



Prediction of Rainfall Categories in Kerala, India, Using Satellite Imagery and a YOLOv8 Classifier

Asha Jayasree^{1*} , Santhosh Kumar Sasidharan² , Rishidas Sivadas³ , Jayan A. Ramakrishnan¹ 

¹ Department of Electronics & Communication Engineering, Government Engineering College Thrissur, APJ Abdul Kalam Technological University, Thiruvananthapuram 695016, India

² Department of Electronics & Communication Engineering, Government Engineering College, Idukki 685603, India

³ Department of Electronics & Communication Engineering, Government Engineering College, Barton Hill 695035, India

Corresponding Author Email: ashaj@gectcr.ac.in

Copyright: ©2025 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/ts.420235>

ABSTRACT

Received: 26 August 2024

Revised: 12 December 2024

Accepted: 24 February 2025

Available online: 30 April 2025

Keywords:

confusion matrix parameters, EUMETSAT, Kerala, prediction, rainfall category, ResNet-50, YOLOv8 classifier

The accurate prediction of daily rainfall has become imperative in mitigating hydro meteorological disasters. This paper presents a novel and reliable method for predicting daily rainfall category over selected geographical locations in Kerala state of India using the satellite imagery data obtained from EUMETSAT. Satellite images over Kerala and the Arabian Sea near to Kerala are used as input to train the YOLOv8 classifier and predict the rain fall category as output. In this work, rainfall is categorized into three classes-No Rain (NR), Very Light Rain to Light Rain (VLR-LR) and Medium Rain to Heavy Rain (MR-HR) based on the actual rainfall values. The trained network is tested with satellite images outside the training set and performance parameters are computed for test images for evaluation. The proposed best model gives a precision of 95% for NR, 89% for VLR-LR and 100% for MR-HR. Comparison of YOLOv8 with ResNet-50 shows that YOLOv8 is effective in predicting the daily rainfall category prediction.

1. INTRODUCTION

India is a geographically diverse country with a variety of climatic regions that have an important effect on its agricultural methods and natural landscapes. The nation can be largely classified into five main climate zones: tropical, arid, semi-arid, temperate, and alpine. Kerala is located at the southwestern corner of India, with the Arabian Sea on its western side. It shares borders with two other Indian states: Karnataka to the north and northeast, and Tamil Nadu to the east [1]. Kerala state of India lies between the northern latitude of 8°17'30" N and 12°47'40" N and east longitudes 74°27'47" E and 77°37'12" E [2]. According to the Köppen climate classification, Kerala state falls under the tropical monsoon climate type [3]. Kerala receives an average rainfall of more than 3000mm, and is mostly obtained from the seasonal monsoon. Unlike in the past years, Kerala has been witnessing severe floods during recent years. The flood of 2018 has been named as the “flood of the century”. It has claimed over 400 lives, affecting the lives of millions of people. The Indian Meteorological Department (IMD) has classified the rain received during those days as “large excess”. Subsequently, in 2019 and 2020 also, the state has witnessed floods and landslides. Kerala is experiencing significant fluctuations and unpredictability in its weather conditions, particularly in the timing and volume of rainfall. In this context, accurate and timely prediction of meteorological parameters and daily rainfall has got broad implications from enhancing crop yields to minimizing casualties and property destruction.

The majority of early techniques for forecasting daily rainfall relied on statistical models to find trends and connections in historical meteorological data. Time Series Analysis can accurately simulate the temporal dependencies found in rainfall data, methods such as Autoregressive Integrated Moving Average (ARIMA) models is widely used. These techniques were developed by Box and Jenkins [4] and have been extensively utilized in meteorology. Regression models, both linear and nonlinear, are frequently used to forecast rainfall based on wind speed, temperature, and humidity. For example, multiple regression models were used by Kidson and Thompson [5] to enhance short-term weather forecasts. Complex patterns in rainfall data can be captured with tremendous potential by neural networks, particularly architectures like Long Short-Term Memory (LSTM) networks. For instance, Shi et al. [6] used convolutional LSTM networks to show how well deep learning works for precipitation prediction. Prediction accuracy can be increased by combining machine learning techniques with conventional statistical methods. The combination of Singular Spectrum Analysis (SSA), Auto Regressive Integrated Moving Average (ARIMA), and Artificial Neural Networks (ANN) into a hybrid model (SSA-ARIMA-ANN) for generating reliable daily rainfall forecasts in a river catchment [7]. Ridwan et.al. has forecasted rainfall using two approaches: (1) Autocorrelation Function (ACF) based on past rainfall data, and (2) Projected Error based on both past and projected rainfall data. Various algorithms, including Bayesian Linear Regression, Boosted Decision Tree Regression, Decision

Forest Regression, and Neural Network Regression, were employed to identify the optimal predictions for rainfall across different time horizons. The results showed improved accuracy when cross-validation was applied, particularly with Boosted Decision Tree Regression, after tuning its parameters [8].

