Journal of
Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.98549438

RAINFALL PREDICTION USING MODERN DEEP LEARNING ALGORITHMS WITH TIME-SERIES DATA

E Valli 1, S. Nirmala Sugirtha Rajini 2
1. Research Scholar, E-Mail: vallielumalai2795@gmail.com
2. Professor, E-Mail:nirmalasugirtharajini.mca@drmgrdu.ac.in
Department of Computer Applications, Dr.MGR. Educational and Research Institute, Maduravoyal, Chennai-95, India

Abstract:- Rainfall forecasting has earned significant studies importance in popular years due to its complexities and rainfall prediction is currently more challenging than ever due to significant changes in the climate. By extracting hidden patterns from previous weather observations, machine learning techniques can predict rainfall.Previous models implement complex statistical models that are often too expensive, both numerically and financially, or inappropriate for downstream applications. In this study of predicting future hourly rainfall high volume using time-series data, models based on Stacked LSTM, XGBoost, and an ensemble, Radial Basis Function Networks (RBFNs), were compared. Climate datasets rom five major cities in the UK cities were used from 2000 to 2020. The models' effectiveness was evaluated using the accuracy, Precision, recall, analysis metrics Loss, the root mean squared error, R-square values,Mean Absolute Error, and Root Mean Squared Logarithmic Anomaly. The predicted resultsshow tetter performance comparing Radial Basis Function Networks (RBFNs) and Stacked-LSTM Networks as rainfall prediction models. The conclusion suggests that models for based on Radial Basis Function Networks (RBFNs) with fewer concealed layers, our approach performs better, highlighting it's appropriate for budget-wise rainfall forecasting applications.

Keywords: - *Modern Deep Learning, Accuracy, Precision, Recall, RBFNs, LSTM, MAE, RMSE, RAE, Pytorch, SARIMA.*

I INTRODUCTION

Rainfall continues to be one of the weather factors that has the biggest impact on many aspects of our daily lives. The effects of the rainy season on the socioeconomic system are significant, and they range from disruptions in transportation in terms of architecturaldamage in the occurrence of a flood. As a result, numerous studies have looked into and recommended temperature and precipitation forecasting methods as a mitigation strategy in advance of any disaster or eventuality. However, these strategies must be effective and timed to improve human mobility, agriculture, and industrial development.

As a result, this study compares three LSTM-Network architecture variants to current machine learning techniques in order to determine which is more appropriate for the reason of forecasting the amount of rain per hour. This study analyses the efficacy of models based on LSTM, Stacked-LSTM, and Bidirectional-LSTM Networks to that of an XGBoost decision tree model and a planned model that would be constructed with an automatic machine learning (AutoML) tool Barrera-Anima,et al.,(2022). [1] is a strategy that provides researchers with a tool for autonomous and data-driven finding the best algorithms and hyperparameter configurations for a certain application. Rainfall forecasting has been used for a long time to analyze the link between rainfall, geographical coordinates (such as latitude and longitude), and other weather-related factors (such as pressure, temperature, wind speed, and humidity).

