STUDY OF TEMPERATURE USING ARTIFICIAL NEURAL NETWORK AND MULTIPLE LINEAR REGRESSION

SARIKA JAIN AND TEJASVANI SHARMA[1]

ABSTRACT. Temperature forecasting has been one of the most important factors considered in climate impact studies on sectors of agriculture, vegetation, water resources and tourism. Moreover it became a usual part of our daily lives to check the temperature or the atmospheric conditions before stepping out. Since the atmospheric soundings are nonlinear and follows a very irregular trend, therefore the method has to be chosen to predict a better technique to find out the structural relationship between the various parameters. The main purpose of this paper is to forecast the temperature with two different methods i.e. first using Artificial neural network (ANN) and second, Multiple linear regression (MLR) and predict the best-fit method with the least root mean square error (RMSE) criteria.

1. INTRODUCTION

Temperature forecasting is the application of science and technology to predict the state of the atmosphere. It is done by collecting the computable data for the current state of the atmosphere. The disorderly nature of the atmosphere, the enormous computational power required to resolve the equations that describe the atmosphere, error involved in estimating the initial conditions. There are a numerous ways for weather forecasting. Forecasts based

on temperature are important for agriculture, and also for traders within commodity markets. Temperature forecasts are also used by utility companies to calculate demand. The outdoor activities are critically reduced by snow and the wind chill, thus forecasts can be used to plan activities around these events, and to plan first and survive them.

There are few studies available that focuses on forecasting temperature using artificial neural networks and multiple linear regressions. Ramos [1] studied artificial neural networks (ANNs) and multiple linear regression (MLR) and established that the Artificial Neural Network method performed better than the linear regression method for monthly and seasonal rainfall. Similar kind of work was done by Chavan [4], studied the Artificial Neural Network (ANN) and Multi Linear Regression (MLR) techniques to find out one day ahead runoff in Amba river basin located at Maharashtra, India.

Another work described by Mathur et al., [2] was about a weather forecasting model using Artificial Neural Network. The weather parameters like maximum temperature, minimum temperature and relative humidity has been forecasted using the attributes extracted by different periods together with the weather parameter time-series. The method used here was feed forward artificial neural networks (ANNs) with back propagation for supervised learning using the weather data collected at Pantnagar station situated in Tarai region of Uttarakhand state, India since April 1996 to March 1999 and is available as weekly average. This data is used first for calculating features over periods of 15, 30 and 45 weeks. The trained ANN was used to predict the future weather conditions. The results are very encouraging and it is found that the feature based forecasting model can make predictions with high degree of accuracy. Alike studies was done by Kumar [3],developed an artificial neural network (ANN) model which can be used to predict weekly mean temperatures in Pantnagar, Uttarakhand, India.

2. Methodologies

2.1. **Data Specifications.** The aim of this paper is to build a pattern that can predict the temperature for a particular station given the atmospheric condition for the previous days (target data) and various parameters on which it depends (input data).

This study considers the data of six months from January 1, 2019 to June 30, 2019 observed at the VIDD (safdarjung airport station Delhi) and collected from the University of Wyoming at two different timings 00Z-12Z (Zulu time zone) where 00Z is 5:30 am and 12Z is 5:30 pm. VIDD station (Delhi) is located at 28.7041° N latitudes, 77.1025° E longitudes approximately 227 meters above mean sea level.

The data comprises of parameters like Relative humidity, Geo potential height, Wind Direction, Wind Speed, Equivalent Potential Temperature, Virtual Potential Temperature, Mixing Ratio, Dew point Temperature, and Atmospheric Pressure. Out of 360 values 330 (5 months and 15 days) has been considered as observed values and the rest 30 i.e. 15 days values are used for forecasting or validation purpose. In ANN technique, single and multiple layers are examined and their outcomes are compared with the 15 days validation values. The best ANN model is taken for further investigation with MLR to compare the observed and modelled values.

