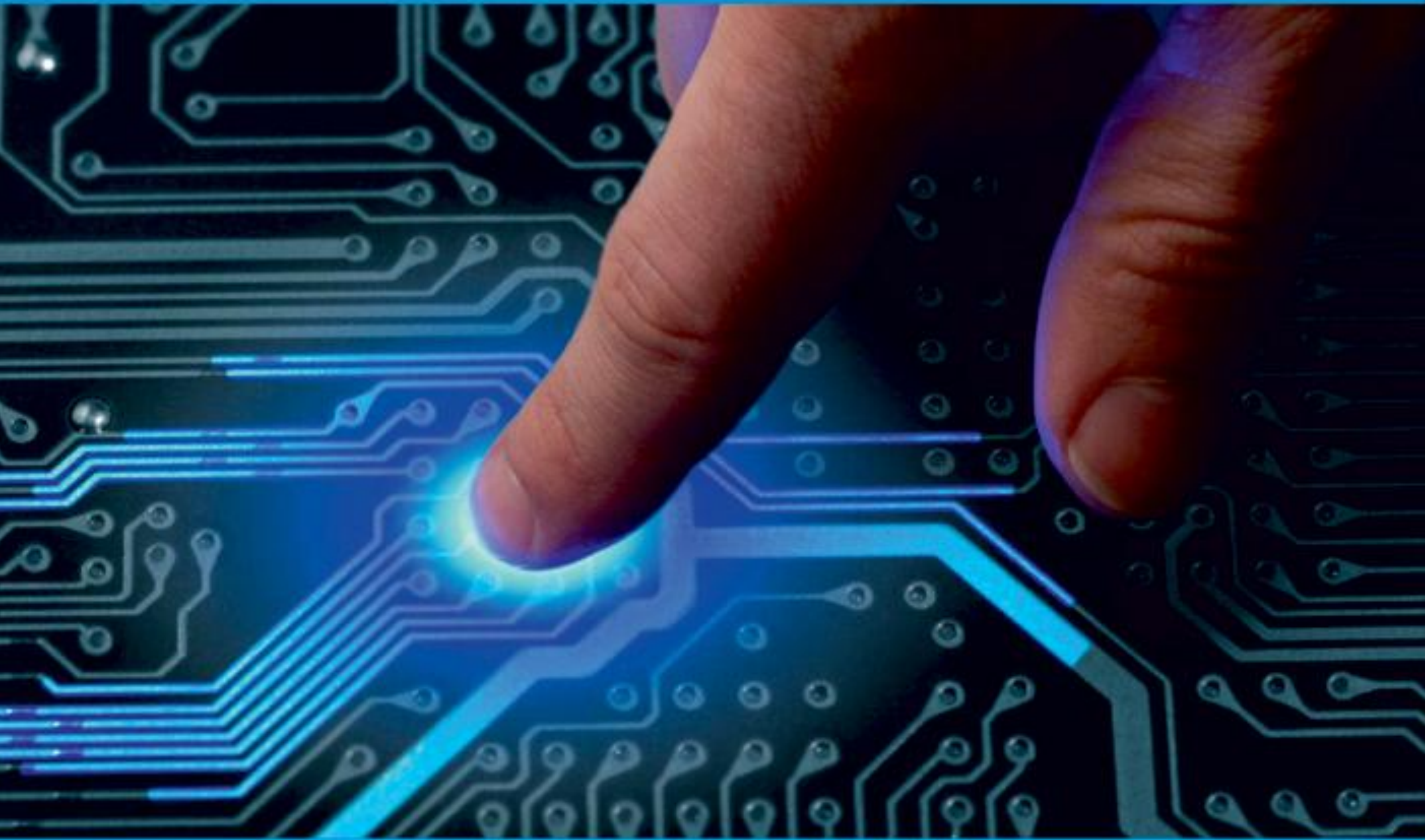
IJIRCCE

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

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.2021.0907193 |

Rainfall Measurement from Radar Data using Hybrid Model

Sahana Manjunath[1],Udayan Mahanta[2], Roshan Kumar Sinha[3], Chetna Kumari[4],Prof.T.P.Manasa[5]