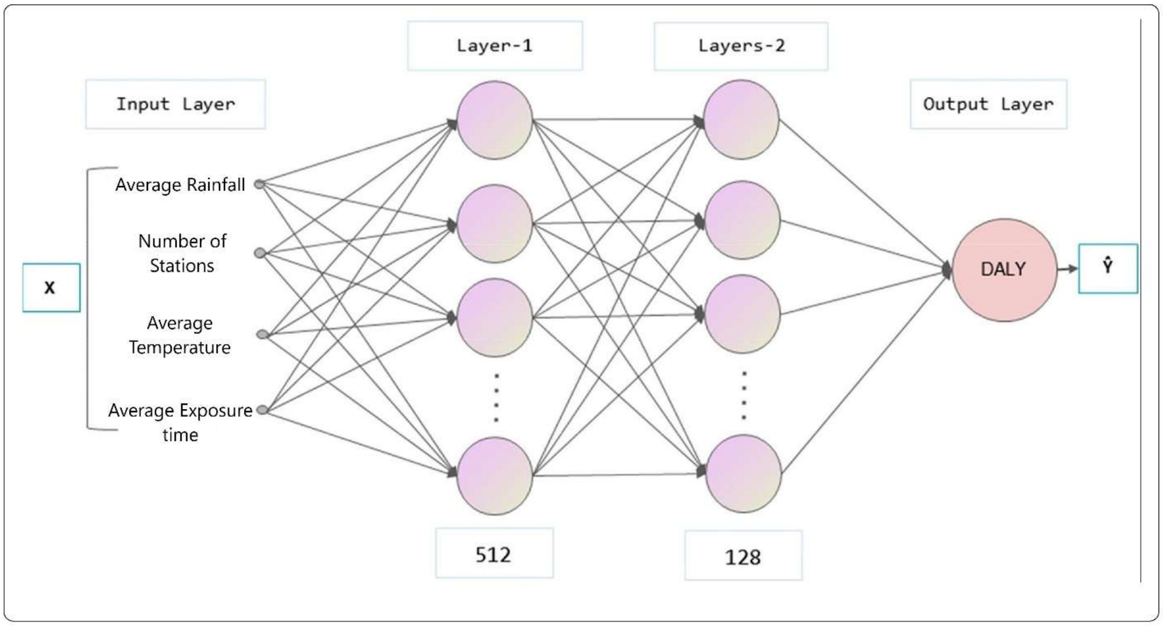

Fig 1: Two-layer RBFN'S Model Structure for Rainfall Predictions

Utilizing data from five different time periods in significant UK cities, a model based on bidirectional Long short-term memory networks was suggested for the task of hourly rainfall forecasting. In order to forecast the amount of rainfall per hour using time-series data from five significant UK cities, models based on LSTM Networks, Stacked-LSTM Networks, Bidirectional-LSTM, Radial Basis Function Networks (RBFNs) Networks, XGBoost, and the final model from Automated ML were tested. However, forecasting precipitation is a trickier task because regional rainfalls vary in space and time. A. Barrera-Animas. Y. Oyedele,et.al.,(2022).

II LITERATURE SURVEY

One of the most significant meteorological variables continues to be rainfall. In a number of areas of our daily lives.This investigation evaluates the performance of models based on LSTM, Stacked-LSTM, and Bidirectional-LSTM Networks to that of an XGBoost decision tree model and a planned model that will be built using an Automated Machine Learning (AutoML) program Barrera-Animas et al.,2022 [1] a methodgives heresearchers a tool to automatically and data-driven search for an algorithm and its hyperparameter values that are best suited for a specific application.Imteaz, M. A., & Hossain, I. (2022) [2] The largest SI value is expected to change in the future, according to this paper. Regarding anticipated water savings from rainwater tanks, more details are provided in the paper. Climate change simulation uses ETank, a daily water balance model that was previously created.

In order to recreate the seasonal rainfall throughout time patterns in Western Australia, this paper compares the effectiveness of linear and non-linear modeling techniques. The linear and non-linear models were developed using commonly used MLR (multiple linear regression) and artificial neural network (ANN) modeling techniques. TheLavenberg-Marquardt algorithm was used in conjunction with the Multilayer Perceptron training rule to build the non-linear ANN models. Iqbal & Rasel,et.al.,(2020).

In this paper, load forecasting is explored using stacked bidirectional long short-term memory (SB-LSTM) recurrent neural networks, a cutting-edge approach for regression analysis of time-series data inside a deep learning framework. The research is carried out on the basis of an instance investigation of Scottish residential demand, for which five-year datasets containing load and weather data recorders are available. The supplied results and analyses allow one to evaluate the precision with which the SB-LSTM approach can make predictions for both day-ahead and week-ahead load forecasting while considering meteorological data meteorological data into account. Mingzhe and colleagues,2019[4]

