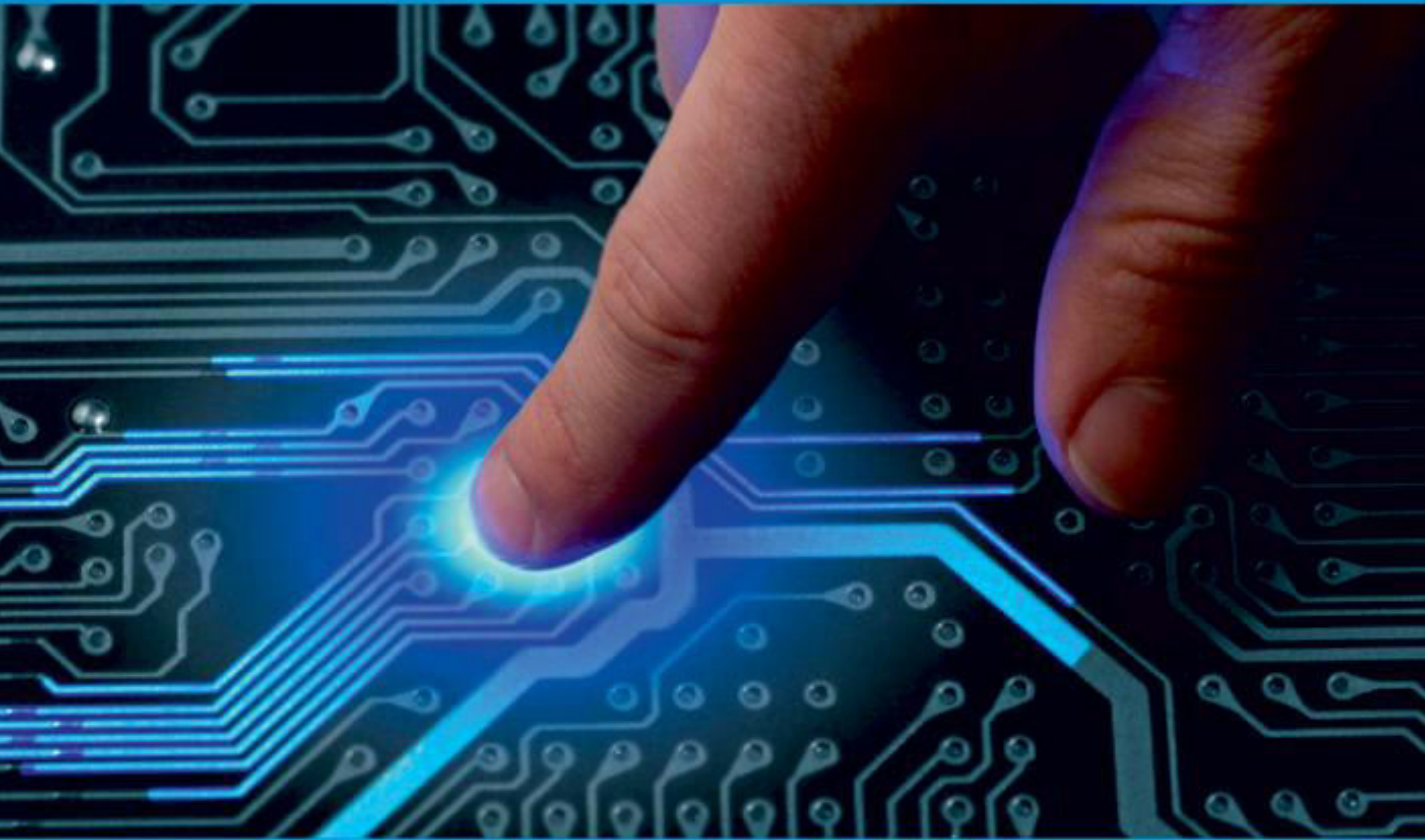
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 7, July 2021

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.542

Topology Preserving Automatic Rainfall Prediction and Statistics Analysis Based on Neural Network

