Turkish Online Journal of Qualitative Inquiry (TOJQI)
Volume 12, Issue 10, October 2021: 416-427

Forecasting Crop Yield Using Discrete Wavelet Transform and Deep Neural Networks [1]

Pankaj Gumber[1],Lal Chand Panwar[2]

Research Scholar, Punjabi University, Patiala, India[1]

Assistant Professor, Punjabi University, Patiala, India[2]

Corresponding Author:pankajgumber06@gmail.com

Abstract

Agriculture is undergoing a metamorphosis due to several environmantal and scoal factors. Due to challenges such as global warming, intermittent rainfall patterns and eroding nutrient values of soil, crop yileds have become more upredictable in the last decade. This has resulted in famines, armer suicides and deaths due to hunger. Thus, one of the key objectives of the world health organization is to provide food security globally and also help the agriculture community as a whole whith special emhpasis on low income group countries. This has made crop yield forecasting extremely important. As the crop yield depends on several factors which are highly uncorrelated in nature, hence machine learing based appraoches have been employed for the purpose. In this paper a deep neural network approach has been proposed along with the discrete wavelet transfrom to forecast crop yields. The wavelet transform has been used as a filtering techniques to remove local disturbances from the data, and deep neural networks have been used for pattern recognition and forecasting. The evaluation of the proposed system has been evaluated in terms of the mean absolute percentage error, accuracy and regression. It has been found that the proposed work outperforms existing baseline techniques in terms of the accuracy of forecasting.

Keywords:crop yiled forecasting, discrete wavelet transform, deep neural networks, mean absolute percentage error, accuracy.

Introduction

One of the main goals of the world health organization (WHO) is the food security programme which aims a proving food to everyone in the world so as to eradicate deaths due to hunger. This is however challenging due to factors such as explosive population increase, global warming, unprecedented climate changes, eroding soil nutrient values, urbanization, conversion of farmland for industrial uses, mass exodus to urban areas to seek livelihood, decling investments in staple crop production, increase in food costs etc. Thus it becomes extremely challenging to ensure food

---

[1]The initial findings of the study were presented at the "Xxx Conference” held in Xxx (Country) on 7-8th December 2012.

[2]Academic title, University, Faculty, Department, e-mail address, ORCID ID

Received: 01.01.2016, Accepted: 01.01.2016

security. As per the statistics of WHO, almost 25,000 people die of hunger each day (Holmes, 2020). It is estimated that a child dies of hunger every 10 seconds and around 3.1 million children die of hunger and malnutirtion each year. The worst hit areas are the Sub-Saharan region in Africa and Asian Countries where the deaths due to hunger are staggeringly large. This leads to the motivation of the WHO to eradicate hunger related deaths by 2030 (Moseley and Battersby, 2020). The situation needs meticulous planning and statistical analysis so as to eradicate hunger related deaths. Crop yield forecasting is one of the key component for the purpose which can render insights into the expected yields therby helping authorities to plan for storgae, distribution and supply of surplus to the needy. Crop yiled forecasting is however channenging due to its dependability on several factors such as time of the year, crop type, amount of rainfall, temperature, type and condition of soil etc (Klompenburg et al., 2020).Attaining maximum crop yield with minimum production cost remains the main goal of crop yield production (Elavarasan and Vincent, 2020) . There happen to be many challenges associated with the crop yield prediction method. The domain of artificial intelligence has helped in understanding and analyzing the agricultural based markets. Early detection of the problems related to crop yield production can help in quick resolution and aid in increasing yield profit. Predictive methods can be implemented to reduce losses under unforeseen circumstances. Moreover, the prediction methods can be utilized to know the favorable time for growing conditions. Different weather conditions have different kinds of impact on the overall crop yield of a particular area (Dang et al, 2020).

There has been a tremendous growth of artificial intelligence and machine learning in the recent years. The agro based systems and industries have also witnessed an increased adoption of these technologies. This domain has been a prominent area of research for accurate prediction crop yield (Nigam et al, 2019). With the use of meteorological data, it is quite efficient to predict the weather and pest impact on the crops. Several factors affect the yield of crops in some or the other way. For farmers, the crop yield and productivity is of vital importance. Weather conditions are one of the key influencers for the crop yield production. Different types of crops have different factors that impact the respective yield. Hence the motivation behind the research is to evaluate crop yield prediction techniques using the concepts of machine learning. In this paper, machine learning based techniques are analyzed and a model employing data pre-processing and ensemble learning is proposed for crop yield forecasting.

