Project Title: **“RAIN-FALL PREDICTION”**

Project Report

1. CHAPTER 1- INTRODUCTION OF PROJECT

The majority of us typically rely on forecasts to inform us of the type of weather to expect. In meteorology, predicting rainfall is a crucial and challenging task. Rainfall is anticipated using special techniques that combine observation, observing, tendencies of information, and patterns. Predicting daily rainfall boosts agricultural output and ensures a reliable supply of food and water to maintain healthy populations. In order to forecast rain, many sorts of research have been done utilising machine learning and data mining methods of information on the environment from many nations. A country with irregular rainfall distribution impacts the agriculture, which is the foundation of the nation's economy. To reduce the issue of the country's drought and food shortage, the country should plan and practise the wise use of rainfall water. Weather forecasting is the most popular service offered by the meteorological department for all nations worldwide. Due to the need for numerous specialists and the uncertainty of all calls, the process is quite hard. Predicting rainfall can assist prevent flooding, save human lives and property. Additionally, it aids in the management of water resources. Prior rainfall data is useful to farmers for improved crop management, which boosts the nation's economy.

Rainfall prediction is a difficult assignment for meteorological scientists because of variations in the timing and amount of the precipitation. Rainfall can be anticipated by using a variety of tools and methodologies. In this study, a feed forward neural network (FFNN) version of an artificial neural network (ANN) is built to predict rainfall. The valuable and alluring gentle computing approach for prediction is called artificial neural networks (ANN). ANN is built entirely on a self-adaptive mechanism, whereby the model learns from historical records, recognises meaningful links between records, and forecasts based on current records. One important factor in managing water resources is making accurate rainfall predictions. The prediction accuracy is measured the usage of confusion matrix and RMSE. The consequences display that the prediction version primarily based totally on ANN shows suited accuracy.

1.1.1 KEY FEATURES

- A user-input-based fuzzy recommendation model.
- Utilization of real-world rainfall data from 1901 to 2015.
- Interactive graphical visualization of results

1.1.2 NEED OF THE SYSTEM

efficient water management, mitigates flood risks, and provides insights for infrastructure planning. By predicting rainfall patterns, the system aids in managing climate impacts, reducing financial losses, and enhancing decision-making for sustainable development.

1.2 PROBLEM STATEMENT

The forecasting of heavy rain presents a significant challenge to the meteorological service because it has a direct impact on both human life and the economy. It is the root cause of yearly natural catastrophes like floods and droughts that affect people all over the world. For nations like India, whose economy is heavily dependent on agriculture, the accuracy of rainfall forecasts is extremely important. By identifying hidden patterns among the readily available elements of historical weather data, Artificial Neural Network (ANN) algorithms may accurately estimate the amount of rainfall. This project makes a contribution by offering a comprehensive study and critique of the most recent data mining algorithms for predicting rainfall. To improve the accuracy of forecasts and climate monitoring, weather parameters should be calculated with a very high level of accuracy, in a timely manner, and under controlled conditions. Since rainfall affects human life, rainfall estimation is crucial. Water consumption and resources.

However, it is exceedingly difficult to predict how much rainfall is affected by geographical and regional differences and features. Predicting rain is one of the more difficult aspects of weather forecasting. Accurate and prompt rainfall forecasting can be very beneficial for planning ahead and implementing adequate security measures for ongoing building projects, transportation operations, agricultural jobs, aviation operations, and flood situations, among other things. By identifying hidden patterns among the readily available elements of historical weather data, machine learning techniques can accurately estimate the amount of rainfall. This study makes a contribution by offering a comprehensive analysis and overview of the most recent machine learning approaches for forecasting rainfall. To improve the accuracy of forecasts and climate monitoring, weather parameters should be calculated with a very high level of accuracy, in a timely manner, and under controlled conditions.

Because of the impact that rainfall has on human life, water resources, and water consumption, rainfall estimation is crucial. However, it is highly challenging to estimate rainfall that is affected by geographical and regional changes and peculiarities.

1.3 Objectives

The objective of this project is to develop a predictive model for rainfall forecasting in India using machine learning techniques. The model aims to analyse historical rainfall data from 1901 to 2015 to predict future rainfall patterns. This information can be crucial for agriculture, water management, and disaster preparedness.

Outline of the report

CHAPTER 1: INTRODUCTION OF PROJECT

CHAPTER 2: LITERATURE SURVEY

CHAPTER 3: SYSTEM ARCHITECTURE

CHAPTER 4: RESULTS AND DISCUSSION

CHAPTER 5: CONCLUSION

2. CHAPTER 2- LITERATURE SURVEY

[1] Chalachew Muluken Liyew and Haileyesus Amsaya Melese proposed machine learning techniques to predict rainfall amounts. This paper explained Utilizing machine learning approaches, determining the pertinent atmospheric characteristics that contribute to precipitation and forecast daily rainfall intensity. Relevant environmental factors were chosen using the Pearson correlation method and utilized as an input for the model for machine learning. The local meteorological station provided the dataset, which was in Bahir Dar, Ethiopia, to evaluate the effectiveness of three machine learning methods (Multivariate Linear Regression, Random Forest, and Extreme Gradient Boost). The performance of the machine learning model was evaluated using Mean absolute error and Root mean squared error approaches. The study's findings

showed that Extreme Gradient Boosting outperformed other machine learning algorithms.

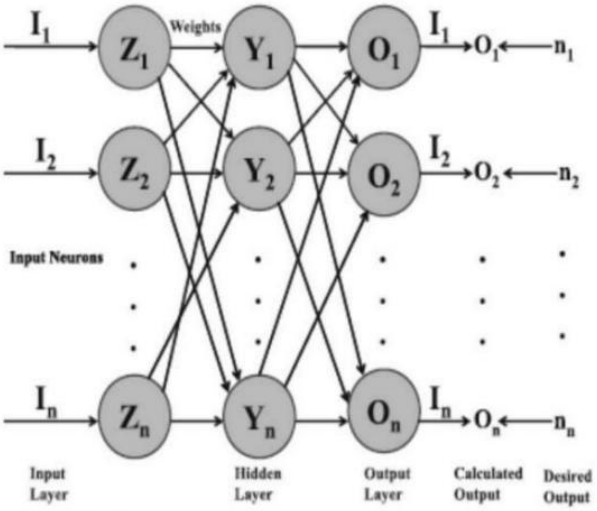
[2] This paper makes a specialty of strategies; (1) Forecasting rainfall with the use of Autocorrelation Function (ACF) primarily based totally on the ancient rainfall records and (2) Forecasting rainfall with the 9 use of Projected Error primarily based totally on ancient and projected rainfall records. Both strategies use extraordinary algorithms inclusive of BDTR, DFR, BLR and NNR to become aware of the optimal prediction for rainfall and extraordinary time horizons. The outcomes provided that for M1, the end result is higher with cross-validation with BDTR and tuning its parameter. The greater the entry covered to the version; the greater the correct version can perform. The quality regression advanced for ACF is BDTR because it has the very best coefficient of determination, R2 (daily: 0.5525075, 0.8468193, 0.9739693; weekly: 0.8400668, 0.8825647, 0.989461; 10 days: 0.8038288, 0.8949389, 0.9607741, 0.9894429; and monthly: 0.9174191, 0.6941756, 0.9939951, 0.9998085) that means the higher rainfall prediction for the future. For approach 2, a variant of the end result when the use of extraordinary normalization strategies and suggests the use of LogNormal normalization with BDTR and DFR offers the quality version.

[3] Aakash Parmar, Kinjal Mistree, Mithila Sompura proposed a review on how the Machine Learning Techniques are useful for the rainfall Prediction. This paper presented review of different methods used for rainfall prediction and problems one might encounter while applying different approaches for rainfall forecasting. Every statistical model has some drawbacks. So, the alternative option of Artificial Neural Network is used. ANN contains a big number of interconnected neurons, which mostly operate in parallel, and are well structured. Categories of neural networks are either single layer or multi-layer. Layer between input layer and output layer is called as hidden layer.