Numerous methods using satellite imagery were investigated and tested to reliably predict the daily rainfall. Due to the highly unpredictable nature of rainfall, sufficient prediction accuracy could not be achieved. Boonyuen et al. [9] used the satellite images of Thailand and categorized images into “Rain” and “No Rain” categories. They used the Convolutional Neural Network method to predict rainfall category and obtained prediction accuracies of about 77.42% for current day rainfall and one day ahead rainfall, 67.47% for two days ahead and 58.06% for three days ahead prediction. Novitasari et al. [10] have used Himawari-8 IR satellite imagery and meteorological data. Adaptive average brightness thresholding method was employed for segmentation and classification was done using back propagation. Simanjuntak et al. [11] used high spatiotemporal Himawari 8 satellite imagery and employed the multivariate LSTM forecasting method for rainfall prediction. Ionescu et al. [12] proposed DeePsat, a CNN model that predicted future satellite images from the past satellite imagery. Liyew and Melese [13] used meteorological data of Ethiopia to analyze the rainfall data. The Machine learning algorithms used in this work are: Multivariate linear regression, Random Forest and XGBoost. Research work was published for forecasting rainfall on yearly, monthly, daily and even on an hourly basis [14, 15].

From a layman perspective, a linguistic description of rainfall gives more impact rather than the exact quantity of rainfall. This paper aims to predict the category of daily rainfall. In this work the daily rainfall is categorized as: No Rain (NR), Very Light Rain to Light Rain (VLR-LR), and Medium Rain to Heavy Rain (MR-HR). The methodology adapted is the YOLOv8 classifier. YOLOv8 is the recent version of real-time object detection and image segmentation model [16]. YOLO's popularity stems from its ability to achieve a high-level of accuracy while maintaining a compact model size. YOLOv8 leverages state-of-the-art innovations in deep learning and computer vision technologies, delivering exceptional results with regard to both processing speed and detection precision. The work is focused on the YOLOv8 model for rainfall category prediction. For a comprehensive evaluation of its effectiveness YOLOv8 models are compared with ResNet-50, an established deep learning architecture for image classification. This comparison will establish the effectiveness of YOLOv8 models in predicting the rainfall category.

The remaining part of the paper is organized as follows. Section 2 investigates the previous works related to daily

rainfall prediction and YOLO. Section 3 elaborates on the rainfall pattern in Kerala, India. Section 4 includes related work analysis. Section 5 details on YOLOv8 classifier structure, and finally, Section 6 includes results and discussions of the YOLOv8 models and the comparison with the ResNet-50 model.

2. RAINFALL PATTERN IN KERALA

Kerala has the Arabian Sea along its western border, while the Eastern boundary is formed by the Western Ghats mountain ranges. Kerala's climate is greatly influenced by the seasonal heavy rainfall during the monsoon season. Kerala has an average of around 120-140 rainy days per year. The unique geographical positioning of Kerala, with its distinctive topographical features, makes it the first recipient of Southwest monsoon rainfall in India. The South West monsoon season, which usually begins in June and continues till September, is the primary rain season in Kerala. Moisture laden air from the Arabian Sea is carried over the Indian subcontinent, bringing rainfall to the state. The Western Ghats, running parallel to the western coast of the state, plays a vital role in determining the amount of rainfall. The moist air ascending the mountains cools, condenses and precipitates, leading to a substantial amount of rainfall.

During the October to December months, Kerala experiences the Northeast monsoon, which is also referred to as the retreating monsoon. The direction of the wind reverses and it moves from the North East to the South West. The main feature of North West monsoon is heavy rains during the afternoon accompanied by lightning and thunder. Other than these monsoon rains, Kerala also receives rain due to cyclonic activities over the Arabian Sea and the Bay of Bengal.

