Prediction of Streamflow using Deep learning and Machine Learning models

A Project Report

Submitted by

Mr GOVIND S R

REG NO : TKM20MEAI08

SEMESTER : IV

In partial fulfillment for the award of the degree of

MASTER OF TECHNOLOGY
IN
Mechanical Engineering (Artificial Intelligence)

Under the guidance of
Dr. ADARSH S

Thangal Kunju Musaliar College of Engineering
Kerala

JULY 2022

DECLARATION

I undersigned hereby declare that the project report "Prediction of Streamflow using Deep learning and Machine Learning models", submitted for partial fulfillment of the requirements for the award of degree of Master of Technology of the APJ Abdul Kalam Technological University, Kerala is a bonafide work done by me under supervision of Dr. Adarsh S . This submission represents my ideas in my own words and where ideas or words of others have been included, I have adequately and accurately cited and referenced the original sources. I also declare that I have adhered to ethics of academic honesty and integrity and have not misrepresented or fabricated any data or idea or fact or source in my submission. I understand that any violation of the above will be a cause for disciplinary action by the institute and/or the University and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been obtained. This report has not been previously formed the basis for the award of any degree, diploma or similar title of any other university.

Place: Kollam
Date: GOVIND S R

Thangal Kunju Musaliar College of Engineering
Centre for Artificial Intelligence

C E R T I F I C A T E

This is to certify that, this report titled ***Prediction of Streamflow using Deep learning and Machine Learning models*** is a bonafide record of the **Project** presented by **GOVIND S R (TKM20MEAI08)**, under our guidance and supervision, in partial fulfillment of the requirements for the award of the degree, **M.Tech in Mechanical Engineering (Artificial Intelligence)** in **APJ Abdul Kalam Technological University** .

Internal Supervisor	Project coordinator	Head of the Department
Dr. Adarsh S Professor Dept of Civil Engg TKMCE	Prof. Sumod Sundar Assistant Professor Centre for Artificial Intelligence TKMCE	Dr. Imthias Ahamed T.P. Professor & HOD Centre for Artificial Intelligence TKMCE

Internal Examiner	External Examiner

ACKNOWLEDGEMENT

A successful project is a fruitful culmination of efforts by many people, some directly involved and some others indirectly, by providing support and encouragement. Firstly I would like to thank the almighty for giving me the wisdom and grace for making my project a successful one. I thank him for steering me to the shore of fulfillment under his protective wings

I express my sincere gratitude to **Dr. T A Shahul Hameed**, Principal of TKMCE, and **Dr. Imthias Ahamed T.P.**, Professor and Head of the Department, Centre for Artificial Intelligence, TKMCE, for their constant support and encouragement throughout the project work.

With a profound sense of gratitude, I would like to express my heartfelt thanks to my guide **Dr. Adarsh S**, Associate Professor, Department of Civil Engineering, TKMCE, for his expert guidance, cooperation and immense encouragement. I would like to express my heartfelt thanks to our project coordinator **Prof.Sumod Sundar** Assistant Professor Centre for Artificial Intelligence, TKMCE,for his constant support and encouragement throughout the project work. I would like to thank **Dr.Santhi Natarajan**, Honorary Professor, for her immense encouragement. I also extend my thanks to the entire faculty and staff members of the Centre for AI, TKMCE, who has encouraged me throughout this work.

I also express my thanks to my loving parents, brother and friends, for their support and encouragement in the successful completion of this project work.

GOVIND S R

Abstract

Monthly streamflow prediction will give us a better idea about for flood warning, hydropower operation, reservoir operations and environmental protection. The current work look in to the prediction and evaluation capability of a Deep learning method such as Long Short Term Memory (LSTM) model, for monthly streamflows of Kidangoor, Pattazhi and Perumannu stations in Kerala. Prediction accuracy of LSTM method is compared with other Machine Learning models, i.e. Random Forest (RF) and Support Vector Regression (SVR). Predicted results of the three stations shows that LSTM model gives better accuracy compared to other models. For improving prediction accuracy of the models various kernel types are tried and provide the best results for the stations. The LSTM outperforms the other methods in almost every stations. It is also found that data preprocessing considerably improves the prediction accuracy in estimation streamflows. The overall results indicate that the LSTM method could be successfully used in predicting and estimating monthly streamflow in Kerala. Best value of R-squared values is shown in Kidangoor data around 0.96 for LSTM. SVR and gives 0.92 and RF gives 0.84 R-squared value. Similarly for every other inputs LSTM provide best performance. While looking into other stations also LSTM outperforms both SVR and RF models. It conclude that LSTM is better around these three models.

Contents

List of Figures

List of Tables

Chapter 1

INTRODUCTION

1.1 General

There are many difficulties in the field of hydrology, river flow and reservoir information forecasting is essential to the management of water resources. Forecasting is an essential tool for accurate and dependable management in reservoir operation, water optimization and allocation, hydropower generation, supplying water to industry, agriculture, or municipalities, and drought management. The ability to estimate the availability of water in the future can be significantly improved by reliable streamflow forecasting at various time scales. Hydrologic time series forecasting has a significant attention recently from numerous studies.The livelihoods of the residents are significantly impacted by the water level. For instance, changes in water mixing and other things of the bottom sediment can result from increases in water levels and discharge rates, which can affect river flow. Therefore, it becomes more and more important to forecast the discharge . The Institute of Water and Flood Management (IWFM), for instance, suggests that more steps be taken to create methods for streamflow and prediction. As a result, numerous models have been introduced during the past ten years to predict discharge[1].

Machine learning models present good relationship for streamflow but introducing a deep learning model will provide better results .In this work found that Long short term memory (LSTM) can be one of the best deep learning model that should be introduced for time series forecasting of monthly streamflow[2]. Then the generated LSTM model is compared with machine learning models like Random forest (RF) and Support Vector Regression (SVR).Performance of every model is compared using R-squared values and different error measures. These approaches will show that a deep learning model may produce greater accuracy compared to conventional machine learning models.

Physically based models require large amount of data and specific assumptions for validation of stream flow. Mean gauge and Discharge of three different stations in Kerala are the datasets used for prediction of runoff. Different combinations of lags are applied for improving the performance of the models. By applying different combinations of lag for each data will assure better results for streamflow prediction.

LSTM is generally used for time better performance model for time series data . Because of its feedback connections and it can deal with vanishing gradient problems. So it produces a better accuracy for time series predictions. Vanishing gradient problem of Recurrent neural network (RNN) can overcome by using LSTM.

Support vector machines (SVM) is generally used for classification problems. Support Vector regression is following the same technique of SVM but it is used for regression problems. SVR is also a machine learning technique and it can used for both linear and non-linear problems. Basic idea behind SVR is similar to SVM which separates the data into different classes using a hyperplane . Here we have two types of data points so SVR can easily separate them into two classes and it makes the problem solving more conveniently[3].
Random Forest (RF) is also a machine learning technique which can be used for both classification and regression problems. Ensemble learning is the method that RF follows. Working of RF is based on creating decision trees on the training period .It is also best for time series data[4].

1.2 Objectives

The main objectives of this projects are,

- To develop a deep learning model (LSTM) and machine learning models (SVR and RF) for the prediction of monthly stream flow in three different stations in Kerala (Pattzhy,Perumannu and Kidangoor).
- To compare the performance of monthly stream flow prediction between deep learning model (LSTM) and machine learning models (SVR,RF).

1.3 Organization Of The Work

All the methods and steps used to complete this work are explained in this report. The report is organized into six chapters. Chapter 1 is giving an introduction of the work and its objectives. Chapter 2 is giving a knowledge about the literature reviews used to do the work. Chapter 3 is deals with the proposed methodology for this work. Chapter 4 deals with the idea of study area and data used for the work. Chapter 5 provide the results obtained from the work.Chapter 6 is used to conclude the work

Chapter 2

LITERATURE REVIEWS

In this section, several studies about streamflow prediction using deep learning and machine learning models are discussed.

