# Prediction Of Indian Rainfall Events Using Satellite data


Mrs. Shweta Thakur
*Apex Institute of Technology*
*Chandigarh University*
Mohali, Punjab, India
e12791@cuchd.in

Abhishek Rana
*Apex Institute of Technology*
*CSE(Hons)- Big Data Analytics*
*Chandigarh University*
Mohali, Punjab, India
7009807156
abhirana1402@gmail.com

Sumit Chaudhary
*Apex Institute of Technology*
*CSE(Hons)- Big Data Analytics*
*Chandigarh University*
Mohali, Punjab, India
7006061015
sumitcoudhary94191@gmail.com

Ashutosh Kumar Singh
*Apex Institute of Technology*
*CSE(Hons)- Big Data Analytics*
*Chandigarh University*
Mohali, Punjab, India
89651680582
ashutoshkumarsingh8968@gmail.com



***Abstract*— The frequency and intensity of heavy rain events have increased globally, posing significant challenges to various sectors including agriculture, infrastructure, and disaster management. Timely prediction of these events is crucial for effective preparedness and mitigation strategies. Satellite data has emerged as a valuable tool for monitoring and predicting weather patterns, including heavy rain events. This study focuses on the development of predictive models utilizing satellite data to forecast heavy/high impact rain events. By using traditional models such as linear regression, SVM with comparison with more advance techniques like RNN, we aim to verify whether Neural Networks perform better than the traditional approaches. By comparing different models, we identify RNN to performed better than other models with MAE (Mean Absolute Error) of 85.50177457045439 which is significantly less than other traditional approaches, hence is capable of providing better insight for rainfall prediction and better accuracy, helping in high impact rainfall preparedness, agriculture, and disaster management.**



***Keywords****—Heavy rain events, High impact, Satellite data, Remote sensing, Feature selection, Environmental variables, Early warning systems, Rainfall Prediction , disaster Management*


## I. Introduction

In Recent times, the frequency and intensity of heavy rain events have increased in recent decades, this rise has been partly related to climate variability and change glaciers melting. These occurrences, which are distinguished by high rates of precipitation over brief periods of time, present serious threats to a number of industries, including infrastructure, human settlements, and agriculture.[2] Predicting rain occurrences in a timely and precise manner is essential to successful disaster response, preparation, and mitigation.

Conventional weather forecast techniques mostly rely on numerical weather models and ground-based observations, and statical approach both of which from past observation have displayed frequently having problems in correctly describing localised episodes of intense rain.[1] In contrast, satellite data, which offers regular observations and a broad spatial coverage, has proven to be an invaluable tool for tracking and predicting weather trends, helping therefore in building the disaster prevention system formed by the responsible bodies of different states.
Recent advances in remote sensing technologies and machine learning Algorithm have made it possible to anticipate heavy rain episodes using satellite data.

Estimates of precipitation acquired from satellites, using sensors like radar and infrared devices, provide important information on the location and amount of rainfall, which Is then used to train the model to provide possible forecasting of weather especially on heavy rain fall events. The goal of this ML project is to create various prediction models for predicting rain events with high potential impacts by utilising satellite data and machine learning approaches. By training various models, we aim to compare and contrast the models that are best suited for namely, Country wide, state wise and district wise rainfall analysis.

This research is important because it can improve tactics for catastrophe risk reduction and early warning systems. Decision-makers may more effectively allocate resources, put preparatory measures in place, and lessen the negative effects on vulnerable populations and vital infrastructure by receiving early and accurate forecasts of heavy rain occurrences. While providing the accuracy for the weather to plan prevention accordingly in case of hazardous rain events.

The models will not only provide independent results, but can also be evaluated by forming ensemble to neutralize the drawbacks of each approach used to by independent and, further increase the rate of accuracy of heavy rainfall events.

## II. EXISTING SYSTEM

While the currently used prediction model suffice the basic need for predicting the weather on normal days while occasionally providing wrong prediction. There is a need for better prediction models which can predict the heavy rainfall events at high accuracy.