UG Student, Dept. of Computer Science Engineering, Bangalore Institute of Technology, Karnataka, India[1,2,3,4]

Assistant Professor, Dept. of Computer Science Engineering, Bangalore Institute of Technology, Karnataka, India[5]

ABSTRACT: The measurement of precipitation is one of the most significant problems that has been pursued since the earliest time in civilization. Majority of the segments of domestic and international economies are impacted by the precipitation distribution. Large infrastructure around the world has been incorporated and installed over a period of time to measure rainfall using rain gauges or with remote sensing instrumentation such as radar. Radar rainfall measurement has been a very active field that has simultaneously seen great progress and challenges. The deterministic precipitation models are time consuming, it becomes very challenging to efficiently use the large volume of radar data in hand. Machine learning methods are proven to be good replacements for traditional mathematical approaches in rainfall measurement. This project presents an approach of measuring the rainfall from radar data by using Hybrid model of LSTM/RNN, XGBoost and Ridge regression to get more accurate results.

KEYWORDS: Hybrid model, LSTM, RNN, XGBoost and Ridge regression.

I. INTRODUCTION

Rainfall varies highly across space and time, making it very tricky to measure. Rain gauges can be considered an effective measurement tool for a specific location, but it is impossible to deploy them in every location. In order to have widespread coverage, data from weather radars is used to measure rainfall nationwide. Managing the challenging inputs from the radar to measure rainfall is a tedious task. This can be achieved by usage of mathematical and machine learning models. The main purpose of this project is to propose a system which predicts the hourly rainfall levels from the radar data using a hybrid model of RNN, XGBoost and Ridge regression with high accuracy.

The prediction of cumulative values from variable-length sequences of vectors with a time component is highly reminiscent in machine learning which demonstrates the power of recurrent neural networks (RNN) in learning long-term dependencies. RNN/LSTM models considers time series information into consideration which proves it to be best suitable for our problem. In order to decrease the error rate we also ensembled Ridge regression and XGBoost. This project acquires data from a polarimetric radar that transmits radio wave pulses with both horizontal and vertical orientations.

Rainfall measurement system has high importance in studies of climatology, hydrology and meteorology. It can be used to specify more precisely the precipitation events that lead to floods. Recording rainfall can help us plan accordingly as we learn nature’s patterns. What's important is that, we learn when to plant seeds to allow plants to grow and produce without artificial irrigation. It can provide a precise and accurate evaluation of the rainfall responsible for runoff and it is often used for rainfall/runoff analysis.

Main Objectives of this Paper is 1) To employ the power of a hybrid model of RNN, XGBoost and Ridge regression to predict the rainfall measurement 2) To maximize the accuracy of the model compared to the other existing systems.

II. RELATED WORK

In [1] the authors have estimated the optimizing parameters for Z-I relationship (Z-radar reflectivity in mm^6/mm^3,I – momentary rainfall intensity in mm/h) based on the set-pair analysis. The maximum of the degree of linkage for judgement criterion of Z-I relationship is given by the following equation.

$$\max \mu_{H-G} = \frac{S_1}{N} + \frac{F_1}{N} i + \frac{P_1}{N} j$$

$$S_1 = \sum D_{K_1} ; \quad F_1 = \sum D_{K_2} ; \quad P_1 = \sum D_{K_3}$$

Hi stands for sequence of Z-I relationship-based rainfall measurements.

| 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.2021.0907193 |

Gi stands for sequence of actual rainfall measurements.
$Di=Ci^2+Ci$.
$Ci=|Hi-Gi|$.
k1 stands for number of identity items(S).
k2 stands for number of different items(F).
k3 stands for number of opposite items(P).
i=05; j=-1.
The research indicates that the above equation provides good estimates of radar rainfall measurements.

