Estimation of Rainfall in Karnataka Using Maching Learning Techniques

Mithuna Vasan R[1], Smt. Anitha R[2]

[1]M.Tech (IT) (Batch of 2020)
[2]Associate Professor
Dept. of Computer Science & Engineering
The National Institute of Engineering, Mysuru, Karnataka, India

---***---

Abstract - *Weather forecasting and rainfall prediction in particular is one of the key areas people from all walks of life always look for. Economy, agriculture and drinking water are a few core dependent zones of rainfall and hence accurate forecasting is vital. This paper presents a set of approaches and experiments to build models using machine learning techniques to predict how much precipitation each month would get for the given year in North interior, South interior and Coastal regions of Karnataka. The research is also focused on data pre-processing, modelling approaches and analysis and comparison of three algorithms Linear Regression (LR), Support vector regression (SVR) and Artificial neural network (ANN). Apart from models that use precipitation as inputs, the algorithms are trained and tested on another dataset that contains meteorological parameters [2] in this study. The outcome of this work includes predicted precipitation in mm, a process for assessing accuracy of models using Mean absolute error between predicted and actual values and comparison between algorithms.*

Keywords: Precipitation, Meteorological Parameters, Machine learning, Mean absolute error.

1. INTRODUCTION

With the advent of technology, machine learning algorithms contribute significantly in developing rainfall prediction systems. Indian Meteorological Department (IMD) has been forecasting the monsoon for over four decades. In spite of maintaining more than 6000 observatories by IMD and state governments, making accurate predictions for tropical weather has been challenging due to its unpredictability nature. The State of Karnataka is situated at 11°30' North and 18°30' North scopes and 74 East and 78°30' East longitude [1].

This study is an effort to analyze the performance of three machine learning algorithms in prediction of monthly and annual rainfall for all three regions namely Southern interior, North interior and Coastal Karnataka. Historical meteorological data has been downloaded from IMD website and Indiawaterportal.org websites to identify rainfall patterns and prediction.

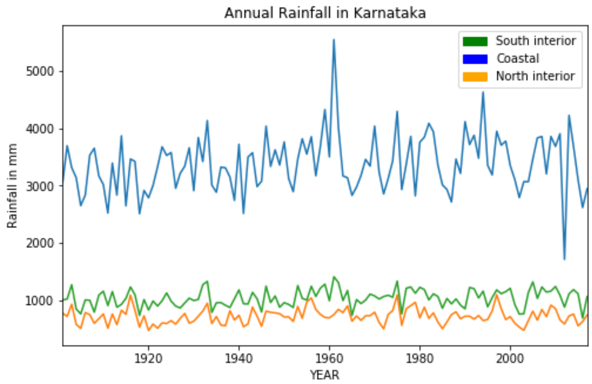

Fig.1 Annual Rainfall in Karnataka

In this paper, three algorithms are analyzed and experimented based on three different set of inputs for each region separately so as to evaluate the accuracy and suitability of the models. Precipitation and other rainfall parameters for this work are collected, trained and tested to achieve the sustainable results through algorithms. The monthly rainfall predictions obtained after training and testing are then compared with actual data. High level outcome is that models built on monthly precipitation, 12 models for 12 months, are successful in predicting the monthly precipitation for the given year. This approach combined with SVR and ANN helped in minimizing errors and maximizing precision. Support vector regression presented maximum precision with minimum error through the comparison between the actual data and predicted outcome data than ANN. This has also validated findings of Sumi et al. [2] that SVR produce more accurate forecast than ANN. Mean Absolute Error between predicted precipitation of a month and actual is as low as 0.1mm and 0.2mm. Since SVR and ANN outperformed LR, LR could be used for learning purpose than for advanced forecasting. Datasets comprised of meteorological data downloaded from IMD and Indian water portal websites are used for this work.

2. LITERATURE SURVEY

There are different techniques used for the prediction of rainfall such as Regression analysis, clustering and

Artificial Neural Networks (ANN). Fundamentally, two approaches are used for predicting rainfall. One is Empirical approach [3] and the other is Dynamical approach [4]. The empirical approach [5] is based on analysis of historical data of the rainfall and its relationship to a variety of atmospheric and oceanic variables over different parts of the world. The most widely used empirical approaches, which are used for climate prediction, are regression, artificial neural network, fuzzy logic and group method of data handling.