1. *Genetic Algorithm* is a good approach for localized weather prediction in India, while providing us with automated feature selection. It is to be noted that Genetic Algorithm and comparatively slow and can struggle to provide accurate weather alerts before time resulting in delayed counter measure to tackle the heavy rainfall.[9]

2. *Using ML and AI* to assist can be a good alternative to this problem statement, but without proper data handling and cleaning the models and become bias towards certain feature and inert towards other which and throw off the accuracy of the model even if minor discrepancy occur in recorded and training data.[10]

3. *LSTM (Long Term Short Term) Networks:* LSTM networks being an alternate version of recurrent neural network (RNN) paradigm provides us with capability of learning long-term features. They have been applied to rainfall prediction by analysing historical rainfall data to forecast future precipitation patterns.

4. *CNNs (Convolutional Neural Networks)*: CNNs are used for feature extraction from rainfall-related dataset based on satellite imagery or radar data. These System effectively capture spatial patterns and correlations in the data, which can be useful for predicting rainfall.

5. *Random Forest:* Random forest is used as an ensemble learning technique that constructs a huge number of DT (decision trees) during training phase and produces the mean prediction for numeric data (regression) or mode of the classes (classification) for each individual tree. It has been used to forecast rainfall by utilizing characteristics taken from past meteorological records.

6. *Deep Belief Networks (DBNs):* Is made up of several different layers of stochastic latent variables, DBNs are probabilistic generative models. In order to model intricate data connections and provide forecasts based on past rainfall data, they have been used in rainfall prediction jobs.

ARIMA (The Auto-regressive Integrated Moving Average) is a statistical technique that is frequently employed for time series forecasting, which includes rainfall prediction, even though it is not exactly related to machine learning. It simulates the connections between a single observation and several lag observations.

## III. PROPOSED SYSTEM

- *Deep Learning Techniques*: Deep neural networks, termed to be another type of artificial neural network consisting of several layers, this technique has shown promise in capturing complex relationships within weather data. They can potentially be deployed to learn intricate patterns from vast datasets and improve prediction accuracy.

- *Hybrid Approaches*: This approach combines traditional statistical methods like ARIMA with artificial learning models. This can construct to take advantage of the strengths of both approaches – statistical models for capturing trends and machine learning for identifying non-linear relationships.

Real-time Data Integration: Including real-time data from radars and satellites alongside historical data can provide a more dynamic picture of weather patterns and potentially lead to more accurate short-term forecasts.

It's important to note that these proposed systems are constantly evolving as researchers explore new algorithms and techniques. The goal is to develop more accurate, efficient, and reliable rainfall prediction systems for various applications.

- *Recurrent neural networks (RNNs)* are a type of machine learning algorithm that can be used to predict rainfall because they can capture effectively the temporal contextual information along time series data. RNNs are able to learn previous-term dependencies, which are essential for predicting weather patterns. RNNs are also flexible and can exploit time-series data, which has made them widely used in research areas such as weather forecast, machine translation, sentiment analysis, and speech recognition.[12]

One study uses RNNs to build highly accurate localized weather predictions model, which are suitable for downstream BPS applications such as energy simulations. Another study uses a hybrid VMD-RNN model to provide a highly reliable prediction, which on analysis provides better performance in predicting high and low values in contrast to the pure LSTM model without any decomposition.

- *Support Vector Machines (SVMs) for Rainfall Prediction*

Core Idea: Unlike decision trees that create a series of branching rules, SVMs focus on identifying an best hyperplane in a high-dimensional feature spatial area that best differentiates the data points that represents rainy and non-rainy days.

Feature Mapping: Similar to decision trees, SVMs use historical weather data as features (temperature, humidity, etc.). However, SVMs can also perform non-linear transformations of these features, essentially mapping them into a higher dimensional space. This enables the SVM to record more complex relationships between features that might not be apparent in the original data space.[11]

Finding the Optimal Hyperplane: The SVM algorithm identifies the hyperplane that maximizes the margin (distance from closet point vector) between the two classes (rainy & non-rainy) in high-dimensional space. The margin refers to the space between the hyperplane and the nearest data points of each class, called support vectors.

Classification and Prediction: Noval unseen data points are plotted on the same high-dimensional space and classified on the basis on which side of the hyperplane they lie on. This allows the SVM to predict whether a new data point represents a rainy or non-rainy day.