Ms. P. Saranya, Ashwik Acharya, Dhanush Rajprasad, K. Sri Gnanesh

Assistant Professor, Dept. of CSE, Velammal Engineering College, Chennai, India

UG Student, Dept. of CSE, Velammal Engineering College, Chennai, India

ABSTRACT: Rainfall forecasting is of utmost importance as even a slight increase or decrease in rainfall could lead to disasters and also affect the production of crops and grains, forecasting acts as an early warning and helps to overcome and improve practices to manage such calamities. The existing systems can not learn the relationship between various factors such as humidity, dew point, solar radiation has on rainfall and is prone to errors. The proposed system effectively aims to improve prediction accuracy and reduce the consumption of hardware resources

A significant part of the proposed system involves predicting rainfall using our hybrid neural network model from past data. The performance of the results is measured with MSE (mean squared error), correlation coefficient, coefficient of efficiency and MAE (mean absolute error).

KEYWORDS: Deep Learning, Neural Networks, Numpy, Matlab, Data Frame

I. INTRODUCTION

Two of the most common issues are Overfitting and Computation time. Overfitting occurs when we achieve a good fit of our model on the training data, while it does not generalize well on new, unseen data. This occurs primarily due to the added layers of abstraction, which allow them to model rare dependencies in the training data.
Many training parameters need to be considered by the Deep Neural Network such as the number of layers, number of units per layer, the learning rate and initial weights. Tuning to choose a set of optimal hyperparameters is usually not feasible due to the cost of time and computational resources.
Deep Learning models require a large quality data set to become adept in making predictions it is noticed that when random noise is added to the data set to increase the data set size, the quality of the model declines drastically.

II. RELATED WORKS

1. "Rainfall Rate Prediction for Propagation Applications: Model Performance at regional Level Over Ireland":
Three global rainfall rate prediction methods are evaluated in their ability to estimate local precipitation statistics, which are the key to predicting the impact of rain on the propagation of electromagnetic waves through the atmosphere. Specifically, the International Telecommunication Union-Radiocommunication Sector (ITU-R) P.837-6, a model for rainfall statistics estimation (MORSE), and the ITU-R P.837-7 prediction methods are tested against long-term rainfall data collected in 19 sites in Ireland. The results indicate that the ITU-R P.837-7 prediction method delivers the best performance and that both the ITU-R P.837-6 prediction method and MORSE exhibit a positive bias, likely due to the overestimation of the yearly rain amount in the maps used as input to such models.

2. " A Data-Driven Approach for Accurate Rainfall Prediction ":

This paper proposes a systematic approach to analyze various parameters that affect precipitation in the atmosphere. Different ground-based weather features such as Temperature, Relative Humidity, Dew Point, Solar Radiation, PWV along Seasonal and Diurnal variables are identified, and a detailed feature correlation study is presented. While all features play a significant role in rainfall classification, only a few of them, such as PWV, Solar Radiation, Seasonal, and Diurnal features, stand out for rainfall prediction. Based on these findings, an optimum set of features are used in a data-driven machine learning algorithm for rainfall prediction. The experimental evaluation using a 4-year (2012–

2015) database shows a true detection rate of 80.4%, a false alarm rate of 20.3%, and an overall accuracy of 79.6%. Compared to the existing literature, our method significantly reduces the false alarm rates.

3. " Multimodel Prediction of Monsoon Rain Using Dynamical Model Selection":

In this paper, the dynamical-model-selection-based multimodel ensemble (DMS-MME) technique is developed for skill improvement of monsoon rain prediction in the medium range (i.e., 24–120 h ahead). The data set consists of 24–120 h daily precipitation forecasts from five state-of-the-art global circulation models (GCMs), i.e., European Centre for Medium-Range Weather Forecasts (Europe), National Centres for Environmental Prediction (USA), China Meteorological Administration (China), Canadian Meteorological Centre (Canada) and U.K. Meteorological Office (U.K.). The DMS-MME forecasts are constructed during the monsoon months (JJAS) for the years 2008–13 over the Indian mainland. For training and verification purposes, the India Meteorological Department rainfall is used. The forecast skill of the DMS-MME model has been compared with the performance of individual models and regression-based MME models. Further, to remove the nonnormality of rainfall distribution, square-root and logarithmic transfer functions are used for normalizing the precipitation data. The impact of these transfer functions on the forecast skill of the DMS-MME model has been assessed. The forecast skill of the proposed model is evaluated using the standard statistical measures. DMS-MME forecasts carry higher skill in terms of verification scores compared with the MME forecasts up to 120 h. It has been found that using the DMS-MME approach with a square-root transfer function (SDMS-MME) gives the best results. SDMS-MME outperforms the operational models and the regression-based MME at all forecast steps.