Poul AK et al.(2019) In this study, artificial neural networks (ANN), adaptive neuro-fuzzy inference systems (ANFIS), and multi-linear regression (MLR) as statistical techniques are used to predict the monthly flow in the St. Clair River between the US and Canada. K-nearest neighbours (KNN) is a non-parametric regression technique. Six scenarios for input combinations are specified in the proposed methodologies to examine the impact of various input data on the results. Statistical indicators are used as the performance criteria to assess the model performances. According to the results, significantly more flow, temperature, and precipitation lag times added to the inputs enhance prediction accuracy[1]

Adnan et al.(2019) introduced a pruned extreme learning machine (OP-ELM) model for forecasting and evaluation of the daily streamflows at the Fujiangqiao and Shehang stations along the Fujiang River. Cross validation method is used to compare the prediction accuracy of the OP-ELM method with that of other machine learning models, including the M5 model tree (M5Tree), multivariate adaptive regression splines (MARS), and adaptive neuro-fuzzy inference system-particle swarm optimization (ANFIS-PSO). The OP-ELM and ANFIS-PSO are two of the best models for calculating daily streamflows upstream and downstream, respectively, according to prediction results from both stations. The linear, linear+sigmoid+Gaussian, and linear+sigmoid kernel types produce the greatest results for both stations when trying to increase the prediction accuracy of the OP-ELM approach [9].

Thapa et al.(2020) This study is for predicting snowmelt-driven discharge in a Himalayan basin, they created a deep learning long short-term memory (LSTM) model. They created the Gaussian process regression (GPR), support vector regression (SVR), and nonlinear autoregressive exogenous model (NARX) models for comparison. The models inputs included the snow area produced from remotely sensed meteorological data and moderate resolution imaging spectroradiometer (MODIS) snow photos. The right input combination for the models was determined using the Gamma test. The shallow LSTM model with a hidden layer outperformed the deeper LSTM models with many hidden layers in terms of performance. Adamax emerged as the most suitable optimizer for this study out of the seven optimizers that were tested[8].

Chapman et al.(2020) In this study, they created and evaluated machine learning models employing programmatically calculable scalar image characteristics to fill data gaps in stream gauge records after automating the processing of time-lapse imagery from a single camera at a single site. From 2012 to 2019, features were retrieved from more than 40,000 daylight photographs shot at hourly intervals. Too dark photographs from dawn and dusk were eliminated by the algorithms during feature extraction. Based on image capture times and USGS timestamps, the image features were combined with stage and discharge data then, using a randomly chosen training set of 30 percentage of the images and the remaining 70 percentage for the test set, they created a methodology to find an appropriate feature set to build machine learning models. Multi-layer Perceptron (MLP), Random Forest Regression (RFR), and Support Vector Regression (SVR) models were used to produce predictions[7].

Parisouj et al.(2020) introduced to find out the prediction ability of three well-known machine learning algorithms For the monthly and daily streamflows of four American rivers,they are Support Vector Regression (SVR), Artificial Neural Network with backpropagation (ANN-BP), and Extreme Learning Machine (ELM) were used. Three major predictor variables (P, Tmax, and Tmin) and their predecessor values are taken into account while developing the model. The best predictor variable was chosen using the SVM-RFE feature selection method. Four calibration statistics were used to assess how well the generated models performed and the evaluation period is from 2014 to 2019 [6].

Dong et al.(2020) suggests a dynamic sliding window method is used to validating the streamflow's varying timing and periodicity features over the year. The long-short term memory (LSTM) is used to pick the ideal window, and then the data with the ideal window size is used for verification. Initially, several datasets of different months are generated using a dynamic window. Using the hydrological data from Zhutuo Hydrological Station, the proposed technique was evaluated (China) [5].

Malik et al.(2020) states Support vector regression (SVR) was optimized in this study using six meta-heuristic algorithms to predict daily streamflow in the Naula watershed, State of Uttarakhand, India: Ant Lion Optimization (SVR-ALO), Multi-Verse Optimizer (SVR-MVO), Spotted Hyena Optimizer (SVR-SHO), Harris Hawks Optimization (SVR-HHO), Particle Swarm Optimization (SVR-PSO), and Bayesian Optimization (SVR-BO) Gamma Test was used to extract the important inputs and parameter combinations for hybrid SVR models before processing. Using performance indicators such as root mean square error (RMSE), scatter index (SI), coefficient of correlation (COC), Willmott index (WI), and visual inspection (time-series plot, scatter plot, and Taylor diagram), the results produced by hybrid SVR models during calibration (training) and validation (testing) periods were compared against observed streamflow [2].

Rabbi et al.(2021) introduced three different machine learning models such as K-nearest neighbour algorithms, decision trees, coupled with the random forest regressor. There are so many hyperparameters presented which have been discovered to be more effective the additional machine learning techniques. Daily Maximum velocity and water level were calculated. With addition to water discharge, as explanatory factors a response variable was used [3].

Meshram et al.(2021) introduced that three AI techniques—ANFIS, GP, and ANN—have been applied in the current work to anticipate streamfow into the Shakkar watershed in India's Narmada Basin. In order to produce a suitable time series model for streamfow forecasting, the models have been utilized taking into account prior streamfow and cyclic terms in the input vector. RMSE, MAE, CORR, and CE were used to assess the performance of the model [4].

2.1 Summary

Prediction of streamflow is very important in Kerala because in Kerala flood is recently causing many problems in Kerala very badly. By studying the related works can find out that prediction of streamflow in kerala using LSTM is a new set of work.So that this gap can be eliminated by generating a LSTM model using Kerala datasets. Datasets used in this work are Pattazhy, Kidangoor and Perumannu in Kerala.

Chapter 3

METHODOLOGY

3.1 Work flow

Collected three daily datasets in Kerala (Kidangoor, Pattazhy and Perumannu). Daily datasets of three stations in Kerala are converted into monthly data. Autocorrelation of discharge and correlation of mean gauge with discharge are find out. According to the correlation best set of lagging is applied for the data. Three different combinations of lag is applied for three station's data then, it is given as the input to three models such as Long short term memory(LSTM), Support vector regression(SVR) and Random forest(RF).After that developed a LSTM model for these all streamflow datas. SVR and RF models are also used for the performance evaluation. LSTM model is developed using python and SVR and RF models are developed using WEKA software. The performances of these three models are compared using the R-Squared value and error measures. (Figure 3.1).

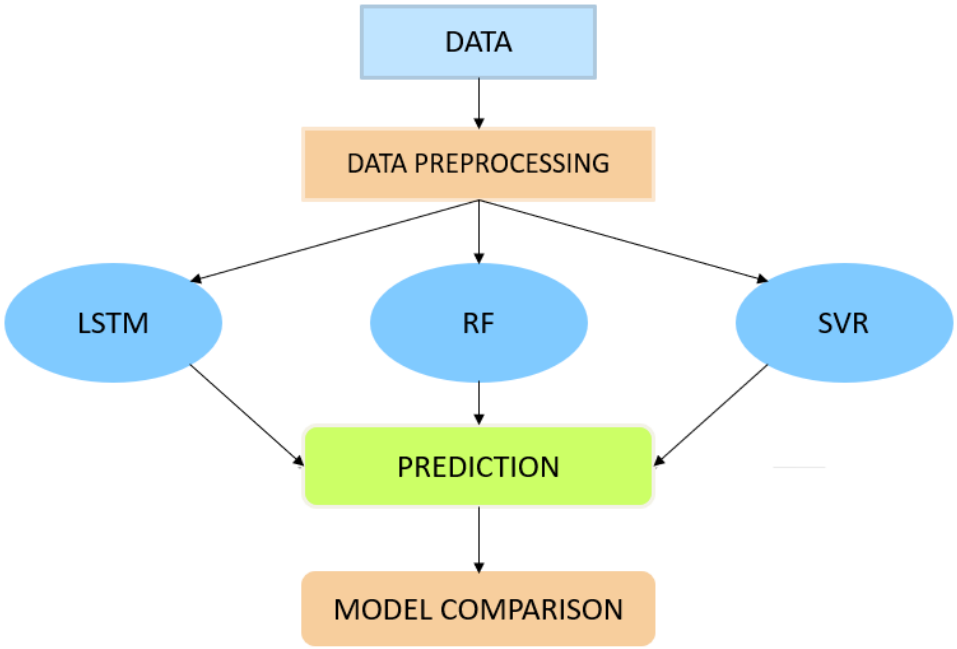

Figure 3.1: Work Flow

3.2 Proposed models

Tree models are used in this study they are,

- Long Short Term memory (LSTM)
- Support vector regression (SVR)
- Random forest (RF)

3.2.1 Long Short Term memory (LSTM)

LSTM is the most advanced version of RNN. RNN suffer from short-term memory due to a vanishing gradient problem. To avoid this problem LSTM is developed, which can work with longer data sequences' and can solve the vanishing gradient problem of RNNs.LSTM can preserve important information from the earlier part of the sequence and carry it forward.LSTM recurrent unit is more complex than that of RNN, they improve learning but required more computational resources. The key elements in the LSTM are cell, input gates, output gates and forget gates.LSTM architecture is shown in(Figure 3.2)

- Cells: it is also known as the memory state, which is like an information highway It is a chain structure, that contains four neural networks and many memory blocks called cells. Information is retained by cells and memory manipulation is done by the gates.
- Forget gates: forget gates are gates which can manipulate memory and retain by cells In LSTM the information that is not useful in the cell state is removed by the forget gate. the unwanted or no longer used are ignored.
- Input gates: additional information for the cell state is given by Input gates. they decide what information is given to the cell. The information is regulated using the sigmoid function and filters.
- Output gate :It is generally used for extracting useful informations from the model and gives the required output.