## IV. METHODOLOGY

Developing a robust rainfall prediction system requires a structured methodology encompassing data preparation, model selection, training, evaluation, and validation. Initially, historical weather data, including rainfall measurements, is collected from reliable sources. Subsequently, the data undergoes preprocessing steps, such as cleaning to handle missing -values, wrongly formatted data, outliers and feature extraction to obtain relevant features like temporal patterns and meteorological variables.

For model selection, three approaches are considered: *Linear Regression (Elastic Net), SVM (Support Vector Machine) and RNN (Recurrent Neural Network)*. Each model is trained using the training dataset.

### 4.1 *Recurrent Neural Network (RNN)*

Recurrent Neural Networks (RNNs) represent a class of artificial neural networks particularly well-suited for sequential data processing and time-series analysis. Unlike traditional feedforward neural networks, which process input data in a fixed, one-directional manner, RNNs possess connections that form closed loop feedback cycles, enhancing the capability for them to exhibit dynamic temporal behaviour.[13]

Incorporating suitable RNN for rainfall prediction, along with appropriate layer configurations and activation functions to capture temporal dependencies effectively. The model is trained using the training dataset retrieved by satellite sensing, with parameters initialized, loss functions defined (typically Mean Squared Error), and optimization algorithms selected (e.g., Adam or RMSprop).

### 4.2 *Support Vector Machine (SVM)*

Support Vector Machine being a supervised machine learning model will provide us with the control model, as being used mostly in current model prediction, while ensembled with other standard statistical approach.

Using SVM our goal is to compare and contrast the different approaches used by us, and infer whether out Neural Net model perform better than the traditional Machine Learning Model.

SVM's sub model SVR can be used in rainfall prediction for providing a metrics to compare the traditional approach with Neural Network approach, and using MSE (Mean Squared Error) for formulating the accuracy of contrasting models for better comparison among the three selected models.

### 4.3 *Linear Regression*

Linear Regression is supervised machine learning approach to find the co-relation between dependant and independent variable. In case of rainfall prediction, we are using elastic net (linear regression's ensembled approach).

The ensembled model contains two regression approaches Lasso Regularization (L1 Regularization) and Ridge Regression (L2 Regularization).

L1 Regularization uses the method which penalizes the sum of absolute values of coefficients. It not only prevents overfitting but also is beneficial in performing variable selection.

$$\boxed{L_{lasso}(\hat{\beta}) = \sum_{i=1}^{n}(y_i - x_i'\hat{\beta})^2 + \lambda \sum_{j=1}^{m} |\hat{\beta}_j|.}$$

L2 Regularization uses a slightly different approach of penalizing the sum of squared coefficients, but it is not capable of performing variable selection, all variable provided to the mode remain in the model during and after the training.

$$\boxed{L_{ridge}(\hat{\beta}) = \sum_{i=1}^{n}(y_i - x_i'\hat{\beta})^2 + \lambda \sum_{j=1}^{m} \hat{\beta}_j^2 = ||y - X\hat{\beta}||^2 + \lambda||\hat{\beta}||^2.}$$

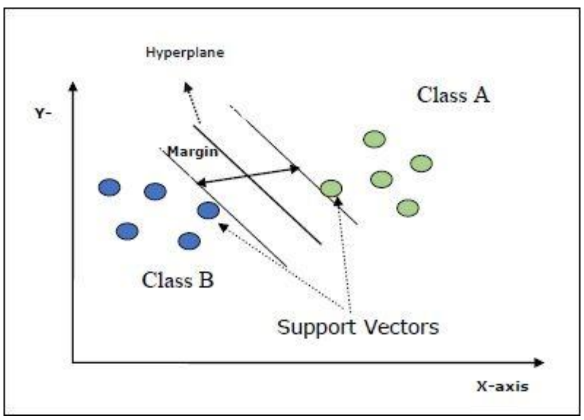


*Fig 1: 2-Dimensional Spatial domain SVM Model diagram*

Model: "model"

| Layer (type) | Output Shape | Param # |
|---|---|---|
| input_1 (InputLayer) | [(None, 3, 1)] | 0 |
| conv1d (Conv1D) | (None, 3, 64) | 192 |
| conv1d_1 (Conv1D) | (None, 3, 128) | 16512 |
| flatten (Flatten) | (None, 384) | 0 |
| dense_21 (Dense) | (None, 128) | 49280 |
| dense_22 (Dense) | (None, 64) | 8256 |
| dense_23 (Dense) | (None, 32) | 2080 |
| dense_24 (Dense) | (None, 1) | 33 |