Back-Propagation Neural Network:

BPNN is made of MLFF neural network which contains one input layer, hidden layers and one output layer. The ultimate goal of BPNN is to decrease the calculated error obtained from the difference between the calculated output and desired output of the

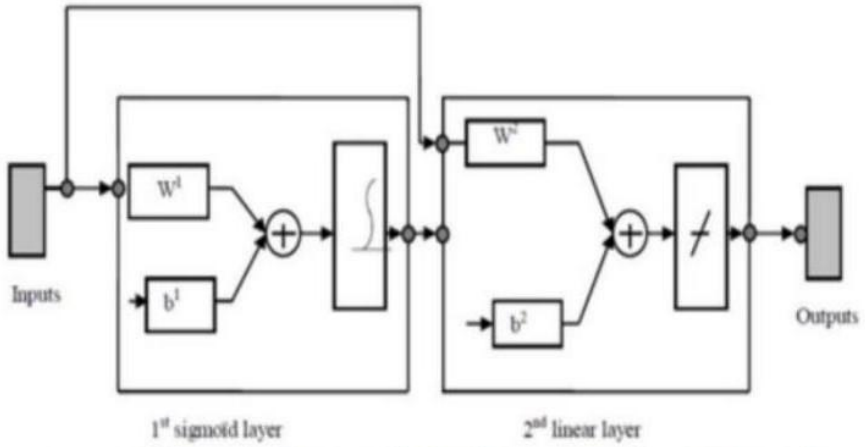
neural network by adjusting the weights after each iteration. So, in BPNN, each information is propagated in backward direction until the calculated error is very small or zero. Alternate option to make the neural network generalize enough is by doing small changes in the number of layers and neurons in the inputs, without changing the output components. Solution is to keep the architecture of neural network relatively simple and small, because complex architectures are much more prone to overfitting.



A BPNN architecture with one hidden layer

Cascade Forward Back Propagation Network:

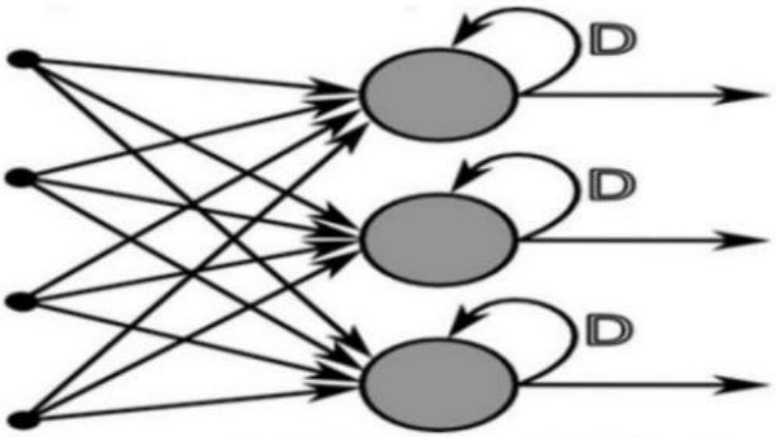
It is used for the prediction of new output data. All the layers in networks are not only connected with its previous layer but also connected with input. Inputs are provided to each layer in network.



Cascade Forward Back Propagation Network

Layer Recurrent Network:

RNNs can use their internal memory to process arbitrary sequences of inputs. RNN are neural networks with a feedback loop. The previous processes of hidden layer and functional outputs are fed back into the network as part of the input to the next hidden layer processes.



Layer Recurrent neural Network

[4] In this article Rahman A-u, Abbas S, Gollapalli M, Ahmed R, Aftab S, Ahmad M, Khan MA, Mosavi A proposed that the Precipitation in any form—such as rain, snow, and hail—can have an effect on dayafter-day outside activities. Precipitation prediction is one in all the difficult tasks in meteorology method. Correct precipitation prediction is currently tougher than before thanks to the intense climate variations. Machine learning techniques will predict precipitation by extracting hidden patterns from historical weather knowledge. Choice of associate acceptable classification technique for prediction could be a troublesome job. This analysis proposes a unique period precipitation prediction system for good cities employing a machine learning fusion technique. The projected framework uses four wide used supervised machine learning techniques, i.e., call tree, Naïve Thomas Bayes, K-nearest neighbors, and support vector machines. For effective prediction of precipitation, the technique of symbolic logic is incorporated within the framework to integrate the prophetical accuracies of the machine learning techniques, additionally called fusion. The results mirror that the projected machine learning fusion-based framework outperforms alternative models.

[5] In this paper Jinle Kang, Huimin Wang, Feifei Yuan, Zhiqiang Wang, Jing Huang and Tian Qiu combinedly proposed that Precipitation could be an important input for hydrologic simulation and prediction, and is wide used for agriculture, water resources management, and prediction of flood and drought, among different activities. Few studies tried deep learning strategies like the progressive for Recurrent Neural Networks (RNNs) networks in meteoric sequence statistic predictions. They deployed Long remembering (LSTM) network models for predicting the precipitation based on meteoric knowledge from 2008 to 2018 in Jingdezhen town. When characteristic the correlation between meteoric variables and also the precipitation, 9 vital input variables were elite to construct the LSTM model. Then, the chosen meteoric variables were refined by the relative importance of input variables to reconstruct the LSTM model. Finally, the LSTM model with final selected input variables was employed to predict the precipitation and also the performance is compared with other classical applied math algorithms and also the machine learning algorithms. The experimental results show that the LSTM is appropriate for precipitation prediction. The RNN models, combined with meteorological variables, may predict the 12 precipitation accurately in Jingdezhen town and supply sufficient time to arrange ways against potential connected disasters.

[6] For various time series, machine learning techniques have been used to find accurate and smooth prediction models. The effectiveness of the models varies based on the trends and peculiar forms of the dataset. Due to the state of Meghalaya's diverse climate and unique geographic location, this study has concentrated on a case study of rainfall and other climate indicators in this region. Five machine learning techniques, including Linear Regression (LR), Regression Trees (RT), Gaussian Process Regression (GPR), Support Vector Machines (SVM), and Ensembles of Trees (ET), were combined to create the ideal machine learning model for predicting rainfall in Meghalaya. The effectiveness of each model was evaluated using its root mean squared error (RMSE). The study examined how well the parameters were optimised before being used as predictors in the training dataset. Almost all datasets from various districts in Meghalaya were found to have extremely good result-oriented models in LR, SVM, and GPR, according to the results. But usually always, LR and SVM algorithms provide the most accurate predictions. Wet day frequency was identified by the studies as the most

sensitive and effective single parameter as a predictor, which might result in the production of prediction models to run the study for a more accurate forecast.

[7] The benefits of rainfall forecasting are illustrated in this study using regression, ensemble RF, and SVR. The available meteorological dataset variables in this ensemble model can be used to forecast daily and monthly rainfall. To evaluate the new model's performance in relation to the chosen area for measuring rainfall parameters, NB, DT, LR, and RF regressions are carried out. It is therefore a useful tool for forecasting rain in a past situation. Using the simulation variables, a 15.46-millimetre root mean square error (RMSE) with 96.34 percent accuracy may be projected. The anticipated model can be used for prediction because it has a lower error rate than RF and linear regression. This is especially helpful in areas with a data shortage. The projected model performs better than the ones now in use and offers a better value for the money. Python is the sole language that can be used to run the simulation. Daily/monthly rainfall magnitudes are more accurately documented with the suggested model than they are with the individual approaches. As a result, it is found that the SVR-RF Ensemble Model is the most effective technique for predicting extreme rainfall events. Because of the autonomous feature picks, deep learning techniques can be utilised to make future predictions. The great level of precision of this model is an aggle vnal benefit.

[8] Pre diction of rainfall gives awareness to people and know in advance about rainfall to take certain precautions to protect their crop from rainfall. Many techniques came into existence to predict rainfall. Machine Learning algorithms are mostly useful in predicting rainfall. Some of the major Machine 13 Learning algorithms are ARIMA Model (Auto-Regressive Integrate d Moving Average), Artificial Neural Network, Logistic Regression, Support Vector Machine and Self Organizing Map. Two commonly used models predict seasonal rainfall such as Linear and Non-Linear models. The first models are ARIMA Model. while using Artificial Neural Network (ANN) predicting rainfall can be done using Back Propagation NN, Cascade NN or Layer Recurrent Network. Artificial NN is same as Biological Neural Networks.

The following table summarizes different methods offered by various techniques for rainfall prediction.