In [2] The authors have deeply analyzed the primary causes of error in radar measurements in Z-R relationship caused by the microphysical and kinematics processes that affect the drop-size distribution and drop-fall speeds.Therefore, they emphasize the need for careful calibration procedures to be followed and that an independent check of system biases be made by comparing radar estimates with rain gaugemeasurements. Data users should be cognizant of those conditions in which the radar rainfall estimates may be erroneous or have little value, such as in the vicinity of ground targets, within shadow areas created by obstacles blocking the radar beam, at excessive distances where the beam becomes exceptionally large and elevated, and when there is extreme variability in cloud processes and, consequently, in Z-R relationships.

In [3] The authors have used the C-band dual polarization radar features like Differential Reflectivity(ZDR), Specific Differential Phase (KDP) and Horizontal Reflectivity (ZH) to measure the precipitation. The following are the C-band rain estimate equations

$$R(K_{DP}) : R = 18.77K_{DP}^{0.769}$$

$$R(K_{DP,} Z_{DR}) : R = 22.4K_{DP}^{0.77}10^{-0.072ZDR}$$

$$R(Z_{H,} Z_{DR}) : R = 0.015Z_{H}^{0.82}10^{-0.290ZDR}$$

In [4] The authors have used regression analysis to measure rainfall from an Orid catchment in Oman. Hourly data from 36 rainfall–runoff events are presented. Two events-one extreme high flow outlier and one with unacceptable rainfall errorswere excluded from the regression. Stepwise least square regression of peak flow, flow volume to rainfall volume and runoff coefficient was performed to obtain the total rainfall data.Out of 36 events 13 events were predicted accurately.

In [5] The authors have employed a deep learning methodology to measure rainfall.A Back Propagation Artificial Neural Network (BPANN) model based on adaptive variable number of hidden layer's nodes was used to estimate the rainfall in Hubei province. ANN models were compared with the multiple linear regression models. The comparisons showed good agreement with ANN estimations.

The mathematical approaches used did not show satisfactory results as the constants used in the equations might vary geographically and periodically. Theother deep learning and machine learning methods did not take the sequential data into consideration which lead to increase in the error rate which can be overcome by the usage of LSTM model, which is used in the current paper.It combines the advantages of both regression and LSTM models towards the accurate precipitation measurement.

III. PROPOSED METHODOLOGY

The system proposes the idea of a hybrid model of XGBoost, Ridge Regression and LSTM to determine the hourly rainfallmeasurement from variable length sequences of radar data which can outperform other models. We first explored the solution to the problem statement by building the regressor models like XGBoost and Ridge Regression which gave us good rainfall estimates but could not predict the outliers well. There was another drawback of regressor models that is it did not consider the sequential data which leads to increase in the error rate. Then we employed a LSTM model which remembers the previous inputs which is best suitable in measuring the precipitation from variable sequences of the radar data. Building a hybrid model by combining the advantages of both Regressor and LSTM model gave a huge decrease in error rate. This systemconsists of three major modules,

A. *Dataset Acquisition and Preprocessing.*

The Dataset consists of radar features from Dual Polarizationradar. DualPolarization radars can transmit radio waves horizontally and vertically which provides us more detailing regarding the reflectivity compared to the data from unipolar radars. This helps in determining the precipitation measurement more accurately. The training data consists of NEXRAD and MADIS data collected over Midwestern corn-growing states in US. It consists of 23 features recorded

| 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.2021.0907193 |

from radar that helps in the precipitation measurement. The test data consists of data from the sameradars and gauges over the remaining days. This datasetposes a lot of challenges. It is cleaned by imputing the missing values that is the values that had NaN for all the ids in Reflectivity column was removed from the observation and rest of the columns were filled with zeros, handling the extreme and unpredictable outliers that is the expected rainfall values above 73mm were discarded since rainfall rate cannot exceed 73mm in an hour, and Normalizing the data using Standard Scaler.Feature Selection and Extraction techniques have been implemented for the Regression models. The variable sized data of hourly readings are converted into fixed length by padding zeros in the end before feeding it into the LSTM model.

B. *Build and Train Hybrid Model.*

This module consists of implementation of three models namely XGBoost, Ridge Regression and LSTM.Our first attempt was to build a XGBoost model.