Total params: 76353 (298.25 KB)
Trainable params: 76353 (298.25 KB)
Non-trainable params: 0 (0.00 Byte)

*Fig 2: Model Summary for complied Neural Network*

## V. MODEL EVALUATION

Model evaluation is a pivotal stage in the development of any machine learning system, particularly in applications like rainfall prediction where accuracy is crucial. In this context, we aim to assess the performance of three distinct models: Linear Regression, RNN and SVM. Each model brings its own strengths and weaknesses to the table, and evaluating them rigorously is essential to make an informed decision about their suitability for the task at hand.

Before delving into the evaluation process, it's imperative to establish a structured methodology. The first step entails data transformation, wherein the available dataset is transformed by cleaning imputing missing value.

It is followed by division of dataset into training dataset, validation and test sets. The training set is employed to train the models, while the validation dataset helps in feature selection, hyperparametric tuning and model selection. The test set remains unutilized until the final evaluation stage, serving as an unbiased measure of each model's generalization performance. Typically, a common split ratio like 70-10-20 (train-validation-test) is employed, although adjustments may be necessary based on factors such as dataset size and complexity.

SVM, a powerful tool for classification and regression tasks, utilizes optimization algorithms to find the hyperplane that best separates data points. Linear Regression, a simple yet effective method, learns to model the relationship between input features and output targets using linear approximation. Following the training phase, the models proceed to validation. Here, the aim is to fine-tune model hyperparameters to enhance performance on unseen data. For RNN, hyperparameters such as the number of layers, hidden units, and learning rate play pivotal roles in shaping model behaviour. SVM's performance is heavily influenced by kernel type, regularization parameter (C), and kernel coefficient (gamma). Linear Regression may benefit from regularization techniques like Ridge or Lasso regression, as well as feature scaling or transformation.

Once the models are trained and validated, the evaluation metrics come into play. For regression tasks like rainfall prediction, common metrics include Mean Absolute Error (MAE), Standard Deviation (SD). MAE represents the average magnitude of errors between predicted and actual values, providing a straightforward measure of model accuracy. SD is a statistical measure that quantifies the amount of variation or dispersion in a dataset. In the context of model evaluation, standard deviation provides valuable insights into the consistency and reliability of model predictions.

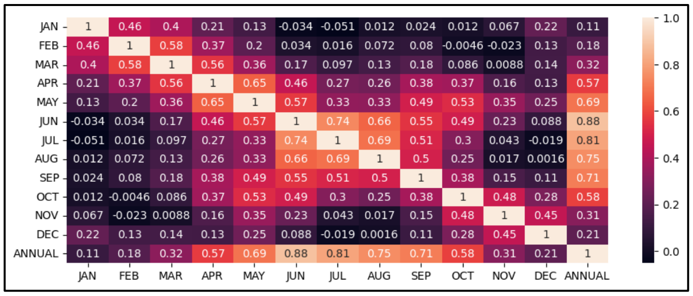


*Fig 3: Heat Map of monthly rainfall amount aiding in finding corelation between features*

## VI. CONCLUSION

The evaluation of rainfall prediction systems utilizing Linear Regression, RNN and SVM reveals valuable insights into their performance and applicability in this critical domain. Each model brings distinct strengths and limitations, which are essential considerations for stakeholders seeking to implement effective forecasting solutions.

In contrast, SVMs offer robustness and versatility, particularly in high-dimensional feature spaces. Their ability to handle non-linear relationships and outliers makes them valuable contenders for rainfall prediction tasks. Nevertheless, SVMs may exhibit scalability issues with large-scale datasets and can be computationally demanding, especially when using kernel methods with high-dimensional input features.

Linear Regression, despite its simplicity, proves to be a reliable baseline for rainfall prediction. Its interpretable nature and computational efficiency make it an attractive choice for initial modelling efforts or situations where transparency and ease of implementation are paramount. However, Linear Regression's linear assumption may limit its ability to capture complex interactions and non-linear dependencies present in rainfall data, leading to potential performance degradation in comparison to more sophisticated models like RNNs and SVMs.

The evaluation process, encompassing data splitting, hyperparameter tuning, and model comparison using metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE), provides valuable insights into each model's predictive performance. Additionally, standard deviation analysis offers critical information about prediction variability and uncertainty, enhancing the robustness and reliability of model evaluations.