The LSTM model proposed in this work is about a multivariate case where mean gauge and discharge are inputs given to the LSTM to predict the corresponding discharge corresponding to the timescale. This work used one layer of LSTM which consists of 1000 cells. A dropout layer is included between the two hidden layers for regularization. It will randomly exclude 50 percentage of the activations of the previous layer from propagating to prevent overfitting. The root mean square (RMS) loss is reduced using the Adam optimizer which can handle sparse gradients on the noisy dataset and little memory is enough, therefore using adam gives more memory efficiency.LSTM model with three datasets is checked with tree different combination of lags.

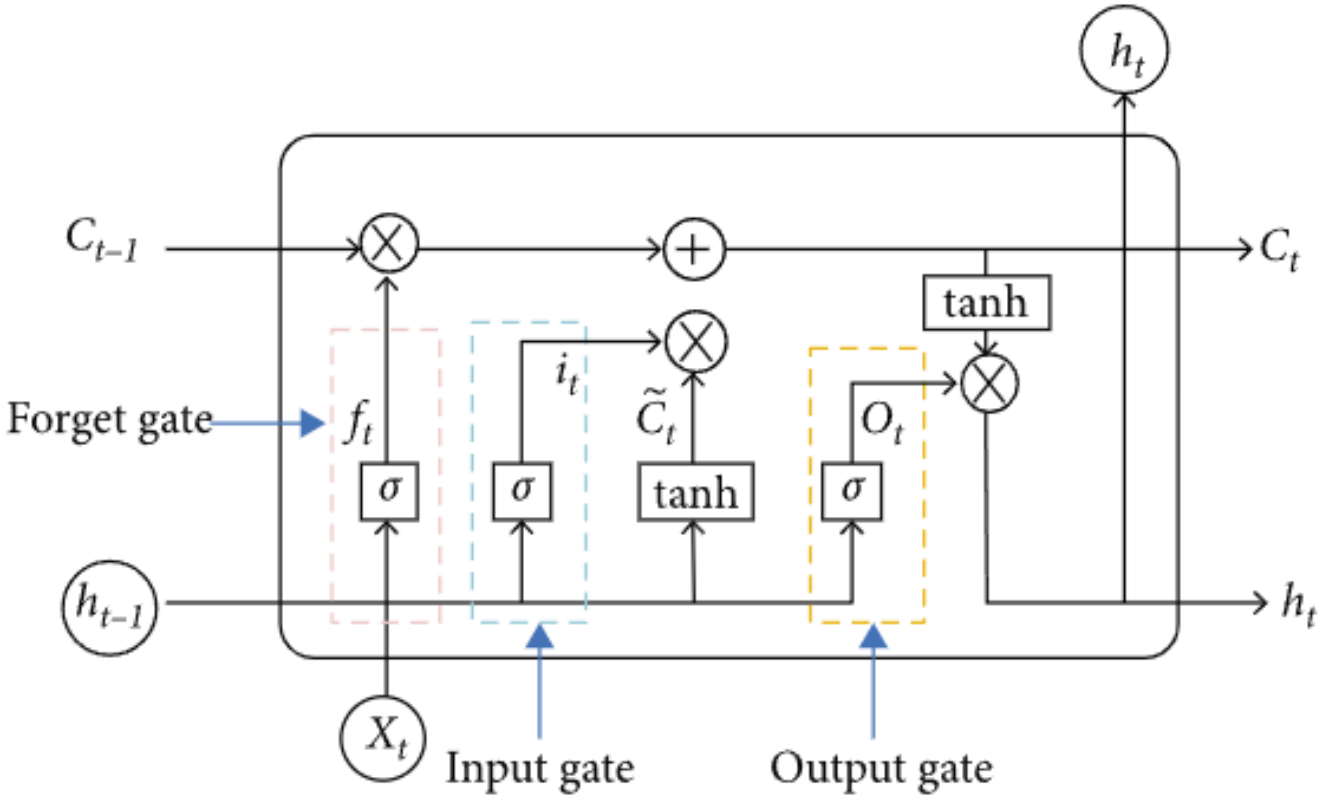

Figure 3.2: Long Short Term Memory (LSTM) Structure

3.2.2 Support Vector Regression (SVR)

Support Vector Regression is a popular machine learning model. SVR is associated with learning algorithms that analyse data for classification and regression analysis. SVR can be said that it is a combination of SVM with regression. They can be used in classification problems or assigning classes .when the given data are not linearly separable. In other words SVR can be said that SVMs that solve the regression problems are called SVR. Commonly used three kernels in SVR are Linear kernel, Polynomial kernel, and Radial basis function (RBF) kernel . Radial basis function kernel shows better results as compared to other kernels. Support Vector Regression proposed in this work is used RBF as the kernel, which is the best and most efficient kernel than the linear kernel and the polynomial kernel.It is very much popular because of its similarity to the K-nearest neighbourhood algorithm. RBF can overcome the space complexity problem as RBF kernel SVR just needs to store the support vectors during training and not the entire dataset. The proposed SVR model in this work uses a Batch size of 100 and then used a c-value(complexity parameter) of 1.0 (figure 3.3).

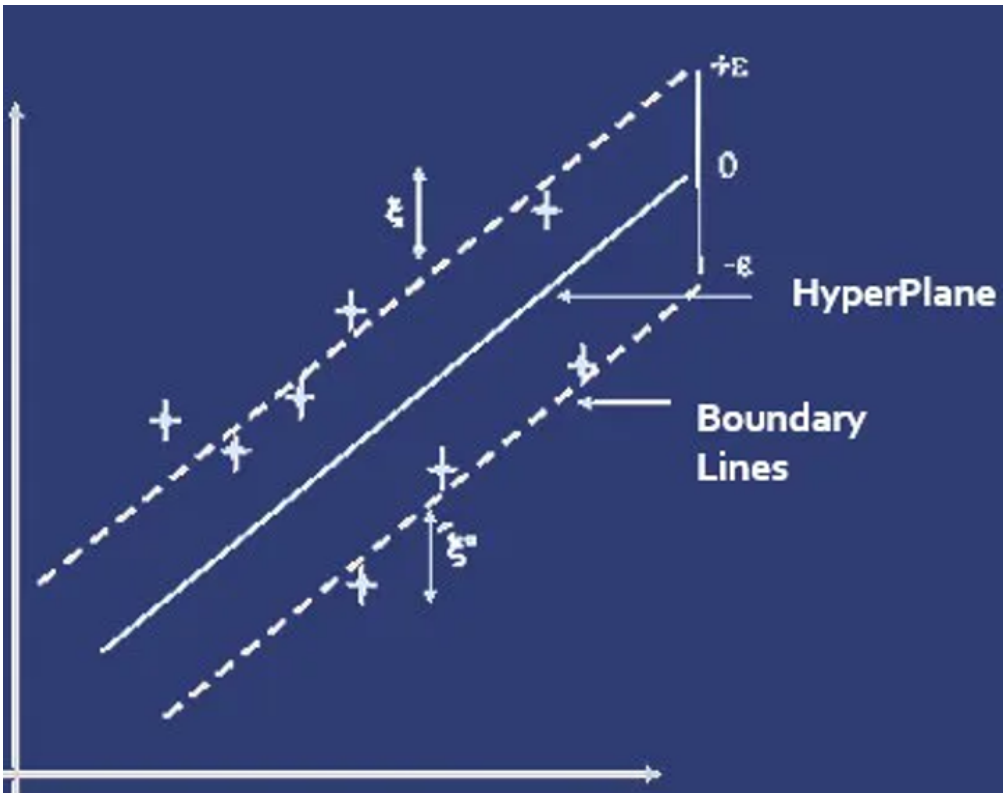

Figure 3.3: Support Vector Regression (SVR) Decision Boundary

3.2.3 Random Forest (RF)

The random forest (RF) method is an ensemble learning technique. It has been successfully used in dealing with various prediction problems. It is a machine-learning algorithm that combines a large set of decision trees to improve the prediction performance of classification and regression trees (CART) method. Each decision tree of RF is grown by using a randomly selected bootstrap sample from the original data set, and the final outcome of RF is the average result of all the trees. Compared to the regression methods, the number of parameters needed to be defined in the RF is very few. There are only two necessary parameters, including the number of variables used in each tree-building process and the number of trees built in the forest. The number of trees built in the forest has significant influence on the result of RF. The insufficient number of trees would result in poor forecasting performance, while the excessive number of trees may lead to complicated predictors (figure 3.4).
RF is a supervised learning model that uses ensemble learning method for regression. It includes a combination of multiple model tree algorithms to make a more accurate prediction. It usually performs great on many problems, including features with non-linear relationships.