Methodology

The main challenge pertaining to the forecasting of crop yield lies in the fact that crop yields are affected by several variables which often show a very little correlation (Gopal and Bhargavi, 2019). Hence it is necessary to design a forecasting which can identify the patterns in the seemingly random data, be able to remove the noisy component around the baseline and forecast the crop yield with high accuracy and low error (Hird J and McDermid, 2009). To attain the objective of high forecasting accuracy and low or moderate number of training iterations, it is necessary to focus on two fundamental aspects:

1) Pre-Process the data so as to remove the noisy component along the baseline.
2) Design an appropriate machine learning alorithm which can find patterns in complex time series data.

Thus the first part of the methodology focusses on the pre-processing part to remove the noisy part and filter the data so as to facilitate training.

A. *Data Pre-Processing*

The pre-processing is done employing the discrete wavelet transform which acts as a recursive filter to filter out local disturbances and noisy nature of the raw data. This step helps in pattern recognition (Khandelwal et al, 2015). The recursive filtration using the wavelet transform for 'ith' level scaling factor can be expressed as (Nury et al, 2017):

$$Sc_{i,t} = \sum_{t=0}^{L-1} k_l S_{i-1,(2t+1-l) mod\ N_{i-1}} \quad (1)$$

The co-efficient value $C_{i,t}$ for an ith level can be expressed as (Madan and Mangipudi, 2018).

$$C_{i,t} = \sum_{i=0}^{1} k_l S_{i-1,(2t+1-l) mod\ N_{i-1}} \quad (2)$$

Here,

k_l and m_l can be expressed as:

$$k_l = (-1)^l m_{L-l-1} \quad (3)$$

$$m_l = (-1)^{l+1} k_{L-l-1} \quad (4)$$

$$for\ l = 0, \dots \dots, L-1 \quad (5)$$

t is the time metric

S is the data stream to which needs to be filtered

$Sc_{i,t}$ is the scaling metric

The wavelet behaves like a multi-level recursive filter which decomposes the data acting like a combination of low and high pass filters (Hajiabotorabi et al., 2019). It can be concluded from existing work that the low pass filtering operation typically contains the baseline data while the noisy component and disturbances are contained by the high pass filtering data. The data can thus be filtered as:

$$S_L \xrightarrow[DWT,L]{} LPF_L + HPF_L \quad (6)$$

Here,

S_L is the data to be filtered up-to 'L' levels

LPF_L is the low pass filtering operation at level 'L' of decomposition.

HPF_L is the high pass filtering operation at level 'L' of decomposition.

DWT stands for the discrete wavelet transform.

The filtering can be used to estimate the noise floor in the data and further filter it using the co-efficient values of the data. The co-efficient representation of the data is given by:

$$S_L \xrightarrow[DWT,L]{} C_{A,L}, C_{D,L} \qquad (7)$$

Here,

$C_{A,L}$ are the approximate co-efficient values of decomposition level 'L'

$C_{D,L}$ are the detailed co-efficient values of decomposition level 'L'

Retaining the approximate co-efficient values and discarding the detailed co-efficient values helps in data filtration. An iterative process to retain $C_{A,L}$ and discard $C_{D,L}$ for each decomposition level L generates a decomposition tree. The validation of the fact that the raw data is filtered can be obtained by observing the decomposition parameters of the raw and filtered data. The decomposition metrics which are commonly chosen for data analysis are (Rhif et al, 2019):

a. Mean:

$$\hat{s} = \frac{1}{N}\sum_{i=1}^{N} s_i \qquad (8)$$

Here,

s is the original data

$\hat{s}$ is the mean

N is the total number of samples

b. Standard Deviation:

$$s.d. = \sqrt{\frac{\sum_{i=1}^{N}(s_i - \hat{s})^2}{N}} \qquad (9)$$

Here,

s.d. is the standard deviation

N is the number of samples

s is the original data

$\hat{s}$ is the mean

c. Median:

$$median\ (x) = \frac{1}{n}\left[x_{\lfloor (n+1)/2 \rfloor} + x_{\lceil (n+1)/2 \rceil}\right] \ (10)$$

