Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 9, No. 3, September : 2023

TO PREDICT THE RAINFALL USING EMPIRICAL AND DYNAMIC APPROACH

Mrs. C. Vijayalakshmi, Research Scholar, Dept. Of Computer Science, Quaid-E-Millath Government College for Women.vijiglory@qmgcw.edu.in
Dr. M. Pushpa, Research Supervisior, Dept. Of Computer Science, Quaid-E-Millath Government College for Women. Dr.PushpaM@qmgcw.edu.in

Abstract
Forecasting rainfall causes serious anxiety and has concentrated on raising awareness among businesses, governments, the scientific community, and risk management organizations. Rainfall is an atmospheric phenomenon that affects a variety of human actions, including the production of crops, the generation of electricity, tourism, and the management of forests, among others. To this scope, rainfall forecasting is important; while this is a maximum correspondence for undesirable inherent occurrences, such as flooding, mudflow, avalanches, and mass wasting, it has changed the world. For that reason, a suitable approach for rainfall prediction composes of obtaining preventative and improvement measures designed for these natural incidents. Using empirical and dynamical methodologies for various climate zones, one may predict rainfall in this regard.

Keywords:
Rainfall, Prediction, Multiple Linear Regression

I. Introduction
At the regional and national levels, weather forecasting uses a variety of techniques to predict rainfall. The two main approaches to predicting rainfall are dynamical methods and empirical methods. The Empirical technique is based on an analysis of historical rainfall data and its correlation with numerous oceanic and atmospheric variables throughout various locations of the earth.
Regression, Artificial Neural Networks and fuzzy logic techniques are the most often used empirical methods for predicting climate To make predictions about how the world's climate system will evolve in response to the initial atmospheric conditions, a dynamical approach use physical models that are based on equation systems. Utilising a numerical method for rainfall forecasting, dynamic approaches are put into practise.
Empirical models have drawbacks since, in contrast to dynamic models, they can only produce one result. They are vulnerable to problems with predictor variable co-linearity and make the erroneous assumption that historical associations would persist in the future [1]. Dynamic models examine how a system evolves over time by simulating the various processes that take place in it. Compared to empirical models, they can be thought of as being more physiologically realistic. They often apply to one site (or numerous homogeneous sites) and demand a lot of input. A variety of Dynamic models are used and discovered that the more specific the model, the more precise the outcomes[2]. The foundation of machine learning approaches for dynamic models is found in several machine learning algorithms [3, [4,] [5]. An artificial neural network (ANN) [6] is a sort of machine learning algorithm that has a potent ability for learning and replicates the functioning of the human brain. However, conventional neural networks frequently contain a single hidden layer with a straightforward topology and shallow depths. A multi-layer neural network called Deep Neural Network (DNN) overtakes all other learning networks in ANN[7, 8]. The best machine learning methods now use DNNs because of their exceptional capacity for learning. Recently, DNN has been employed in the fields of image recognition, audio recognition, and natural language processing [9], [10], and [11]. Compared with traditional ANN, DNN includes Multi Layer Perceptron(MLP) and has stronger learning ability. It can

discover complex relationships between data, which has been shown to be a successful strategy for prediction and categorization.

II. Data Analysis

The empirical techniques of Artificial Neural Networks, fuzzy Logic Sets, and data mining are used to analyse the rainfall forecasts. Making predictions requires us to predict when rain will occur in a certain location. Regression is used to determine the number of millimetres (mm) of rainfall in a particular area using the Karl Pearson Correlation Coefficient. The multiple linear regression method must be used to forecast the amount of rain that will occur in the upcoming years. The empirical data used in the current study were gathered from Chennai's metrological department of Tamil Nadu. The winter months of Tamil Nadu, September, October, and November, are shown in Fig. 1 as interpolated data for five years. The Pearson correlation coefficient was applied to this data in order to calculate the amount of rainfall that fell at the ground level. The computed and graphed Pearson correlations between the data from various months are presented. The graph demonstrates that there is no discernible relationship between the three months and the monthly rainfall during the winter. The Pearson correlation coefficient between two variables is calculated as the covariance of two variables divided by the total of their standard deviations. To determine the linear relationship between two quantitative variables, one uses the Pearson correlation coefficient.