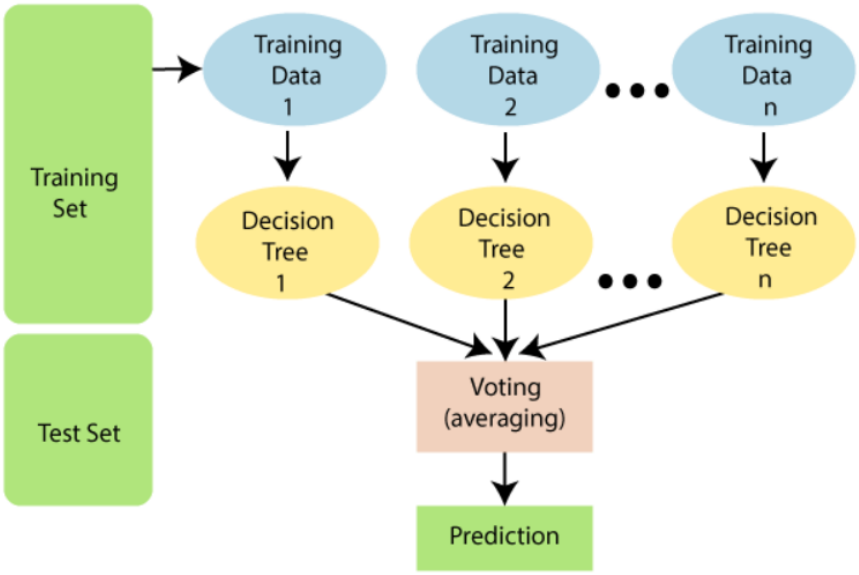

Figure 3.4: Random Forest (RF)

Chapter 4

STUDY REGION AND DATASETS USED

There are datasets of three different stations in Kerala is included in this study they are,

- Kidangoor
- Pattazhy
- Perumannu

Each station contain 2 features. They are,

- Mean gauge
- Discharge

4.0.1 Kidangoor

Kidangoor is located in Kottayam district (Kerala).The dataset consists of mean gauge readings and discharge from 1987 to 2015 time period. 360 monthly datasets are used in this study.

4.0.2 Pattazhy

Pattazhy is in Kollam district (Kerala), this dataset contains data from 1979 to 2015. 384 monthly datasets are used in this study.

4.0.3 Perumannu

Perumannu is in Kannur district (Kerala), this dataset contains 1986 to 2015. 360 monthly datasets are used in this study.

Stations such as Kidangoor, Perumannu and Pattazhy are some of the flood affected area in Kerala. Flood can be evaluated easily using discharge of water in those areas. There for the prediction of discharge is a the key factors in these areas. Flood affected times in these stations are,

- Kidangoor : 9 August 2019
- Perumannu : 19 August 2018
- Pattazhy : 14 November 2021

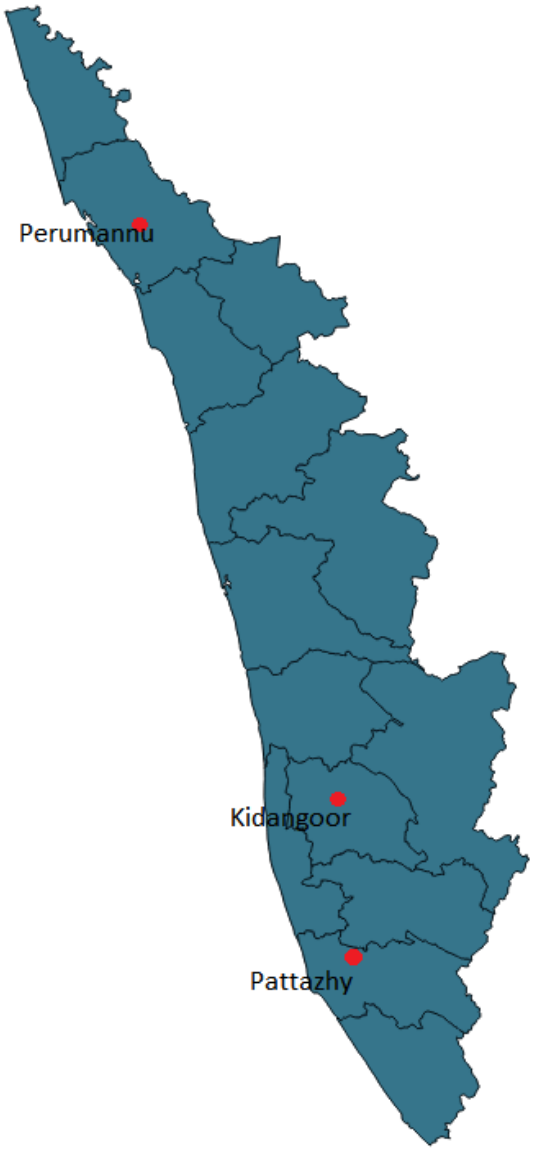

Figure 4.1: Selected stations

4.1 Data preprocessing

4.1.1 Kidangoor

Daily dataset of Kidangoor is converted to monthly then find out the autocorrelation (ACF) of discharge so that relation of discharge with discharge can find out. This gives an idea

about the perfect lag to be taken for the discharge which will give us the best performance for the Kidangoor dataset. By finding out the autocorrelation function the ideal lag taken is six and cross-correlation (CCF) of mean gauge and discharge of Kidangoor station is generated. This gives the idea about the relation of mean gauge to discharge and find out the best lags which will increase the accuracy of the models. By validating of the cross-correlation function the best lag should take for mean gauge is also six. Then convert the dataset with lag 6 for both Mean gauge and Discharge (Figure 4.2). Tree combinations of lag is applied they are,

- Mean gauge lag 6
- Discharge lag 6
- Combination of mean gauge lag 6 and discharge lag 6

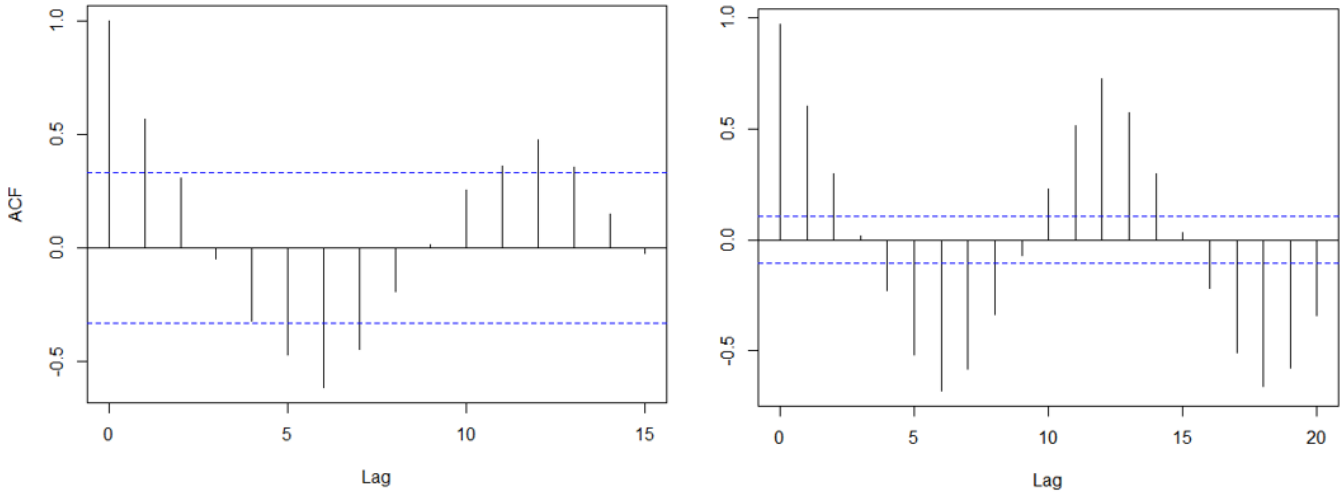

Figure 4.2: ACF and CCF of Kidangoor station

4.1.2 Pattazhy

Same as Kidangoor dataset Pattazhy station is also converted to monthly. Then find out the Autocorrelation (ACF) of discharge and cross-correlation (CCF) of mean gauge and discharge for Pattazhy then Find out the best lags, According to ACF Discharge lag is find out as 2 and mean gauge lag is 1, then converted the dataset according to the lag (Figure 4.3). The combinations of lag is applied they are,