Refonaa et al. [6] used Linear regression technique to classify the images and predict the rainfall. Deo and Shain [7] predicted Drought Index in eastern Australia using ML based on rainfall and mean, minimum and maximum air temperatures. Sen [8] has presented long-range summer monsoon rainfall forecast model based on power regression technique with the use of El Nino, Eurasian snow cover, north west Europe temperature, Europe pressure gradient, 50 hPa Wind pattern, Arabian sea SST, east Asia pressure and south Indian ocean temperature in previous year. The experimental results showed that the model error was 4%.

Kashiwao et al. [5] explained and compared the ANN models used for rainfall prediction experiment. Quiet a good number of researchers used ANN algorithm [5, 9-15], produced useful results and motivation for including Neural networks for this study. Hasan et al. [16] combined different kernel functions of Support Vector Machine (SVM) and an exclusive data preprocessing technique windowing to predict rainfall.

Botsis et. al [17] compared support vector regression (SVR) and multilayer feed-forward neural network (MFNN) models with respect to their forecasting capabilities. The results of the study demonstrate that the SVR model is more effective than MLP in simulating the rainfall–runoff relationship and the SVR algorithms have better generalization capability than the conventional artificial neural networks.

3. Methodology

A. Approaches

With modelling inputs taken from two datasets, monthly/annual precipitation for Karnataka for the period of 1901-2017 (Dataset1), and meteorological parameters data for Mysore District for the period 1901-2002 (Dataset2), following approaches are implemented.

Approach 1: Precipitation of previous three months are used as predictors for next month precipitation. Whole Dataset1 is converted into two arrays, first array is predictor variable X, consists of triplets of consecutive three months and second array is dependent variable Y, consists of fourth months. For example, Precipitation of January, February and March are predictors for April. For predicting January precipitation, last three months of previous year are used.

Approach 2: Dataset1 is divided into 12 arrays for 12 months and vertical consecutive three months are predictors for next month. For example, Precipitation of April 2001 is predicted using April 1998, April 1999 and April 2000.

Approach 3: Dataset2 consists of meteorological parameters Cloud cover, Average temperature, Wet day frequency, Vapor pressure and Precipitation for each month. Precipitation for any month is predicted by other four parameters of the same month. As there are four predictors and data are huge, this approach is experimented only for Mysore district.

B. Data Preprocessing

Data pre-processing involves transforming raw data into consistent, complete, error-free and compatible as per requirements. Replacing null values with respective mean values and extending categorical variables, extraction, merging are a few steps performed as part of pre-processing for this study. For Approach 1, last three months of previous year are prepended to next year. For Approach 3, meteorological parameters extracted from five files, merged and arranged in such way that each month in a row gets five values.

C. Algorithms

Machine learning algorithms implemented for this study are explained in this section.

1. Linear Regression

Linear Regression comes under Supervised Learning. Prediction of dependent variable is thru independent variable by making linear relationship between them. Graph in Fig 2 shows the equation line Y=mX+b when values are plotted. Linear regression is chosen for this study as it is Regression is best suitable for linearly separable data and is being used in many mathematical applications.

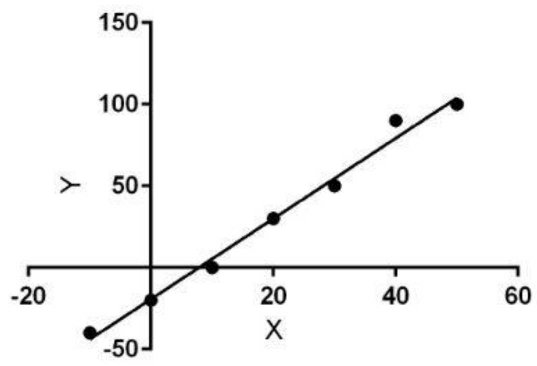

Fig.2 Linear Regression

2. Support Vector Regression

SVR [18] is used on continuous values for regression in which the linear function is extended to non-linear mapping. A linear regression function is constructed as

$$(\boldsymbol{x}) = \langle \boldsymbol{w}, \boldsymbol{x} \rangle + b \text{ where } \boldsymbol{w}, \boldsymbol{x} \in \mathbb{X} \text{ and } b \in \mathbb{R}.$$

Objective is to find a function $(\boldsymbol{x})$ with maximum ϵ deviation. To make $(\boldsymbol{x})$ flat

$$\text{minimize } \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{l}(\xi i + \xi i')$$

$$\text{subject to } \begin{cases} y - \langle \omega, \Phi(x) \rangle - b \le \varepsilon \\ \langle \omega, \Phi(x) \rangle + b - y \le \varepsilon \end{cases}$$