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

Where,

r = Pearson Correlation Coefficient

x_i = x variable samples $\quad$ y_i = y variable sample

$\bar{x}$ = mean of values in x variable $\quad$ $\bar{y}$ =mean of values in y variable

It assesses the degree and direction of the association between two variables and has a range of -1 to +1.

Scale of correlation coefficient	Value
$0 < r \leq 0.19$	Very Low Correlation
$0.2 \leq r \leq 0.39$	Low Correlation
$0.4 \leq r \leq 0.59$	Moderate Correlation
$0.6 \leq r \leq 0.79$	High Correlation
$0.8 \leq r \leq 1.0$	Very High Correlation

Regression Co-efficient using Co-relation and Standard Deviation is as follows:

$$byx = r \cdot \frac{\sigma y}{\sigma x}$$

Where σ is standard deviation and r is co-relation .Five years of data have been used to compute Pearson coefficient and then derived the rainfall using regression approach. Centimeters (cm) are used to measure rainfall on earth. The graph's x and y axes are used to plot the measurement and the years. We provide an output using the input data as soon as we can. As a result, the output predicts ground-

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 9, No. 3, September : 2023

level rainfall during the next three years. Because the data employs a regression approach to estimate rainfall for the following three years and uses some dependant variables to assume some values from actual values, the outcome will be approximate.

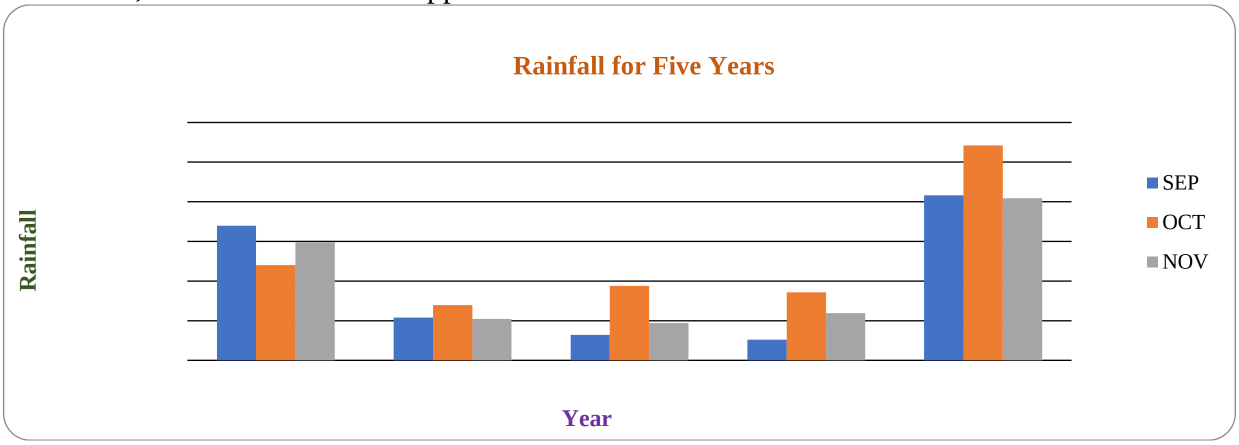


Fig 1:Input Rainfall for FIVE Years

III. Multiple Linear Regression Model

The Multiple Linear Regression (MLR) approach is used to estimate the average summer monsoon rainfall in a particular year using the monthly rainfall data from the summer monsoon of the previous year. The MLR equation is set as $\hat{Y}=a_0+a_1X_1+a_2X_2+a_3X_3$ after computation. Where the regression coefficients a_0, a_1, a_2, and a_3

X_1 = September precipitation for year y
X_2 = October precipitation for year y
X_3 = November precipitation for year
y = Average precipitation for year y+1

This method estimates the amount of rain that will occur in future years using the regression coefficients' mean values for rainfall data from September, October, and November, as illustrated in the following Fig 2.