1.*XGBoost model*

The features that had high correlation greater than 0.9 were dropped.The dataset is grouped by id and mean of each feature is considered. Feature extraction was performed using Principal Component Analysis to extract relevant features and to reduce the training time. Hyper Parameter Optimization was performed using RandomisedSearchCV.The model was trained on the best hyperparameters obtained which are as follows.

gamma: 0.4
learning_rate: 0.1
max_depth: 10
min_child_weight: 1
n_estimators: 100

2. *Ridge Regression*

We implemented our own algorithm for Ridge Regression. The main reason to choose Ridge Regression is to avoid the overfittingand provide extra bias so that the model predicts well with other test data.The following are the steps to implement the Ridge Regression model

Step 1 - We implement a class and define init method that expects alpha as parameter.

Step 2 - Definition of fit method.

Arguments - x(matrix containing features)
y(matrix containing target)

- Adding an intercept column to x along second axis.
- Make the intercept accessible even after fitting.
- Creating an identity matrix of dimension equal to the no of columns of x including the intercept.
- Adding bias by making the first diagonal element as zero. A[0,0]=0
- Creating a bias term corresponding to alpha for each column of x not including the intercept.
- Computing the slope and associate it with the object using eq(1)

$$\hat{\Theta} = (X^T X + A)^{-1}(X^T y) \quad \text{eq(1)}$$

- Return the object

Step 3 - Definition of predict method

Arguments – x(matrix containing test features)

- Acquire the slope associated with the object.
- Create a predictor column to x along second axis.
- Find the dot product of predictor variable and slope.
- Return the object.

3. *LSTM model*

A Masked LSTM model is built which predicted the outliers and the non-linearities in the rainfall patterns based on the sequences of data. The network architecture is as follows

Step 1 - Subclass of the class Layer is defined to remove masking before Flatten.

Step 2 – A fully connected dense layer with 16 hidden neurons is created.

Step 3 - A BatchNormalisation layer is added.

Step 4 - A Masking layer is added to mask off the padded timesteps.

| 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.2021.0907193 |

Step 5 - A Bidirectional LSTM wrapper layer with 64 units is added .
Step 6 - A Time Distributed wrapper layer is used around the Dense output layer which gives 64 output units .
Step 7 - Steps 3,4,5,6 are repeated for 128 units.
Step 8 - A Non Masking layer is defined before the Flatten layer.
Step 9 - A Dropout layer of dropout rate 0.5 is defined to avoid overfitting.
Step 10 - A final Dense layer is created which gives the output neuron.
Step 11 - The model is optimized by Nadam optimizer and the loss function is mean absolute error.

C. *Integration of models and prediction of rainfall*.

The three models are saved in an appropriate file. An important insight we got from Exploratory DataAnalysis is that the values of rainfallabove 5mm in an hour is an indication for outliers.Since LSTM predicted the outliers and normal rainfall readings very well we consider it as a threshold to compare and combine the results from the three models. The average of XGBoost and LSTM results are considered for readings below 5mm in LSTM since the Ridge Regression overestimated the rainfall values below 5mm. For the values greater than 5mm the results from all the three models are considered.

IV. EXPERIMENTAL RESULTS

As we see in Table 1, the error analysis performed on the test set. It clearly proves that the Mean Absolute Error(MAE) and Mean Squared Error (MSE) is least for theHybrid model. Hybrid model performs exceptionally well compared to the applications of LSTM, XGBoost and Ridge Regression individually. Fig 1,2,3,4 shows the visualization of the first 100 ids and their corresponding predicted and actual rainfall measurements using Ridge regression, XGBoost, LSTM and Hybrid Model respectively.

