% of its total population lives in rural areas and 90% of the rural population is directly or indirectly dependent upon agriculture which is the main source of their income [2].

Rice is the main agricultural crop and the primary staple food of Bangladesh; almost 90% of the population depend on rice for their carbohydrate demand and is widely cultivated all over the country throughout the year. Rice accounts for 94% of all the food grains produced annually-about 75% of the total cropped area and more than 80% of the total irrigated area is planted to rice [3]. Bangladesh has acquired a respectful position among major rice producing and consuming countries-it is the fourth highest rice producer and consumer globally. Despite the lack of resources and only a few potential initiatives, the country has successfully achieved self-sufficiency in rice production [4]. Therefore, rice production has a strong interconnection with its sustainable economic development and overall national progress.

In Bangladesh, there are three major rice varieties namely, Boro, Aman and Aus cultivated in three different seasons, their respective contribution to the total annual rice production is 55%, 39%, and 6%, respectively [1]. Boro is the dominant rice variety, accounting for more than half of the nation's total rice production. Based on the agro-climatic properties of different rice crop along with the seasonal variability of temperature, moisture, and rainfall, the cultivation season of Boro rice is from December to May, Aman rice cultivation season is between July and December, whereas the cultivation season of Aus rice is April to August [5]. Rice plant is a type of grass that

* Corresponding author:

kawsar0109@gmail.com (Kawsar Akhand)

Published online at <http://journal.sapub.org/ijaf>

Copyright © 2018 Scientific & Academic Publishing. All Rights Reserved

requires high temperature and adequate moisture for its growth and development. It is best suited for an area with high humidity and temperature, prolonged sunshine hours as well as an assured supply of water [6]. The favourable temperature and rainfall required throughout the life span of a rice crop ranges from 21 to 37°C and 1000-1500 mm, respectively. Bangladesh has witnesses three main seasons from a climatic standpoint— a hot and humid summer from March to May, a hot and humid monsoon season from June through October, and a cooler and drier winter from November through February. The annual average rainfall varies from 1,500 mm to 5,000 mm; similarly, the temperature varies from 12 to 35°C, whereas humidity varies from 65 to 90% [7]. Therefore, there is a strong correlation between weather conditions for rice cultivation and Bangladeshi weather. Importantly, the climate of Bangladesh is favourable for rice cultivation.

Boro rice season starts during winter particularly when the duration of days is shorter during the month of December-February and ends during summers in the month of April-May. It is a two season-crop (winter and summer). To facilitate production, planting should ideally be completed before 7TH January [5]. During Boro rice growing season from January to May, the monthly average maximum temperature, minimum temperature and average humidity ranges from 25.2°C to 33.2°C, 12.5°C to 24.5°C and 71% to 79%, respectively. Every crop has an optimum temperature range for its growth. For Boro rice crop that facilitates maximum photosynthesis under Bangladesh's climatic condition ranges from 25 to 30°C [8]. Temperature has a significant impact on Boro rice production. For example, overly hot and overly cold conditions are known to adversely affect crop production [9]. The life cycle of Boro rice crop is approximately 140-160 days, depending on the varieties of rice. It is divided into three phases: vegetative, reproductive and ripening phase. The vegetative phase assumes significance because it is characterized by the development of tillers and plant population. This phase denotes a critical stage of Boro rice life cycle because tillers are developed in the months of January and February when the temperature remains below 20°C, which is less than minimum optimum temperature and causes the plants to suffer from cold injury. Because of low temperature and cold injury, the emergence and development of tillers takes place slowly. However, 90% of the Boro rice production depends on the 1st and 2nd steps of tillers [5]. The period of vegetative stage is from January to March. January and February represent the cold and dry winter seasons in Bangladesh, but March is a hot and humid summer season. After overcoming the cold injury, in March Boro rice plants commence the rapid development of tillers and leaves, resulting gradual increase in plant height also getting preparation for the reproductive phase. Therefore, February and March are critical months for healthy and dense vegetation development in addition to better production of Boro rice.

The weather condition of an area such as temperature,

humidity, precipitation, sunshine and atmospheric carbon dioxide plays an important role in determining its agriculture production [8]. Healthy vegetation is paramount for better rice production, which is closely related to both photosynthesis and Normalized Difference Vegetation Index (NDVI). Notably, NDVI is widely used as a measure of plant productivity, an indicator of the environmental resources and a determinant of ecosystem potentiality [10]. As the most widely used vegetation index that reflects vegetation properties, NDVI is calculated from the visible and near-infrared light reflected by vegetation. Healthy vegetation absorbs most of the visible light reaching it, reflecting a large portion of the near-infrared light. Therefore, NDVI values are much higher for vegetation as compared to bare soil, cloud and water [10]. NDVI and Brightness Temperature (BT) characterize healthy or unhealthy vegetation. The Vegetation Condition Index (VCI) and the Temperature Condition Index (TCI) characterize the moisture and thermal conditions of a particular area, respectively, derived from NDVI and BT [11]. For a given area, the VCI and TCI values have a very strong correlation between these indices and agricultural crop yields during the critical period of crop growth [12]. A high value of VCI corresponds to unstressed and healthy vegetation whereas a low TCI value corresponds to unhealthy vegetation due to thermal stress [13]. For the purpose of this study, we selected VCI data for weeks 5-8 within the month of February because during this period vegetation is more responsive to moisture and optimize plant population. Meanwhile we chose TCI data for weeks 9-12 within the month of March since this period is important for the formation of healthy vegetation; during this period, vegetation responds more favorably to temperature. Moreover, the data selection period (February & March) is known as the critical period of Boro rice life cycle since better production mainly depends on this period.