Based on the rainfall data between 2001 and 2021, Kerala is showing an increasing trend in heavy rainfall. 2007 and 2018 are reported to have received the maximum amount of rainfall under the heavy rainfall category, i.e., above 64.4mm. The state has witnessed floods and landslides during the years 2018, 2019, 2020 and 2021. Events of cloud bursts have triggered torrential rain at various places in Kerala. It was reported that the flooding of August 2019 was a result of a Mesoscale cloud burst which is not usually observed in Kerala. Forecasts of these heavy spells of rainfall will warn the government and people to take necessary precautions against the effects of heavy downpour.

The Indian Meteorological Department (IMD) has classified and defined the daily rainfall, based on its amount, into various categories as shown in Table 1 [17]. In this work these rainfall categories have been regrouped and are categorized into three major categories- NR, VLR-LR and MR-HR.

Table 1. Categorisation of daily rainfall

IMD Categorization of Daily Rainfall [17]		Rainfall Categorization for This Study	
Rainfall Category	Daily Rainfall(mm)	Rainfall Category	Daily Rainfall(mm)
NR	0	NR	0
Very Light Rain (VLR)	0.1-2.4	Very Light Rain-Light Rain (VLR-LR)	0.1-15.5
Light Rain (LR)	2.5-15.5		
Moderate Rain (MR)	15.6-64.4		
Heavy Rain (HR)	64.5-115.5	Moderate Rain to heavy Rain (MR-HR)	15.6 and above
Very Heavy Rain (VHR)	115.6-204.4		
Exceptionally Heavy Rain (EHR)	>244.5		

3. PROPOSED PREDICTION MODEL USING YOLOv8 CLASSIFIER

In this work, YOLOv8 is used for predicting the category of rainfall. For that, the YOLOv8 is trained using the satellite images of the previous day. Meteosat IODC Airmass images of Southern Asia, obtained from EUMETSAT are used for training, validation and testing of the model NVIDIA GeForce RTX 3050 8GB GPU in the Linux environment used for developing the prediction model.

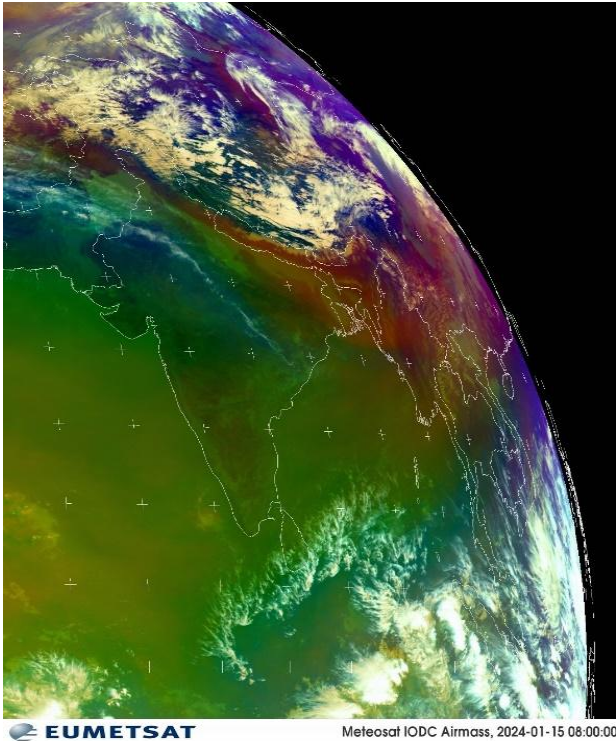
- The following are the main features of the proposed model:
- (1) The proposed model utilizes daily rainfall data sourced from the Indian Meteorological Department (IMD) website for classifying days into NR, VLR-LR, and MR-HR categories.
 - (2) Satellite images sourced from EUMETSAT are employed in the proposed model. Airmass images from the preceding day are utilized for training, enabling the model to predict the next day's rainfall category when tested with the satellite image from the previous day.
 - (3) The proposed system uses the YOLOv8 classifier to achieve high prediction accuracy. There is no existing literature that reports using YOLOv8 for the prediction of daily rainfall category.

YOLOv8 is the latest and cutting edge YOLO approach which can be used for object detection, image classification as well as segmentation [18]. It was released by Ultralytics in the year 2023. The architecture of YOLOv8 builds upon previous versions of YOLO algorithms. This is the first study that has employed the YOLOv8 algorithm for rainfall category prediction using satellite images. The proposed approach provides improved accuracy, versatility and reduced false classification. YOLO has gained popularity due to its ability to achieve a good balance between accuracy and model size [18]. YOLO is designed to detect and classify objects in images and video frames. YOLO uses a grid based approach, where the image is divided into a grid and each grid cell detect objects contained within itself.