MODELS	MAE	MSE
RIDGE REGRESSION	1.13	2.38
LSTM	0.95	2.26
XGBOOST REGRESSOR	0.77	2.18
HYBRID	0.61	1.65

Table 1 Error analysis on test set

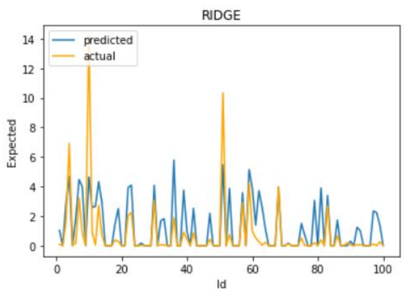


Fig 1. Ridge regression Visualization

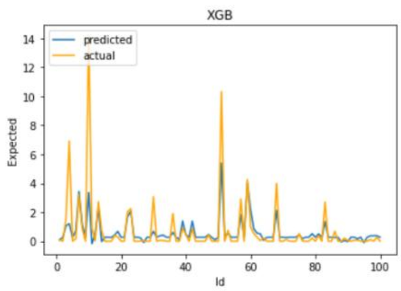


Fig 2. XGBoost Visualization

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.2021.0907193 |

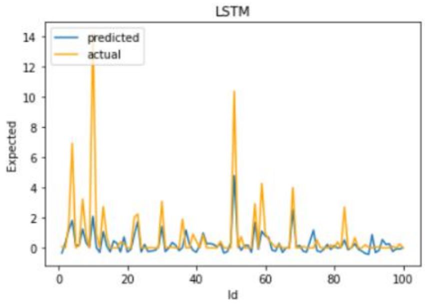


Fig 3. LSTM Visualization

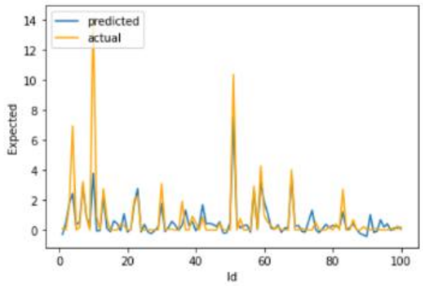


Fig 4. Hybrid Model Visualization

V. CONCLUSION AND FUTURE WORK

This project presents an approach that shows the potential of the Hybrid model (LSTM, XGBoost and Ridge Regression) in determining the rainfall measurements from radar data. The model has outperformed the other deep learning models like ANN and CNN. The model's performance is proven to be better than separate application of LSTM, XGBoost, Ridge regression individually. In future work, we can focus on predicting the extreme outliers which are a bit under estimated by the current approach.

REFERENCES

1. Peng, Luo &Xingyuan, Song &Wenwen, Tian. (2009). A Study on Quantitative Radar Rainfall Measurements by the Method of Set-pair Analysis. 385 - 388. 10.1109/GCIS.2009.375.
2. Wilson, James W., and Edward A. Brandes. "Radar Measurement of Rainfall— A Summary." Bulletin of the American Meteorological Society, vol. 60, no. 9, 1979, pp. 1048–1058. JSTOR, www.jstor.org/stable/26219220. Accessed 4 Aug. 2021. S.K. Dhurandher, S. Misra, M.S. Obaidat, V. Basal, P. Singh and V. Punia,'An Energy-Efficient On Demand Routing algorithm for Mobile Ad-Hoc Networks', 15th International conference on Electronics, Circuits and Systems, pp. 958-9618, 2008.
3. Paulitsch, H., Teschl, F., and Randeu, W. L.: Dual-polarization C-band weather radar algorithms for rain rate estimation and hydrometeor classification in an alpine region, Adv. Geosci., 20, 3–8, https://doi.org/10.5194/adgeo-20-3-2009, 2009.
4. Neil McIntyre , Aisha Al-Qurashi& Howard Wheater. Regression analysis of rainfallrunoff data from an arid catchment in Oman,Hydrological Sciences Journal 52:6, 1103-1118, DOI: 10.1623,2007
5. Keshtkar, Mohammad &Merad, Lotfi& Sidi Mohammed, Meriah&Miraoui, A. & BOUSAHLA, M. &Hanaeia, Hengameh&Lafdani, Elham & Moghaddam Nia, Alireza &Pahlavanravi, Ahmad & Ahmadi, Azadeh& JAJARMIZADEH, MILAD. Daily Rainfall-Runoff Prediction and Simulation Using ANN, ANFIS and Conceptual Hydrological MIKE11/NAM Models. International Journal of Engineering & Technology Sciences. 1, 2013.

INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 7.542

doi crossref

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

Scan to save the contact details