S.no	Year	Title	Authors	Methodologies used	Pros	Cons
1	2021	Machine learning aggle ve to predict daily rainfall amount.	Chalachew Muluken Liyew,Haile yesus Amsaya Melese	➢ Linear regression. ➢ Data-driven machine learning algorithm. ➢ Multivariate Linear Regression (MLR). ➢ Random Forest (RF). ➢ XGBoost gradient descent.	➢ Able to quantify how one or more predictor variables influence the outcome in relation to each other. ➢ Missing values can be handled automatically using Random Forest. ➢ Suitable for big data sets.	➢ Complexity. ➢ Long Training Period. ➢ When dealing with noisy data and deep trees, tree algorithms like XGBoost and Random Forest may overfit the data.
2	2021	Rainfall forecastin g model using machine learning	Wanie M. Ridwan, Michelle Sapitang, Awatif Aziz, Khairul Faizal Kushiar, Ali	➢ **METHOD-1:** Forecasting rainfall using Autocorrelation function ➢ **METHOD-2:**	➢ **Method-1** is the best regression developed for ACF is BDTR since it has	➢ The accuracy level of **ML** models used in predicting

		methods: Case study Terengganu, Malaysia	Najah Ahmed, Ahmed El-Shafie	Forecasting rainfall using Projected error based on historical and projected rainfall data. ➢ Both methods use different algorithms such as BDTR, DFR, BLR and NNR	highest coefficient of determination ➢ **Method-2**, overall model performances show that normalization using LogNormal is preferably giving a good result of each category.	rainfall based on historical data has been one of the most critical concerns in hydrological studies
3	2017	Machine Learning Techniques for Rainfall Prediction	Aakash Parmar, Kinjal Mistree, Mithila Sompura	✓ Two Widely used methods ✓ Statistical methods and Numerical Weather Prediction Model ✓ Artificial Neural Network 1. Back-Propagation Neural Network 2. Cascade Forward Back Propagation Network 3. Layer Recurrent Network	Due to nonlinear relationships in rainfall data and ability of learning from the past makes Artificial Neural Network a preferable approach from all available approaches	Frequency, intensity and amount are main characteristics for time series rainfall. These values can be varied from one position on earth to other position of earth and from one time to other time. Every statistical model has some drawbacks

4	2022	Rainfall Prediction System Using Machine Learning Fusion for Smart Cities	Rahman A-u, Abbas S, Gollapalli M, Ahmed R, Aftab S, Ahmad M, Khan MA, Mosavi A	➢ Meteorology ➢ machine learning fusion technique ➢ Cleaning ➢ Normalization ➢ splitting	temperature prediction in order to efficiently utilize clean solar energy. Enhanced the accuracy of rain prediction system	It has only one disadvantage if due to any reason, the data which will be used for prediction is compromised, then the prediction cannot be trusted.
5	2020	Prediction of Precipitation Based on Recurrent Neural Networks	Jinle Kang, Huimin Wang, Feifei Yuan, Zhiqiang Wang, Jing Huang and Tian Qiu	• RNN • LTSM • Statistical methods like correlation	- appropriate for precipitation prediction. -The RNN models, combined with meteorological variables, may predict the precipitation accurately -supply sufficient time to arrange ways against potential connected disasters.	-storage issues due to more usage of memory -performance may reduce due to the merging of methods.
6	2021	Prediction of Rainfall using Machine Learning Algorith	Shabbir Ahmed Osmani A, Foysol Mahmud B, and Md.Abu Zafor C.	➢ Linear Regression (LR), ➢ Regression Trees (RT), ➢ Gaussian Process Regression (GPR),	➢ Supervised machine learning algorithm are very consistent to pattern	When different models were developed using a range of parameters then a particular parameter or

		ms for Different Districts of Meghalaya		➢ Support Vector Machines (SVM) ➢ Ensembles of Trees (ET)	recognition. ➢ Out of seven aggle v e wet day frequency is achieved as the best parameter alone to make a consistent regression model for prediction.	group of parameters could be more sensitive to produce an efficient model.
7		Rainfall Prediction Using Machine Learning Based Ensemble Model	Veera Ankalu. Vuyyuru; Giduturi. Apparao; S. Anuradha	• Liner Regression • Ensemble learning • Back Propagation Neural Network, Back Propagation Neural Network	• Deep learning methods can be used to forecast the future because of the aggle ve feature selections. Another advantage of using this model is that it has a high level of accuracy.	• Storage issues • A particular parameter or combinati on of parameters may be more sensitive when various models are constructe d utilising a variety of

					• This model is determined to be the best method for forecasting extreme rainfall events.	parameters , leading to the creation of an aggle ve model.
8		Rainfall Prediction Using Machine Learning & Deep Learning Techniqu es	CMAK Zeelan Basha, Nagulla Bhavana, Ponduru Bhavya, Sowmya V	➢ ARIMA MODEL(AutoRegr essive Integrated Moving Average) ➢ ARTIFICIAL NEURAL NETWORK ➢ Support Vector Machines	➢ The accuracy can be measured by the MSE and RMSE comparing with the other models. In circumstanc es of water resource and management , human being life and the climate they possess, precipitation prediction is of huge importance ➢ Because of nonlinear relationships	➢ The accuracy level of ML models used in predicting rainfall based on historical data has been one of the most critical concerns in hydrological studies

					in rainfall datasets and the ability to learn from the past, Artificial Neural Network makes a superior solution to all approaches available.	