Here,

X is the ordered list

N is the number of samples

$\lfloor . \rfloor$ represents the floor function

$\lceil . \rceil$ represents the ceiling function

d. Mean Absolute Deviation (MAD):

$$M.A.D.(s) = \frac{1}{N}\sum_{i=1}^{N}|s_i - \hat{s}| \qquad (11)$$

Here,

M.A.D. represents the mead absolute deviation

N is the number of samples

s_i is the individual sample value of s

s is the mean value od s

e. L1 Norm:

$$|X|_{L1} = |A1| + |A2| \qquad (12)$$

Here,

$|X|_{L1}$ is the L1 norm

A1 and A2 are the comprising vectors

f. L2 Norm

$$|X|_{L2} = \sqrt{x_1^2 + x_2^2 + \cdots \ldots \ldots \ldots + x_n^2} \qquad (13)$$

The stopping condition for the machine learning algorithm to reach convergence is the successive stability of the cost function or objective function which is considered as the mean square error in this case. The mean square error (mse) is defined as:

$$mse = \frac{1}{n}\sum_{I=1}^{N}(p_i - a_i)^2 \qquad (14)$$

Here,

p denotes the predicted value.

a denotes actual value.

i the number of samples.

If the decomposition values of the approximate co-efficients are identical to the raw data and those for the detailed co-efficients are non-identical, then such a decomposition implies that the noisy part has contained in the detailed co-efficients and can be removed or filtered by discarding the detailed co-efficient values (Fernandez-Ordoñez et al., 2017).

B. *Training*

The next critical stage is training in which the data pre-processed data needs to be applied to a machine learning model for pattern recognition. In this case, an ensemble deep neural network has been used to forecast crop yield. The output of the neural network is given by (Tealab, 2018):

$$y = f(\sum_{i=1}^{i=n} X_i W_i + \Theta) \qquad (15)$$

Here,

X represents the inputs

Y represents the output

w represents the weights

f represents the activation function.

θ represents the activation function

To analyze the data in this case, a deep neural network with 10 hidden layers is designed. Each of the modules of the ensemble neural network is fed with the C_A and C_D values of decomposition. The weighted sum of each of the modules is summed up to obtain the final output (Bhoslae et al., 2018)

The summation of the individual outputs is given by:

$$Y_{tot} = \sum_{i=1}^{N} y_i \qquad (16)$$

Here,

Y_{tot} denotes the total output of the ensemble neural network.

y_i denotes the individual outputs of the ensemble modules.

N is the number of modules in the ensemble.

The number of modules has been taken as 4 corresponding to the 3 detailed and 1 approximate co-efficient values. The training rule employed is the back propagation based gradient descent with the objective function taken as the mean square error (Islam et al, 2018). The training rule is given by:

$$w_{t+1} = w_t - \alpha \frac{\partial e}{\partial w} \qquad (17)$$

Here,

w_{t+1} is the weight of $(t+1)^{st}$ iteration.

w_t is the weight of iteration number t.

e stands for error in iteration 't'

α is the learning rate

The mean absolute percentage error (MAPE) for the system has been computed as (Huang et al., 2017):

$$MAPE = \frac{100}{N} \sum_{i=1}^{N} \frac{|p_i - a_i|}{a_i} \qquad (18)$$

p denotes the predicted value.

a denotes actual value.

i the number of samples.