The architecture of YOLOv8 consists of two primary components: a backbone network that extracts features and a head network that processes these features for detection tasks [19]. YOLOv8 is an improvement over YOLOv5 that combines the Cross-Stage partial idea [19], feature fusion method [19] and PPF module. The backbone of YOLOv8 is a modified version of CSPDarknet53 (Cross-Stage Partial) architecture. Architecture has 53 convolutional layers and uses cross-stage partial connections [18]. The head of YOLOv8 consists of multiple convolutional layers followed by a series of fully connected layers. An essential enhancement in the YOLOv8 model involves integrating a self-attention mechanism within the network's head [20]. This allows the model to focus on various parts of the image. The Feature pyramid network consisting of multiple layers detects objects at different scales [21]. The different model sizes of YOLOv8 are: YOLOv8-n (nano), YOLOv8-s (small), YOLOv8-m (medium), YOLOv8-l (large) and YOLOv8-xl (extra-large) [20]. The size of the model is proportional to mean average Precision (mAP) and is inversely related to inference time. Yolov8 classifier identifies and outputs a single class based on the probability values. The result of the classification task includes both a class index and a confidence score.

The proposed model predicts the next day's rainfall category using the YOLOv8 model. For this, the satellite images are categorized into NR, VLR-LR and MR-HR, depending on the amount of rainfall received on the

subsequent day. The YOLOv8 model is trained using these categorized satellite images. YOLOv8 with its deep neural network architecture, performs well in classification tasks. The dataset for the model consists of satellite images obtained from EUMETSAT. Figure 1 shows a sample satellite image. The satellite image has been trimmed to match the dimensions of the area of interest, such that it includes the total of Kerala state and the nearby Indian Ocean region. The 41.5° Indian Ocean Data coverage (IODC) Southern Asia's Airmass images are obtained from Meteosat-9 satellite.



a) Airmass image of Southern Asia



b) Cropped image

Figure 1. Airmass and cropped views of Southern Asia

Table 2. Dataset description

Class	Training	Validation	Testing
NR	600	50	50
VLR-LR	500	50	50
MR-HR	500	50	50

A total of 1900 images were used for the prediction model. A dataset of Airmass images for South Asia has been meticulously compiled from the EUMETSAT website, with daily downloads initiated in 2019 and continuing through 2023. This dataset serves as the resource for the current research work, consisting of a total of 1,900 satellite images. The images are categorized primarily into three groups: the majority fall under the NR category, and the lowest number of images are MR-HR.

The limited number of images in the MR_HR category reflects the variability in precipitation patterns across the region, which can impact data availability. However, the dataset's current size of 1,900 images provided a substantial foundation for the research work. As data collection continues over subsequent years, there is significant potential to expand this dataset further. Increasing the number of satellite images will improve the predictive accuracy of the proposed model. The number of images used for training, validation and testing for each class is depicted in Table 2.

The training process for the rainfall prediction model involves utilizing satellite imagery captured on the preceding day as input data. Subsequently, during the testing phase, the model is provided with the satellite image from the prior day as input, and it performs classification to predict the category of rainfall for the current day.

Within the YOLOv8 framework, various model sizes are available, denoted as yolov8-n-nano, yolov8-s-small, yolov8-m-medium, yolov8-l-large, and yolov8-x-extra-large. The relationship between model size, mean average precision (mAP), and inference time in YOLOv8 is characterized by linear proportionality to mAP and inverse proportionality to inference time [22]. Larger models exhibit higher mAP but demand more inference time for accurate object detection. Conversely, smaller models offer faster inference times but tend to have lower mAP. Opting for bigger models is advantageous when dealing with limited data.

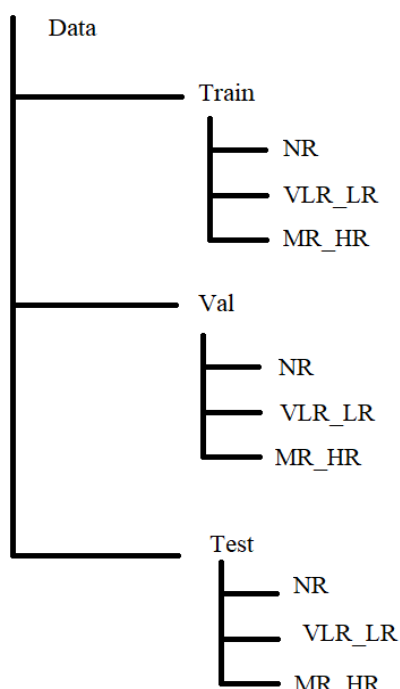


Figure 2. Data folder arrangement

This research trains multiple YOLOv8 variants along with a ResNet-50 model on the customised dataset, followed by