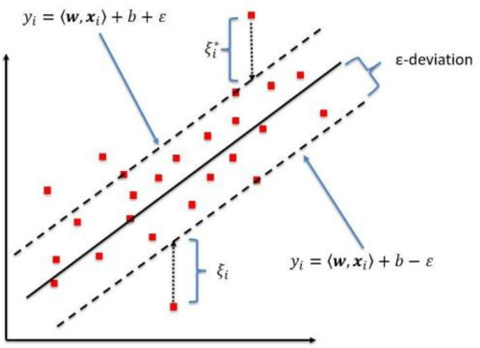

Fig.3 Support Vector Regression

For this study, Grid-search is used to optimize the parameters associated with SVR [18] and RBF kernel function for mapping lower dimensional data to higher dimension. Grid search helps in tuning hyper parameters C and gamma so as to improve performance of model. As in Fig.3, boundary at epsilon distance from the original hyper plane is decided in such a way that data points closest to the hyper plane or the support vectors are within that boundary line.

3. Artificial Neural Network

A typical supervised learning algorithm ANN is mainly comprised of input, output and hidden layers. Input transformed by hidden layers are used by output layer. The interconnected computational nodes also called as neurons are basically tunable units. They learn from input and pass the signal to another unit so as to optimize the output [19]. The number of layers, their types, and the way they are connected to each other is called the network architecture.

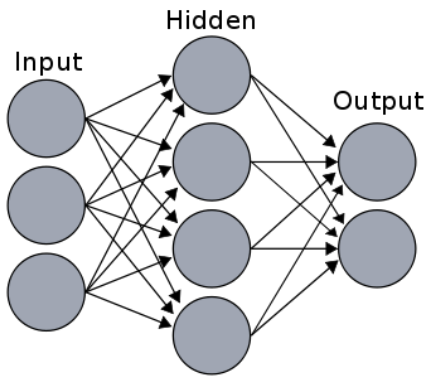

Fig.4 ANN Structure

The structure of ANN is shown in Fig 4. Except for the input nodes, each node is a neuron that uses a nonlinear activation function. Weight indicates the strength of a particular node and bias value makes the activation function up or down. Neurons in hidden layers use activation function to transform input signal to output which is used as input to nodes in the next layer in the network. Activation function calculates weighted sum and adds bias to produce nonlinear output. ANN uses the technique of error back propagation to train the network configuration [2] and given below is the formula used:

$$x_{t+T+(m-1)T}^{F}$$

$$= f\,(X_t, \omega, \theta, m, h)$$

$$= \theta_0 + \sum_{i=1}^{m} \omega_j^{out} \emptyset\left(\sum_{i=1}^{m} \omega_{ji} x_{t+(i-1)T} + \theta_j\right)$$

In order to improve efficiency by tuning parameters, a variation of ANN topology called Convolutional neural network is used for this study. CNNs are generally applied for image classification problems, where the model learns an internal representation of a two-dimensional input, in a process referred to as feature learning. They possess proven ability in extracting complex relationships in input features [20]. Layers are comprised of convolution units and each convolution unit receives its input from multiple units from the previous layer which together form a proximity. Therefore, the input units that form a small neighborhood share their weights [21]. Neurons in CNNs [19] self-optimize through learning. They perform operations such as scalar product followed by a nonlinear function which is basis of any ANN. Neurons learn from data through its hierarchical structure of layers and enhances the learning of complex relationships in input features. Python library is used for experiments with following parameter features:

Activation function - ELU Exponential unit
Conv1D - one-dimensional convolution unit
Optimizer - admax
Loss function – Mean absolute error
Epochs (iterations) is set to 150. For each epoch,

- Processing records of the training data case,
- Comparing actual values to predicted values,
- Calculating the loss function,
- Backpropagating error through the layers and
- Adjusting network weights and finally validation

are taking place till better loss value is obtained.

4. Experimental Results

Various experiments conducted for this study are explained in this section. Table 1 shows predictor variables (X) used for each experiment, modelling inputs and algorithm.
Fig 5 shows forecasting procedure followed for this study.