III. DESIGN AND IMPLEMENTATION

3.1 METHODOLOGIES

Linear Regression algorithm tries to find the relationship between Humidity and Temperature based on the linear equation :

$$y = a\_0 + a\_1 * x$$

Linear regression algorithms find the best value for a_0 and a_1 .

Decision Trees are a nonparametric supervised learning method used for classification and regression . The goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features. A tree can be seen as a piecewise constant approximation.

Logistic regression uses the logistic function at the core , the logistic function is a s type curve that takes any real value number and maps it to a value between 0 and 1 .

$$1 / (1 + e^{-value})$$

3.2 RAINFALL PREDICTION ARCHITECTURE

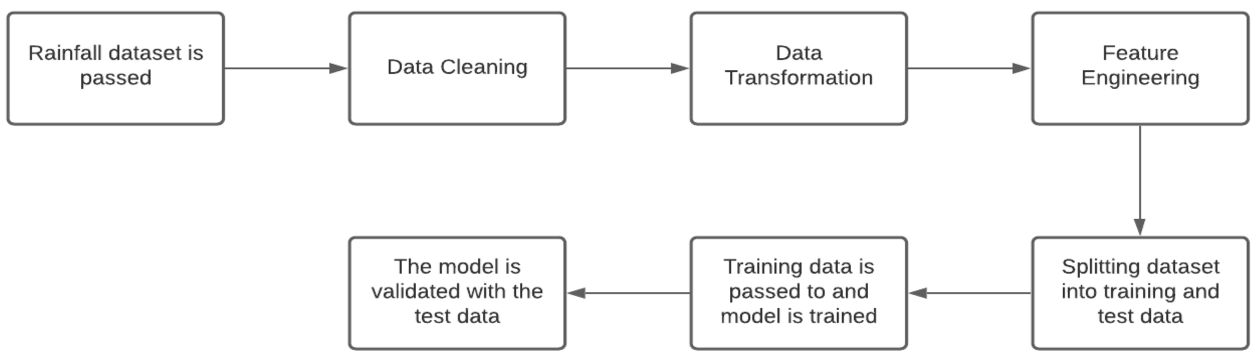


Fig 3.1 Rainfall Prediction Architecture

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907203 |

3.2.1 DATASET PASSED

Dataset of a particular region is taken into account.This dataset has many information like temperature.,humidity,pressure,wind speed, dew point,visibility etc.The dataset is in a raw form which will be corrected in the following data cleaning and preprocessing step.We use this data that was obtained to train the algorithm for the desired results based on the information provided.

Formatted ...	Summary	Precip Type	Temperatu...	Apparent ...	Humidity	Wind Spee...
2006-04-01 00:00:00.000 +0200	Partly Cloudy	rain	9.472222222222221	7.388888888888875	0.89	14.1197
2006-04-01 01:00:00.000 +0200	Partly Cloudy	rain	9.355555555555558	7.227777777777776	0.86	14.2646
2006-04-01 02:00:00.000 +0200	Mostly Cloudy	rain	9.377777777777778	9.377777777777778	0.89	3.9284000000000003
2006-04-01 03:00:00.000 +0200	Partly Cloudy	rain	8.2888888888888889	5.944444444444446	0.83	14.1036
2006-04-01 04:00:00.000 +0200	Mostly Cloudy	rain	8.755555555555553	6.977777777777779	0.83	11.0446
2006-04-01 05:00:00.000 +0200	Partly Cloudy	rain	9.222222222222221	7.1111111111111	0.85	13.9587
2006-04-01 06:00:00.000 +0200	Partly Cloudy	rain	7.733333333333334	5.522222222222221	0.95	12.3648
2006-04-01 07:00:00.000 +0200	Partly Cloudy	rain	8.77222222222222	6.527777777777778	0.89	14.1519

Fig 3.2 Weather History Dataset

3.2.2 DATA CLEANING

The dataset obtained undergoes a process called data cleaning in which the inaccurate or the outlier records will be detected and corrected It assures and increases the accuracy of the model.The data has few records where the required factors were not recorded and the rainfall in centimeters was marked as “T” if there was trace precipitation.