Bangladesh's weather stations network is currently not sufficient to adequately collect timely, accurate weather information for efficient crop monitoring and yield well before the harvesting period. It is very difficult to generate useful weather information about agriculture i.e., crop field development, greenness, and other healthy activities only for 35 surface observatories. Meanwhile remote sensing technology makes it easier to acquire information about an object or an area without being direct contact or physically interfering with that object or area. In addition, remote sensing technique makes it possible to collect real-time weather data from dangerous and inaccessible areas. Moreover, satellite data are available in a cost-effective, regular, accurate and timely manner. Therefore, remote sensing technology has been widely used in making predictions relating to crop yield. Remote sensing data are also available for applications and will continue to provide Earth observations. Furthermore, remote sensing techniques are capable of identifying and monitoring crop growing stages, crop vigor, biomass, crop classes and crop yield predictions, also making inferences about the status of

plants as well as their environmental conditions on an almost real-time basis [14].

The performance of a particular model depends on a number of factors, such as the volume of data, selection of model, the methods of validating results and the measures used for comparison. Statistical methods have traditionally been used for making prediction and classification. To this end, regression models are not perfect in forecasting complicated situations such as non-linear and extreme values data because of the linear relationship between the parameters. On the other hand, ANN is highly applicable on complex relations and scenarios where there is a strong nonlinearity between different parameters and crop production variables. Therefore, Neural Networks (NNs) methods have emerged as a very important tool for a wide variety of applications including making forecasts about crop production in areas where traditional statistical techniques have previously been used [15]. ANN refers to a mathematical model that processes information using a way in which the biological neural system such as the brain processes information [16]. Owing to its intellectual characteristics, ANN can be used to develop a model to forecast crop yield. In addition, ANN entails self-learning, compatibility and flexibility capabilities. The study aims to develop a reliable prediction model in order to predict Boro rice yield before harvest using ANN and AVHRR sensor based real time acquired satellite data. The model delivers a great performance based on the error of prediction and accuracy.

2. Materials and Method

2.1. Data

Boro rice yield statistical data and remote sensing satellite data for the same period (1988-2011) were used in this research.

2.1.1. Boro Rice Yield Data

Boro rice yield data were collected from Bangladesh Bureau of Statistics' (BBS), which published 'the yearbook of Agricultural Statistics', and the 'Rice Database' of Bangladesh Rice Research Institute (BRRI) of the People's Republic of Bangladesh [1, 17]. BBS is a governmental authority that conducts core statistics of agricultural crop and is also responsible for accumulating, compiling and disseminating crop statistics for the entire national system. On the other hand, BRRI engages with the research and development of rice production. The BBS regularly publishes official statistical data on significant aspects of grain production such as crop type, cultivation area, production, yield etc. in their publication "the Yearbook of Agricultural Statistics" to address the users' needs. Therefore, it can be used as a reliable reference guide by policy-makers, planners, economists, researchers, international agencies and educationists in equal measure. There are two types of agricultural statistics: structural and

annual statistics are conducted in Bangladesh. According to the Food and Agriculture Organization (FAO) guidelines, structural statistics is generated by collecting data through a full count/sample census, normally at an interval of 10 years. Similarly, annual agricultural statistics are generated by collecting data through annual/seasonal sample surveys. The annual statistical Boro rice yield data from 1988-2011 were used for this research and were calculated by dividing total Boro rice production (in tons) by the sown area (hectares) [1].

2.1.2. Remote Sensing Satellite Data

National Oceanic and Atmospheric Administration (NOAA) operational polar orbiting environmental satellite mounted Advanced Very High Resolution Radiometer (AVHRR) sensor-based remote sensing satellite data were used in this study. AVHRR is an instrument (sensor) flown on satellite that measures solar energy reflected/emitted from the Earth's surface in several visible as well as infrared bands. Satellite data were presented by weekly Normalized Difference Vegetation Index (NDVI) and Brightness Temperature (BT) collected from the NOAA Global Vegetation Index (GVI) dataset from 1998 to 2011. The GVI products are widely used as satellite data products for monitoring the vigor and density of green vegetation, estimating crop acreage, detecting plant stress and classifying land cover. The GVI was developed from AVHRR-observed reflectance/emission in visible (VIS, 0.58-0.68 μ m, Ch1), near infrared (NIR, 0.72-1.00 μ m, Ch2), and infrared (IR, 10.3-11.3 μ m, Ch4) wavelengths. Meanwhile the GVI dataset was developed by sampling 4 km² Global Area Coverage (GAC) data to 16 km² spatial resolution along with daily observations to seven day composite [12, 18]. NDVI was calculated by pre- and post-launch calibrated VIS and NIR reflectance as follows:

$$NDVI = (NIR - VIS) / (NIR + VIS) \quad (1)$$

The IR measurements in 10.3-11.3 μ m wavelengths were converted into BT. NDVI and BT have a high frequency noise due to clouds residual, sun-sensor angular effects and bidirectional reflectance etc. which makes it difficult to use them. In order to remove noise the time series values of NDVI and BT were smoothed by applying statistical techniques [12].

2.1.3. Development of Vegetation Health Indices

Vegetation health indices were developed from the fundamental properties of green vegetation, the reflection and emission of solar radiation in visible (VIS), near-infrared (NIR) and infrared (IR, ch4) of spectral channels. Healthy vegetation reflects less radiation due to high chlorophyll absorption of solar radiation within the VIS channel and reflects more radiation due to the scattering of light by leaf internal tissues and water content in the NIR channel. In addition, it emits less thermal radiation due to the transpiring canopy is cooler in the IR spectral band [5]. As a result, NDVI becomes large and BT remains small for healthy

vegetation; this scenario reverses for unhealthy vegetation [19].