Ultimately, the selection of the most suitable model for rainfall prediction hinges on a nuanced understanding of the trade-offs between predictive accuracy, computational complexity, and interpretability. Stakeholders must carefully weigh these factors against their specific requirements and constraints to make informed decisions. Moreover, ongoing research and advancements in machine learning present opportunities to refine existing models and develop innovative approaches tailored to the evolving challenges of rainfall prediction and environmental forecasting. By embracing a data-driven and iterative approach to model development and evaluation, stakeholders can foster continuous improvement in rainfall prediction systems, thereby enhancing their effectiveness in mitigating risks and informing decision-making processes.

## VII. RESULTS AND OUTPUTS

By training multiple models, we were able to compare the different approaches, some were traditional like linear regression using elastic net (encompassing both L1 and L2 Regularization approach), while other used concepts of Neural Networks to build systems which employ concepts such fixed recurrent feedback loop.

Moreover, we were able to predict the rainfall pattern using different models, due to the dataset containing continuous values we compared them by using MAE (Mean Absolute Error) as the Performance metrics.

From the Results we discovered, that models like linear regression are prone to biases due to overfitting the prediction curve, whereas SVR handled overfitting by using hyperplane but the margin between vectors were small, resulting in unreliable separation resulting in ~127 MAE.

Finally, RNN perform we with MAE of ~85, which was better than Elastic Net and SVR. RNN performed well due its recurrent nature helping in feature selection and mitigating overfitting. The RNN uses the Neurons which are trained on data multiple times in the hidden layers hence reducing the occurrence of skewed prediction.

The Result of MAE for each approach used in this study is given below:

TABLE 1: PREDICTION OBSERVATION FOR INDIAN RAINFALL

| ***Algorithm*** | ***MAE*** |
|---|---|
| ***Linear Regression*** | ***96.32435229744083*** |
| ***SVR*** | ***127.1600615632603*** |
| ***Recurrent Neural Network*** | ***85.50177457045439*** |

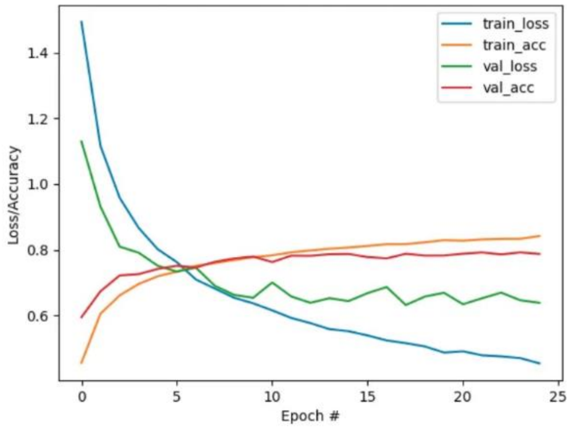


*Fig 4: Training Loss and Accuracy [Epoch #45]*

## VIII. FUTURE SCOPE

In the realm of rainfall prediction systems, the future holds considerable promise for advancements that can revolutionize the field. Future research endeavours could focus on integrating developing technologies such as machine learning, deep learning and big data analytics to improve the accuracy and reliability of predictions.

By leveraging these innovative approaches, researchers can delve deeper into the intricate relationships between meteorological variables and develop more sophisticated models capable of capturing complex rainfall patterns. Moreover, the integration of high-resolution satellite imagery, ground-based sensor networks, and advanced radar systems offers opportunities to improve the spatial and temporal resolution of rainfall data, enabling more precise predictions at local and regional scales.

Additionally, there is scope for exploring interdisciplinary collaborations to incorporate socio-economic factors and environmental indicators into prediction models, thereby enhancing their utility for disaster prevention/preparedness, agricultural tasks, water resource management, and infrastructure planning.

Overall, the future of research in rainfall prediction systems lies in leveraging cutting-edge technologies and interdisciplinary approaches to develop robust, accurate, and actionable forecasting tools that can mitigate the impacts of extreme weather events and support sustainable development initiatives.

## IX. REFERENCE


[1] Duan, S. B., Li, Y. L., Wang, L. J., Wang, D. W., & Chen, J. (2020). Rainfall prediction using ensemble deep learning models: A case study in China. Journal of Hydrology, 587, 124946.