The expansion of high-performance computing, the development of ensemble mathematical climate and weather prediction, the growing interest in shifting away from deterministic decision-making and towards risk-based decision-making that takes forecast uncertainty into account, and the efforts of communities such as the (HEPEX), which focuses on advancing relevant combination forecasting, all have contributed to the significant momentum th Rebecca Emerton et al., 2020[5]

Meteorological data, ensemble theories, classification, and regression models including Neural Networks, support vector machine (SVM), decision trees, Nave Bayes, J48, CART, and ID3 are the primary topics discussed in this study. However, the main thrust area in geographical data sciences has been the deployment of several neural networks models such as BPNN, FFNN, GWLM-NARX, RNN, and TDNN. Sheikh Amir Fayaz and others [6]. The proposed GWLR uses the GWO to obtain the best possible regression coefficients, which it then applies to forecast rainfall based on time-series meteorological data input. So, from 1901 to 2015, the state of Jammu and Kashmir and India's accumulated data are used to detect rainfall. In comparison to other existing methods, the proposed GWLR is found to be the most effective with the least MSE value of 0.005 and the highest PRD value of 1.700 percent. Razeef Mohd et al.,2022[7]

III PROPOSED METHODOLOGY

The objective of this investigation is to predict the rainfall collection of volume data using time-series data from some more and few major UK cities. The Radial Basis Functional Networks (RBFNs) created with the PyTorch Tool and the on Radial Basis Function Networks (RBFNs) were found to outperform all other models tested. This indicates that models based on Radial Basis Function Networks (RBFNs) with fewer hidden layers perform better for this approach.In particular, this study aims to compare forecast models based on Stacked-LSTM

Networks, XGBoost, andRadial Basis Function Networks (RBFNs) an AutoML-based algorithms