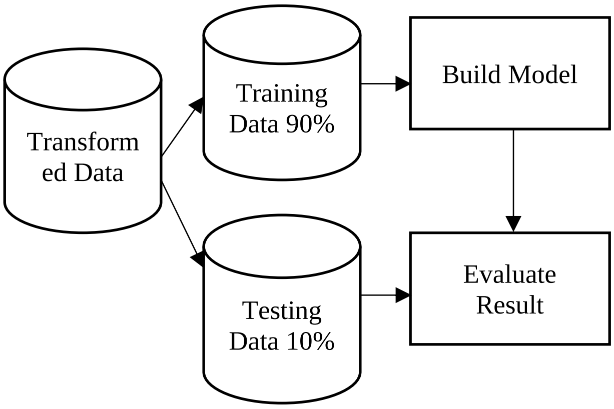

Fig.5 Procedure for forecasting

Mean absolute error between actual precipitation and predicted precipitation is calculated to find out the accuracy of model

	Predictor	Location	Period	Output variable Y (Precipitation)	Model
Approach 1	Precipitation	South interior Karnataka	Previous three months	For each month	LR
	Precipitation	South interior Karnataka	Previous three months	For each month	SVR
	Precipitation	South interior Karnataka	Previous three months	For each month	ANN
	Precipitation	North interior Karnataka	Previous three months	For each month	LR
	Precipitation	North interior Karnataka	Previous three months	For each month	SVR
	Precipitation	North interior Karnataka	Previous three months	For each month	ANN
	Precipitation	Coastal Karnataka	Previous three months	For each month	LR
	Precipitation	Coastal Karnataka	Previous three months	For each month	SVR
	Precipitation	Coastal Karnataka	Previous three months	For each month	ANN
Approach 2 Separate model for each month	Precipitation	South interior Karnataka	Same month of previous 3 years	For each month	LR
	Precipitation	South interior Karnataka	Same month of previous 3 years	For each month	SVR
	Precipitation	South interior Karnataka	Same month of previous 3 years	For each month	ANN
	Precipitation	North interior Karnataka	Same month of previous 3 years	For each month	LR
	Precipitation	North interior Karnataka	Same month of previous 3 years	For each month	SVR
	Precipitation	North interior Karnataka	Same month of previous 3 years	For each month	ANN
	Precipitation	Coastal Karnataka	Same month of previous 3 years	For each month	LR
	Precipitation	Coastal Karnataka	Same month of previous 3 years	For each month	SVR
	Precipitation	Coastal Karnataka	Same month of previous 3 years	For each month	ANN
Approach 3	Cloud cover, Average temperature, Wet day frequency, Vapor pressure	Mysore District	For each month	For each month	LR
		Mysore District	For each month	For each month	SVR
		Mysore District	For each month	For each month	ANN

Table 1. Predictors, Location and Period used in experiments

	South Interior Karnataka				North Interior Karnataka				Coastal Karnataka			
		Linear	SVR	ANN		Linear	SVR	ANN		Linear	SVR	ANN
Approach 2	Overall	27.32	26.91	29.58	Overall	22.65	21.35	24.17	Overall	81.53	79.59	97.56
	YEAR 2001	27.48	19.87	29.3	YEAR 2001	26.14	4.8	23.45	YEAR 2001	40.54	34.18	31.89
	YEAR 2010	25.5	12.88	21.25	YEAR 2010	21.48	3.57	20.52	YEAR 2010	68.07	43.78	39.97
	YEAR 2017	38.4	35.82	35.53	YEAR 2017	23.72	22.27	24.39	YEAR 2017	35.41	44.65	38.59
		Linear	SVR	ANN		Linear	SVR	ANN		Linear	SVR	ANN
Approach 1	Overall	44.72	35.97	37.36	Overall	36	32.31	31.23	Overall	229.47	146.46	129.99
	YEAR 2001	41.25	40.13	37.11	YEAR 2001	31.76	24.61	28.21	YEAR 2001	163.24	50.97	45.98
	YEAR 2010	50.53	36.44	41.88	YEAR 2010	33.92	33.35	33.7	YEAR 2010	203.99	67.22	118.08
	YEAR 2017	47.87	53.26	47.77	YEAR 2017	41.94	33.63	34.26	YEAR 2017	130.88	66.1	51.56

Table 2. Mean Absolute Errors of experiments conducted

Results for all experiments using Approaches 1 and 2 are shown in Table 2. Significant difference in performance among Linear, SVR and ANN are observed. Approach 2 performed better than Approach 1. SVR has outperformed other two algorithms as Mean absolute error is quite less. A sample output Fig clearly indicates prediction by SVR is close to actual.

Sample output prediction for all three regions of Karnataka is shown for the year 2010 in Fig 6., Fig 7. and Fig 8. by SVR model using Approach 2. It is observed that prediction is almost matching with actual. Test cases for prediction are years 2001, 2010 and 2017.