[2] Gado, A. A., Rajasekhar, C. P., & Deo, R. C. (2020). Rainfall prediction using LSTM model with satellite-based observations over India. Journal of Hydrology: Regional Studies, 36, 102339.

[3] Gao, Y., Zhu, X., Fu, C., & Smethie, W. B. (2020). Prediction of heavy rainfall events using multi-source remote sensing data and a deep learning model. Remote Sensing of Environment, 247, 112002

[4] Liu, Y., Ouyang, T., Guan, X., Zhou, Y., & Zhang, H. (2020). Convolutional neural network for predicting high-impact rainfall events using satellite-based precipitation estimates. Advances in Space Research, 65(2), 575-587.

[5] Nogueira, D. S., de Brito, R. R., de Carvalho Júnior, O. A., & de Oliveira, R. A. L. (2020). Convolutional neural network based approach for rain intensity estimation in satellite imagery. Remote Sensing, 12(10), 1683.

[6] Singh, R. K., & Rani, R. (2019). Rainfall prediction using machine learning techniques: A review. Environmental Monitoring and Assessment, 191(6), 377.

[7] Wang, Y., Wang, Y., Zhang, H., & Zhang, L. (2020). Prediction of extreme rainfall events using a deep learning model combining LSTM and attention mechanism. Remote Sensing, 12(13), 2124.

[8] Xu, Z., Zhu, X., Fu, C., Smethie, W. B., & Liang, S. (2022). A machine learning ensemble framework for predicting extreme rainfall events using satellite observations. Remote Sensing, 14(11), 2773.

[9] Jae-Hyun Seo, Yong Hee Lee, Yong-Hyuk Kim, "Feature Selection for Very Short-Term Heavy Rainfall Prediction Using Evolutionary Computation", Advances in Meteorology, vol. 2014, Article ID 203545, 15 pages, 2014. https://doi.org/10.1155/2014/203545

[10] S. Manandhar, S. Dev, Y. H. Lee, Y. S. Meng and S. Winkler, "A Data-Driven Approach for Accurate Rainfall Prediction," in IEEE Transactions on Geoscience and Remote Sensing, vol. 57, no. 11, pp. 9323-9331, Nov. 2019, doi: 10.1109/TGRS.2019.2926110.

[11] E. Hussein, M. Ghaziasgar and C. Thron, "Regional Rainfall Prediction Using Support Vector Machine Classification of Large-Scale Precipitation Maps," 2020 IEEE 23rd International Conference on Information Fusion (FUSION), Rustenburg, South Africa, 2020, pp. 1-8, doi: 10.23919/FUSION45008.2020.9190285.

[12] Poornima, S.; Pushpalatha, M. Prediction of Rainfall Using Intensified LSTM Based Recurrent Neural Network with Weighted Linear Units. Atmosphere 2019, 10, 668. https://doi.org/10.3390/atmos10110668

[13] F. Karim, S. Majumdar, H. Darabi and S. Chen, "LSTM Fully Convolutional Networks for Time Series Classification," in IEEE Access, vol. 6, pp. 1662-1669, 2018, doi: 10.1109/ACCESS.2017.2779939.

[14] Benlan He, Yong Shi, Qian Wan, Xi Zhao, Prediction of Customer Attrition of Commercial Banks based on SVM Model, Procedia Computer Science, Volume 31,2014, Pages 423-430, ISSN 1877-0509,https://doi.org/10.1016/j.procs.2014.05.286.(https://www.sciencedirect.com/science/article/pii/S1877050914004633)

[15] J. W. Kelly, A. D. Degenhart, D. P. Siewiorek, A. Smailagic and W. Wang, "Sparse linear regression with elastic net regularization for brain-computer interfaces," 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, San Diego, CA, USA, 2012, pp. 4275-4278, doi: 10.1109/EMBC.2012.6346911.

[16] Penghui Liu, Jing Liu, Kai Wu, CNN-FCM: System modelling promotes stability of deep learning in time series reduction, Knowledge-Based Systems, Volume 203,2020,106081,ISSN0950-7051,https://doi.org/10.1016/j.knosys.2020.106081.(https://www.sciencedirect.com/science/article/pii/S0950705120303592)