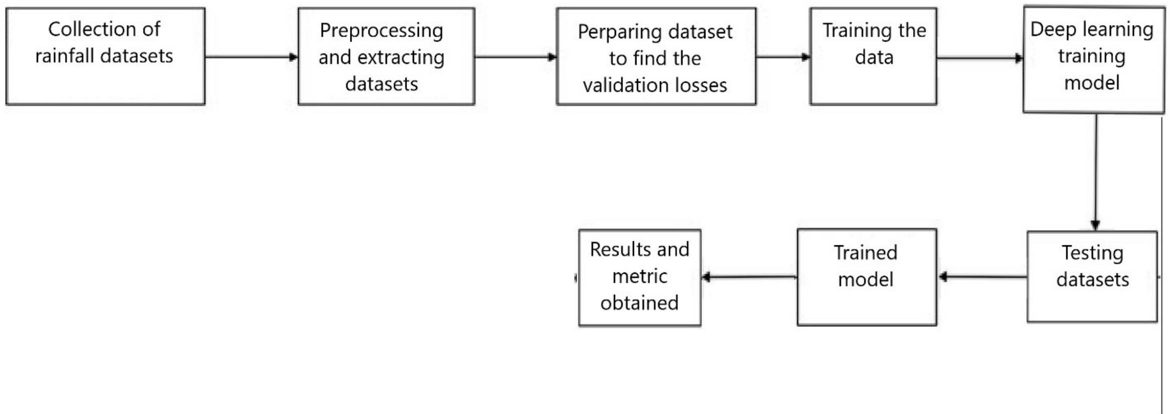

Fig 2: Block Diagram for Testing, Training And Validation Phases

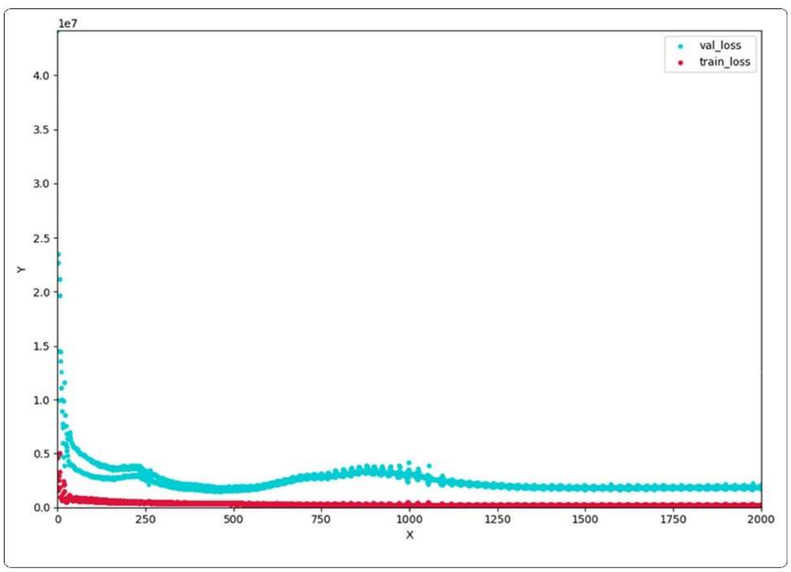

Fig 3: Loss function of RBFN's model

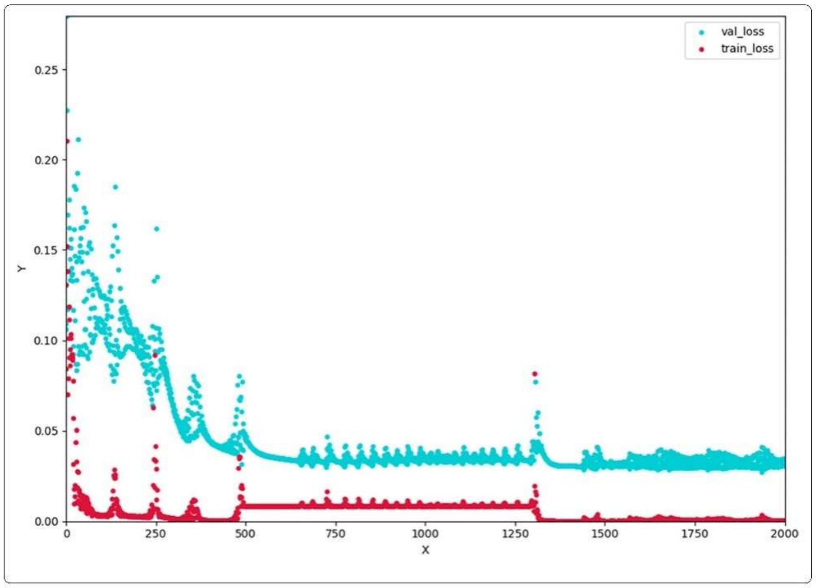

Fig 4: Loss function of LSTM model

Stacked-LSTM models:

In a stacked LSTM model, multiple layers of LSTM cells are stacked on top of each other. Each layer takes the output of the previous layer as input, allowing the network to learn increasingly complex representations of the input data.Each layer's LSTM cells use gating mechanisms to selectively recall or forget information, making them very good at modeling long-term dependencies in sequential data. Overall, stacked LSTM models are an effective tool for modeling sequential data and have generated modern results in a variety of applications.

XGBoost model:

XGBoost (eXtreme Gradient Boosting) is a popular gradient boosting framework that is used for building machine learning models.It is meant to be highly scalable and efficient, and it is able to handle classification and regression problems. XGBoost works by adding new models to an ensemble iteratively, with each model attempting to correct the errors of the previous model. It engages gradient descent optimization to determine the best model parameters and incorporates normalization techniques to avoid overfitting.XGBoost also includes features such as missing value imputation and built-in cross-validation to improve model accuracy.It has been successfully used in a variety of uses, including image classification, natural language processing, and financial forecasting. XGBoost is widely regarded as one of today's most powerful and flexible deep learning frameworks.