In this algorithm we cannot work with alphabets hence we need to clean the data to have them as numericals.The values are converted to numericals and then used for the further processes

3.2.3 DATA TRANSFORMATION

After the process of data cleaning and preprocessing the values are evaluated for valid records . If the records are not valid then they are detected and corrected based on their values.The values need to be of more value if not then they are removed from the table . For example if a table has more number of null values which gives no values and of no use then they are removed . It reduces the space and also helps in finding the data even more accurate . The table with less number of null values is evaluated and the average value is taken into account .

Unused and repeated columns are removed in this possession to increase the data accuracy . Rounding Off the numerical values also help and using the correct data types increases the efficiency of the process . All kinds of times stamps are removed to handle the time and date anomalies of the record . It also involves converting categorical data into numerical values ; many machine learning models require categorical data to be in a numerical format, requiring conversion of values such as yes or no to 1 or 0 . Be cautious not to accidentally create order to unordered categories such as converting mr, miss, and mrs to 1, 2 and 3.

3.2.4 FEATURE ENGINEERING

Feature engineering is the process of using domain knowledge of the data to create features that make machine learning algorithms work . If feature engineering is done correctly , it increases the predictive power of machine learning algorithms by creating features from raw data that help facilitate the machine learning process .

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907203 |

3.2.5 SPLITTING OF DATA

Here the dataset is considered and splitted into two forms namely the training data and the testing data . It helps to train the model with the specific data values from the data set and the model will be able to get validated from the testing process.The splitting of the data is necessary because the values have to be splitted in order to achieve the training and testing.

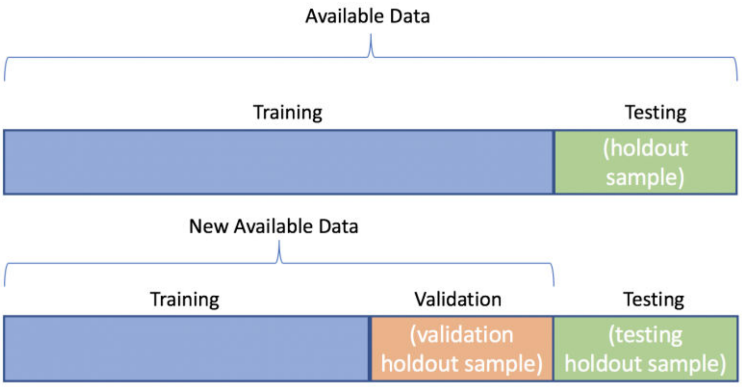


Fig 3.3 Data Splitting

3.2.6 TRAINING

Here the model is fed with the training data set that allows the model to get trained and get the required knowledge to perform the testing . The model uses the machine and deep learning algorithms to acquire the knowledge from the dataset.We consider the dataset that contains the weather report of a particular region which includes the information such as humidity , wind speed,temperature,pressure and dew point etc . These features help the model to get equipped with the knowledge needed and will be able to evaluate and predict the results . The particular data set of a region will be helpful in predicting the forecast of the complete state . Once the training is completed the model will be able to work on its own then the model's efficiency and accuracy will be evaluated.

3.2.7 VALIDATION WITH TEST DATA

After the training process the model will be ready to evaluate its own data set and produce the results . The model will be fed with the testing data set then

The model will use all the knowledge that it has gained and predict the result according to the data set provided.Later the results will be validated and the model will be evaluated for efficiency .

3.3 MODULES

A)Data Visualization

B) Preprocessing

C)Prediction

3.4 MODULE DESCRIPTION

A module is a separate unit of software or hardware. Characteristics of modular components include portability and interoperability which allows them to function in another system with the components of other systems.

3.4.1 DATA VISUALISATION

Data Visualisation is common in day to day life. Various charts and graphs are used to illustrate the practical approach towards the classification of rainfall with the help of data visualisation methods. Since it was impossible to analyze the large datasets earlier, the data visualisation techniques have made it easier to plot the graphs for the

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907203 |

better understanding of the weather. With the help of data visualisation patterns such as the highest, lowest and average rainfall in the States/Union Territories the weather of India has been visualised.