Simulation Results

The simulations have been performed on MATLAB 2020a with an i5 9300H CPU with a clock speed of 2.4GHz and available RAM of 8GB. The first part of the experiment entails data pre-processing so as to remove the noise and disturbance effects form the raw data. For the purpose the wavelet transform has been employed. A three level decomposition of the raw data has been performed in which the detailed co-efficient values are discarded and approximate co-efficient values are retained so as to filter out the noise effects. The approximate co-efficient values along with the detailed co-efficient values are used to train an ensemble neural networks. The data is split in the ratio of 70:30 for training to testing. The parameters used for training are time, rainfall, moisture, humidity, temperature and soil type. The data has been acquired from Kaggle.

data - Copy.xlsx

dataCopy

	A	B	C	D	E	F	G	H	I	J	K
	Moisture Number	rainfall Number	AverageHu... Number	MeanTemp Number	maxTemp Number	Mintemp Number	alkaline Number	sandy Number	chalky Number	clay Number	YieldMT Number
1	Moisture	rainfall	Average Hu...	Mean Temp	max Temp	Min temp	alkaline	sandy	chalky	clay	Yield (MT)
2	12.8017	0.0124	57	62	71	52	0	1	0	0	1.0647e+03
3	12.8517	0.0042	57	58	73	43	0	1	0	0	1.1042e+03
4	12.7768	0	56	58	69	46	0	0	1	0	1.0385e+03
5	12.9420	0.0317	62	56	70	43	0	1	0	0	1.0015e+03
6	12.9847	0	65	56	70	42	0	0	0	1	1.0798e+03
7	12.9645	0.0272	65	58	70	46	1	0	0	0	1.0419e+03
8	12.7380	0.0268	61	56	70	42	0	0	0	1	1.1419e+03
9	12.8194	0.0103	58	57	72	42	0	0	0	1	1.1360e+03
10	12.8839	0.0205	63	60	76	45	0	0	1	0	1.0589e+03
11	12.7845	0.0601	62	59	71	47	0	1	0	0	1.0037e+03
12	12.9688	0.0841	56	58	69	46	0	1	0	0	1.1007e+03
13	12.7844	0	63	56	70	42	1	0	0	0	1.1256e+03
14	12.9446	0	67	58	72	43	0	1	0	0	1.1457e+03
15	12.9253	0.1245	58	60	75	46	0	0	0	1	1.0085e+03
16	12.8211	0.0745	59	58	68	49	1	0	0	0	1.0675e+03
17	12.9392	0.0986	53	60	68	53	0	0	0	1	1.0874e+03
18	12.8159	0.2228	62	60	69	50	1	0	0	0	1.1030e+03
19	13.0568	0.1286	59	56	67	45	0	1	0	0	1.1079e+03
20	12.8980	0.1134	58	54	68	40	0	1	0	0	1.0975e+03
21	13.0346	0.0801	58	53	68	38	1	0	0	0	1.1090e+03
22	13.0240	0.0777	66	55	63	47	1	0	0	0	1.0561e+03
23	12.8621	0.0942	71	57	68	46	0	0	1	0	1.0872e+03
24	12.8558	0.0893	72	53	63	43	1	0	0	0	1.0174e+03
25	12.9398	0.1106	65	53	66	40	1	0	0	0	1.0086e+03
26	12.9462	0.1222	84	55	60	50	0	0	0	1	1.1470e+03
27	12.8317	0.1797	73	56	69	43	0	0	0	1	1.0427e+03

Fig. 1. Importing data from XL file to Matlab workspace.

Figure 1 depicts the data in the Matab workspace after loading the data. The data is accessible in the Matab workspace for analysis.

The dependent variables (feature) along with the target variable (yield) is decomposed to 3 levels of DWT. This would mean an approximate co-efficient value denoted by ‘a’ and three detailed co-efficient values ‘d1’, ‘d2’, ‘d3’ would be obtained through the decomposition. The metrics of decomposition are chosen as maximum value of variables, minimum value, mean, median, standard deviation, mean absolute deviation, L1 norm and L2 norm. The analysis of the parameters helps us in understanding the effect of the wavelet decomposition on the data cleaning process. ‘S’ corresponds to the original data stream, C_A corresponds to the approximate co-efficient values while C_D corresponds to the detailed co-efficient values. Tables 1, 2 and 3 enlist the values of decomposition. An evaluation of the decomposition metrics and insights into the data are presented subsequently.