comprehensive performance assessment and comparison to determine the optimal model for daily rainfall category prediction. The latest version of YOLOv8 library, YOLOv8.0.232, is used to develop the proposed prediction model. To construct the prediction model, a custom dataset is created, comprising satellite images representing distinct rainfall categories. The dataset creation process involves establishing a "data" folder, within which three subfolders named "train," "test," and "val" are created. Each of these subfolders further encompasses three folders, each designated with the name of a specific rainfall category, containing the respective satellite images. The folder arrangement of the dataset is shown in Figure 2.

4. RESULTS AND DISCUSSIONS

4.1 YOLOv8 models

The heavier models of YOLOv8 are composed of more layers and parameters and hence they are capable of achieving more precise learning, but they require longer computation time for training and prediction. Table 3 shows the number of layers and parameters for each YOLOv8 model.

There are around 30 user configuration hyperparameters for YOLO model and the training results vary with the hyperparameter values. The key hyperparameters in YOLOv8 are: i) Learning rate; ii) Batch size; iii) Epochs; iv) Input image size; v) Optimizer type and parameters. These hyperparameters are tuned to optimize the models performance. Table 4 show the results of average precision of training for different learning rates. The epochs with best accuracy is chosen for prediction. The first parameter chosen is the learning rate. If the learning rate is set too low, the training process becomes inefficient and time-consuming. On the other hand, an excessively high learning rate can cause instability in the model's learning process, potentially reducing overall performance by preventing proper convergence. So it is necessary to fix the learning rate that produces highest accuracy and then fix the other parameters. Subsequently other hyperparameters are fixed based on the accuracy.

As per Table 4, the learning rate is fixed to be 0.002 for YOLOv8s and is 0.001 for all the other YOLOv8 models. After fixing the learning rate the number of epochs are fixed for each YOLOv8 model. Table 5 shows the relation between the number of epochs and accuracy different YOLOv8 models.

From the table it is clear that maximum accuracy is obtained for 50 epochs for YOLOv8l, YOLOv8m and YOLOv8s while it is 70 epochs for YOLOv8n. Table 6 gives the values of hyper parameters used for training. After fixing the epochs, learning rate and batch size based on the accuracy values, optimizer is fixed. An optimizer can significantly affect the model's convergence speed and accuracy. The different optimizers are Stochastic Gradient Descent (SGD), Adaptive Moment Estimation (Adam) and Adam with Weight Decay (AdamW). Table 6 summarises the accuracy values for different optimizers.

Table 3. Layers and parameters for YOLOv8 model

Category	YOLOv8n	YOLOv8s	YOLOv8m	YOLOv8l
Layers	225	256	295	365
Parameter (in million)	3	11.1	25.9	43.6

Table 4. Accuracy on varying base learning rate for different YOLOv8 models

YOLOv8 Model	Learning Rate	Accuracy
YOLOv8l	0.0005	0.922
	0.0008	0.928
	0.001	0.934
	0.002	0.933
	0.003	0.933
	0.004	0.921
YOLOv8m	0.0005	0.911
	0.0008	0.916
	0.001	0.915
	0.002	0.912
	0.003	0.912
	0.004	0.910
YOLOv8s	0.0005	0.918
	0.0008	0.919
	0.001	0.921
	0.002	0.922
	0.003	0.918
	0.004	0.916
YOLOv8n	0.0005	0.920
	0.0008	0.925
	0.001	0.926
	0.002	0.925
	0.003	0.925
	0.004	0.921

Table 7 shows the optimized values of different hyperparameters. After the completion of training, run directory is created where model metrics like confusion metrics and results like training and validation loss are obtained. In validation mode,

model is evaluated on a validation set which measures its accuracy and performance. The trained model is tested on a set of test images. In this work training, validation and testing is done using four different YOLOv8 models viz YOLOv8n, YOLOv8s, YOLOv8m and YOLOv8l. Their performances are compared and the best model is identified. Figure 3 shows the training and validation loss of all the four YOLOv8 models.