2.2. Artificial Neural Network.

2.2. **Artificial Neural Network.** Artificial neural networks are one of the useful tools used in machine learning. As the word suggests "neural", these are the brain-stimulated systems which are considered to copy the way that we humans learn things. Neural networks consist of input and output layers, as well as (in most cases) a hidden layer consisting of units that convert the input into something that the output layer can use. These are magnificent tools for detecting patterns which are far too complex or numerous for a human programmer to extract and teach the machine to recognize.

Moreover, neural networks (also called "perceptrons") have been discovered around 1940s, it is only in the last several decades where they have become an important part of artificial intelligence. This is because of the arrival of a new technique called "back propagatio", which is an algorithm for supervised learning of artificial neural networks using gradient descent.

There are several types of neural network, each of which has their own levels of complexity. The most basic type of neural net is something called a feed forward neural network, in which information travels in only one direction from input to output as shown in Figure 1. A typical feed forward with back propagation network should have at least three layers- an input layer, a hidden

layer, and an output layer. Suitable selection of number of hidden layers and the number of neurons in each of those needs trial.
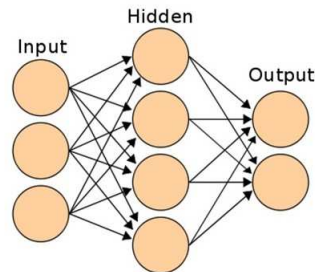

FIGURE 1. Construction of multilayer feed forward neural network

When researchers or computer scientists set out to train a neural network, they typically divide their data into three sets- first a training set, which helps the network create the various weights between its nodes. Secondly, they fine-tune it using a validation data set. Finally, they'll use a test set to check if it can successfully turn the input into the desired output.

After a number of iterations the neural network is trained and the weights are saved. The test set of data is compared to the trained neural network to test the result of the neural network. Moreover, the following factors also affect the neural network's result:

(1) Increasing the number of neurons/layer also increases the performance as the MSE decreases.
(2) Tan-Sigmoid function is our best option for hidden layers due to its very fast learning rate and sensitivity towards change in number of samples and neurons/layer.

2.2.1. *Single Layer neural network.* A single-layer neural network constitutes the simplest form of neural network, in which there is only one layer of input nodes that send weighted inputs to a subsequent layer of receiving nodes as in Figure 2.

2.2.2. *Multilayer neural network.* A multi-layer neural network contains more than one layer of artificial neurons. These differ widely in their design. It is important to note that while single-layer neural networks were useful early
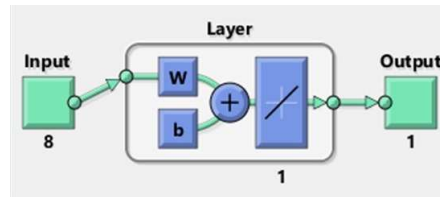
FIGURE 2. Single layer model

in the evolution of AI, the multi layer networks are used broadly in today's scenario. The multilayer neural network comprises of more than one hidden layer, as shown in Figure 3 (a neural netwok with ten hidden layer).
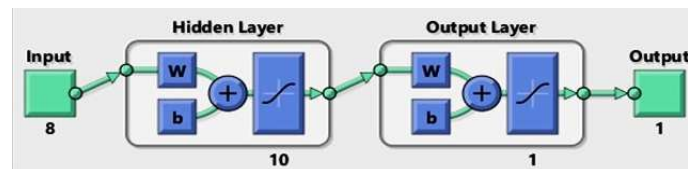

FIGURE 3. Multilayer model

2.2.3. *Multiple linear regressions (MLR)*. Multiple linear regression (MLR), also called as multiple regression, is a statistical technique that uses various discrete variables to predict the outcome of a respective variable. The aim of multiple linear regressions (MLR) is to develop the correlation between the discrete (dependent) variables and respective (independent) variable.