Table 1. *Statistical analysis of raw data ‘s’*

S.No.	**Parameter**	**Value**
1.	Maximum	1024
2.	Minimum	978.2
3.	Mean	998.8
4.	Median	998.2
5.	Standard Deviation	5.798
6.	Medium Absolute Deviation	3.05
7.	L1 Norm	2.16×10^7
8.	L2 Norm	1.515×10^5

Table 2. *Statistical Analysis of Approximate Co-efficients, ‘Ca’*

S.No.	**Parameter**	**Value**
1.	Maximum	2892
2.	Minimum	2775
3.	Mean	2825
4.	Median	2823
5.	Standard Deviation	16.14
6.	Medium Absolute Deviation	8.849
7.	L1 Norm	7.128×10^6
8.	L2 Norm	1.419×10^5

Table 3.*Statistical Analysis of Detailed Co-efficients, ‘Cd’*

S.No.	**Parameter**	**Value**
1.	Maximum	8.613
2.	Minimum	-14.84
3.	Mean	0.09748
4.	Median	2.377
5.	Standard Deviation	1.793
6.	Medium Absolute Deviation	3.05
7.	L1 Norm	4854
8.	L2 Norm	119.5

The statistical analysis for the approximate value indicate the following:

1) The approximate co-efficient values contain the majority of the information of the data stream

2) Retaining the approximate co-efficient values help in retaining the maximum statistical information of the data.

3) The iterative process of retaining the approximate co-efficient values makes the histogram approach the actual data.

Moreover, the significance of the analysis lies in the fact that it tells about the statistical information contained in the data stream, the decomposed approximate values, the decomposed detailed values and the synthesized data stream. The histogram analysis implies the fact that as the number of levels increases, the local disturbances in the data are eventually removed. From table I, II and III, it can be clearly seen that as the number of decomposition levels increase, the approximate co-efficient values tend to align towards the original data stream in terms of statistical characteristics

1) The detailed co-efficient values however tend to deviate from the actual data stream as the number of levels increase. This clearly indicates that if the approximate co-efficient values are retained and the detailed coefficient values are discarded, then the amount of local variations and disturbances can be removed.

2) The wavelet transform can be used to maintain monotonicity in local intervals so as to make the training more effective for the neural network.

The above three points indicate that the wavelet decomposition along with the time series prediction can enhance the accuracy of prediction and also increase the regression for large data sets.

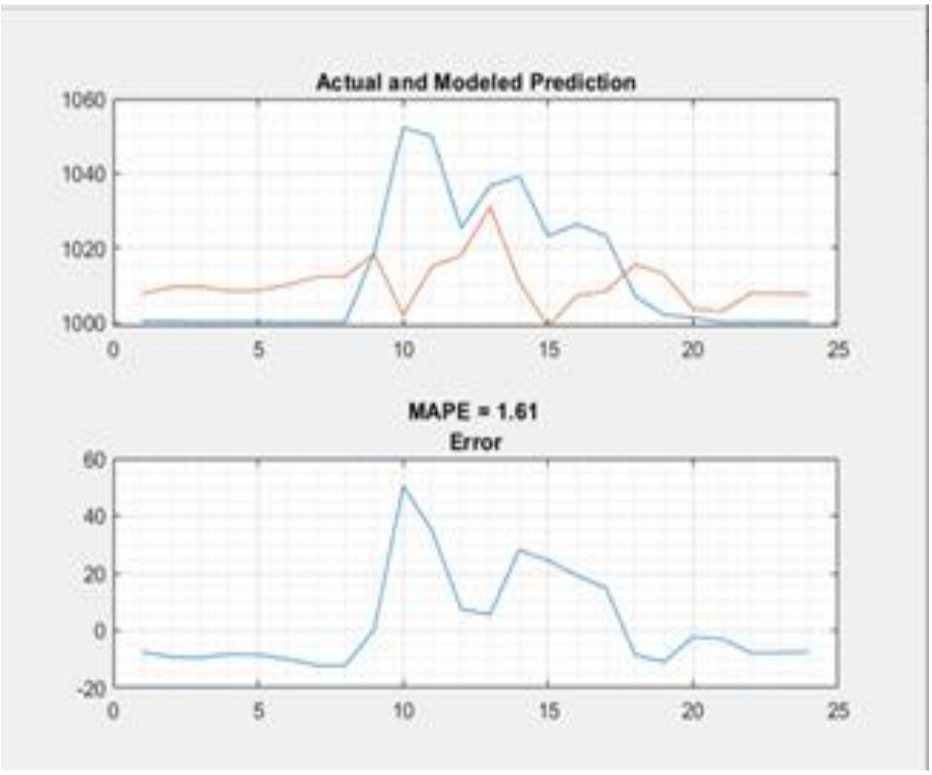

Fig. 2. Forecasted and Actual Yield.