Exploratory data analysis is generally cross-classified in two ways. First, each method is either non-graphical or graphical. And second, each method

is either univariate or multivariate (usually just bivariate). Non-graphical methods generally involve calculation of summary statistics, while graphical methods obviously summarize the data in a diagrammatic or pictorial way. Univariate methods look at one variable (data column) at a time, while multivariate methods look at two or more variables at a time to explore relationships. Usually our multivariate EDA will be bivariate (looking at exactly two variables), but occasionally it will involve three or more variables. It is almost always a good idea to perform univariate EDA on each of the components of a multivariate EDA before performing the multivariate EDA

3.4.2 PREPROCESSING

A pre-processing technique used to reduce the effects of minor observation errors. The sample is divided into intervals and replaced by categorical values. Indicator variables: This technique converts categorical data into boolean values by creating indicator variables. If we have more than two values (n) we have to create n-1 columns. Centring & Scaling: We can centre the data of one feature by subtracting the mean to all values. To scale the data, we should divide the centred feature by the standard deviation.

Labels represent string labels of both ordinal and nominal features in relation to categorical feature set. Some labels may have order associated with them (ordinal features) while others may not have any orders associated with them (nominal features). It is an important part of data preprocessing to encode labels appropriately in numerical form in order to make sure that the learning algorithm interprets the features correctly. In the following section, you will see how you could use the LabelEncoder class of sklearn.preprocessing module to encode labels of categorical features. Label encoding refers to transforming word labels into a numerical form so that algorithms can understand how to operate on them

3.4.3 PREDICTION

Firstly the input is given to the NN. The weights which are given to neural networks are random numbers which are untrained. The value of weights which are given to the NN must have values in the middle of -1 and +1. Weight training is used to decrease the error function in neural networks. By doing all This it gives the output. If the output is incorrect then adjusts the weights on the NN. By doing so on up to the resemblance output is given by the neural network. Then stop adding the weights and consider the weights and check it for another dataset. The rainfall has been predicted using deep learning techniques. Two deep learning techniques which were used are Multilayer Perceptron and Auto-Encoders. Auto -Encoders are responsible in time series forecasting by performing the feature extraction.

The weights considered in the next data in the data goes wrong then the process repeats. Hidden layer neural networks are calculated by the sigmoid function. It is having one neuron in the output layer. All these neurons are interrelated with each layer. In Backpropagation NN input signals are captured by input neurons and output signals are captured by output neurons.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907203 |

SNAPSHOTS :

HEAT MAP THAT SHOWS CORRELATION OF ALL PARAMETERS:

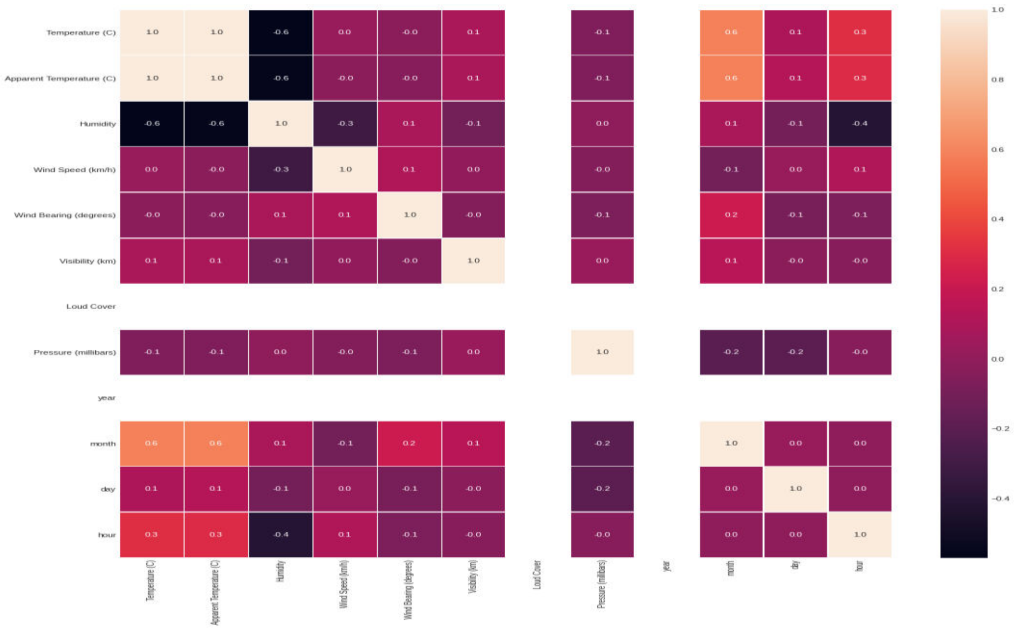


Fig 1 Correlation Heat Map

BAR GRAPH OF SUMMARY OF WEATHER:

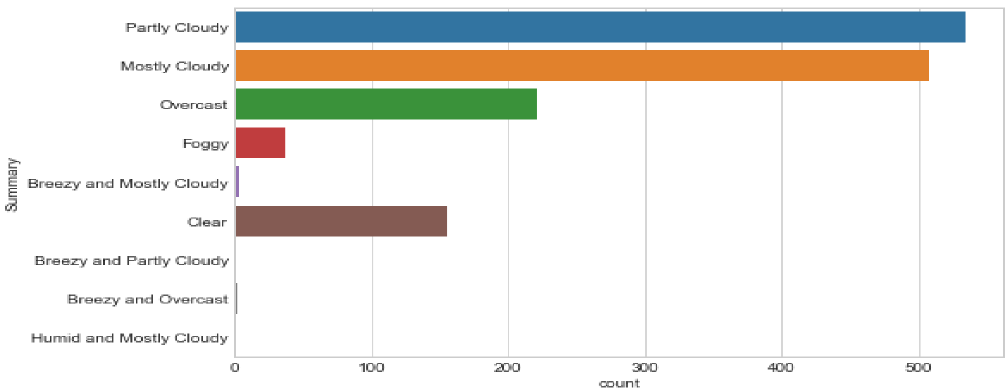


Fig 2 Weather Summary Bar Graph

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907203 |

BAR GRAPH AFTER APPLYING CATEGORY TO EACH CATEGORICAL VALUE:

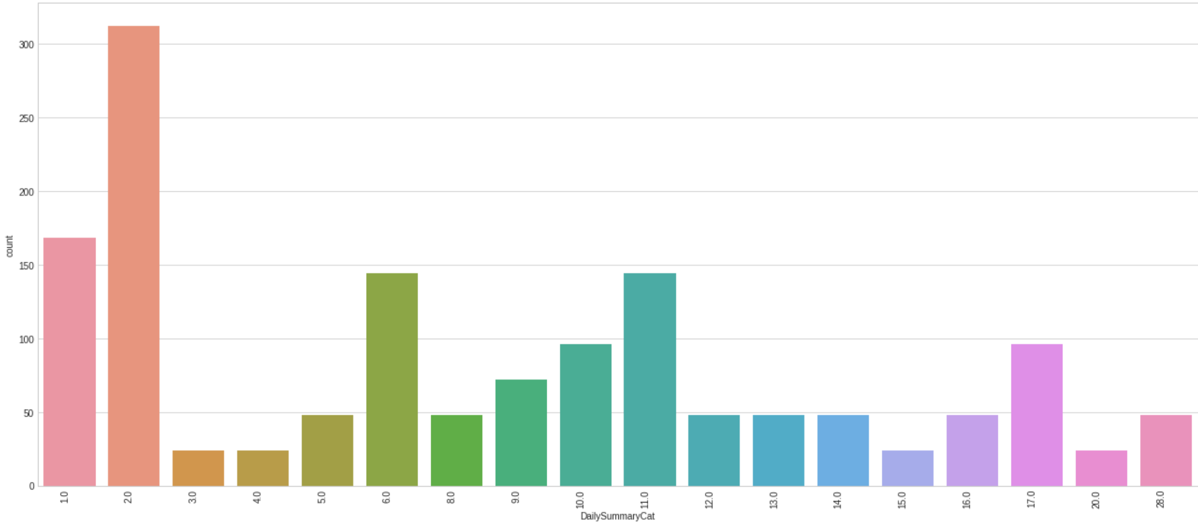


Fig 3 Categorical Value Bar Graph

IV. CONCLUSION

The correspondence between the observed and the predicted rainfall is better than the correspondence between the observed and the simulated rainfall magnitude. This can be considered as the potential of the hybrid NN in improving the rainfall prediction . We developed a different model: using a rainfall-flow pattern based on historical rainfall and flood flow data for real-time predictions of short-term flood stream-flow. Our models predict the flood process line in real-time using hydrological feature extraction and spatial-temporal metrics for similar rainfall flow patterns. The experimental results based on the datasets of various models in wet and drought-ridden watersheds show that the proposed models offer considerable advantages in accurately predicting the peak time of rains in real time

The future research work includes Rainfall prediction algorithm and model design and Rainfall prediction practice based on the feature set, algorithm, and model .

REFERENCES

[1] F. Fotovatikhah, M. Herrera, S. Shamshirband, K.-W. Chau, S. F. Ardabili, and M. J. Piran, ''Survey of computational intelligence as basis to big flood management: Challenges, research directions and future work,'' Eng. Appl. Comput. Fluid Mech., vol. 12, no. 1, pp. 411–437, Jan. 2018.

[2] J. Reynolds, S. Halldin, C. Xu, J. Seibert, and A. Kauffeldt, ''Sub-daily runoff predictions using parameters calibrated on the basis of data with a daily temporal resolution,'' J. Hydrol., vol. 550, pp. 399–411, Jul. 2017.

[3] Z. M. Yaseen, A. El-Shafie, O. Jaafar, H. A. Afan, and K. N. Sayl, ''Artificial intelligence based models for stream-flow forecasting: 2000– 2015,'' J. Hydrol., vol. 530, pp. 829–844, Nov. 2015.

[4] J. Feng, L. Yan, and T. Hang, ''Stream-flow forecasting based on dynamic spatio- temporal attention,'' IEEE Access, vol. 7, pp. 134754–134762, 2019.

[5] C. P. Chen and C.-Y. Zhang, ''Data-intensive applications, challenges, techniques and technologies: A survey on big data,'' Inf. Sci., vol. 275, pp. 314–347, Aug. 2014.

[6] H. Gao, W. Huang, and X. Yang, ''Applying probabilistic model checking to path planning in an intelligent transportation system using mobility trajectories and their statistical data,'' Intell. Automat. Soft Comput., vol. 25, no. 3, pp. 547–559, 2019.

[7] H. Gao, W. Huang, Y. Duan, X. Yang, and Q. Zou, ''Research on cost driven services composition in an uncertain environment,'' J. Internet Technol., vol. 20, no. 3, pp. 755– 769, 2019.

[8] H. Gao, Y. Duan, L. Shao, and X. Sun, ''Transformation-based processing of typed resources for multimedia sources in the iot environment,'' Wireless Netw., pp. 1–17, Nov. 2019.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907203 |

[9] H. Mislan, S. Hardwinarto, and M. A. Sumaryono, ''Rainfall monthly prediction based on artificial neural network: A case study in Tenggarong Station, East Kalimantan- Indonesia,'' Procedia Comput. Sci., vol. 59, pp. 142–151, Jan. 2015.
[10] Dubey, ''Artificial neural network models for rainfall prediction in pondicherry,'' Int. J. Comput. Appl., vol. 120, no. 3, pp. 30–35, Jun. 2015.
[11] N. Sharma, M. Zakaullah, H. Tiwari, and D. Kumar, ''Runoff and sediment yield modeling using ANN and support vector machines: A case study from Nepal watershed,'' Model. Earth Syst. Environ., vol. 1, no. 3, p. 23, 2015
[12] S. Das, R. Chakraborty, and A. Maitra, ''A random forest algorithm for nowcasting of intense precipitation events,'' Adv. Space Res., vol. 60, no. 6, pp. 1271–1282, Sep. 2017.
[13] Y. Zaman, ''Machine learning model on rainfall—A predicted approach for Bangladesh,'' Ph.D. dissertation, United Int. Univ., Dhaka, Bangladesh, 2018.