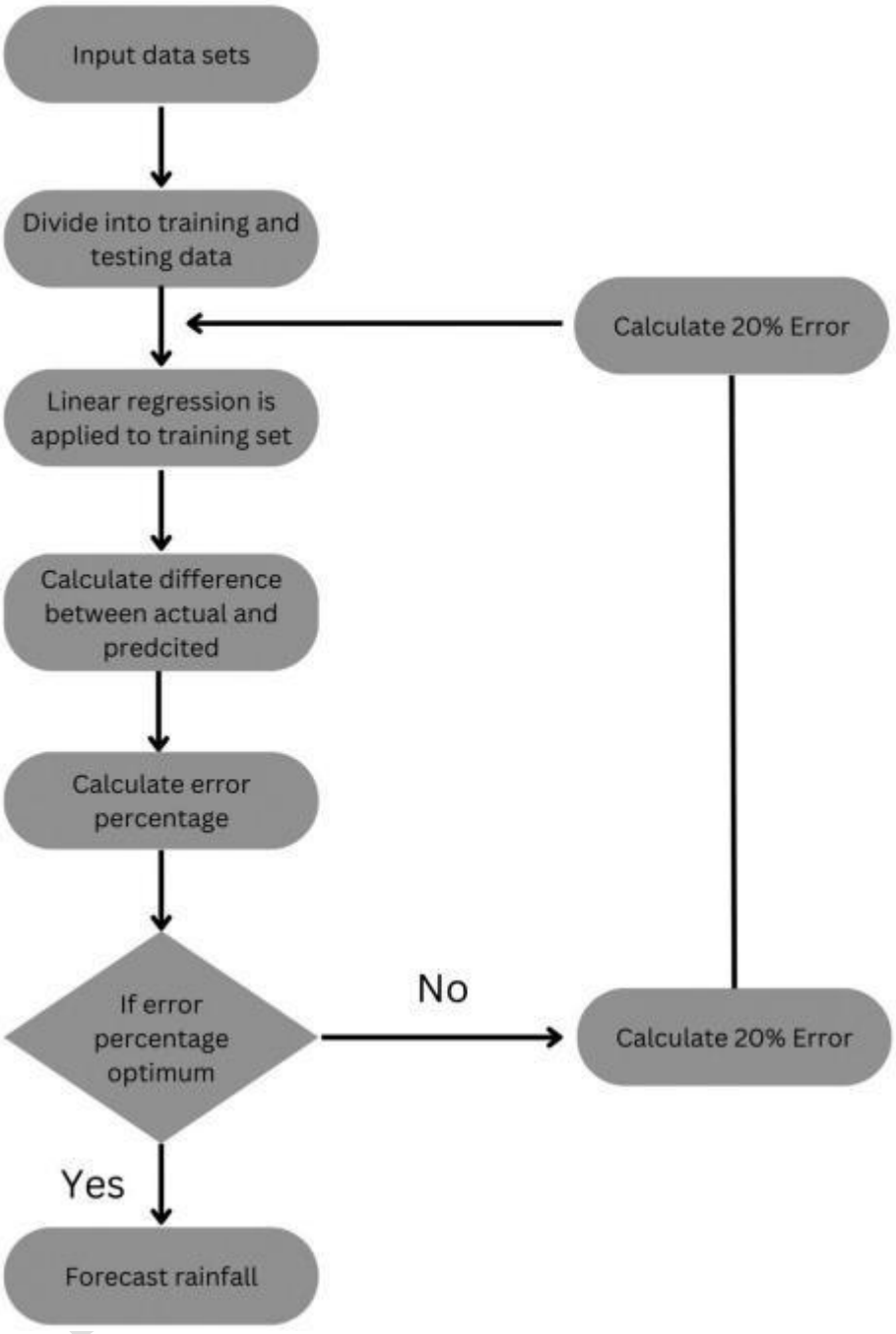
3. CHAPTER 3 – SYSTEM ARCHITECTURE

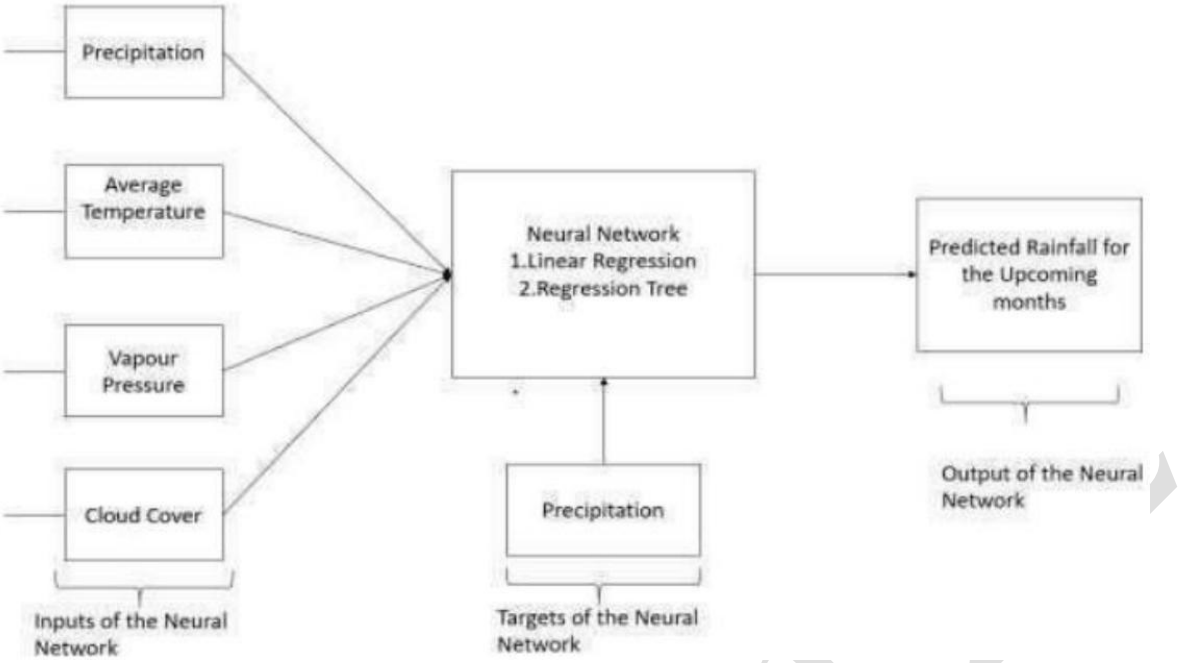
3.1 Introduction

The system architecture for the rainfall prediction model consists of the following components:

- Data Collection Layer: Collects historical rainfall data from reliable sources.
- Preprocessing Layer: Cleans and preprocesses the dataset, handles missing values, and scales features.
- Feature Engineering Layer: Generates features relevant for prediction, such as seasonal trends and historical averages.
- Model Training Layer: Applies machine learning algorithms, including Lasso Regression and SVM with RBF Kernel.
- Prediction Layer: Predicts future rainfall based on input data.
- Evaluation Layer: Calculates performance metrics like MAE, MSE, RMSE, and Accuracy Score.
- Visualization Layer: Displays results through graphs and charts.

3.2 System Architecture





3.2.1 Explanation of the Diagram

DATA COLLECTION:

The process of gathering, measuring, and analysing precise insights for research using accepted, established methods is known as data collection. It includes the area name, information, year, month, etc. Depending on the type of data needed, different disciplines of research require different approaches to data gathering. The average rainfall in India from 1901 to 2021 is recorded in the dataset of rainfall that was collected from the Kaggle website.

DATA EXPLORATION AND PRE-PROCESSING:

Data exploration, also known as exploratory data analysis (EDA), offers a basic collection of exploration tools that integrate real-time data into data analytics at a fundamental level. Combining human techniques with automated tools, like data visualisation, charts, and preliminary reports, can be used for data exploration. Data pre-main processing's goal is to change the input variables to make them more closely match the expected result. The data conversion, missing value 7 management,

categorical encoding, and separation of the dataset into training and testing datasets were all included in the data pre-processing step. The target variable's missing values were eliminated, and the other features were filled using the data's mean because the raw data had missing values and incorrectly encoded values.

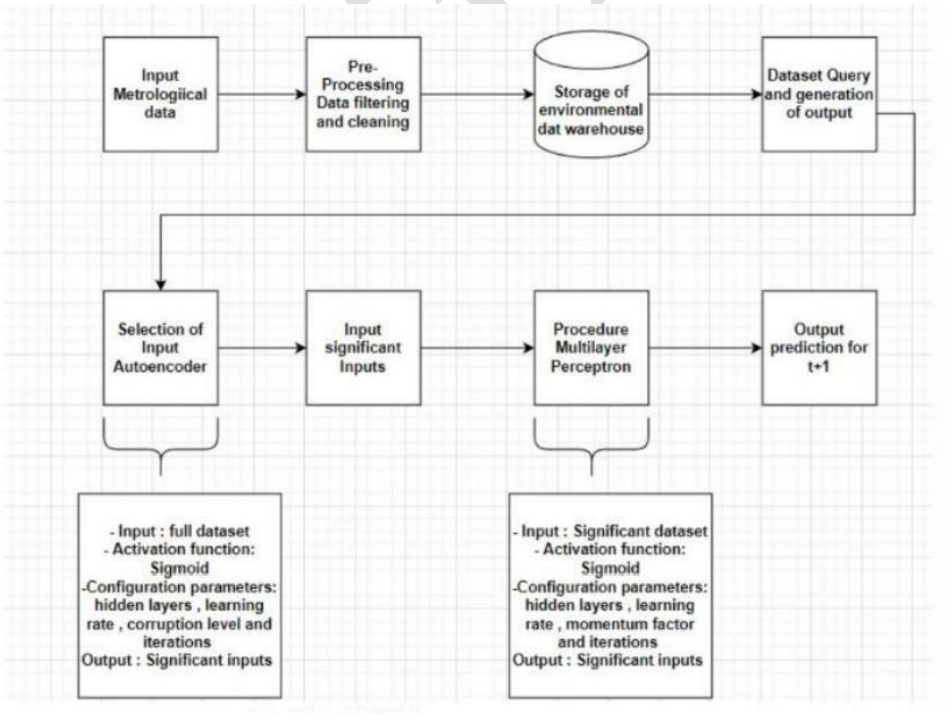
DATA VISUALIZATION:

The graphic display of information and data is known as data visualisation. Data visualisation tools offer an easy approach to observe and analyse trends, outliers, and patterns in data by utilising visual elements like charts, graphs, and maps. To analyse vast volumes of data and make data-driven decisions, data visualisation tools and technologies are crucial in the world of big data.

MODELING:

The Modelling entails teaching a machine learning algorithm to infer labels from features, finetuning it for the needs of the business, and testing it with holdout data. A trained model that may be used for inference, or generating predictions on fresh data points, is the result of modelling

3.3 Work-Flow Design



Algorithm Used:

The rainfall prediction system utilizes Lasso Regression, Support Vector Machine (SVM) with RBF Kernel, and Fuzzy Logic algorithms:

- **Lasso Regression** is used for feature selection and regularization, reducing model complexity and preventing overfitting by shrinking coefficients of less important features to zero. It efficiently handles high-dimensional datasets.
- **Support Vector Machine** (SVM) with RBF Kernel captures complex, nonlinear relationships in the data by mapping input features to a higher-dimensional space. This approach improves prediction accuracy even in challenging scenarios.
- **Fuzzy Logic** is incorporated for handling uncertainty and imprecision in rainfall prediction. It mimics human reasoning by processing vague or ambiguous inputs and generating reliable outputs. By modelling rainfall conditions with fuzzy sets, it improves decision-making in uncertain scenarios.

Additionally, GridSearchCV is used for hyperparameter tuning to optimize the performance of these algorithms.

Inputs:

Historical Rainfall Data (1901-2015): Monthly and annual rainfall data from 36 sub-divisions in India.

Features:

- Year, Month
- Previous Month's Rainfall
- Sub-division data

Outputs

- Predicted rainfall values for future months and years.
- Comparative analysis of actual vs. predicted rainfall.

3.2.2 MODELS

LINEAR REGRESSION MODEL:

A machine learning algorithm based on supervised learning is linear regression. It executes a regression operation. Regression uses independent variables to model a goal prediction value. It is mostly used to determine how variables and forecasting relate to one another. The task of predicting a dependent variable's value (y) based on an independent variable is carried out using linear regression (x). Therefore, x (the input) and y (the output) are found to be linearly related by this regression technique (output). Thus, the term "linear regression" was coined.

FUZZY LOGIC MODEL:

The rainfall prediction project uses a Fuzzy Logic Model to handle uncertainties in weather forecasting. It processes inputs like temperature, humidity, and historical rainfall data through fuzzification, transforming them into fuzzy sets. Using IF-THEN rules, the model evaluates inputs and aggregates results to generate fuzzy outputs. The final prediction is obtained through defuzzification, converting fuzzy values back to crisp outputs. This approach mimics human reasoning, handles imprecise data effectively, and enhances prediction accuracy, making it suitable for real-world scenarios with inherent uncertainties.

LASSO MODEL:

Shrinkage is used in the linear regression method known as lasso regression. When data values shrink toward a middle value, such as the mean, this is called shrinkage. Simple, sparse models are encouraged by the lasso approach (i.e., models with fewer parameters). When you want to automate certain steps in the model selection process, such as variable selection and parameter removal, or when your models exhibit a high degree of multi-collinearity, this particular sort of regression is a good choice.