Vegetation health indices Vegetation Condition Index (VCI) and Temperature Condition Index (TCI) of a particular area were developed from its NDVI and BT using some processing and algorithms. The details of these methods and algorithms to calculate NDVI and BT from VIS, and NIR as well as IR spectral bands and GVI from NDVI, BT, are presented in Kogan [12, 20, 21]. This paper briefly mentioned some methods which include: (a) elimination of temporal high frequency noise (clouds, Sun and sensor angular effects, etc.) from NDVI and BT annual time series; (b) calculation of the mean annual cycle; (c) calculation of climatology from multi-years of data; and (d) estimation of medium-to-low frequency fluctuations NDVI and BT (departure from climatology) associated with weather variations. VCI characterizing moisture condition and TCI characterizing thermal conditions in response to weather impacts were calculated using the following equations:

$$VCI = 100(NDVI - NDVI_{min}) / (NDVI_{max} - NDVI_{min}) \quad (2)$$

$$TCI = 100(BT_{max} - BT) / (BT_{max} - BT_{min}) \quad (3)$$

where NDVI, NDVI_{min}, and NDVI_{max} represent the smoothed weekly NDVI, its multi-year absolute minimum and maximum, respectively; BT, BT_{min}, and BT_{max} denote similar values for BT. VCI and TCI algorithms were developed using NDVI and BT which separates the weather component and ecosystem component by Max-Min criteria [12, 19]. On the other hand, VCI and TCI indices are numeric values denoted in the range of 0 to 100. The numeric values of VCI changes from 0 to 100, reflecting the changes in moisture condition from severe bad to favorable whereas TCI value changes from 0 to 100, reflecting the changes in thermal condition from extremely stressed (dryness) to favorable (healthy and unstressed vegetation).

2.2. Artificial Neural Network

An Artificial Neural Network (ANN) simulation tool was used to develop a prediction model for this study. A nonlinear auto regressive with exogenous (external) input, or NARX time series, Neural Network was used for early predictions and the model development. An ANN refers to an interconnected group of “nodes” known as “neurons”, “neurodes”, “processing elements” or “units” - the basic building block to form a network which functions in the same manner as human central nervous systems. Fundamentally, ANNs are referred to as a “black box” approach. Similar as ‘black box’ technology, ANN has the ability to learn input–output correlation by training the input to produce the expected output [22]. NN training is accomplished in order to perform a particular task by adjusting the weight and bias values between the elements. As a consequence of adjusting or training the network, a particular input (target) leads to a specific output [23]. Figure 1 demonstrates that neural network performs adjustment process until the output matches the target based on a comparison between them.

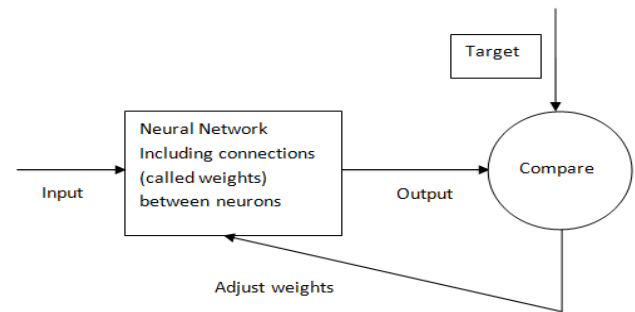


Figure 1. Artificial Neural Network (ANN) basic principle diagram

Training signifies a key feature of neural network simulation software. Therefore, the selection of an appropriate training function is an important issue for NN toolbox software and for developing a NN based model. For this model ‘trainlm’ training function is used which is the fastest training function for back-propagation algorithm, in addition to, it is the default training function for the feed-forward network that updates weight and bias values in accordance to Levenberg-Marquardt optimization. According to a pre-defined percent for data division, the input data are randomly divided into training, validation and test data sets while performing the network training. For this research, data divisions were selected 70% for training, 15% for validation and 15% for testing. The network uses training data set to adjust the network based on the error between target (input) and predicted (output) value, validation data set to measure network generalization, and test data set to measure the performance of an individual network. Validation data set has an impact on the accuracy of network because training stops automatically when generalization ceases to improve, which is indicated by an increase in the mean square error (mse) of the validation samples. Every attempt of training will generate a different result because a different set of data are selected for training, validation and testing on each occasion [24].

The weight and bias denote two important parameters for NN training; they are the adjustable parameters of the network processing unit (neuron). The weight and bias values are adjusted to reduce error and obtain the optimum output [25]. Network input gets connected to the hidden layer neuron through input weight; on the other hand, the hidden layer neuron is connected to the output layer neuron through line weight. As a processing unit, each neuron has input(s), a bias, a summer, a transfer function and an output. The actual output of a neuron depends on the particular transfer function. For our proposed model, the log-sigmoid transfer function was chosen for hidden layer neurons, which takes the input value between plus and minus infinity and squashes the output into a range between 0 & 1. A linear transfer function, which produces its input as the output, was chosen for the output layer neuron.

2.3. Proposed Model

The proposed model is a multi-layer feed-forward back-propagation ANN used in order to develop a Boro rice

yield prediction model for Bangladesh, as illustrated in figure 2.

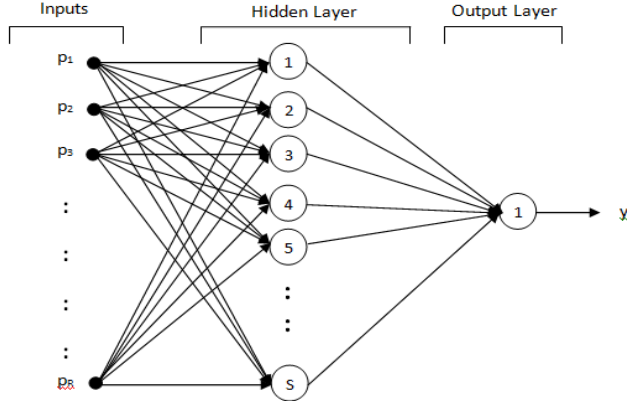


Figure 2. Feed-forward back propagation ANN with layers and connections

In this study, we used the satellite data product weekly VCI and TCI data as input and statistical Boro rice yield data as the target for the same duration of 24 years from 1988-2011. Week 1 denotes the first week of the year (1st week of January), while week 52 denotes the last week of the year (last week of December). We used VCI for weeks 5-8 and TCI for weeks 9-12 from 1988-2011 (24 years) as the input data, which signifies a 1×24 cell array of 8×1 matrices, representing the dynamic data of 24 time steps of 8 elements. We used same time period (1988-2011) for Boro rice yield data as target data which represents the dynamic data of 24 time steps of 1 element. Each element of the input vector (vector \mathbf{p}) is connected to every neuron of the hidden layer through input weight matrix \mathbf{IW} ; each hidden layer neuron is also connected to the output neuron through line weight matrix \mathbf{LW} .