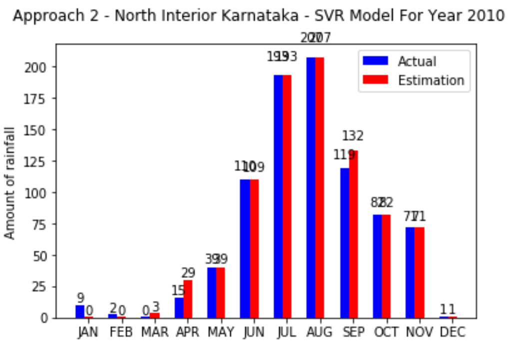

Fig 6. North interior

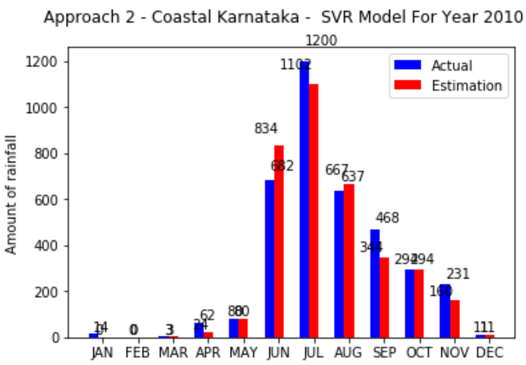

Fig 7. Coastal Karnataka

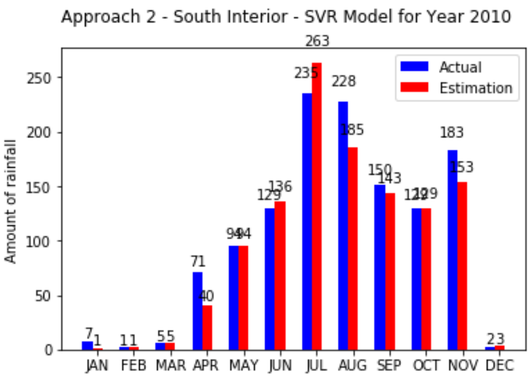

Fig 8. South Interior

Comparison between algorithms is shown in Fig 9. to indicate precision in rainfall prediction by SVR.

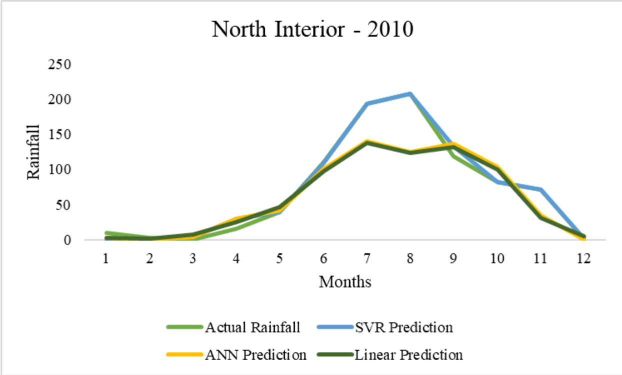

Fig 9. Comparison graph

Approach 3, MAE results by all algorithms are shown in Table 3 for Mysore District based on meteorolgical parameters for sample years 1990 and 2001. Predictions are not so accurate as other two approaches. They tend to show negative predicted values in spite of parameter tuning and hence this needs further study. Modeling inputs are meteorological parameters measured in the same month for which precipitation is predicted.

Table 3. Rain parameter-based model **-** Mysore District

	Linear	SVR	ANN
Overall	70.78	53.59	53.87
YEAR 1990	34.91	20.62	24.33
YEAR 2001	46.51	21.01	19.27

5. Conclusion

This paper explores the performance of three algorithms in monthly and annual rainfall prediction with three different approaches for the regions of Karnataka and particularly suggests to implement SVR model with underlying Grid search technique for super accuracy. Second point is that monthly vertical models on historical precipitation data used in Approach 2 produce high precision results. Separate model implementation for each region is necessary due to varying rainfall patterns across Karnataka. Prediction based on Precipitation of consecutive months used in Approach 1 shows average performance only. Meteorological parameters-based prediction modeling needs further study as the mean absolute error on higher side and hence future scope of forecasting process may include finding correlation between individual rain parameter and precipitation by creating separate models.