RIDGE REGRESSION MODEL:

Ridge Regression is a method for analysing multi-collinearity-affected multiple regression data. Although least squares estimates are unbiased when multi-collinearity

occurs, their enormous variances make it possible that they are far from the true value. Ridge regression lowers the standard errors by biassing the regression estimates to some extent. It is envisaged that as a result, estimations will be more trustworthy.

SVC MODEL:

An SVM classifier, or support vector machine classifier, is a type of machine learning algorithm that can be used to analyse and classify data. A support vector machine is a supervised machine learning algorithm that can be used for both classification and regression tasks. The Support vector machine classifier works by finding the hyperplane that maximizes the margin between the two classes. The Support vector machine algorithm is also known as a max-margin classifier. Support vector machine is a powerful tool for machine learning and has been widely used in many tasks such as hand-written digit recognition, facial expression recognition, and text classification. Support vector machine has many advantages over other machine learning algorithms, such as robustness to noise and the ability to handle large datasets.

RANDOM FOREST MODEL:

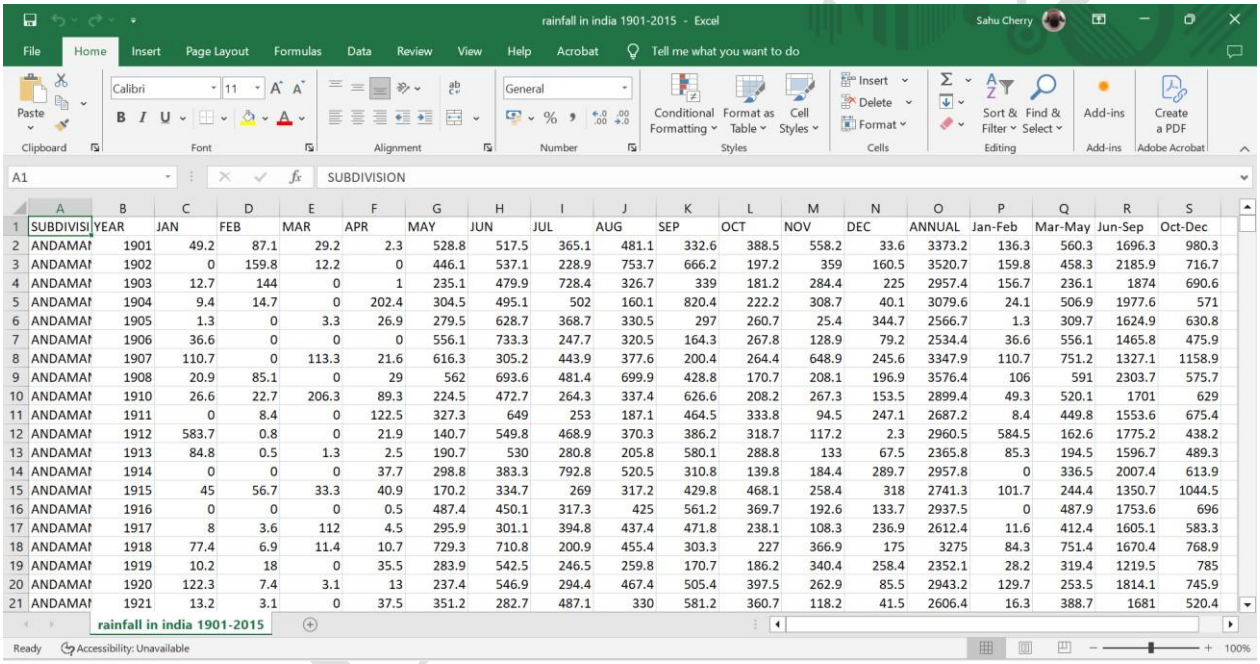
Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model. As the name suggests, “Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and 21 takes the average to improve the predictive accuracy of that dataset.” Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output. The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting.

4. CHAPTER 4 – RESULTS AND DISCUSSION

4.1 Dataset description

- Data set is taken from KAGGLE.
- Our data set has state wise rainfall data information.
- The data is monthly data for 115 years from 1901-2015
- This data set contains monthly rainfall detail of 36 meteorological sub-divisions of India.
- Rainfall unit: mm

  https://www.kaggle.com/datasets/rajanand/rainfall-in-india



rainfall in india 1901-2015 - Excel

SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL	Jan-Feb	Mar-May	Jun-Sep	Oct-Dec
ANDAMAN	1901	49.2	87.1	29.2	2.3	528.8	517.5	365.1	481.1	332.6	388.5	558.2	33.6	3373.2	136.3	560.3	1696.3	980.3
ANDAMAN	1902	0	159.8	12.2	0	446.1	537.1	228.9	753.7	666.2	197.2	359	160.5	3520.7	159.8	458.3	2185.9	716.7
ANDAMAN	1903	12.7	144	0	1	235.1	479.9	728.4	326.7	339	181.2	284.4	225	2957.4	156.7	236.1	1874	690.6
ANDAMAN	1904	9.4	14.7	0	202.4	304.5	495.1	502	160.1	820.4	222.2	308.7	40.1	3079.6	24.1	506.9	1977.6	571
ANDAMAN	1905	1.3	0	3.3	26.9	279.5	628.7	368.7	330.5	297	260.7	25.4	344.7	2566.7	1.3	309.7	1624.9	630.8
ANDAMAN	1906	36.6	0	0	0	556.1	733.3	247.7	320.5	164.3	267.8	128.9	79.2	2534.4	36.6	556.1	1465.8	475.9
ANDAMAN	1907	110.7	0	113.3	21.6	616.3	305.2	443.9	377.6	200.4	264.4	648.9	245.6	3347.9	110.7	751.2	1327.1	1158.9
ANDAMAN	1908	20.9	85.1	0	29	562	693.6	481.4	699.9	428.8	170.7	208.1	196.9	3576.4	106	591	2303.7	575.7
ANDAMAN	1910	26.6	22.7	206.3	89.3	224.5	472.7	264.3	337.4	626.6	208.2	267.3	153.5	2899.4	49.3	520.1	1701	629
ANDAMAN	1911	0	8.4	0	122.5	327.3	649	253	187.1	464.5	333.8	94.5	247.1	2687.2	8.4	449.8	1553.6	675.4
ANDAMAN	1912	583.7	0.8	0	21.9	140.7	549.8	468.9	370.3	386.2	318.7	117.2	2.3	2960.5	584.5	162.6	1775.2	438.2
ANDAMAN	1913	84.8	0.5	1.3	2.5	190.7	530	280.8	205.8	580.1	288.8	133	67.5	2365.8	85.3	194.5	1596.7	489.3
ANDAMAN	1914	0	0	0	37.7	298.8	383.3	792.8	520.5	310.8	139.8	184.4	289.7	2957.8	0	336.5	2007.4	613.9
ANDAMAN	1915	45	56.7	33.3	40.9	170.2	334.7	269	317.2	429.8	468.1	258.4	318	2741.3	101.7	244.4	1350.7	1044.5
ANDAMAN	1916	0	0	0	0.5	487.4	450.1	317.3	425	561.2	369.7	192.6	133.7	2937.5	0	487.9	1753.6	696
ANDAMAN	1917	8	3.6	112	4.5	295.9	301.1	394.8	437.4	471.8	238.1	108.3	236.9	2612.4	11.6	412.4	1605.1	583.3
ANDAMAN	1918	77.4	6.9	11.4	10.7	729.3	710.8	200.9	455.4	303.3	227	366.9	175	3275	84.3	751.4	1670.4	768.9
ANDAMAN	1919	10.2	18	0	35.5	283.9	542.5	246.5	259.8	170.7	186.2	340.4	258.4	2352.1	28.2	319.4	1219.5	785
ANDAMAN	1920	122.3	7.4	3.1	13	237.4	546.9	294.4	467.4	505.4	397.5	262.9	85.5	2943.2	129.7	253.5	1814.1	745.9
ANDAMAN	1921	13.2	3.1	0	37.5	351.2	282.7	487.1	330	581.2	360.7	118.2	41.5	2606.4	16.3	388.7	1681	520.4

4.2 Performance Metrics

MAE (Mean Absolute Error): Measures average absolute errors.

MSE (Mean Squared Error): Measures average squared errors.