During network training, the simulation window shown in figure 3 indicates the random division of target data set: out of the 24 samples, only 16 (70%) were considered for training; 4 (15%) for validation; and the remaining 4 (20%) for testing.

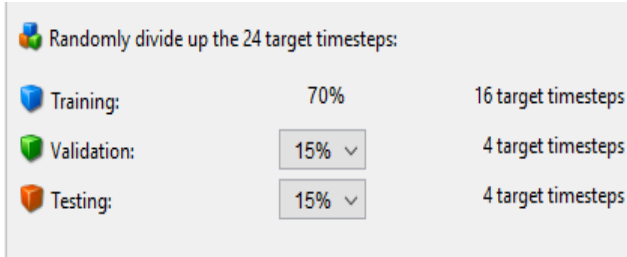


Figure 3. Simulation window of Data division for training, validation and testing

In figure 2 model diagram ‘y’ denotes the predicted value or the simulated result for Boro rice yield from the model, whereas S represents the number of neurons in the hidden layer. For our proposed model, we selected 8 (eight) neurons for the hidden layer using the trial and error method and one neuron for the output layer because output layer neuron is selected based on the number of category of element on the

target vector. From network simulation, it was found that the total number of weight and bias values comes up to 161. Of the 161, there were 9 bias values, because for our proposed model we used 8 neurons for the hidden layer and 1 (one) neuron for the output layer, with each neuron having only one bias value. The remaining 152 values are weight values. The matrices for weights and biases from matlab simulation are illustrated below, where \mathbf{b} refers to the bias, \mathbf{IW} is the input weight and \mathbf{LW} denotes the line weight.

$$\begin{aligned} \mathbf{b} &= [8 \times 1 \text{ double}] \\ &= [0.0047] \\ \mathbf{IW} &= [8 \times 16 \text{ double}] \quad [8 \times 2 \text{ double}] \\ \mathbf{LW} &= \begin{bmatrix} & \\ & \\ & \\ & \\ & \\ & \\ & \\ & \end{bmatrix} \quad \begin{bmatrix} \\ \\ \\ \\ \\ \\ \\ \end{bmatrix} \\ &= [1 \times 8 \text{ double}] \quad [1 \times 1 \text{ double}] \end{aligned}$$

The selection of neurons for the hidden layer is one of the main challenges in designing an ANN. The number of neurons in the hidden layer impacts the learning capability, complexity, generalization, and accuracy of the problem. Therefore, the selection of an appropriate number of neurons for the hidden layer suffices to ensure the complexity of the problem. However, if this becomes too large, it may lead to the over-fitting of the training set and loss of generalization ability. Hence, trial and error method is the ideal way of selecting hidden neurons. The proposed model, which reflects a time series prediction model, is composed of two types of inputs: (a) weekly VCI & TCI time series data and (b) Boro rice yield time series data. Here, $y(t)$ in equation ‘4’ refers to the Boro rice yield time series whereas $x(t)$ signifies the weekly VCI & TCI time series. The values of Boro rice yield time series are forecasted using this time series as well as VCI & TCI time series. This type of prediction is called nonlinear auto regressive with exogenous (external) input, or NARX. It can be written as:

$$y(t) = f(y(t-1), y(t-2), \dots, y(t-d), x(t-1), x(t-2), \dots, x(t-d)) \quad (4)$$

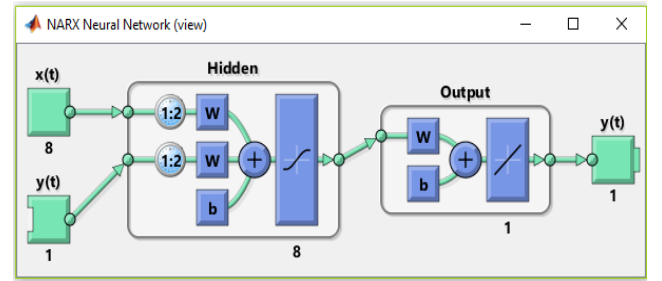


Figure 4. NARX Neural network Boro rice yield prediction model simulated diagram

Figure 4 illustrates the simulated NARX Neural Network diagram of the proposed model. It can be clearly seen that the two-layer feed-forward network with two sets of data ($x(t)$ & $y(t)$) is applied as inputs, log-sigmoid transfer function in the hidden layer and linear transfer function in the output layer. There are 8 neurons in the hidden layer and 1 neuron in the output layer; the output (predicted value) is $y(t)$, which denotes Boro rice yield. For this study, trial and error method

was selected to determine the appropriate number of hidden neurons in order to design the proposed model. Accordingly, we tested ANN for 6, 8, 10, 12, 15, 20, 25, and 30 neurons in the hidden layer. In order to select the best model, comparisons were made between the results to determine the best performance from the proposed ANN model, such as the errors of prediction. It was found that optimum performance was achieved for the 8 hidden neurons.