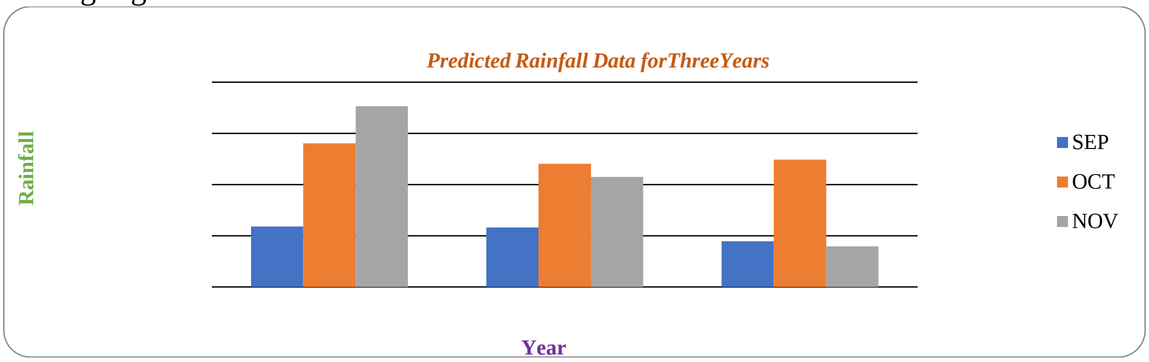


Fig 2: Predicted Rainfall Data for Three Years

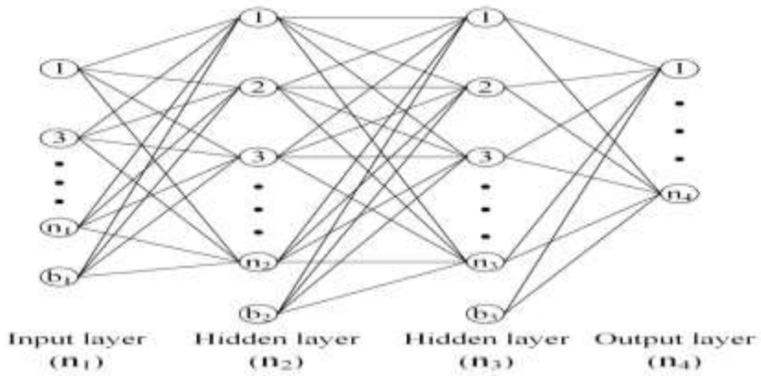
The MLR equation is used to produce the statistics, which are then compared to Pearson correlation coefficient graphic results. The three-month input value is used as the basis for the computed values and is a reliable predictor. Once the number has been determined by mathematics, It must fall within, be higher than, or be lower than the values shown by the graphical notations.All computed results are discovered to be lower than the Pearson correlation coefficient graphic values. This shows that none of the months can reliably predict the average year monsoon rainfall.

IV. Multi-Layer Perceptron

Multiple-Layer Perceptrons An artificial neural network (ANN) is an operational model composed of a huge number of interconnected nodes (neurons). Each node has its own output function, also known as an activation function. Every pair of nodes transmits a connection signal, which is represented by a connection, and each connection carries a weighted value called a weight. An ANN has a weight-equivalent memory. The MLP [12] structure is an example of a DNN, It is a subtype of ANN with a more complex structure and deeper depth. There are several layers in the MLP; the first layer is the input layer, the last layer is the output layer, the middle layers are referred to as hidden layers, and there are numerous neurons in each layer. The MLP neurons exhibit both full inter-layer and connectionless intra-layer connectivity [13]. Figure 3 depicts the MLP schematic, which has two hidden layers. The N_1 nodes in the input layer, which indicate the input data's dimension, make up the entire layer. N_2 and N_3 neurons are present in the hidden layer 1 and layer 2, respectively. N_4 neurons in the output layer represent the output data's dimension. The only layer that doesn't have a bias node is the output layer.