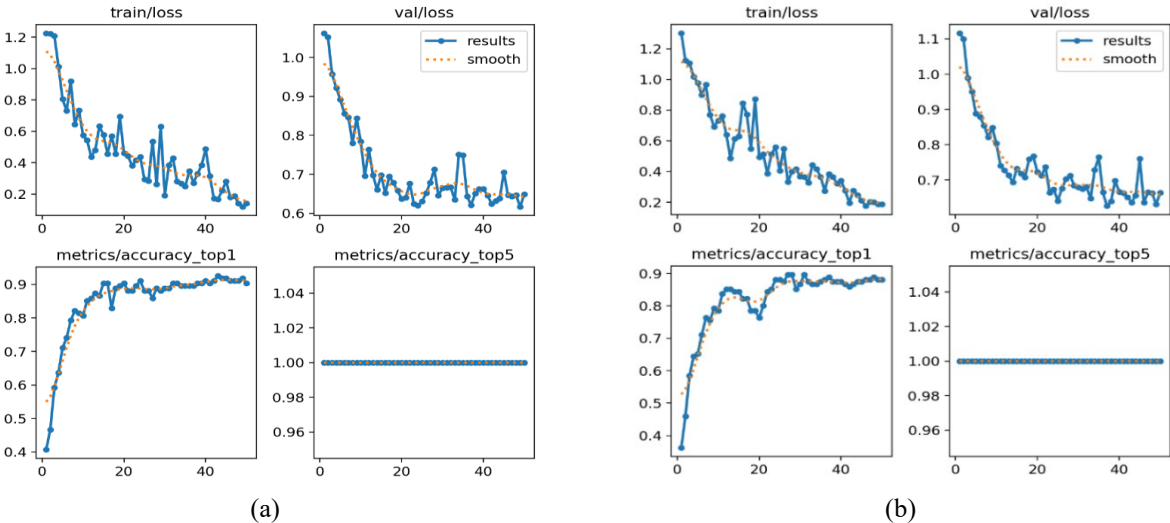
Figure 4 normalized confusion matrix obtained for YOLOv8l model.

Table 5. Accuracy on varying epochs for different YOLOv8 models

YOLOv8 Model	Epochs	Accuracy
YOLOv8l	10	0.928
	30	0.933
	50	0.941
	70	0.933
	100	0.934
YOLOv8m	10	0.908
	30	0.916
	50	0.926
	70	0.920
	100	0.915
YOLOv8s	10	0.922
	30	0.934
	50	0.944
	70	0.941
	100	0.916
YOLOv8n	10	0.920
	30	0.925
	50	0.925
	70	0.928
	100	0.921

Table 6. Accuracy on different optimizers for various YOLOv8 models

YOLOv8 Model	Optimizer	Accuracy
YOLOv8l	AdamW	0.942
	Adam	0.928
	SGD	0.934
YOLOv8m	AdamW	0.922
	Adam	0.910
	SGD	0.899
YOLOv8s	AdamW	0.920
	Adam	0.910
	SGD	0.845
YOLOv8n	AdamW	0.928
	Adam	0.915
	SGD	0.886



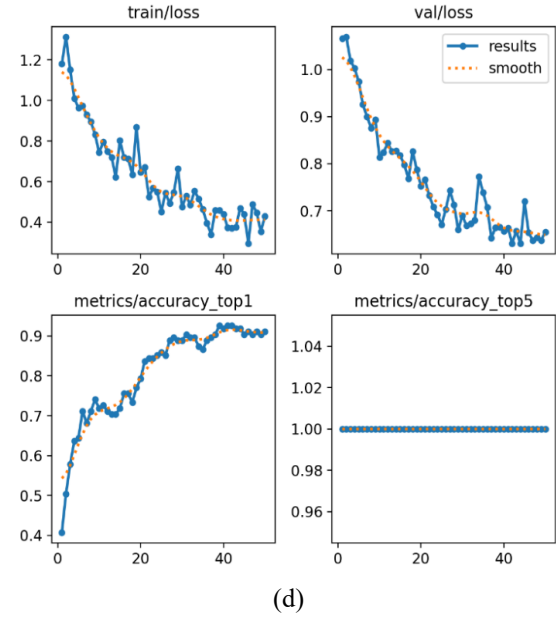
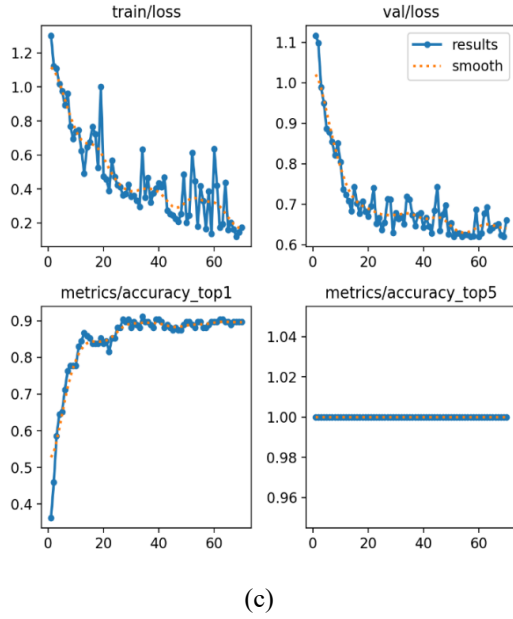