Figure 2 represents the curves for the predicted and the forecasted values. The red curve depicts the forecasted or predicted values. The blue curve depicts the actual values. It can be clearly concluded that the value of the regression is clearly related to the accuracy of prediction for the system. A comparative analysis of the previous and proposed work in terms of the evaluation parameters is given in table 4.

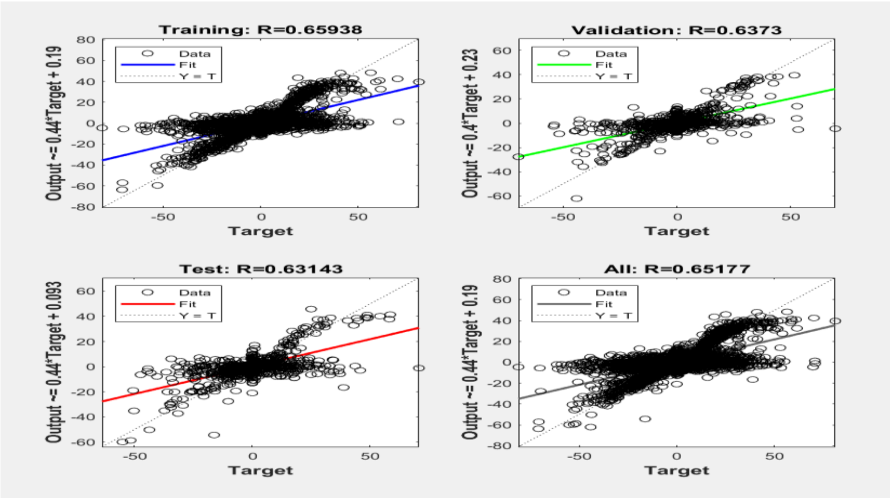

Fig. 3. Regression Analysis of Forecasting.

Figure 3 depicts the regression for training, testing, validation and overall cases. A summary of the obtained results is presented in table 4..

Table 4.*Summary of Simulation Results.*

S.No.	**Parameter**	**Value**
1.	Machine Learning Model	Neural Net
2.	Architecture	Back Propagation
3.	Hidden Layers	10
4.	Training Epochs	25
5.	Time to convergence	2 mins, 4secs
6.	MSE at convergence	37.3
7.	Validation Checks	6
8.	Regression Training	0.65938
9.	Regression Testing	0.63143
10.	Regression Validation	0.6373
11.	Regression Overall	0.65177
12.	μ at convergence	0.100
13.	MAPE	1.61%
14.	Accuracy Proposed Work	98.39%

Table 5. *Comparison with existing techniques.*

S.No.	**Author and Approach**	**Forecasting Accurac**
1.	Elavarasan et al. Deep Reinforcement Learning	93%

2.	Dang et al. Support Vector Regression.	85%
3.	Nigam et al. Random Forests.	67.80%
4.	**Proposed Approach,** **DWT + Gradient Boost Based** **Ensemble Deep Neural Network**	98.39%

A comparative analysis of the proposed work with contemporary techniques shows that the proposed technique outperforms the existing techniques in terms of accuracy of prediction. This can be attributed to the combined data filtration and ensemble learning approach adopted in this work.

Conclusion

It can be concluded that crop yield prediction is a critically important forecasting problem trying to address food security in the world. However, it is challenging to accurately forecast crop yields since the data is generally random and complex and the yield depends on multiple parameters. The proposed system uses a two step approach in which the data is first filtered and secondly a deep neural network employing back propagation is used for pattern recognition. The performance of the system has been evaluated in terms of the mean square error, mean absolute percentage error, regression and accuracy. From the results, it can be observed that the system trains in low number of iterations and also achieves low MAPE value. A comparative analysis with respect to previous work also shows that the proposed technique outperforms the existing technique in terms of prediction accuracy.