Table 1 shows the predicted value (simulated result) generated by the proposed model for a different number of hidden neurons. It was found that the hidden layer with 8 neurons generated the best result. Based on the error of prediction between actual and simulated (predicted) values derived from table 2, it was found that the error of prediction in each year for these 8 hidden neurons is reasonable and was less as compared to other hidden neurons. Therefore, we selected 8 hidden neurons for our proposed prediction model due to the fact that we got the optimal prediction result from

year 1990 to 2011.

3. Results and Discussion

3.1. Performance Analysis (Validation)

The performance of the proposed feed-forward ANN prediction model is evaluated and selected using the error of prediction for different neurons within the hidden layer, as illustrated in table 2. With a view to determine the best performance, the Mean Square Error (MSE) - which is the average squared error between the actual (y_a) and the predicted (y_p) data - is calculated for different hidden neurons using the following equation and the result shown in table 3.

$$F = MSE = \frac{1}{N} \sum_{i=1}^N (y_a - y_p)^2$$

Table 1. Boro rice yield output (simulated result) for different neurons in the hidden layer

Year	Target/Actual yield (Ton/Hectare)	Predicted yield (Ton/Hectare) for different no. of neurons in hidden layer							
		6	8	10	12	15	20	25	30
1988	2.39								
1989	2.46								
1990	2.49	2.5658	2.6322	2.4880	2.6013	2.5835	2.6814	2.2997	2.6648
1991	2.58	2.5162	2.5847	2.5774	2.6460	2.6650	2.6794	2.6669	2.6271
1992	2.53	2.5152	2.5335	2.5273	2.5819	2.7183	2.6029	2.6363	2.4340
1993	2.62	2.4653	2.6221	2.6125	2.6519	2.7509	3.2248	2.6398	2.6967
1994	2.45	2.3270	2.6509	2.4603	2.6555	2.7228	2.7856	2.5618	2.5284
1995	2.62	2.6577	2.5735	2.6111	2.6699	2.6572	2.5250	2.7180	2.6921
1996	2.68	2.7325	2.7017	2.6772	2.6490	3.1905	3.1701	2.6612	2.7893
1997	2.82	2.8290	2.8500	2.8149	2.8560	2.9256	3.0681	2.7937	2.9133
1998	2.99	3.1196	2.9677	3.5259	2.9178	3.0629	2.7059	2.9645	2.4976
1999	3.02	3.1166	3.0133	3.2846	2.9678	3.0893	2.7591	3.201	3.0101
2000	3.17	3.4789	3.3155	3.5034	3.1651	3.0878	3.2577	3.1874	3.1966
2001	3.12	3.4844	3.1196	3.1244	3.1245	3.2213	3.5208	3.1183	2.4130
2002	3.18	3.0968	3.1267	3.2039	3.1067	3.2625	3.3374	3.3264	3.9845
2003	3.26	3.2651	3.2722	3.3740	3.2717	3.0899	3.5399	3.2432	3.3909
2004	3.40	3.3147	3.3914	3.3924	3.3985	3.3132	3.2166	3.3563	3.5267
2005	3.44	3.3940	3.3893	3.5358	3.4317	3.4637	3.4430	3.4284	3.4132
2006	3.52	3.4899	3.5211	3.5178	3.5191	3.2487	3.3847	3.4776	3.5440
2007	3.85	3.7546	3.8268	3.8263	3.8319	3.8195	3.2120	3.4445	3.3223
2008	3.78	3.6696	3.6195	3.7801	3.7886	3.4995	3.5224	3.7416	3.7402
2009	3.84	3.8113	3.8283	3.8307	4.2243	4.0307	4.0060	3.7715	3.7377
2010	3.90	3.9071	3.8855	3.9041	3.9028	4.0602	3.8696	3.8831	3.6759
2011	3.90	3.9384	3.8939	3.8983	3.5310	3.5884	3.7237	3.7460	3.8349

Table 2. Error of prediction of Boro rice yield for different neurons in the hidden layer

Year	Target/Actual yield (Ton/Hectare)	Error of prediction for different no. of neurons in hidden layer							
		6	8	10	12	15	20	25	30
1988	2.39								
1989	2.46								
1990	2.49	-0.0758	-0.1422	0.0019	-0.1113	-0.0935	-0.1914	0.1902	-0.1748
1991	2.58	0.0637	-0.0047	0.0025	-0.0660	-0.0850	-0.0994	-0.0869	-0.0471
1992	2.53	0.0147	-0.0035	0.0026	-0.0519	-0.1883	-0.0729	-0.1063	0.0959
1993	2.62	0.1546	-0.0021	0.0074	-0.0319	-0.1309	-0.6048	-0.0198	-0.0767
1994	2.45	0.1229	-0.2009	-0.0103	-0.2055	-0.2728	-0.3356	-0.1118	-0.0784
1995	2.62	-0.0377	0.0465	0.0088	-0.0499	-0.0372	0.0949	-0.0980	-0.0721
1996	2.68	-0.0525	-0.0217	0.0027	0.0309	-0.5105	-0.4901	0.0187	-0.1093
1997	2.82	-0.0090	-0.0300	0.0050	-0.0360	-0.1056	-0.2481	0.0262	-0.0933
1998	2.99	-0.1296	0.0223	-0.5359	0.0721	-0.0729	0.2840	0.0254	0.4923
1999	3.02	-0.0966	0.0067	-0.2664	0.0521	-0.0693	0.2608	-0.1813	0.0098
2000	3.17	-0.3089	-0.1455	-0.3334	0.0048	0.0821	-0.0877	-0.0174	-0.0266
2001	3.12	-0.3644	0.0004	-0.0044	-0.0045	-0.1013	-0.4008	0.0016	0.7069
2002	3.18	0.0831	0.0533	-0.0239	0.0732	-0.0825	-0.1574	-0.1464	-0.8045
2003	3.26	-0.0051	-0.0122	-0.1140	-0.0117	0.1700	-0.2799	0.0167	-0.1309
2004	3.40	0.0852	0.0086	0.0075	0.0014	0.0867	0.1833	0.0436	-0.1267
2005	3.44	0.0459	0.0507	-0.0958	0.0082	-0.0237	-0.0030	0.0115	0.0267
2006	3.52	0.0300	-0.0011	0.0021	0.0008	0.2712	0.1352	0.0423	-0.0240
2007	3.85	0.0953	0.0232	0.0236	0.0180	0.0304	0.6379	0.4054	0.5276
2008	3.78	0.1103	0.1605	-0.0001	-0.0086	0.2804	0.2575	0.0383	0.0397
2009	3.84	0.0286	0.0117	0.0092	-0.3843	-0.1907	-0.1660	0.0684	0.1022
2010	3.90	-0.0071	0.0145	-0.0041	-0.0028	-0.1602	0.0303	0.0168	0.2240
2011	3.90	-0.0384	0.0061	0.0016	0.3689	0.3115	0.1762	0.1538	0.0650