Figure 3. Training and validation loss of a) YOLOv8 large; b) YOLOv8 medium; c) YOLOv8 small; d) YOLOv8 nano models

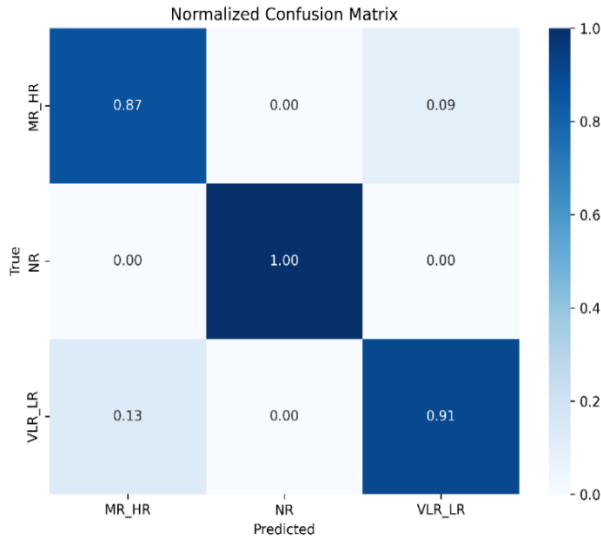


Figure 4. Confusion matrix obtained for the validation data for YOLOv8 large models

Table 7. Hyperparameter values for different YOLOv8 models

Parameter	YOLOv8l	YOLOv8m	YOLOv8s	YOLOv8n
No. of Epochs	50	50	70	50
Optimizer	AdamW	AdamW	AdamW	AdamW
Learning rate	0.001	0.001	0.002	0.001

4.2 ResNet-50 prediction model

ResNet-50, short form of Residual Network with 50 layers, is a deep learning architecture introduced by Microsoft Research in 2015 [23]. It is highly effective and useful in computer vision due to its ability to efficiently train very deep neural networks. The architecture solves the degradation problem, where increasing the network's depth reduces its

accuracy rather than improving its performance.

ResNet-50, which has undergone preliminary training on the extensive ImageNet dataset containing millions of diverse images across 1000 categories, is employed as a feature extraction mechanism to identify and analyze spatial patterns within satellite imagery.

The model is fine-tuned to adapt it to the specific task of rainfall category prediction by adding custom dense layers on top of the pre-trained base. Data Augmentation is done using the Image Data Generator class from Keras. The softmax activation function is applied to obtain the probability distribution across the three rainfall categories. A categorical cross-entropy loss function is used for predicting rainfall categories. Adam Optimizer used here to minimise the loss. Adam adjusts learning rates dynamically, facilitating faster convergence during training. Figure 5 shows the confusion matrix obtained using the ResNet-50 prediction model.