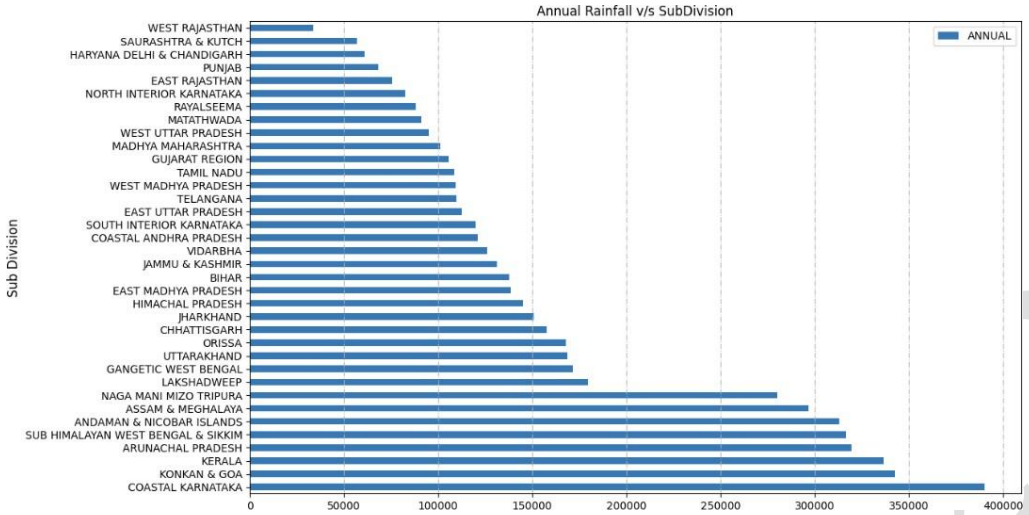
RMSE (Root Mean Squared Error): Measures error magnitude.

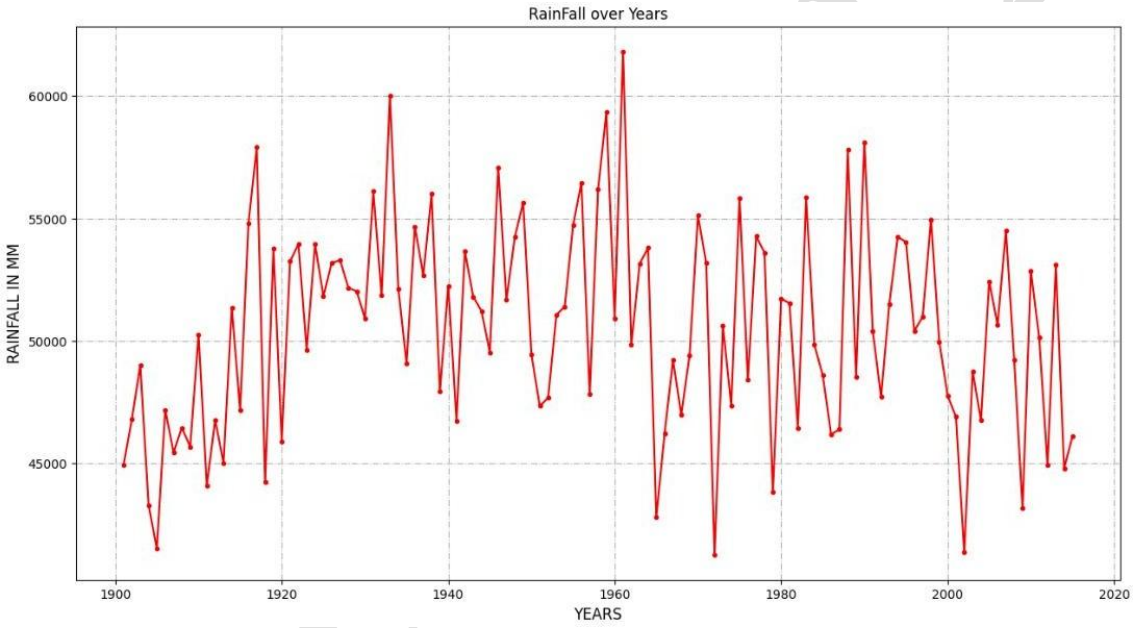
Accuracy Score: Measures model accuracy.

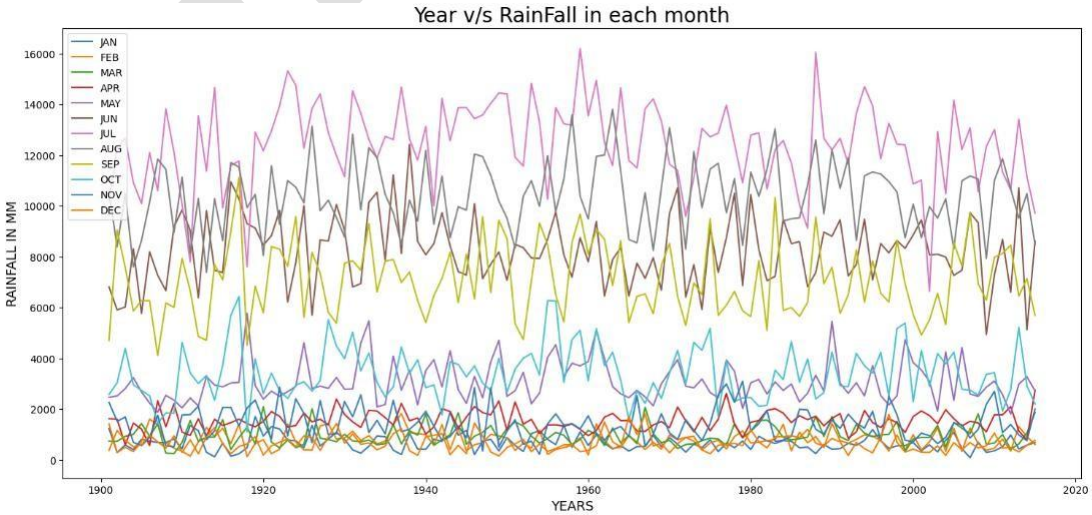
4.3 RESULT

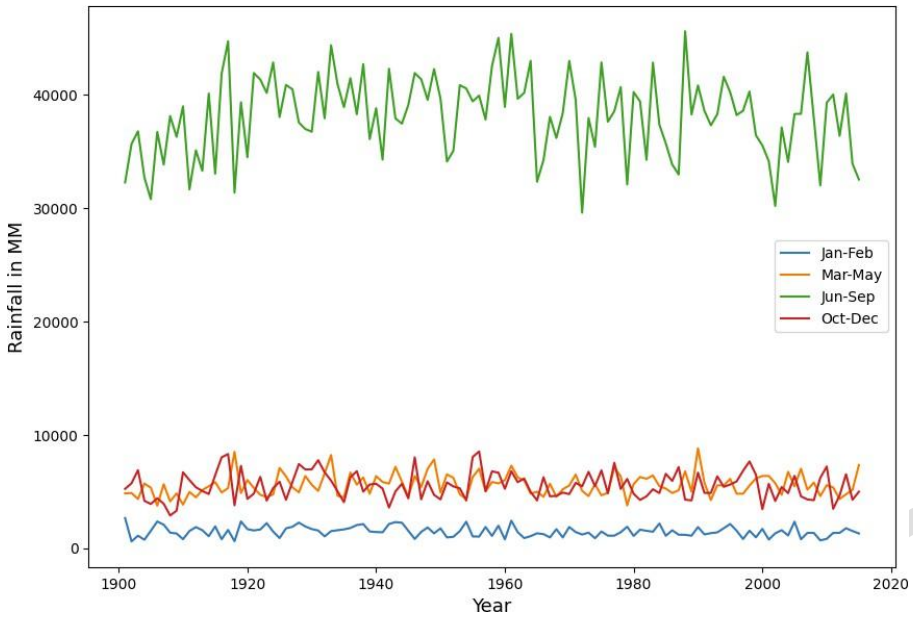
CLASSIFICATION OF ALGORITHMS	TEST DATD	TRAIN DATA
Linear Regression	MAE: 28.3642562333521 MSE: 8648.51533116080 RMSE: 92.997394217046	MAE: 37.3961076801073 MSE: 38548.2150041563 RMSE: 196.33699346826
Fuzzy Logic	MAE: 781.034125042683 MSE: 1328262.50920933 RMSE: 1152.5027154889	MAE:781.034125042683 MSE:1328262.509209333 RMSE:1152.50271548891
Lasso Regression	MAE: 29.4773255221194 MSE: 8719.71659874927 RMSE: 93.379422780124	MAE: 38.5101549995708 MSE: 38564.0742614723 RMSE: 196.37737716313
Ridge Model	MAE: 28.3643117676505 MSE: 8648.50794923844 RMSE: 92.997354528171	MAE: 37.3961779941596 MSE: 38548.2150055971 RMSE: 196.33699347193
SVC Model	MAE: 415.020483010218 MSE: 476593.307214778 RMSE: 690.35737644699	MAE: 387.909558662399 MSE: 434979.956460826 RMSE: 659.53010277077
Random Forest Model	MAE: 85.7316941596728 MSE: 28933.3978576626 RMSE: 170.09820063029	MAE: 64.0631937554805 MSE: 28187.6055778304 RMSE: 167.89164832662