References

1. Bhosale S, Thombare R, Dhemey P, Chaudhari A (2018). Crop Yield Prediction Using Data Analytics and Hybrid Approach, *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA),* 1-5.
2. Dang C, Liu Y, Yue H, Qian J (2021).Autumn Crop Yield Prediction using Data-Driven Approaches:-Support Vector Machines, Random Forest, and Deep Neural Network Methods, *Canadian Journal of Remote Sensing, Taylor and Francis,* 47(2), 162-181.
3. Elavarasan D, Vincent P (2020).Crop Yield Prediction Using Deep Reinforcement Learning Model for Sustainable Agrarian Applications, *IEEE Access 2020,* 8, 86886-86901.
4. Fernandez-Ordoñez Y, Soria-Ruiz J (2017), Maize crop yield estimation with remote sensing and empirical models, *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS),* 3035-3038.
5. Gopal P, Bhargavi R (2019). A novel approach for efficient crop yield prediction, *Computers and Electronics in Agriculture*, Elsevier, 165, 104968.
6. Hajiabotorabi Z, Kazemi K, Samavati F, Ghaini F (2019). Improving DWT-RNN model via B-spline wavelet multiresolution to forecast a high-frequency time series, *Expert Systems with Applications, Elsevier* 2019, 138, 112842.
7. Hird J, McDermid G (2009). Noise reduction of NDVI time series: An empirical comparison of selected techniques, *Remote Sensing of Environment, Elsevier,* 113(1), 248-258.
8. Holmes J. (2020). *Losing 25,000 to Hunger Every Day*. Retrieved from UN Chronicle: https://www.un.org/en/chronicle/article/losing-25000-hunger-every-day
9. Huang X, Huang G, Yu C, Ni S, Yu L (2017). A multiple crop model ensemble for improving broad-scale yield prediction using Bayesian model averaging, *Journal of Field Crops Research, Elsevier* 211, 114-124.

10. Islam T, Chisty T, Chakrabarty A (2018). A Deep Neural Network Approach for Crop Selection and Yield Prediction in Bangladesh, *2018 IEEE Region 10 Humanitarian Technology Conference (R10-HTC),* 1-6.
11. Khandelwal I, Adhikari R, Verma G (2015). Time series forecasting using hybrid ARIMA and ANN models based on DWT decomposition, *Procedia Computer Science, Elsevier*, 48, 173-179.
12. Klompenburg T, Kassahun A, Catal C (2020). Crop yield prediction using machine learning: A systematic literature review. *Computers and Electronics and Agriculture, Elsevier 2020*, 177, 105709.
13. Madan R and Mangipudi P (2018). Predicting Computer Network Traffic: A Time Series Forecasting Approach Using DWT, ARIMA and RNN, *2018 Eleventh International Conference on Contemporary Computing (IC3)*, 1-5.
14. Moseley W, Battersby J (2020). The vulnerability and resilience of African food systems, food security, and nutrition in the context of the COVID-19 pandemic. *African Studies Review, Cambridge Publications,* 63(3), 449-461.
15. Nigam A, Garg S, Agrawal A, Agrawal P (2019). Crop Yield Prediction Using Machine Learning Algorithms,*2019 Fifth International Conference on Image Information Processing (ICIIP)*, 125-130.
16. Nury A, Hasan K, Alam M (2017). Comparative study of wavelet-ARIMA and wavelet-ANN models for temperature time series data in northeastern Bangladesh, *Journal of King Saud University – Science, Elsevier*, 29(1), 47-61.
17. Rhif M, Abbes A, Farah I, Martínez B, Sang Y (2019). Wavelet transform application for/in non-stationary time-series analysis: a review, *Applied Sciences*, *MDPI*, 9(7), 1-22.
18. Tealab A (2018). Time series forecasting using artificial neural networks methodologies: A systematic review, *Future Computing and Informatics Journal*, *Elsevier*, 3(2) 334-340.