Performance Analysis of Activation Functions for Wheat Crop Yield Prediction

1Shital H. Bhojani, Dr. 2Nirav Bhatt
1Research Scholar, RK University & Assistant Professor, AAU, Anand-Gujarat
2Associate Professor, MCA Department, RK University, Rajkot-Gujarat

shitalbhojani@aau.in, nirav.bhatt@rku.ac.in

Abstract. Activation function (AF) plays a vital role in the neural network. This research study focuses on the performance evaluation of the various AFs in multilayer perceptron (MLP) algorithm for the wheat crop yield prediction at the regional level. Nowadays researchers are publishing the new AFs and also proposing the improvement in existing AFs for more accurate results. Sigmoid AF is the default function used in the MLP algorithm. In this research we have applied the various AFs in MLP using the WEKA java libraries and performance analysis has been carried out for the selected districts of the Gujarat state. Based on the mean absolute percentage error (MAPE) performance measure output is discussed. Therefore scientists and researchers can select the appropriate AF to improve the prediction accuracy and hence improve the performance of the network as per the research data.

Abstract. Neural Network, Activation Function (AF), MLP, Yield Prediction

1. Introduction
Crop yield is affected by many factors such as weather parameters, rainfall, soil type and its fertility, atmospheric conditions, management practices, etc. Data mining is the process to extract significant knowledge, patterns and trends from datasets. Data mining and machine learning techniques are making their roots stronger in the field of agriculture. Neural networks can be considered as an approach to the problem of computation. We can apply the neural network approach to diverse fields such as speech and text recognition, image processing, various scientific, engineering and commercial disciplines, prediction, modeling etc. Therefore, it is essential to implement the new technologies in farming and have to follow the scientific farming methods for better output. This research paper was focused on various activation functions and its assessment used for the wheat crop yield prediction.

The WEKA time series forecasting package was used with the help of WEKA JAVA libraries. Various activation functions were applied for forecasting using the MLP algorithm of classification technique.

2. Activation functions
The activation function, also referred to as the transfer function, is applied on the inner product in the neural network. To improve the performance of the neural network selection of activation function is very much important. Activation functions manipulate the data through gradient descent and generate
an output for the neural network. We can categorise the activation function as linear or nonlinear. When we have the linear input we get the linear output. So activation functions are applied to the outputs of the linear models to produce the non-linear outputs. The input and output parameters of the neural network decide the type of activation function to be used for further processing. When activation function applies to input values are generally a sum of products formed and if bias exists then it is also included for the computing. Bias is an integer value associated with each node in NN as an extra input. In neural network derivatives of AFs are used for updating the curve and to know in which direction and how much to change. The derivative also called a gradient of an AF is really important for training the neural network. In neural network different activation functions exist like sigmoid, tanh, softmax, ReLU, Leaky ReLU (LReLU), ELU, loglog, logsigm (sigmoid square), cloglog etc.

2.1. Sigmoid
The Sigmoid is a non-linear activation function. Its value lies between 0 and 1. This function is majorly used in NN. Sigmoid activation function is providing good results with my research datasets. It is used in binary classification problems. It is mainly used in the feedforward neural network.

2.2. logsigm
This is the square of the sigmoid function and provides good results for the current research study.

2.3. The hyperbolic tangent – tanh
As compared to the sigmoid function values of tanh function lies between -1 and 1. It produces zero centred output so assisting the backpropagation process in the NN. This activation function is not providing a good result with my research datasets.

2.4. Rectified Linear Unit (ReLU)
Some text. This is the widely used activation function and provides the better result in some types of datasets. But for time series prediction it’s not proving the satisfactory results.

3. Review of literature
Jianli et. al. discussed the linear and non-linear AF such as sigmoid, tanh, ReLu and various improvements of ReLu like leaky ReLu, PReLU, RReLu, and ELU. They have compared the characteristics of these AFs and also mention their pros and cons. The selection of AF is dependent on the aim and the area of our research [1]. Leonid Datta discussed the properties and problems faced by the AF. He explained the properties of widely used AFs and shared properties of AFs. He reviewed the vanishing gradient and the dead neuron problems of widely used AFs. He concluded that for the weight initialization performance of the normal initialization beats the performance of Xavier initialization. ReLU gives good performance with the normal initialization in some types of networks [2]. Heny Pratiwi et. al. used the population density dataset of Indonesia for the period 2003-2015 for prediction purposes. They have used the matlab 6 application for training and testing purposes and applied binary sigmoid activation function (logsig) in the network model. They conducted that the best architectural model can be generated using sigmoid AF in backpropagation NNs for the selected dataset [4]. Rahul et. al. provide new theoretical characterizations that support the use of the ReLU, its variants. They stated the different remarks, theorem, proposition and examples in the framework of L-splines and showed that the choice of activation implicitly defines a NN regularizer [5]. Khaki et. al. implemented CNN-RNN model models in the most efficient manner to forecast corn and soybean yields across the entire Corn Belt 13 states in the United States for 3 years 2016 to 2018 using environmental data and management practices. They concluded that the model predicted good results and it could be used in future yield prediction tasks [6]. da S. Gomes [7] and the co-author suggested the new activation function namely log-log, probit and loglog, cloglog, logsigm in NN for time series forecasting. They used the Matlab platform for conducting the different experiments. The experiment
contains a MLP algorithm with one hidden layer, variation in number of nodes. The evaluation was carried out by the two learning algorithms namely LM and CFG and the best model was chosen using values of MAPE for forecasting. Author recommends the new AFs for smaller network structure and financial time series data sets. Raju Prasad Paswan [8] applied the thnh, log sigmoid and linear transfer function in MLP neural network algorithm on maize and rice crop datasets to forecast area and productivity using the Matlab environment. The result of different AFs shows that the predicted values are within ±20% error.