- Mean gauge lag 1
- Discharge lag 2
- Combination of mean gauge lag 1 and discharge lag 2

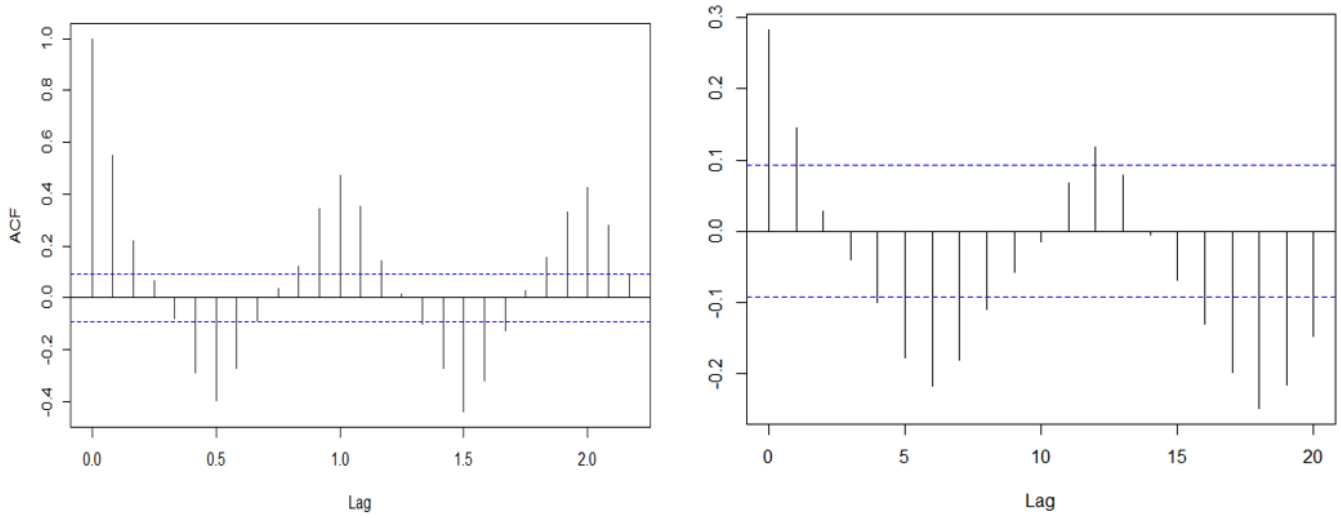

Figure 4.3: ACF and CCF of Pattazhy station

4.1.3 Perumannu

Daily dataset of Perumannu is converted to monthly. Then find out the Autocorrelation (ACF) of discharge and cross-correlation (CCF) of mean gauge and discharge of Perumannu station is generated and find out the best lags which will increase the accuracy of the models. Both ACF and CCF gives lag as 1. Then convert the dataset with lag 1 for both Mean gauge and Discharge (Figure 4.4). Tree combinations of lag is applied they are,

- Mean gauge lag 1
- Discharge lag 1
- Combination of mean gauge lag 1 and discharge lag 1

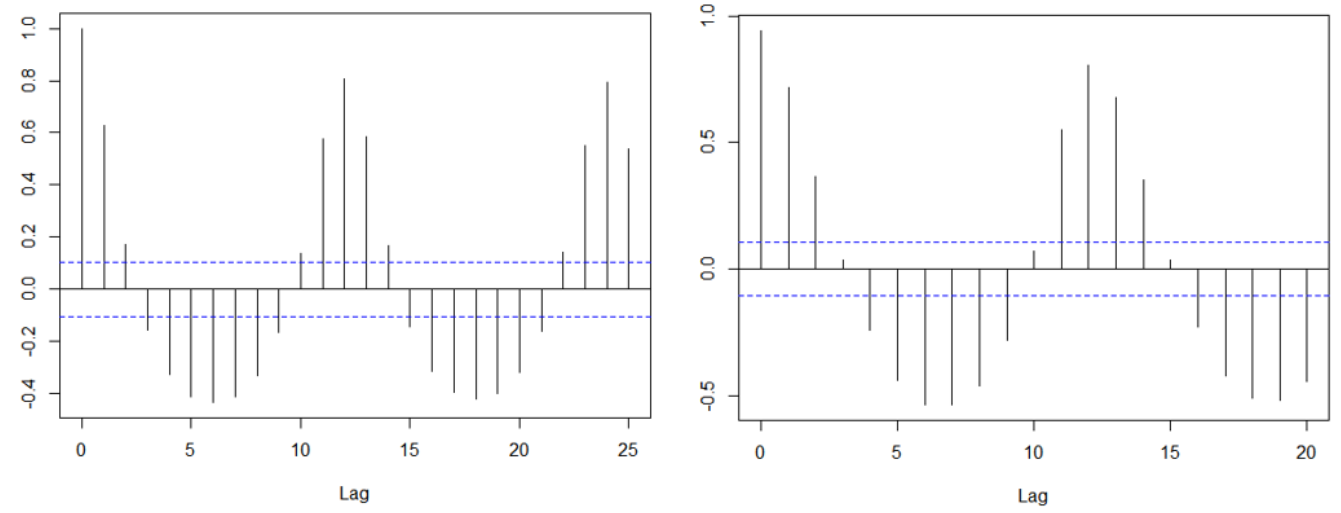

Figure 4.4: ACF and CCF of Perumannu station

Chapter 5

RESULTS AND DISCUSSIONS

In this work the discharge is predicted using LSTM and it is compared with conventional machine learning models like SVR and RF. The experiments were done with three different datasets in Kerala.Kidangoor, Pattazhy and Perumannu are the stations used in this study.

5.1 Experimentation using Long Short Term Memory

LSTM with 1000 neurones are used in this work. A dropout layer is included between the two hidden layers for regularization. It will randomly remove 50 percentage of the activations of the previous layer from propagating to prevent overfitting. The root mean square (RMS) loss is reduced using the Adam optimizer which can also handle sparse gradients on the noisy dataset and little memory is enough, therefore using adam gives more memory efficiency.LSTM model with three datasets is checked with tree different combination of lags. The generated models and model loss are shown in these figures (figures 5.1 to 5.6).

5.2 Comparison of LSTM with Random forest and SVR

Long short term memory (LSTM) R-Squared values and RMSE values are compared with SVR and RF.In support vector regression RBF kernal is used which is the best and most efficient kernel than the linear kernel and the polynomial kernel.For RF the number of parameters needed to be defined is very few. There are only two necessary parameters, including the number of variables used in each tree-building process and the number of trees built in the forest.The comparison of R-squared values and RMSE for testing are shown in tables (Table 5.1 and 5.2).Also the radar plots of performance evaluation is shown in figures (Figure 5.8-5.9).