Radial Basis Function Networks (RBFNs)

Radial Basis Function Networks (RBFNs) are a type of neural network used to predict a variety of activities. The proposed model incorporates an adaptive learning algorithm and a feature selection procedure into a modified RBFN architecture. The proposed model's efficacy is assessed in the thesis using several benchmark datasets from various domains, including finance, energy, and free health care. The results show that the proposed model outperforms a number of other cutting-edge prediction models, which include traditional RBFNs and other deep learning models like LSTM and GRU.

$$\boldsymbol{f(x) = \sum i = 1\ to\ Mw_i * \Psi(||x - c_i) + b} \qquad (1)$$

where:

- f(x) is the network's output for an input data x
- w_i are the internet backbone weights.
- $c_i$are the centers of the radial basis functions.
- ||**x** - c_i|| is the distance between the input **x** and the center **c_i**.
- **φ()** is the radial basis function.
- **b** is the biased term

The radial basis function **φ()** is typically a Gaussian function, which is given by:

φ(r) = exp(-(r/σ)^2) (2)

where **r** is the distance between the input **x** and the center **c_i**, and **σ** is a parameter that controls the width of the Gaussian function.RBFN model consists of an input layer, a hidden layer with radial basis functions, and an output layer.

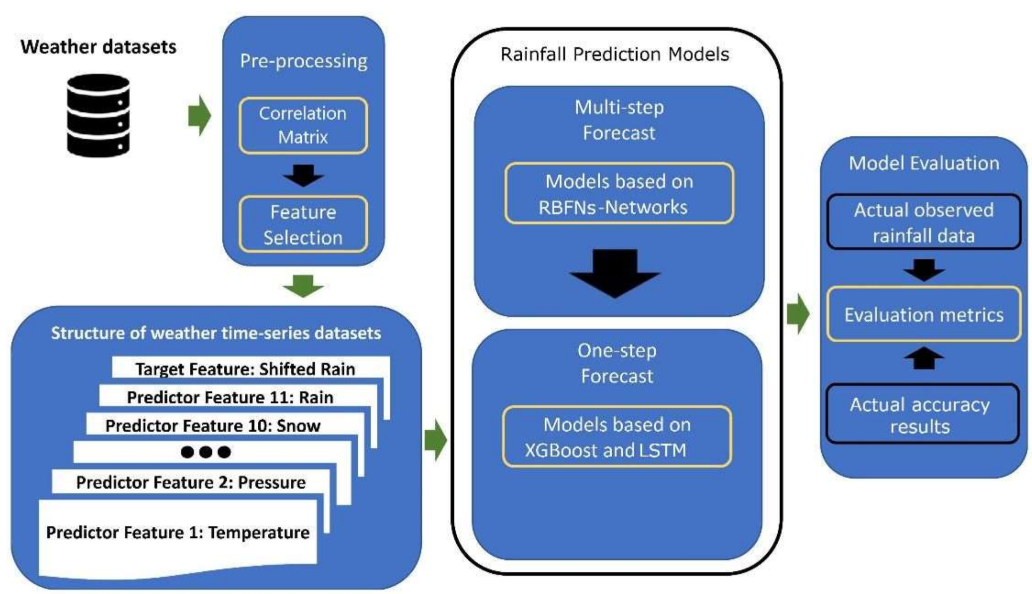

Fig 5: Model Evaluation Methodology

IV RESULT AND DISCUSSION

The outcomes Figure:7 shows that three comparison models, LSTM, XGBoost, and RBNF, achieved exactly equivalent Loss values for all city evaluation metrics. Accuracy, recall, and precision are the five cities' values. According to the performance indicators, the performance of RBFNs is superior to that of LSTM and XGBoost. Notwithstanding, the training and validation loss curves should be evaluated.Figures 3 and 4 are the models with the greatest performance in the evaluation metrics and the best for training and validating loss curves of all models tested in the hyperparameters grid search, similar to the approach that was used present the results of the previous figure 1.