4. Materials and Methods

The dataset for the result study was collected from Directorate of Agriculture, Gandhinagar and from Agro-meteorology department, Anand Agricultural University, Anand. WEKA “Waikato Environment for Knowledge Analysis” JAVA libraries were used for the implementation of various AFs. Dataset includes the weather and yield data. Weather parameters such as basic sun shine hours (BSS), maximum temperature (MAXT), minimum temperature (MINT), relative humidity morning (RH1), relative humidity afternoon (RH2), vapour pressure morning (VP1) and vapour pressure afternoon (VP2) were used for the prediction the wheat yield at regional level. Daily weather data were converted as per Standard Meteorological Week (SMW) to the weekly data for the research study. There are eight agro-climatic zones in Gujarat state. A total of six districts from different agro-climatic zones are selected namely Anand, Sabarkantha, Banaskantha, Bhavnagar, Junagadh and Ahmedabad. Data from 1990–1991 to 2016–2017 were used for the time series forecasting.

Figure 1: Structure and propagation in NN

Processing the data from input layer to the output layer is called propagation. The values from input nodes are sent to the subsequent nodes using some AFs and reach to the output nodes. Figure 4 shows the propagation in the neural network.

Neural network consists of processing units, neurons, and directed, weighted connections between those neurons. In NN there are many neurons that work in correspondence of weights and bias. Weight initialization method plays an important role in training and performance of a neural network AF at a hidden layer is very much important; if it is absent then outputs will be the linear combination of the input parameters. Another name of AF is Transfer Function. Traditional mathematical model-based algorithms have reasonably high clarity in how a model and the input data lead to output decisions. Conventional NN typically have a fixed, bounded continuous nonlinear AF at each neuron, which is the key to the overall nonlinear behaviour of NN[3]. Research work is carried out by WEKA data mining tool and forecasting package. MLP Classification algorithm was used for the wheat yield prediction at district level. Research work is carried out by multiple trials and uses the different values combination. However it was noticed in this study that one hidden layer with 0.3 learning rate and 0.2
momentum generates the better results. In this research it is observed that the number of nodes, lags, and instances are also dependent on each other.

5. Result and Discussion
Experimental results obtained using MLP algorithm and various AFs applied for wheat yield prediction. Various AFs listed below are used for the performance analysis.

Table 1: List of Activation Functions used in this study

| AF1   | Sigmoid       | AF5   | loglog with 15 node |
|-------|---------------|-------|---------------------|
| AF2   | cloglog       | AF6   | loglog with 16 node |
| AF3   | loglog        | AF7   | ReLU                |
| AF4   | logsigm       | AF8   | tanh                |

Results of actual and predicted yield for selected districts were compared and a conclusion was drawn. After comparing the result it was analyzed that when data applied to the ReLU function the error rate of actual and predicted yield was very much higher for most of the districts. Using the different AFs, predicted yields were overestimated and underestimated for selected districts. Table 2 shows the year wise maximum and minimum error rate values for various AFs.

Table 2: Minimum and Maximum Error Rate of Actual and Predicted Yield year wise

| Year     | AF1 | AF2 | AF3 | AF4 | AF5 | AF6 |
|----------|-----|-----|-----|-----|-----|-----|
| 2015-16  | MIN | -0.33 | 1.48 | 1.01 | -3.90 | 1.47 |
|          | MAX | 21.34 | 32.50 | 27.25 | 26.96 | 23.74 | 29.99 |
| 2016-17  | MIN | 10.19 | 0.32 | 8.04 | 6.15 | -2.10 |
|          | MAX | 27.23 | 42.49 | 26.63 | -44.42 | 29.13 | -34.02 |
| 2017-18  | MIN | 0.20 | -1.77 | 0.42 | 1.89 | 2.36 |
|          | MAX | 35.36 | 51.40 | 43.44 | 48.66 | 53.16 | 53.30 |

By analyzing table 2 we can observe that error rate values were much higher. Default sigmoid AF values were also lying between 0.20 and 35.36 which is not acceptable. Same way for each above listed AF the values are higher. Therefore it is not recommended to try and use another AF which provides better performance for the agriculture dataset. Hence, there is a need to propose some new AF and recommend amending the MLP algorithm for crop yield forecasting.