6. References

[1] R Sivajothi, K Karthikeyan, B Sarkar. Analysis of precipitation pattern using various stochastic distributions in Karnataka state, India. Journal of Physics: Conference Series, Volume 1139, International Conference on Applied and

Computational Mathematics 10 August 2018, Tamilnadu, India

[2] Sumi SM , Zaman MZ, Hirose H. A Rainfall Forecasting Method Using Machine Learning Models and its Application to the Fuuoka City Case. International Journal of Applied Mathematics and Computer Science December 2012.

[3] Hashim, R, Roy C, Motemedi S and Smamshirband A . Selection of meteorological parameters affecting rainfall estimation using neuro-fuzzy computing methodology. Atmospheric Research 171, 21–30 (2016).

[4] Wint Thida Zaw, Thinn Thu Naing. Empirical Statistical Modelling of Rainfall Prediction over Myanmar. World Academy of Science, Engineering and Technology, International Journal of Computer and Information Engineering Vol:2, No:10, 2008

[5] Kashiwao T, Nakayama K, Ando A, Ikeda K, Lee M, Bahadori A. A neural network-based local rainfall prediction system using meteorological data on the Internet: A case study using data from the Japan Meteorological Agency. Applied Soft Computing Volume 56, July 2017, Pages 317-330

[6] Refonaa J, Lakshmi M, Abbas R, Raziullha M. Rainfall Prediction using Regression Model. International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8, Issue-2S3, July 2019

[7] Deo, R. C. & Şahin, M. Application of the extreme learning machine algorithm for the prediction of monthly Effective Drought Index in eastern Australia. Atmospheric Research 153, 512–525 (2015)

[8] Sen N “New forecast models for Indian south-west Monsoon season Rainfall”, in Current Science, vol. 84, No. 10, May 2003, pp.1290- 1291.

[9] Litta AJ, Idicula SM, and. Mohanty UC. Artificial Neural Network Model in Prediction of Meteorological Parameters during Premonsoon Thunderstorms. Hindawi Publishing Corporation, International Journal of Atmospheric Sciences Volume 2013, Article ID 525383, 14 pages.

[10] Teschl, R., Randeu, W. L. & Teschl, F. Improving weather radar estimates of rainfall using feed-forward neural networks. Neural networks 20, 519–527 (2007).

[11] French, M. N., Krajewski, W. F. & Cuykendall, R. R. Rainfall forecasting in space and time using a neural network. Journal of hydrology 137, 1–31 (1992).

[12] Yen, M., Liu, D., Hsin, Y. . Application of the deep learning for the prediction of rainfall in Southern Taiwan. *Sci Rep* **9,** 12774 (2019).

[13] Kin C. Luk, J.E.Ball and A.Sharma. An Application of Artificial Neural Networks for Rainfall Forecasting. Mathematical and Computer Modelling 33 (2001) 883-699.

[14] Chaudhari MS, Choudhari NK. Study of Various Rainfall Estimation & Prediction Techniques Using Data Mining. American Journal of Engineering Research (AJER) e-ISSN: 2320-0847 p-ISSN : 2320-0936 Volume-6, Issue-7, pp-137-139

[15] Parmar A, Mistree K, Sompura M. Machine Learning Techniques for Rainfall Prediction: A Review Conference Paper · March 2017. Conference: 2017 International Conference on Innovations in information Embedded and Communication Systems (ICIIECS), At Coimbatore, Volume: 3

[16] Hasan N, Nath NC, Rasel RI. A support vector regression model for forecasting rainfall. Published in 2nd International Conference on Electrical Information and Communication Technologies (EICT) 2015.

[17] Botsis D, Latinopoulos L and Diamantaras K. Rainfall-runoff modeling using support vector regression and artificial neural networks. 12th International Conference on Environmental Science and Technology (CEST2011)

[18] Zhang H, Chen L, Qu Y, Zhao G, Guo Z. Support Vector Regression Based on Grid-Search Method for Short-Term Wind Power Forecasting. Computational Science in Smart Grids and Energy Systems. Volume 2014 |Article ID 835791

[19] O'Shea K, Nash R. An Introduction to Convolutional Neural Networks. ArXiv e-prints, 2015/11/01.

[20] Kleynhans T, Montanaro M, Gerace A, Kanan C. Predicting Top-of-Atmosphere Thermal Radiance Using MERRA-2 Atmospheric Data with Deep Learning. Remote Sensing 9(11):1133 · November 2017.

[21] Haidar A and Verma B, "Monthly Rainfall Forecasting Using One-Dimensional Deep Convolutional Neural Network," in IEEE Access, vol. 6, pp. 69053-69063, 2018, doi: 10.1109/ACCESS.2018.2880044.