The collection of datasets from UK cities contains 8 columns and 37k row. This dataset contains measurements from different stations.For this research work data from the Kaggle dataset: https://www.kaggle.com/josephw20/uk-met-officeweather-data, it contains measurements of various weather metrics from differentpoints in the UK. The dataset is located here:ch3/uk_rainfall_prediction/data/MET_Office_Weather_Data.csv.
Accuracy Analysis:

$$Accuracy = \frac{True\ Positive + True\ Negative}{True\ Positive + True\ Negative + False\ Positive + Flase\ Negative} - (3)$$

The following table 1 shows the accuracy value of the three different models LSTM, RBNF's, and XGBoost based on the number of iterations.

Table 1: Accuracy Comparison of LSTM, XGBoost and RBNF's models

No of Iterations	**Stacked LSTM Accuracy %**	**XGBoost Model Accuracy %**	**RBNF'sModel Accuracy %**
10	85	88.5	89
20	86.4	89.2	89.5
30	87	89.7	90
40	87.2	90	90.7
50	87.6	90.2	91.8

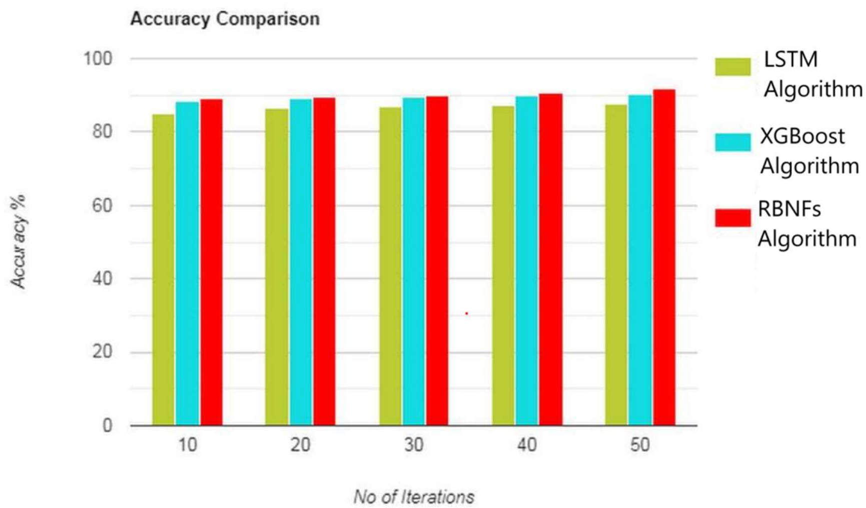

Fig 6: Accuracy Comparison graph of LSTM, XGBoost and RBNF's

Table 2: Hyperparameter values and performance of RBFN's top 5 UK Cities

Model Metrics	**Bristol**	**Bradford**	**Cambridge**	**Lincoln**	**Bath**
No ofestimators	28	30	27	32	30
Subsample ratio of features	0.5	0.7	0.5	0.5	0.9
Subsample rarioof training	0.8	0.5	0.6	1.0	1.5
RMSE	0.29	31.2	0.28	0.33	0.32
MAE	0.04	0.07	0.04	0.05	0.7
RAE	11.0	12.0	14.0	16.0	17.0

Model Comparison:

To contrast and assess the fitting and prediction accuracy of the three models, three performance metrics were used: root mean square error (RMSE), mean absolute error (MAE), and mean of the absolute value (RAE). The better the prediction effect, the lower the values of the three measures. The average prediction error is established by the simplest measure of fitting and prediction accuracy, MAE. By increasing the prediction error, the root mean square error becomes especially sensitive to infrequent errors and can better reflect the accuracy of the prediction results. The following are the specific calculation formulas [4][5][6].

$$\boldsymbol{MAE = \left(\frac{1}{n}\right)\Sigma(i = 1\ to\ n)|y_i - \hat{y}_i} \qquad (4)$$

$$\boldsymbol{RMSE = \sqrt{\textstyle\sum_{i=1}^{N} ||y(i) - \hat{y}(i)||^{2}}} \qquad (5)$$

$$\boldsymbol{RAE = \frac{\sum_{j=1}^{N}|e_j|}{\sum_{j=1}^{N}|a_j - \bar{a}|}} \qquad (6)$$