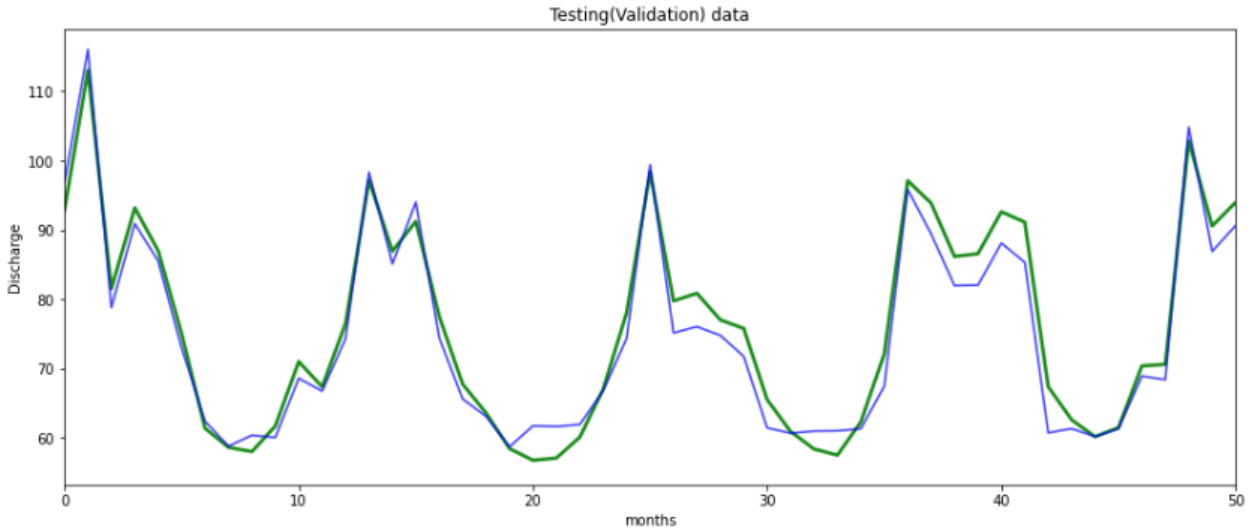

Figure 5.1: Kidangoor LSTM Model discharge lag 6 mean gauge lag 6

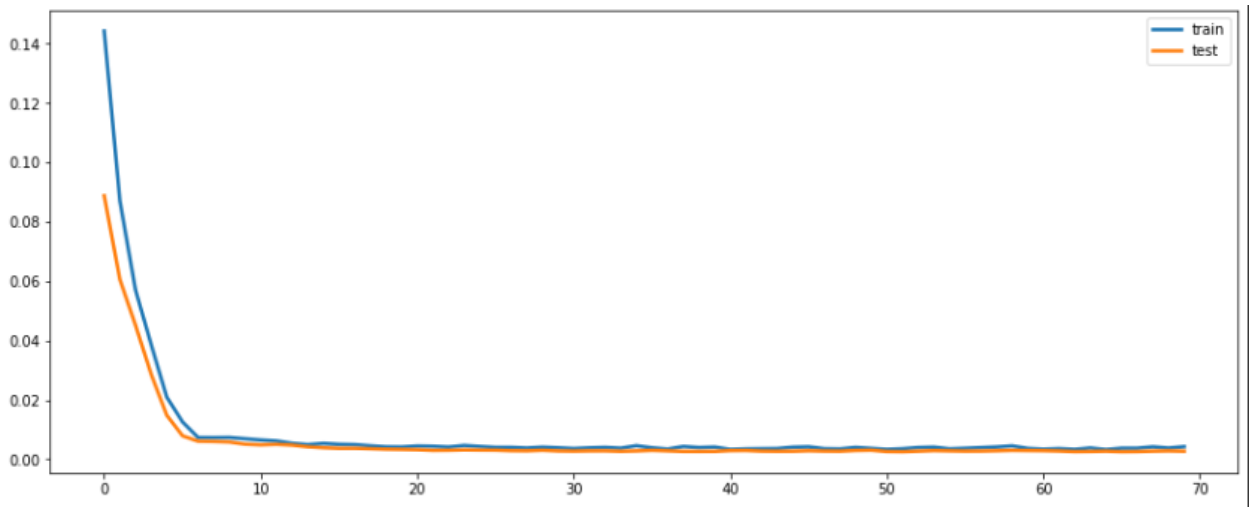

Figure 5.2: Kidangoor LSTM Model loss discharge lag 6 mean gauge lag 6

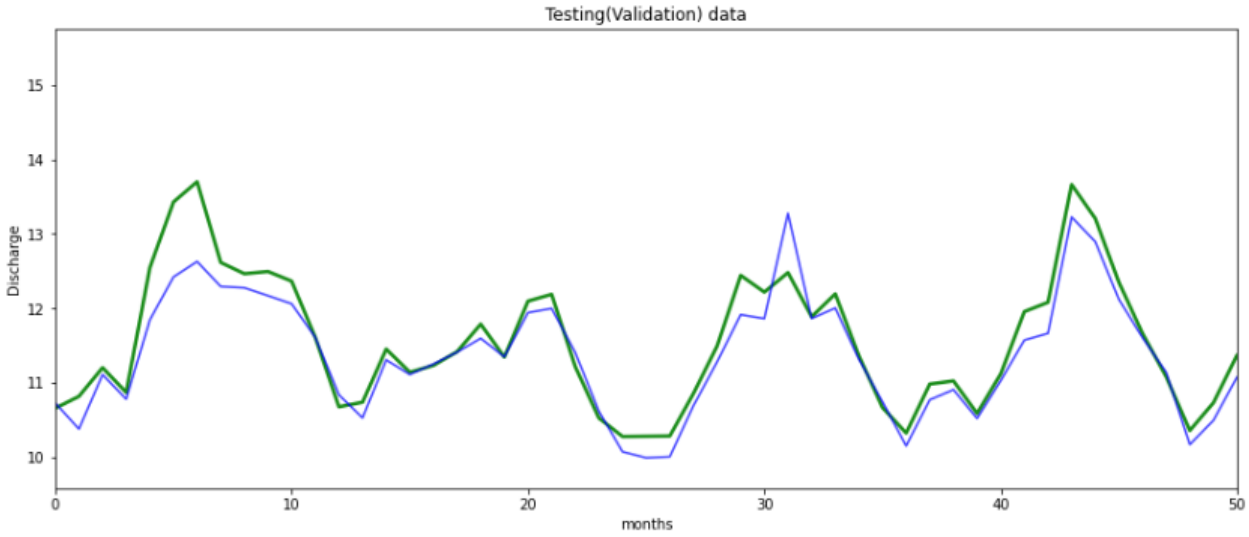

Figure 5.3: Pattazhy LSTM Model discharge lag 2 mean gauge lag 1

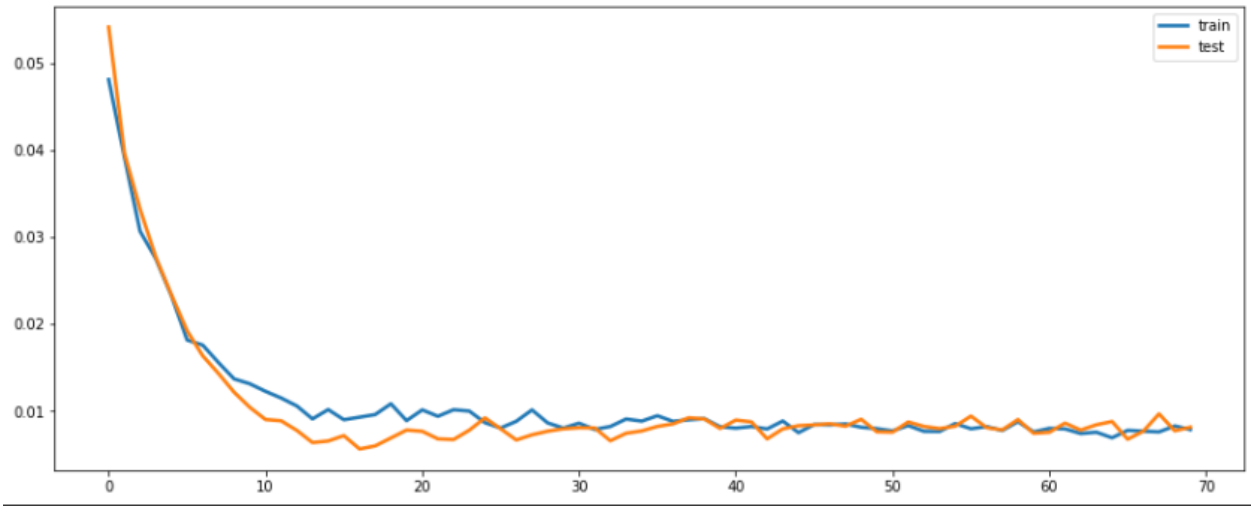

Figure 5.4: Pattazhy LSTM Model loss discharge lag 2 mean gauge lag 1

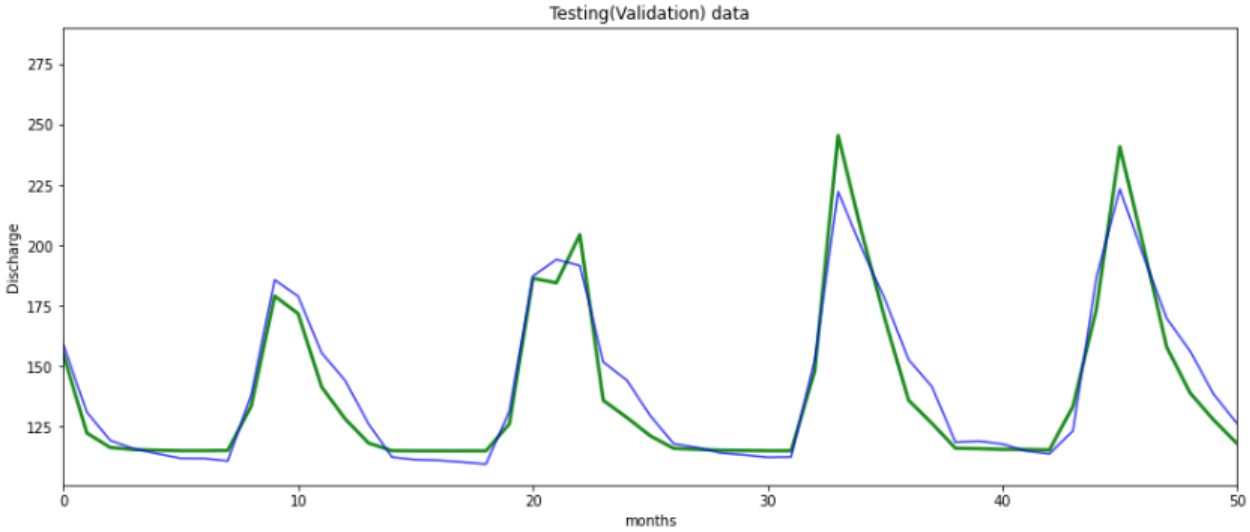

Figure 5.5: Perumannu LSTM Model loss discharge lag 1 mean gauge lag 1

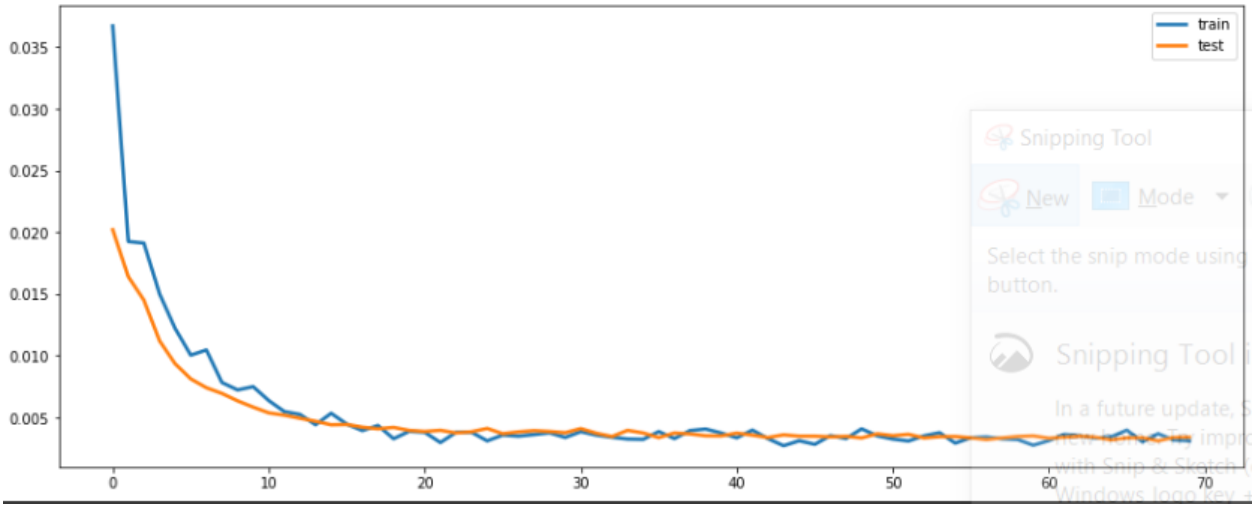

Figure 5.6: Perumannu LSTM Model loss discharge lag 1 mean gauge lag 1

Comparing the figures of three stations the Kidangoor station shows more accurate figures of LSTM model. Also the models loss plot is better for Kidangoor station comparing Pattazhy and Perumannu. Also the model loss plot of Kidangoor LSTM is precise than other stations.It is because of the Kidangoor station have better correlated dataset. Pattazhy station have less correlated dataset compared to other stations. So that its generated model plot and model loss of LSTM are less accurate.

Table 5.1: Comparison of R-squared values of Kidangoor,Pattazhy and Perumannu for training

Stations	Input	LSTM	RF	SVR
Kidangoor	$Q_{(t-6)}$	0.972	0.866	0.941
	$M_{(t-6)}$	0.973	0.738	0.821
	$Q_{(t-6)}M_{(t-6)}$	0.972	0.876	0.946
Pattazhy	$Q_{(t-2)}$	0.792	0.78	0.722
	$M_{(t-1)}$	0.782	0.77	0.681
	$Q_{(t-2)}M_{(t-1)}$	0.790	0.82	0.784
Perumannu	$Q_{(t-1)}$	0.92	0.861	0.891
	$M_{(t-1)}$	0.93	0.873	0.893
	$Q_{(t-1)}M_{(t-1)}$	0.92	0.892	0.921

Table 5.2: Comparison of R-squared values of Kidangoor,Pattazhy and Perumannu for testing

Stations	Input	LSTM	RF	SVR
Kidangoor	$Q_{(t-6)}$	0.964	0.846	0.921
	$M_{(t-6)}$	0.962	0.724	0.676
	$Q_{(t-6)}M_{(t-6)}$	0.964	0.844	0.924
Pattazhy	$Q_{(t-2)}$	0.739	0.601	0.564
	$M_{(t-1)}$	0.691	0.594	0.534