Table 3. Performance of the proposed model for different hidden neurons

Performance of NN prediction model	No. of Neurons in the hidden Layer							
	6	8	10	12	15	20	25	30
MSE	0.0158	0.0053	0.0224	0.0165	0.0363	0.0839	0.0151	0.0840

Table 4. Comparison of regression performance for different hidden layer neurons

	Regression (R) value for different number of neurons in the hidden layer							
	6	8	10	12	15	20	25	30
Training	0.993	0.999	0.999	0.999	0.967	0.910	0.996	0.988
Validation	0.991	0.999	0.968	0.993	0.927	0.622	0.978	0.436
Test	0.871	0.968	0.961	0.977	0.960	0.899	0.884	0.959

where N represents the number of samples 22 (1990-2011) shown in column 1, y_a denotes the actual Boro rice yield shown in column 2, y_p is the predicted Boro rice yield shown in column 3-10 in table 1 and $e_y = (y_a - y_p)$ refers to the error between the actual and predicted Boro rice yield, as shown in column 3-10 within table 2. The performance of NN (based on MSE) for different hidden neurons is illustrated in table 3; it is seen that optimum outcome occurs for 8 hidden neurons which was selected for our proposed model to predict Boro rice yield in Bangladesh. The model performance assumes significance because the final mean square error is small.

Figure 5 displays the graphical representation of simulated

performance plot of our proposed model. It is plotted between Mean Square Error (MSE) and epochs for training, validation, and test performance of the network. The performance plot demonstrates that the MSE for training, validation, and testing are decreased until epoch 3; after that, only training MSE decreases but validation and testing MSE increase until epoch 6 after which, network training was stopped. The plot clearly shows that the best validation performance occurred at epoch 3. Training stopped when generalization ceased to improve, such as an increase in validation error which occurred after epoch 6. No data over-fitting occurred because the validation and test curves

were very similar. At the same time, there was no increase in testing or validation error before epoch 3. In addition, the test curve did not increase significantly before an increase in the validation curve.

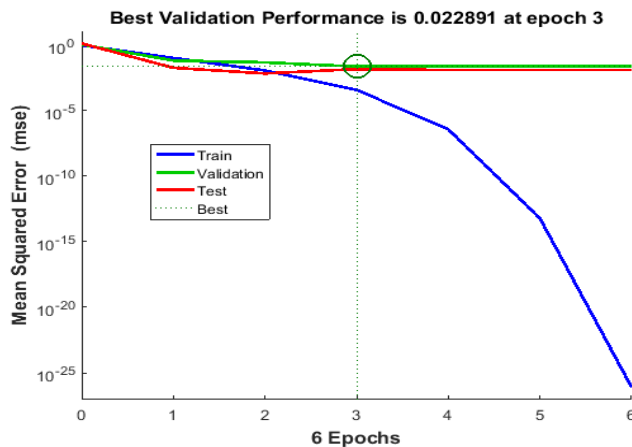


Figure 5. Neural Network training performance plot with best validation performance

3.2. Data Relationship Analysis

Regression value generated by ANN simulation during network training is a good indicator to determine the strength of the relationship between Boro rice yield actual data and the corresponding predicted data. It was ascertained whether network training signified the perfect target and predicted values were the same, but it happened rarely during practice. Regression plot shows the relationship between target and output of the network. In our proposed model, regression plots for training, validation, testing and all data sets represented by four axes are illustrated in figure 6.

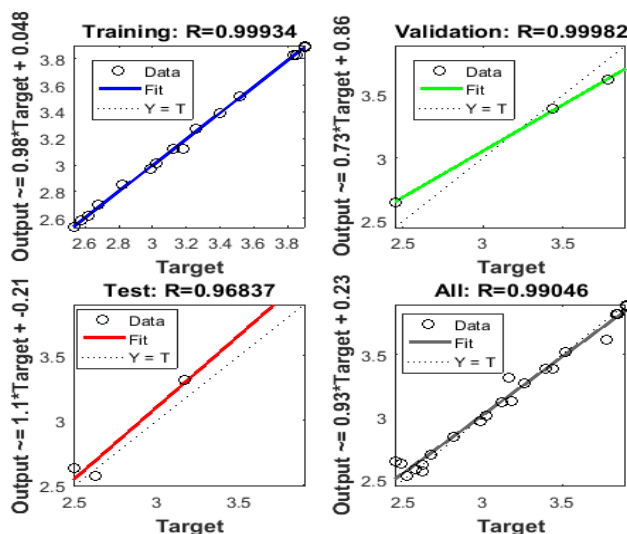


Figure 6. Model regression plots for training, validation, test and all data

Each plot has two lines: the dashed line represents the perfect correlation when the output is exactly equal to the target whereas the solid line represents the model's best fit between the output and target. It is evident that there is a good linear relationship between the predicted value and

target for all four situations.