Fig:7 (a)Comparison results of the RBFN's model; (b) Comparison results of the LSTM model;(c)Comparison results of the XGBoost model; (d) Comparison results of the RMSE in three models; (e) Comparison results of the MAE in threemodels; (f)Comparison results of the RAE in three modelsthree models

V CONCLUSION

The objective of this report was to make a comparison of rainfall forecasting models based on Radial Basis Function Networks (RBFNs) architectures with modern Machine Learning algorithms. To achieve this objective, multiple Stacked-LSTM models,XGBoost (baseline model) and one Radial Basis Function Networks (RBFNs) model were compared to an ensemble model yielded by an Automated Machine Learning approach.To evaluate the performance of the implemented rainfall forecasting models, historical weather data from the UK cities of a non-exhaustive hyperparameters grid search have been used to obtain the best achievement of models-based Radial Basis Function Networks (RBFNs).

REFERENCES:

[1] Barrera-Animas, A. Y., Oyedele, L. O., Bilal, M., Akinosho, T. D., Delgado, J. M. D., & Akanbi, L. A. (2022). Rainfall prediction: A comparative analysis of modern machine learning algorithms for time-series forecasting. Machine Learning with Applications, 7(August 2021), 100204. https://doi.org/10.1016/j.mlwa.2021.100204

[2] Imteaz, M. A., & Hossain, I. (2022). Climate Change Impacts on 'Seasonality Index' and its Potential Implications on Rainwater Savings. Water Resources Management, 0123456789. https://doi.org/10.1007/s11269-022-03320-z

[3] Hossain, Iqbal & Rasel, H. M. & Imteaz, Monzur & Mekanik, Fatemeh. (2020). Long-term seasonal rainfall forecasting using linear and non-linear modeling approaches: a case study for Western Australia. Meteorology and Atmospheric Physics. 132. 10.1007/s00703-019-00679-4.

[4] Zou, Mingzhe & Fang, Duo & Harrison, Gareth & Djokic, Sasa. (2019). Weather-Based Day-Ahead and Week-Ahead Load Forecasting using Deep Recurrent Neural Network. 341-346. 10.1109/RTSI.2019.8895580.

[5] Rebecca Emerton., Qingyun Duan ., Andrew W. Wood, Fredrik&Wetterhall, David E.Robertson(2021):Ensemble Flood Forecasting: Current Status and Future Opportunities https://doi.org/10.1002/wat2.1432.

[6] Sheikh Amir Fayaz, Majid Zaman, Muheet Ahmed Butt(2022):Knowledge Discovery in GeographicalSciences—A Systematic Survey of Various Machine Learning Algorithms for Rainfall Predictionhttps://www.researchgate.net/publication/354289092.

[7] Qiao, S., Chen, D., Wang, B., Cheung, H. N., Liu, F., Cheng, J., Tang, S., Zhang, Z., Feng, G., & Dong, W. (2021). The Longest 2020 Meiyu Season Over the Past 60 Years: Subseasonal Perspective and Its Predictions. Geophysical Research Letters, 48(9), 1–11. https://doi.org/10.1029/2021GL093596.

[8] Lou, H. R., Wang, X., Gao, Y., & Zeng, Q. (2022). Comparison of ARIMA model, DNNmodel and LSTM model in predicting disease burden of occupational pneumoconiosis in Tianjin, China. BMC Public Health, 22(1), 1–15. https://doi.org/10.1186/s12889-022-14642-3

[9] Qi XM, Luo Y, Song MY, Liu Y, Shu T, Liu Y, et al. Pneumoconiosis: current status and future prospects. Chin Med J. 2021;134(8):898–907.

[10] Lou, H. R., Wang, X., Gao, Y., & Zeng, Q. (2022). Comparison of ARIMA model, DNN model and LSTM model in predicting disease burden of occupational pneumoconiosis in Tianjin, China. BMC Public Health, 22(1), 1–15. https://doi.org/10.1186/s12889-022-14642-3.