V. MLP Adjustment and Optimal Algorithm

The MLP is a feed-forward network that covers all middle levels by computing progressively from the input layer to the output layer. There is no interference between the nodes at the same level because they are all calculated simultaneously from their predecessors [14]. Each node's value is equal to the weighted total of all the nodes in the layer below it. The MLP feed-forward process is the name of this computation procedure. A MLP with m hidden layers will have input and output dimensions that are n1 and n_{m+2} correspondingly. Each hidden layer contains n_2, n_3.... n_{m+1} nodes. The following formula is used to determine each node value in this MLP's feed-forward process:
$x_{ij} = f(W_i X_{i-1} + b_{i-1})$ (2) where Xij stands for the j neuron's value in the i layer. In layers i-1 through i, W_i denotes the weight vector of the j neurons. The value vector of every neuron in layer i–1 is represented by X_{i-1}. The bias of the i -1 layer is represented by b_{i-1}, and the activation function is f.



VI. Data Collection and Pre-processing

Principal Component Analysis (PCA) and min-max normalisation make up the bulk of the pre-processing effort. Five high-altitude components and eight earth surface elements make up the model's initial data. An isobaric surface is typically used in place of horizontal height and geopotential height, respectively. As a result, weather information is always shown as an isobaric surface. For instance, 500hPa typically equates to 5.5 km of height. The weather system at 500 hPa frequently regulates the rainfall system. Actual height (X_1), temperature (X_2), Dew point and temperature differences (X_3), winds' direction (X_4), and wind speed (X_5) at 500hPa height are the five altitude elements chosen by this model based on the domain experiences. The direction and speed of the wind have an impact on the movement of the rainfall system. Humidity and temperature dew point differences are directly correlated. The internal energy of the rainfall system is influenced by temperature, the gap between dew point and real height, and temperature. The local atmospheric conditions in the region are represented by the ground surface factor. Different rainfall occurs in different regions due to variations in surface variables. The total amount of clouds (X_6), the rate of ground surface wind (X_7), direction of the ground surface wind (X_8), ground-level atmospheric pressure (X_9), 3 hour pressure change at the ground's surface (X_{10}), temperature of the earth's surface (X_{12}), difference in the ground surface temperature and dew point (X_{11}), and neighbouring areas experienced three hours of rainfall (X_{13}) are the eight surface-level factors that were considered. Every surrounding region creates an MLP with the same forecasting area for that forecasting area. The thirteen factors are listed in Table 1. These elements serve as our model's first input. Min-max Normalisation. Min-max normalisation is one of the most often used techniques for data normalisation. Data between 0 and 1 can be standardised. Data pre-processing is required since the magnitude of many elements varies. When a sequence is processed, its highest value correlates to 1, its lowest value to 0, and its remaining values are proportionally converted between 0 and 1. The following is a representation of the min-max normalization formula:

$x^* = x - x_{min} / x_{max} - x_{min}$

where x denotes a value in the series of primitive variables, and x_{max} and x_{min} represent the highest and lowest values in the variables. After normalisation, PCA is used to shrink the input's dimension. The 99% criterion, which states that the Eigen values of the selected factors together account for more than 99% of all Eigen values, is used to determine how many new factors should be added. More than 99% of the original data can be represented by the total information of the new factors after calculation. The number of components required is not specified by this criterion because it may vary depending on the forecasting area but will not go beyond the initial inputs. However, it does determine the amount of information maintained. Three to eight elements are often required, on average. The amount of computer power needed is drastically decreased following the PCA technique. The following Table 1 enlist the factors considered for the MLP.