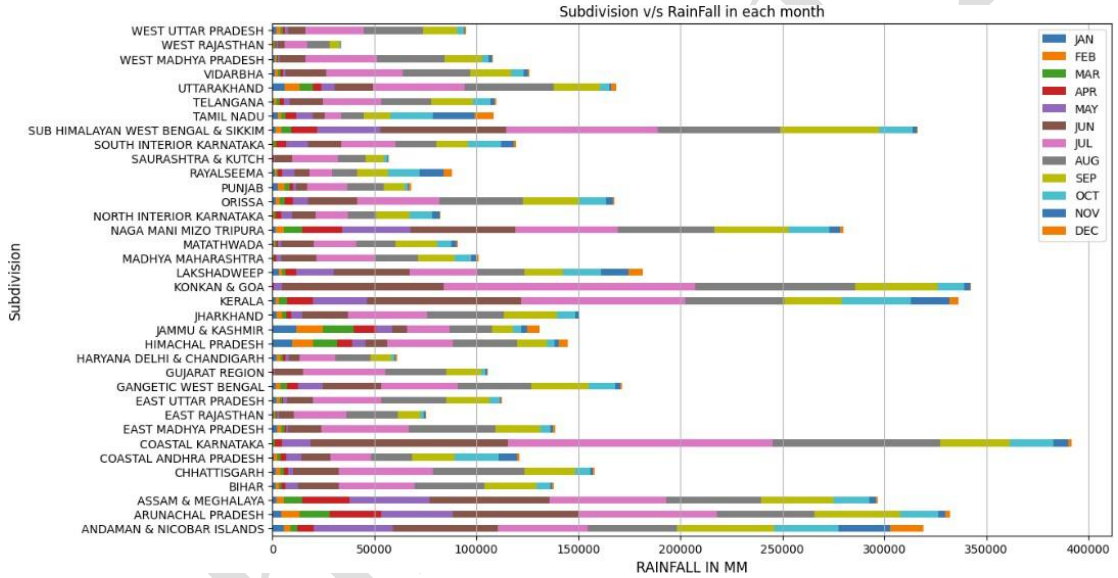
Output Visualisation:

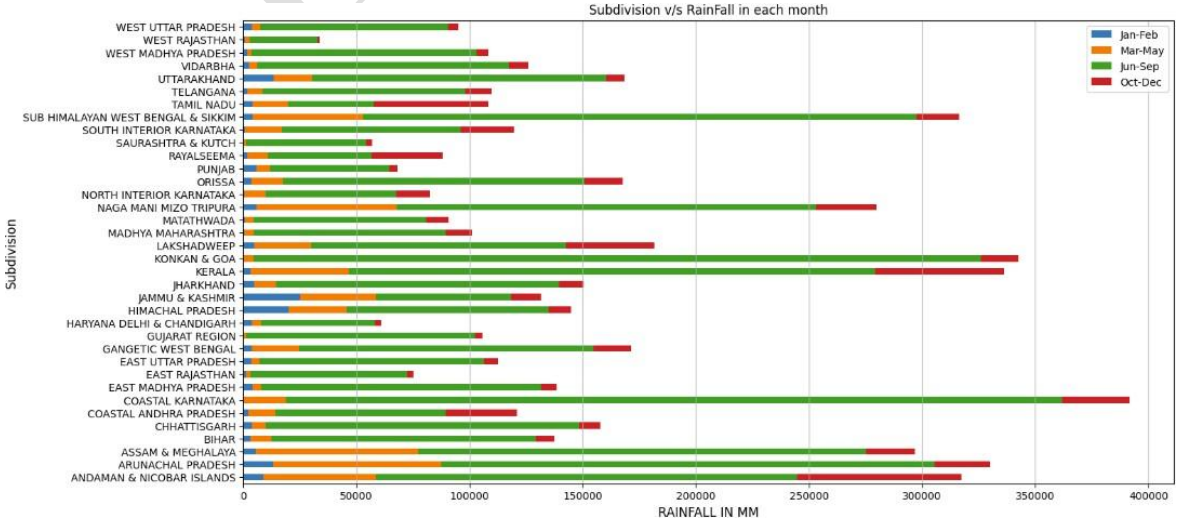


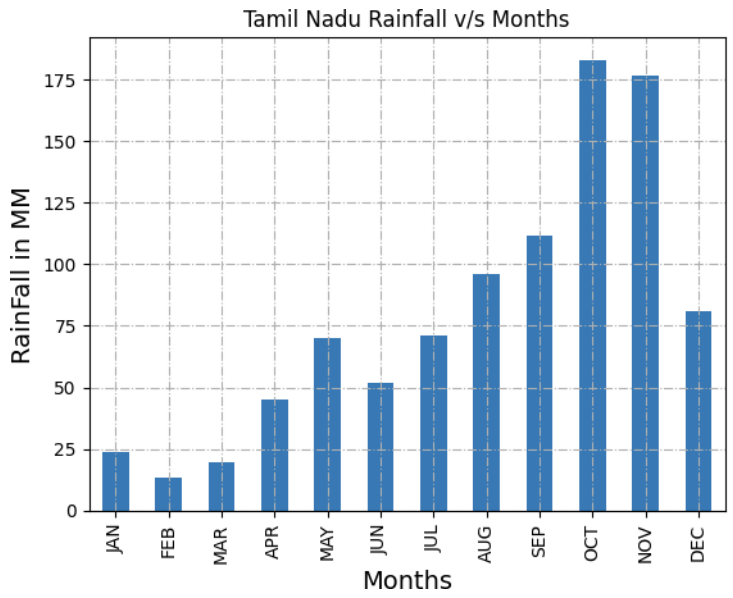


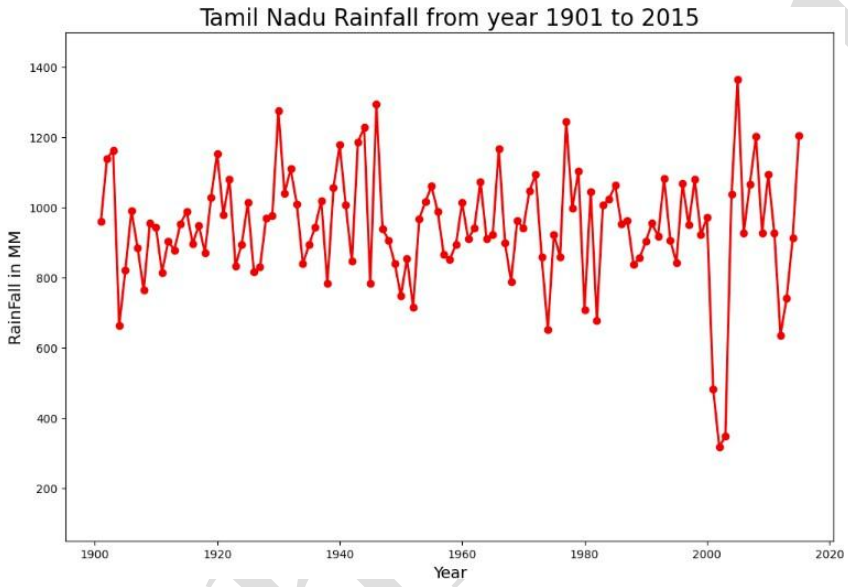


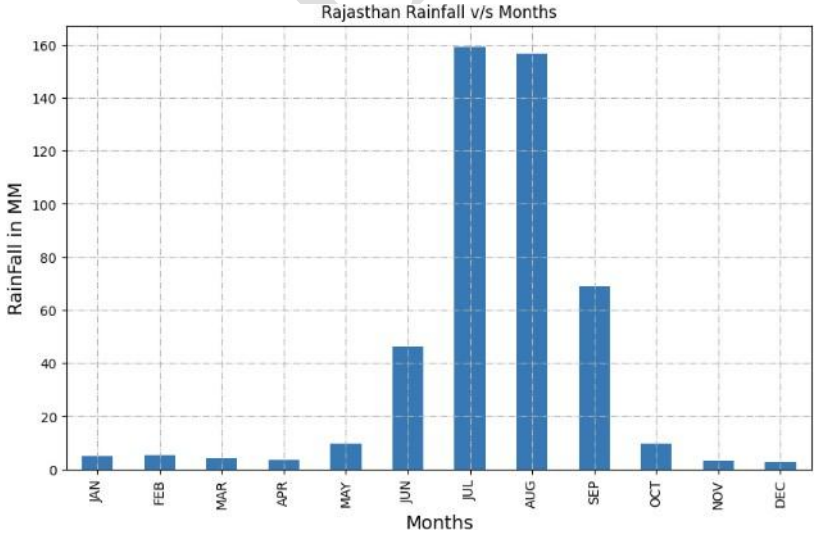

Rainfall in MM
Year
Jan-Feb
Mar-May
Jun-Sep
Oct-Dec

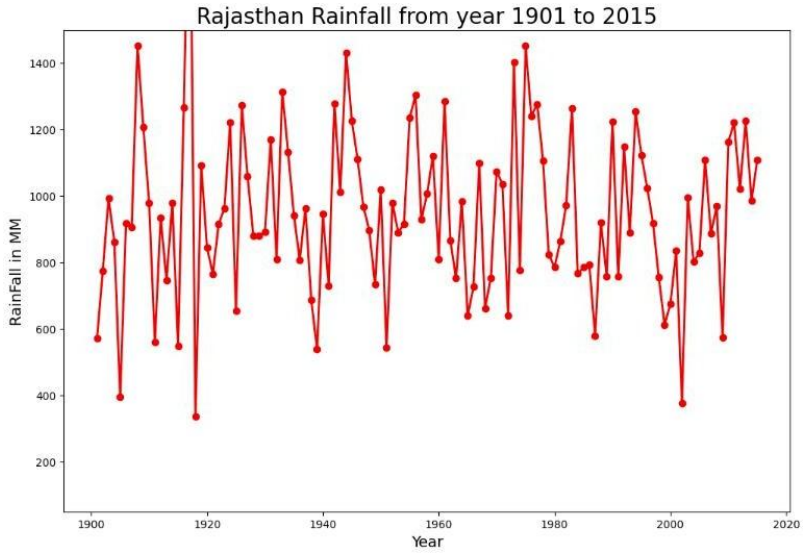

Subdivision v/s RainFall in each month
Subdivision
RAINFALL IN MM
JAN
FEB
MAR
APR
MAY
JUN
JUL
AUG
SEP
OCT
NOV
DEC


Subdivision v/s RainFall in each month
Subdivision
RAINFALL IN MM
Jan-Feb
Mar-May
Jun-Sep
Oct-Dec

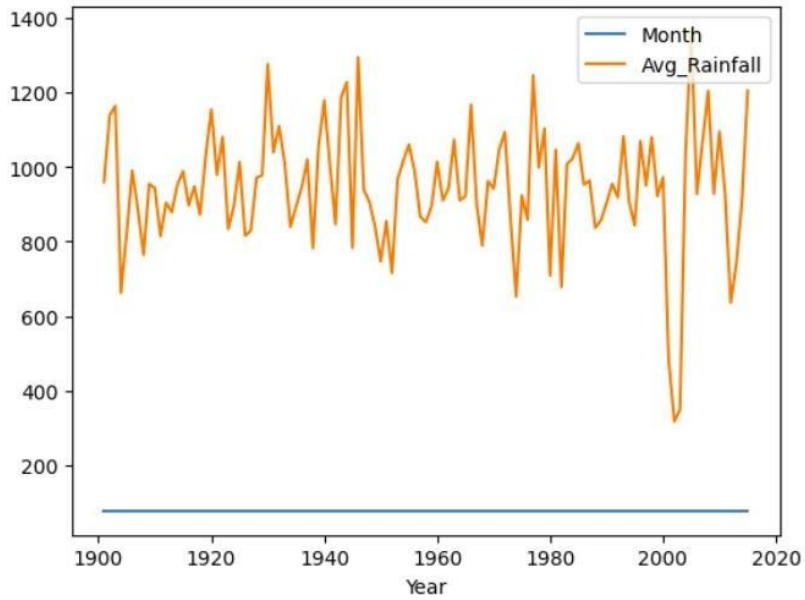

Tamil Nadu Rainfall v/s Months
RainFall in MM
0
25
50
75
100
125
150
175
JAN
FEB
MAR
APR
MAY
JUN
JUL
AUG
SEP
OCT
NOV
DEC
Months


Tamil Nadu Rainfall from year 1901 to 2015
RainFall in MM
200
400
600
800
1000
1200
1400
1900
1920
1940
1960
1980
2000
2020
Year


Rajasthan Rainfall v/s Months
RainFall in MM
0
20
40
60
80
100
120
140
160
JAN
FEB
MAR
APR
MAY
JUN
JUL
AUG
SEP
OCT
NOV
DEC
Months


Rajasthan Rainfall from year 1901 to 2015
RainFall in MM
Year
1400
1200
1000
800
600
400
200
1900
1920
1940
1960
1980
2000
2020


JAN 1 0.46 0.4 0.21 0.13 -0.034 -0.051 0.012 0.024 0.012 0.067 0.22 0.11
FEB 0.46 1 0.58 0.37 0.2 0.034 0.016 0.072 0.08 -0.0046 -0.023 0.13 0.18
MAR 0.4 0.58 1 0.56 0.36 0.17 0.097 0.13 0.18 0.086 0.0088 0.14 0.32
APR 0.21 0.37 0.56 1 0.65 0.46 0.27 0.26 0.38 0.37 0.16 0.13 0.57
MAY 0.13 0.2 0.36 0.65 1 0.57 0.33 0.33 0.49 0.53 0.35 0.25 0.69
JUN -0.034 0.034 0.17 0.46 0.57 1 0.74 0.66 0.55 0.49 0.23 0.088 0.88
JUL -0.051 0.016 0.097 0.27 0.33 0.74 1 0.69 0.51 0.3 0.043 -0.019 0.81
AUG 0.012 0.072 0.13 0.26 0.33 0.66 0.69 1 0.5 0.25 0.017 0.0016 0.75
SEP 0.024 0.08 0.18 0.38 0.49 0.55 0.51 0.5 1 0.38 0.15 0.11 0.71
OCT 0.012 -0.0046 0.086 0.37 0.53 0.49 0.3 0.25 0.38 1 0.48 0.28 0.58
NOV 0.067 -0.023 0.0088 0.16 0.35 0.23 0.043 0.017 0.15 0.48 1 0.45 0.31
DEC 0.22 0.13 0.14 0.13 0.25 0.088 -0.019 0.0016 0.11 0.28 0.45 1 0.21
ANNUAL 0.11 0.18 0.32 0.57 0.69 0.88 0.81 0.75 0.71 0.58 0.31 0.21 1
JAN FEB MAR APR MAY JUN JUL AUG SEP OCT NOV DEC ANNUAL
1.0
0.8
0.6
0.4
0.2
0.0


Month
Avg_Rainfall
1400
1200
1000
800
600
400
200
1900
1920
1940
1960
1980
2000
2020
Year

5. CHAPTER 5 –CONCLUSION

5.1 CONCLUSION

Rainfall prediction is a beneficial but challenging task. Data mining techniques have the ability to predict the rainfall by extracting and using the hidden knowledge from past weather data. In the last decade, many researchers have worked to increase the accuracy of rainfall prediction by optimizing and integrating data mining techniques. Various models and techniques are available today for effective rainfall prediction but still there was a lack of a compact literature review and systematic mapping study which could reflect the current problems, proposed solutions and the latest trends in this domain. The performance was evaluated by comparing the predicted results with the observed (actual) measures. Information retrieval metrics and statistical techniques were used for performance analysis of proposed techniques in comparison with other methods from the literature.

5.2 FUTURE WORK

Improvement in Prediction Accuracy:

- Use ensemble learning techniques (e.g., Random Forest, Gradient Boosting).
- Implement deep learning models like LSTM.

Incorporation of Additional Parameters:

- Include weather factors such as temperature, humidity, and pressure.
- Analyse the impact of climate change.

Extension to Other Geographical Regions:

- Apply the model to other countries with similar climates.
- Create a global rainfall prediction system.

Real-Time Prediction System:

- Implement real-time rainfall prediction using IoT sensors and cloud computing.

- Deploy the system for public access through a web or mobile application.

REFERENCES:

- Chalachew Muluken Liyew,Haileyesus Amsaya Melese, “Machine learning techniques to predict daily rainfall amount”.
- Wanie M. Ridwan, Michelle Sapitang, Awatif Aziz, Khairul Faizal Kushiar, Ali Najah Ahmed, Ahmed El-Shafie, “Rainfall forecasting model using machine learning methods: Case study Terengganu, Malaysia”.
- Aakash Parmar, Kinjal Mistree, Mithila Sompura, “Machine Learning Techniques for Rainfall Prediction”.
- Rahman A-u, Abbas S, Gollapalli M, Ahmed R, Aftab S, Ahmad M, Khan MA, Mosavi A, “Rainfall prediction system using ML Fusion for smart cities”.
- Jinle Kang, Huimin Wang, Feifei Yuan, Zhiqiang Wang, Jing Huang and Tian Qiu, “Prediction of Precipitation Based on Recurrent Neural Networks”.
- Shabbir Ahmed Osmani A, Foysol Mahmud B, and Md.Abu Zafor C, “Prediction of Rainfall using Machine Learning Algorithms for Different Districts of Meghalaya”.
- Veera Ankalu. Vuyyuru; Giduturi. Apparao; S. Anuradha, “Rainfall Prediction Using Machine Learning Based Ensemble Model”.
- CMAK Zeelan Basha, Nagulla Bhavana, Ponduru Bhavya, Sowmya V, CMAK Zeelan Basha, “Rainfall Prediction Using Machine Learning & Deep Learning Techniques”.