Figure 3(a) to 3(f) shows the actual and predicted yield using the different AFs or selected districts.
The performance evaluation was carried out by the MLP algorithm using various AFs, and observed the values of mean absolute percentage error (MAPE) for wheat yield prediction. Figure 4 shows the MAPE of selected districts. As we discussed, the error rate is much higher for each district although the resulting MAPE values were quite satisfactory. For ReLu and tanh the values were in thousands and lakhs for most of the districts so its result was not included in this study. Table 3 shows the year wise maximum and minimum MAPE values for various AFs.

Table 3: Minimum and Maximum MAPE year wise

| Year     | Mean Absolute Percentage Error (MAPE) |
|----------|---------------------------------------|
|          | AF1        | AF2        | AF3        | AF4        | AF5        | AF6        |
| 2015-16  | MIN 0.0008 | 0.004      | 0.0003     | 0.0003     | 0.0001     | 0.0005     |
|          | MAX 0.5432 | 0.8135     | 0.1919     | 0.1719     | 0.179      | 0.5323     |
| 2016-17  | MIN 0.0006 | 0.0031     | 0.0003     | 0.0003     | 0.0001     | 0.0004     |
|          | MAX 0.4477 | 0.751      | 0.1673     | 0.1489     | 0.1821     | 0.5327     |
| 2017-18  | MIN 0.0007 | 0.0032     | 0.0003     | 0.0003     | 0.0001     | 0.0004     |
|          | MAX 0.4473 | 0.798      | 0.2065     | 0.1515     | 0.1874     | 0.5556     |

MAPE is a measure of forecast accuracy percentage which plays an important role to find out the best-fitted model in statistics. It is also noticed that the value of learning rate and momentum is not affecting the result for the selected datasets. Research work is carried out by taking the confidence...
interval at a level of 95%. Figure 4(a) 4(b) and 4(c) shows the MAPE for the year 2015-16, 2016-17 and 2017-18 respectively.

Figure 4(a) MAPE for the year 2015-16 for selected districts

Figure 4(b) MAPE for the year 2016-17 for selected districts

Figure 4(c) MAPE for the year 2017-18 for selected districts

6. Conclusions
In this study, we have discussed the result of various AFs applied for wheat crop yield prediction at district level. MLP provides the error rate of actual and predicted yield between 0.2 and 35, which is not acceptable. During research study it was noticed that the model was not fitting while applying my dataset to MLP algorithm using the ReLu and Tanh activation function as MAPE was much higher, reaching to thousands and lakhs. For sigmoid, loglog and logsigm the model is fitted but with higher
MAPE. So it is concluded that it’s quite difficult to choose the best AF and it’s mainly dependent on the aim of our research and the type of datasets.

Acknowledgments
I am grateful to the Directorate of Agriculture, Gandhinagar, and Agro-meteorology Department, Anand Agricultural University, Anand, Gujarat, for providing the yield and weather datasets respectively for my research work.

References
[1] Jianli Feng, Shengnan Lu 2019 Performance Analysis of Various Activation Functions in Artificial Neural Networks IOP Conf. Series: Journal of Physics: Conf. Series 1237 (2019) 022030
[2] Leonid Datta 2020 A Survey on Activation Functions and their relation with Xavier and He Normal Initialization arXiv:2004.06632v1
[3] Shital H. Bhojani, Dr. NiravV. Bhatt Wheat crop yield prediction using new activation functions in neural network Neural Comput&Applic
[4] Heny Pratiwi, Agus Perdana Windarto, S. Susliansyah, Rinir Restu Aria, Susi Susilowati, Luci Kanti Rahayu, Yuni Fitriani, Agustiena Merdekawati, Indra Riyana Rahadjeng 2020 Sigmoid Activation Function in Selecting the Best Model of Artificial Neural Networks IOP Conf. Series: Journal of Physics: Conf. Series 1471 (2020) 012010
[5] Rahul Parhi and Robert D. Nowak 2020 The Role of Neural Network Activation Functions arXiv:1910.02333v2
[6] Khaki Saeed, Wang Lizhi, Archontoulis Sotirios V. 2020 A CNN-RNN Framework for Crop Yield Prediction Frontiers in Plant Science Volume 10
[7] da S. Gomes, G.S., Ludermir, T.B. & Lima, L.M.M.R., 2011 Comparison of new activation functions in neural network for forecasting financial time series. Neural Comput&Applic
[8] Raju Prasad Paswan1, ShahinAra Begum and BorshaNeog 2018 Comparison of Abilities of Different Activation Functions of Artificial Neural Network to Predict Crop Area and Crop Production. Int. J. Pure App. Biosci. 6 (6): 212-220

Abbreviations
| DM  | Data mining          |
| NN  | Neural network       |
| AF  | Activation function  |
| SMW | Standard Meteorological Week |
| BSS | Basic sun shine hours |
| MAXT| Maximum temperature  |
| MINT| Minimum temperature  |
| RH1 | Morning relative humidity (%) |
| RH2 | Afternoon relative humidity (%) |
| VP1 | Morning vapour pressure |
| VP2 | Afternoon vapour pressure |
| MAPE| Mean absolute percentage error |