Regression (R) value indicates the relationship between the actual and predicted value. R value close to 1 indicates the best linear relationship. If R is equal to 1, it means there is an exact linear relationship. Similarly, the value of R close to zero demonstrated an absence of a linear relationship between output and target, which indicates a random relationship. Simulation results of the regression value based on different neurons within the hidden layer are shown in table 4. In comparison, it is found that the 8 neurons we selected for our model gave the best result. R values for training, validation and test data sets are shown to be greater than 0.96 and for all the response is 0.99. This indicates the accuracy of our model is very high. Therefore, the success of ANN-based Boro rice yield prediction model is adequately proven.

3.3. Model Accuracy Analysis

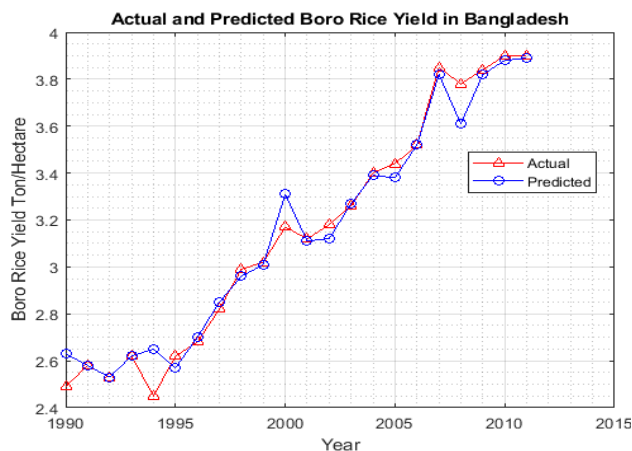
The accuracy of our model was evaluated based on the comparison between actual and predicted data. To select the optimal model, we tested ANN for varying number of neurons (6, 8, 10, 12, 15, 20, 25 & 30) in the hidden layer and calculated the error of prediction as shown in table 2. We selected 8 hidden neurons for our model because the MSE was lowest in comparison to others, as shown in table 3. To determine the accuracy of the selected model, we calculated the percent error of prediction between the actual (target) Boro rice yield statistical data and the predicted data (NN simulated output). We calculated the percentage error of prediction using the following equation, the result shown in table 5.

$$\% \text{ error} = \frac{\text{Actual yield} - \text{Predicted yield}}{\text{Actual yield}} \times 100$$

The accuracy of the model is examined by using data from table 5. Column 5 of table 5 shows that the percentage error of prediction for each of the 22 years (from 1990 to 2011) is very small. The maximum error is 0.2009 (8.2%), whereas the minimum error is 0.004 (0.01%). In addition, the percentage error of prediction for each year is less than 10% and 90% of these relative errors are less than 5%. These interpretations about the actual outcomes indicate that the proposed model has a high level of accuracy for predicting Boro rice yield in Bangladesh. Therefore, we can safely conclude that the ANN based prediction model is highly promising. In order to compare the target and predicted data in a more efficient manner, a graph is plotted using the two types of data. Figure 7 shows a graphical comparison of the predicted and actual Boro rice yield in Bangladesh. This graph represents a high degree of similarity between the predicted and actual values, which clearly indicates a high level of accuracy. Therefore, the credibility of ANN-based Boro rice yield prediction model being a potential simulation tool that can be used to produce a reliable and accurate yield prediction is reinforced; furthermore, this model can be applied for other crops as well. Hence, it is proven that ANN has a strong prediction capability.

Table 5. Target and predicted Boro rice yield and percent error of prediction in Bangladesh

Year	Target/Actual	Predicted	error	% of error
1988	2.39			
1989	2.46			
1990	2.49	2.6322	-0.1422	-5.71%
1991	2.58	2.5847	-0.0047	-0.18%
1992	2.53	2.5335	-0.0035	-0.14%
1993	2.62	2.6221	-0.0021	-0.08%
1994	2.45	2.6509	-0.2009	-8.2%
1995	2.62	2.5735	0.0465	1.77%
1996	2.68	2.7017	-0.0217	-0.81%
1997	2.82	2.8500	-0.0300	-1.06%
1998	2.99	2.9677	0.0223	0.74%
1999	3.02	3.0133	0.0067	0.22%
2000	3.17	3.3155	-0.1455	-4.58%
2001	3.12	3.1196	0.0004	0.01%
2002	3.18	3.1267	0.0533	1.68%
2003	3.26	3.2722	-0.0122	-0.37%
2004	3.40	3.3914	0.0086	0.25%
2005	3.44	3.3893	0.0507	1.47%
2006	3.52	3.5211	-0.0011	-0.03%
2007	3.85	3.8268	0.0232	0.60%
2008	3.78	3.6195	0.1605	4.25%
2009	3.84	3.8283	0.0117	0.30%
2010	3.90	3.8855	0.0145	0.37%
2011	3.90	3.8939	0.0061	0.16%

**Figure 7.** Graphical comparison of the predicted and actual Boro rice yield in Bangladesh

4. Conclusions

The agriculture and rice sector in Bangladesh has remarkable profound impact on the food and livelihood security of its population. Reliable and timely yield prediction is an important source of information that enables the government, entrepreneurs and farmers to take preventive measures, avoid unpleasant situations, minimize

the risks, and make well-informed decisions. This study developed a feasible prediction model to predict Boro rice yield in Bangladesh using ANN as a simulation tool and AVHRR sensor-based vegetation health indices characterizing both moisture condition (VCI) and thermal condition (TCI) as predictor variables. In order to develop an appropriate model, a different number of neurons within the hidden layer were tested, and it was found that the hidden layer with eight neurons generated the optimal prediction value. It was seen that increasing the number of neurons in the hidden layer did not contribute better prediction. This research showed that Boro rice yield prediction can be enhanced using the ANN and satellite data product VCI & TCI at approximately two to three months before the harvest at a prediction error of less than 10%. Therefore, this prediction model can serve as an important tool for planners, policymakers, economists, analysts, academics, educationists, researchers, NGOs as well as other stakeholders. This model is potentially applicable for other rice producing countries worldwide where remote sensing and statistical data are available for use.