The formula for the Multiple Linear Regression is:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots\ldots + \beta_p x_{ip} + \epsilon \,,$$

where, for $i = n$ we have the following observations:

y_i = dependent variable ;

x_i = explanatory variables ;

β_0 = y-intercept (constant term) ;

β_p = slope coefficients for each explanatory variables ;

ϵ = the model's error term (also known as the residuals) .

The multiple regression models are based on the various conjectures given below:

- There is a correlation between the dependent variables and the independent variable.
- Yi observations are selected independently and randomly from the population.
- Residuals should be normally distributed with a mean of 0 and variance σ.

3. RESULTS AND DISCUSSION

3.1. **Validation of the ANNs.** The RMSE value for the single layer ANN model was calculated and found out to be *0.27* which was higher than that of ten layers ANN model with RMSE *0.007*. RMSE calculates residual errors which give an overall idea of the difference between the observed and modelled values. Hence the model with the low RMSE is preferred over the single layer ANN model.
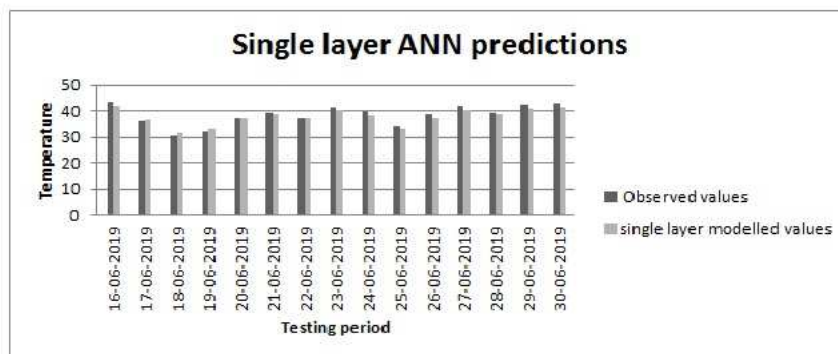

FIGURE 4. Graph of observed and predicted values for single layer ANN model

3.2. **Validations of the MLR.** After using multiple linear regression by taking 5 months and 15 days data as input and predicted the last 15 days data, we found RMSE as *0.066* with the modelled and observed values. Figure 6 describes the observed and predicted values using MLR technique.
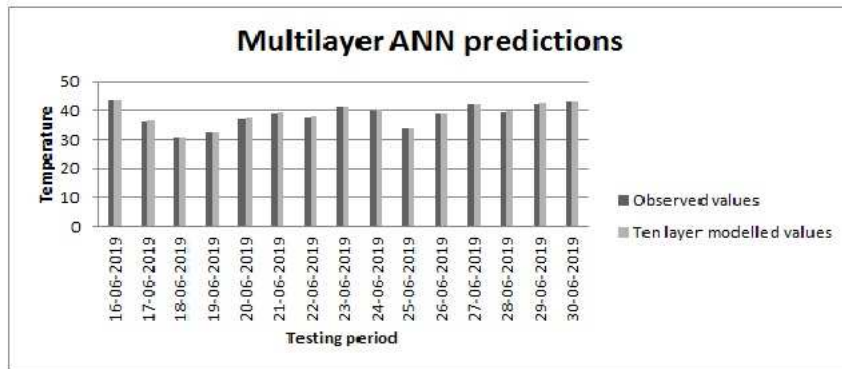
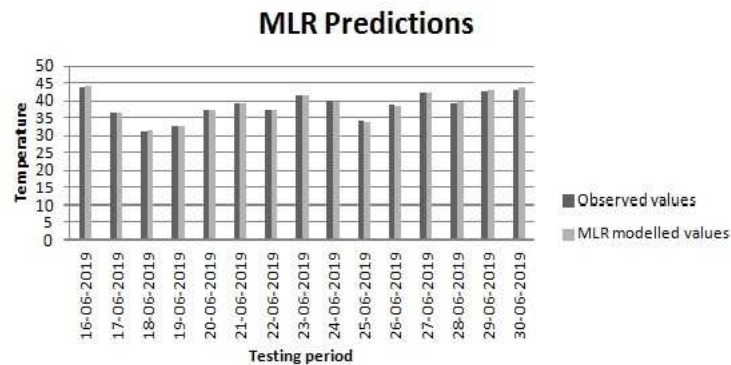FIGURE 5. Graph of observed and predicted values for ten layer ANN model