	$Q_{(t-2)}M_{(t-1)}$	0.744	0.650	0.608
Perumannu	$Q_{(t-1)}$	0.896	0.801	0.861
	$M_{(t-1)}$	0.919	0.834	0.868
	$Q_{(t-1)}M_{(t-1)}$	0.920	0.850	0.885

Table 5.3: Comparison of RMSE values of Kidangoor,Pattazhy and Perumannu for training

Stations	Lag	LSTM	RF	SVR
Kidangoor	$Q_{(t-6)}$	0.042	0.039	0.043
	$M_{(t-6)}$	0.041	0.045	0.0393
	$Q_{(t-6)}M_{(t-6)}$	0.042	0.038	0.043
Pattazhy	$Q_{(t-2)}$	0.079	0.044	0.060
	$M_{(t-1)}$	0.081	0.044	0.067
	$Q_{(t-2)}M_{(t-1)}$	0.079	0.042	0.058
Perumannu	$Q_{(t-1)}$	0.052	0.039	0.036
	$M_{(t-1)}$	0.047	0.038	0.036
	$Q_{(t-1)}M_{(t-1)}$	0.050	0.036	0.042

Table 5.4: Comparison of RMSE values of Kidangoor,Pattazhy and Perumannu for testing

Stations	Lag	LSTM	RF	SVR
Kidangoor	$Q_{(t-6)}$	0.043	0.041	0.043
	$M_{(t-6)}$	0.0437	0.060	0.099
	$Q_{(t-6)}M_{(t-6)}$	0.044	0.041	0.043
Pattazhy	$Q_{(t-2)}$	0.0921	0.041	0.110
	$M_{(t-1)}$	0.0972	0.104	0.113
	$Q_{(t-2)}M_{(t-1)}$	0.0964	0.107	0.117
Perumannu	$Q_{(t-1)}$	0.070	0.040	0.039
	$M_{(t-1)}$	0.0636	0.0405	0.039
	$Q_{(t-1)}M_{(t-1)}$	0.0677	0.040	0.037

Table 5.5: Comparison of MAE values of Kidangoor,Pattazhy and Perumannu for training

Stations	Lag	LSTM	RF	SVR
Kidangoor	$Q_{(t-6)}$	0.034	0.048	0.033
	$M_{(t-6)}$	0.033	0.044	0.051
	$Q_{(t-6)}M_{(t-6)}$	0.034	0.048	0.031
Pattazhy	$Q_{(t-2)}$	0.059	0.064	0.101
	$M_{(t-1)}$	0.061	0.065	0.101
	$Q_{(t-2)}M_{(t-1)}$	0.059	0.062	0.065
Perumannu	$Q_{(t-1)}$	0.040	0.045	0.047
	$M_{(t-1)}$	0.035	0.044	0.046
	$Q_{(t-1)}M_{(t-1)}$	0.039	0.042	0.038

Table 5.6: Comparison of MAE values of Kidangoor,Pattazhy and Perumannu for testing

Stations	Lag	LSTM	RF	SVR
Kidangoor	$Q_{(t-6)}$	0.034	0.048	0.033
	$M_{(t-6)}$	0.033	0.044	0.051
	$Q_{(t-6)}M_{(t-6)}$	0.034	0.048	0.031
Pattazhy	$Q_{(t-2)}$	0.059	0.064	0.101
	$M_{(t-1)}$	0.061	0.065	0.101
	$Q_{(t-2)}M_{(t-1)}$	0.059	0.062	0.065
Perumannu	$Q_{(t-1)}$	0.040	0.045	0.047
	$M_{(t-1)}$	0.035	0.044	0.046
	$Q_{(t-1)}M_{(t-1)}$	0.039	0.042	0.038