TABLE 1

Factor	Input Value
500hPa height(X1)	Forecast area value
500hPa temperature(X2)	Forecast area value
500hPa temperature dew point difference(X3)	Forecast area value
500hPa wind direction(X4)	Forecast area value
500hPa wind speed(X5)	Forecast area value
Total cloud amount (X6)	Neighbouring area value
Surface wind speed (X7)	Neighbouring area value
Surface wind direction (X8)	Neighbouring area value
Surface air pressure(X9)	Neighbouring area value
Surface 3 hour pressure change(X10)	Neighbouring area value
Surface temperature dew point difference(X12)	Neighbouring area value
Rainfall over pass 3 hours(X13)	Neighbouring area value

VII. Conclusions

Using the empirical method, utilizing only three months' worth of input data from five years, we derived values for rainfall fall at ground level utilizing Karl Pearson correlation coefficient and MLR to forecast rainfall at ground level in the next years. Predicted values eventually fall barely short of the actual values. With more Dynamic parameter the predicted values can be more close to the actual values.

VIII. References

[1]D.B. Lobell, M.B. Burke, On the use of statistical models to predict crop yield responses to climate change Agric For. Meteorol., 150 ,pp. 1443-1452,2018.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 9, No. 3, September : 2023

[2] Korhonen, T. Palosuo, T. Persson, M. Höglind, G. Jégo, VanOijen, A.-
stavsson, G. Bélanger, P. Virkajärvi
Modelling grass yields in northern climates – a comparison of three growth models for timothyField Crop. Res., 224 ,pp. 37-47,2018.
[3]N. Bartoletti, F. Casagli, S. Marsili-Libelli, A. Nardi, and L. Palandri,"Data-driven rainfall/runoff modelling based on a neurofuzzyinference system," Environmental Modelling &Software,vol. 106, pp. 35–47, 2018.
[4]L. Cozzi, "Weather models as virtual sensors to data drivenrainfall predictions in urban watersheds," EGU General AssemblyConference Abstracts, vol. 15, pp. 111–115, 2013.
[5]K. Sudheer, A. Gosain, and K. Ramasastri, "A data-driven algorithm for constructing artificial neural network rainfall-runoffmodels," Hydrological processes, vol. 16, no. 6, pp. 1325–1330, 2002.
[6]R. J. Schalkoff, Artificial neural networks, vol. 1. McGraw-Hill NewYork, 1997.
[7]X. Glorot and Y. Bengio, "Understanding the difficulty of trainingdeep feedforward neural networks," vol. 9, pp. 249–256, 2010.
[8] J. Schmidhuber, "Deep learning in neural networks: An overview,"Neural networks, vol. 61, pp. 85–117, 2015.
[9] D. Ciregan, U. Meier, and J. Schmidhuber, "Multi-column deepneural networks for image classification," in Computer vision andpattern recognition (CVPR), 2012 IEEE conference on, pp. 3642–3649,IEEE, 2012.
[10] R. Collobert and J. Weston, "A unified architecture for natural languageprocessing: Deep neural networks with multitask learning,"in Proceedings of the 25th international conference on Machine learning,pp. 160–167, ACM, 2008.
[11] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A.-r. Mohamed, N. Jaitly,A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath, et al., "Deepneural networks for acoustic modeling in speech recognition: Theshared views of four research groups," IEEE Signal ProcessingMagazine, vol. 29, no. 6, pp. 82–97, 2012.
[12] T. Shimobaba, N. Kuwata, M. Homma, T. Takahashi, Y. Nagahama, M. Sano, S. Hasegawa, R. Hirayama, T. Kakue, A. Shiraki, et al., "Deep-learning-based data page classification for holographic memory," arXiv preprint arXiv:1707.00684, 2017.
[13] D. Imseng, P. Motlicek, P. N. Garner, and H. Bourlard, "Impact of deep mlp architecture on different acoustic modeling techniques for under-resourced speech recognition," in Automatic Speech Recognition and Understanding (ASRU), 2013 IEEE Workshop on, pp. 332–337, IEEE, 2013
[14] S. Lange, K. Gehmlich, A. S. Lun, J. Blondelle, C. Hooper, N. D. Dalton, E. A. Alvarez, X. Zhang, M.-L. Bang, Y. A. Abassi, et al., "Mlp and carp are linked to chronic pkcα signalling in dilated cardiomyopathy," Nature communications, vol. 7,2016.