FIGURE 6. Graph of observed and predicted values for MLR

3.3. **Result.** In ANN technique, we found that RMSE with Observed values is *0.007* using multi-layer neural network and with MLR it is *0.066*. So clearly, Multiple Layer ANN technique gives us best fit prediction. Figure 7 shows graph for the Modelled ANN, MLR and observed values.
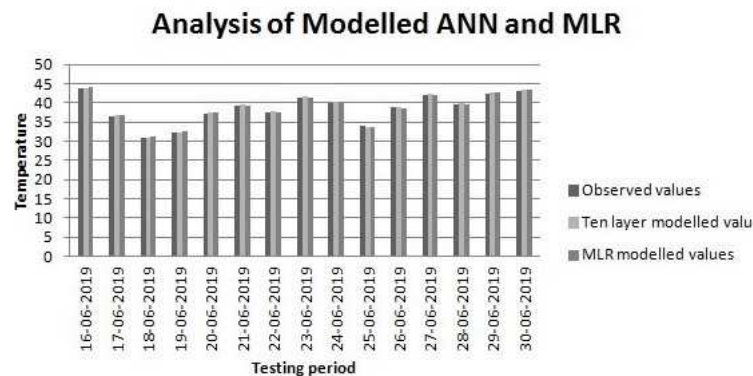
FIGURE 7. Graph of Modelled ANN, MLR and observed values

4. CONCLUSION

This paper examined the appropriateness of artificial neural networks and multiple linear regression analysis by RMSE criteria .Both ANN and MLR methods provided good fit with the observed data for 15 days ,but the ten layers ANN method had the best-fit RMSE among other methods for the testing period. This specified that ANNs are usable method for the modelling of temperature in time series. ANNs showed that the networks are capable as competitive tool. Again, the multi-layer ANN simulations give the best results, and its performance is shown in Figure 7.

In general, the results showed the use of ANNs produced more accurate results than MLR by low RMSE, which can be attributed to the fact that ANNs perform tasks that a linear program is unable to do. Also, one of the superiority of ANNs is their capacity for temporal processing and thus their ability to incorporate various predictive values as inputs without any additional effort.

REFERENCES

[1] A. M. RAMOS: *Spatial disaggregation of rainfall simulated by atmospheric models in Northeast Brazil* Master's thesis, Federal University of Paraiba, Paraiba, 2000.

[2] S. MATHUR, A. KUMAR, M. CHANDRA: *A feature based neural network model for weather forecasting,* World Academy of Science, Engineering and Technology, **34** (2007), 11-22.

[3] P. KUMAR, P. KASHYAP, S. JAVED: *Temperature forecasting using Artificial neural Network*, College of Technology, GB Pant University of Agriculture and Technology, India, 2013.

[4] P. M.CHAVAN, S. K. UKARANDE: *Comparison of Rainfall-Runoff Model Using ANN and MLR*, Dept. of Civil, YTCEM, Raigad ,India, 2017.

DEPARTMENT OF MATHEMATICS
AMITY SCHOOL OF APPLIED SCIENCES
AMITY UNIVERSITY
GURUGRAM, HARYANA, INDIA
E-mail address: Sarika33jain@gmail.com

DEPARTMENT OF MATHEMATICS
AMITY SCHOOL OF APPLIED SCIENCES
AMITY UNIVERSITY
GURUGRAM, HARYANA, INDIA
E-mail address: Samaira886@gmail.com