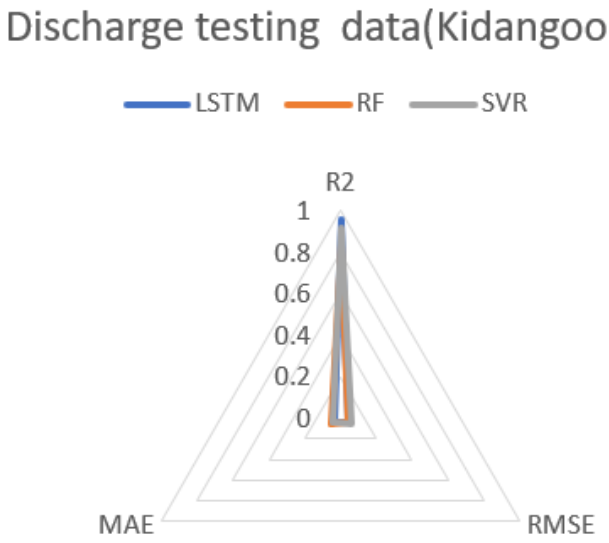

Figure 5.7: Radar Plot of Kidangoor performance evaluation

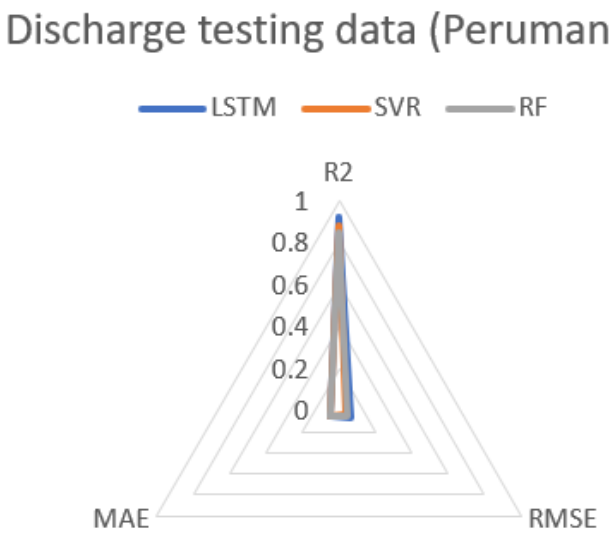

Figure 5.8: Radar Plot of Perumannu performance evaluation

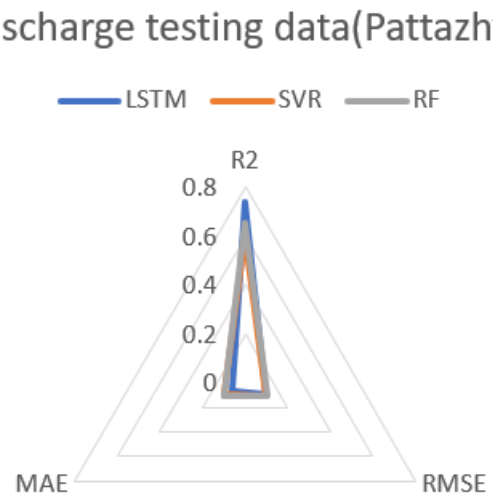

Figure 5.9: Radar Plot of Pattazhy performance evaluation

Tables shows that LSTM possess better performance in every three stations. Both other machine learning models lay under the LSTM in the case of R-Squared values, RMSE values and MAE values. Like we previously discussed due to the less correlation of Pattazhy dataset R-Squared values are much lesser in other stations.While considering R-squared values, RMSE values MAE values, LSTM can consider as the best for streamflow forecasting than SVR and RF.

While looking into the radar plots of Kidangoor Pattazhy and Perumannu (Figure 5.8-5.9), the best results are shown by LSTM than other machine learning models (SVR and Random forest) in every stations. LSTM is near to the R-Squared value in every stations radar plots that means compering to every models LSTM shows best performance. At the lagged value of discharge 6, mean gauge 6 in Kidangoor data possess better performance in LSTM . The table of R-Squared and RMSE also shows that LSTM is better than SVR and RF.Comparing every stations and models the LSTM has the best R-Squared value that is 0.96 in Kidangoor data (Table 5.1,Table 5.2). In every station RMSE value and MAE values are also less for LSTM .That also means that LSTM is far better than both machine learning models (SVR and Random forest).

Chapter 6

CONCLUSION AND FUTURE WORKS

Streamflow prediction is very important in the future of water resource management. This study contains three datasets in Kerala, they are Kidangoor, Pattazhy and Perumannu. The current work sought to determine the best effective models for predicting monthly streamfow while taking into account three AI techniques including LSTM, SVR and RF. Using the same set of datasets for comparison, calibration, and validation applied to all of the models stated above, and three indicators of performance measures such as R-square value, RMSE values and MAE values were used to assess the effectiveness of the various constructed models. Based on the outcomes of three stations such as Kidangoor, Pattazhy and Perumannu, the LSTM model is showing better performance around all selected data. Kidangoor station with mean gauge lag 6 and discharge lag 6 posses best R-squared value of 0.96. Similarly for same combination of input, SVR give 0.92 and RF give 0.84 of R-squared values.Also for other stations(Pattazhy and Perumannu) LSTM has the best R-squared values. Similarly for RMSE values and MAE values LSTM shows better performance than other ml models. This means for every inputs of every stations LSTM shows best results than SVR and RF.

References

[1] Poul AK, Shourian M, Ebrahimi H(2019). A comparative study of MLR, KNN, ANN and ANFIS models with wavelet transform in monthly stream flow prediction. Water Resources Management.;33(8):2907-23.

[2] Malik A, Tikhamarine Y, Souag-Gamane D, Kisi O, Pham QB (2020). Support vector regression optimized by meta-heuristic algorithms for daily streamflow prediction. Stochastic Environmental Research and Risk Assessment.;34(11):1755-73.

[3] Rabbi, M. I. I., Galib, M. M. H., Hasan, M. A., Toma, P. M., Ahamed, M. (2021). Discharge Prediction at Bahadurabad Transit of Brahmaputra-Jamuna Using Machine Learning and Assessment Of Flooding. Journal of Water Resources Pollution Studies, 6.

[4] Meshram SG, Meshram C, Santos CA, Benzougagh B, Khedher KM (2021). Streamflow prediction based on artificial intelligence techniques. Iranian Journal of Science and Technology, Transactions of Civil Engineering. 30:1-1.

[5] Dong, Limei, Desheng Fang, Xi Wang, Wei Wei, Robertas Damasevicius, Rafał Scherer, and Marcin Woźniak. (2020)."Prediction of Streamflow Based on Dynamic Sliding Window LSTM" Water.:12(11): 30-32.

[6]]Parisouj P, Mohebzadeh H, Lee T (2020). Employing machine learning algorithms for streamflow prediction: a case study of four river basins with different climatic zones in the United States. Water Resources Management.;34(13):4113-31.

[7] Chapman, K. W., Gilmore, T. E., Chapman, C. D., Mehrubeoglu, M., Mittelstet, A. R. (2020). Camera-based Water Stage and Discharge Prediction with Machine Learning. Hydrology and Earth System Sciences Discussions, 1-28.

[8] Thapa, S., Zhao, Z., Li, B., Lu, L., Fu, D., Shi, X., ... Qi, H. (2020). Snowmelt-driven streamflow prediction using machine learning techniques (LSTM, NARX, GPR, and SVR). Water, 12(6), 17-34.

[9] Adnan, R.M., Liang, Z., Trajkovic, S., Zounemat-Kermani, M., Li, B., Kisi, O., (2019) Daily streamflow prediction using optimally pruned extreme learning machine, Journal of Hydrology

[10] Sudriani, Y., Ridwansyah, I., Rustini, H. A. (2019). Long short term memory (LSTM) recurrent neural network (RNN) for discharge level prediction and forecast in Cimandiri river, Indonesia. In IOP Conference Series: Earth and Environmental Science (Vol. 299, No. 1, p. 012037). IOP Publishing.

[11] Kratzert, F., Klotz, D., Brenner, C., Schulz, K., Herrnegger, M. (2018). Rainfall–runoff modelling using long short-term memory (LSTM) networks. Hydrology and Earth System Sciences, 22(11), 6005-6022.