REFERENCES

- [1] Yearbook of Agricultural Statistics-2016. Bangladesh Bureau of Statistics (BBS), Statistics and Informatics Division, Ministry of Planning, Government of the people's republic of Bangladesh, <www.bbs.gov.bd>.
- [2] THE WORLD BANK- IBRD-IDA, "Bangladesh's Agriculture: A poverty reducer in need of modernization", <<http://www.worldbank.org>>.
- [3] BRKB, Bangladesh Rice Knowledge Bank. Bangladesh Rice Research Institute, Gazipur, Bangladesh, <<http://www.knowledgebank-brrri.org/riceinban.php>>.
- [4] Raihan, S., "The Political economy of food price policy-The case of Bangladesh," UNU-WINDER. 2013, No. 2, ISBN 978-92-9230-579-6.
- [5] Adhunik Dhaner Chash 2017. Bangladesh Rice Research Institute, Gazipur-1701, Bangladesh.
- [6] Bhagirath, S.C., Khawar, I. and Gulshan, M., "Rice Production Worldwide," Springer, 2017, pp, 52-58; ISBN-3319475169, 9783319475165.
- [7] Rahman, A., Kogan, F. and Roytman, L., "Analysis of malaria cases in Bangladesh with remote sensing data," Am. J. Trop. Med. Hyg., 74, 17–19 (2006).
- [8] Basak, J. K., "Climate change impact on rice production in Bangladesh: Results from a model," Unnayan Onneshan-The innovators, Dhanmondi, Dhaka-1209.
- [9] Toshichika, I. and Navin, R., "How do weather and climate influence cropping area and intensity?," Global Food Security, Vol. 4, p.46-50 (2015).
- [10] Kogan, F.N., "Application of vegetation index and brightness temperature for drought detection," Adv. Space Res., Vol. 15.

- No. 11, pp. 91-100 (1995).
- [11] Salazar, L., Kogan, F. and Roytman, L., "Using vegetation health indices and Partial least square method for estimation of corn yield," *International Journal of Remote Sensing*, Vol. 29, p. 175-189 (2008).
 - [12] Kogan, F.N., "Global drought watch from space," *Bulletin of the American Meteorological Society*, Vol. 78, p. 621-636 (1997).
 - [13] Bhuiyan, C., Saha, A. K., Bandyopadhyay, N. and Kogan, F. N., "Analyzing the impact of thermal stress on vegetation health and agricultural drought- a case study from Gujarat, India," *GIScience & Remote Sensing*, Vol. 54, p. 678-699 (2017).
 - [14] Islam, A. K. and Bala, S. K., "Assessment of Potato Phenological Characteristics using MODIS-Derived NDVI and LAI Information," *GIScience & Remote Sensing*, 45, No. 4, p.1-17 (2008).
 - [15] Paswan, R.P. and Begum, S.A., "Regression and Neural Networks Models for Prediction of Crop Production," *International Journal of Science & Engineering Research*, Vol. 4, Issue 9, 98-108 (2013).
 - [16] Saravanan, S., Kannan, S. and Thangaraj, C., "Forecasting India's Electricity Demand Using Artificial Neural Network," *International conference on Advances In Engineering, Science and Managament (ICAESM-2012)*, Nagapattinam, Tamil Nadu, India, 30-31 March 2012, IEEE, 2012, ISBN: 978-81-909042-2-3 © 2012 IEEE.
 - [17] Rice Database. Bangladesh Rice Research Institute, Government of the People's Republic of Bangladesh, <<http://www.brri.gov.bd>>, last accessed on October 5, 2017.
 - [18] Kidwell, K. B., "Global Vegetation Index user's guide," U.S. Department of Commerce, NOAA, National Environmental Satellite Data and Information Service, National Climatic Data Center, Satellite Data Services Division: Camp Springs: MD, U.S.A., 1997.
 - [19] Jensen, J.R., "Remote sensing of the Environment: An earth resource perspective," Prentice Hall: Chandler, AZ, U.S.A., 2000.
 - [20] Kogan, F.N., "Operational space technology for global vegetation assessment," *Bull Am. Meteorol. Soc.*, 82, 1949-1964 (2001).
 - [21] Kogan, F., Stark, R., Gitelson, A., Jargalsaikhan, L., Dugrajav, C. and Tsooj, S., "Derivation of pasture biomass in Mongolia from AVHRR-based vegetation health indices," *Int. J. Remote Sensing*, Vol. 25, N. 14, p. 2889-2896 (2004).
 - [22] Ismael, M. and Ibrahim, R., "Adaptive neural network prediction model for energy consumption," *ICCRD 2011 3rd International conference*, Shanghai, China, 11-13 March 2011, IEEE, 2011, DOI. 10.1109/ICCRD.2011.5763864.
 - [23] Beale, M., Hagan, M. and Demuth, H., "Neural Network Toolbox™ Getting Started Guide," Matlab, Mathworks, R2017b.
 - [24] Beale, M., Hagan, M. and Demuth, H., "Neural Network Toolbox™ User's Guide," Matlab, Mathworks, R2017b.
 - [25] Mohan, B.T., Vilas, N. G. and Prashant, P.B., "Hourly Load Forecasting Using Artificial Neural Network for a Small Area," *IEEE-International Conference On Advances in Engineering, Science And Management (ICAESM-2012)*, 30-31 March, 2012, ISBN: 978-81-909042-2-3 @ 2012 IEEE.