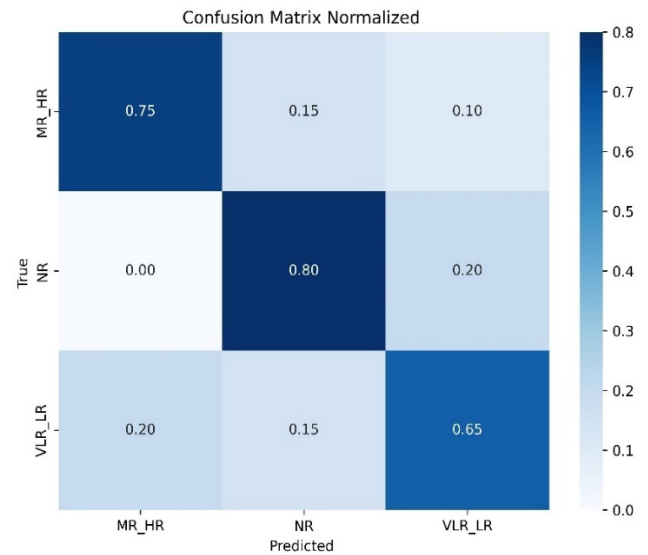


Figure 5. Normalized confusion matrix obtained for ResNet-50

Table 8. Average validation accuracy for different YOLOv8 models

Prediction Model	Average Accuracy (%)
YOLOv8 Large	95.00
YOLOv8 Medium	92.67
YOLOv8 Small	94.00
YOLOv8 Nano	93.33
ResNet-50	73.33

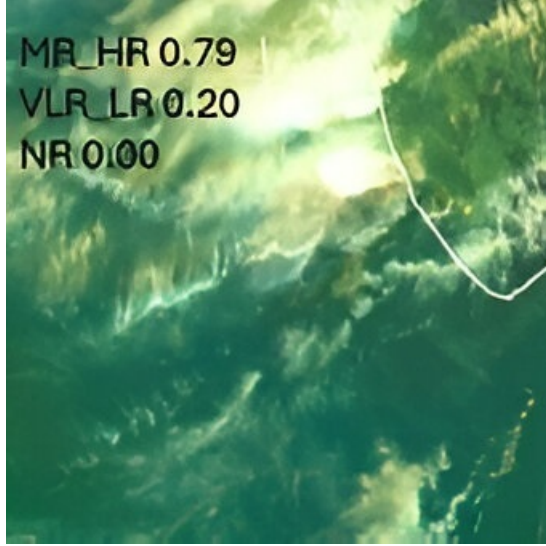


Figure 6. Category probabilities of a tested image using YOLOv8s

Table 8 tabulates the average accuracy for each YOLOv8 model and ResNet model. The results clearly indicate that the YOLOv8l model gives the best results.

After training and validation, the trained model is tested with satellite images of each category. The model generates probability estimates for different rainfall categories of the test image and subsequently predicts the specific rainfall category for the tested image. Figure 6 shows a predicted test image. It is a satellite image belonging to MR_HR and the probabilities predicted are: NR- 0, VLR_LR- 0.21 and MR_HR-0.79. Since the highest probability value is that of MR_HR, the predicted rainfall category is MR_HR.

Based on the testing of images using all the YOLOv8 models, the performance of each class is evaluated based on the following accuracy metrics [24]:

i) **Accuracy:** Accuracy is a metric that evaluates the percentage of correctly categorized instances out of the total number of objects in the dataset [24]. To calculate this metric, divide the number of accurate predictions by the total number of predictions made by the model.

$$Accuracy = \frac{No. of correct predictions}{Total Predictions}$$

ii) **Precision:** In multi-category classification, precision for a particular category is the ratio of images correctly identified as belonging to that specific category to the total images predicted by the model to belong to that category. For example, the Precision for NR category is defined as:

$$Precision_{NR} = \frac{True Positives_{NR}}{True Positives_{NR} + False Positives_{NR}}$$

iii) **Recall:** Recall represents the ratio of images in a particular category that the model has correctly identified out of all images belonging to that class. For example, the Recall for NR category is defined as:

$$Recall_{NR} = \frac{True Positives_{NR}}{True Positives_{NR} + False Negatives_{NR}}$$

iv) **F1 Score:** It is the harmonic mean of precision and recall. For example, the F1 score for the NR category is defined as:

$$F1score_{NR} = \frac{2 * Precision * Recall}{Precision + Recall}$$

The performance metrics for different YOLOv8 models are summarized in the Table 9.

The performance of the various YOLOv8 models and ResNet-50 are compared in the chart shown in Figure 7.

Table 9. Precision, Recall and F1 scores for different models

Model	Rainfall Category	Precision	Recall	F1 Score
YOLOv8 Large	NR	93	91	92
	VLR_LR	84	93	88
	MR_HR	100	91	95
YOLOv8 Medium	NR	95	91	93
	VLR_LR	82	93	88
	MR_HR	98	89	93
YOLOv8 Small	NR	91	91	91
	VLR_LR	84	82	83
	MR_HR	91	91	91
YOLOv8 Nano	NR	93	96	95
	VLR_LR	89	95	92
	MR_HR	89	95	92
ResNet-50	NR	79	75	77
	VLR_LR	73	80	76
	MR_HR	68	65	67

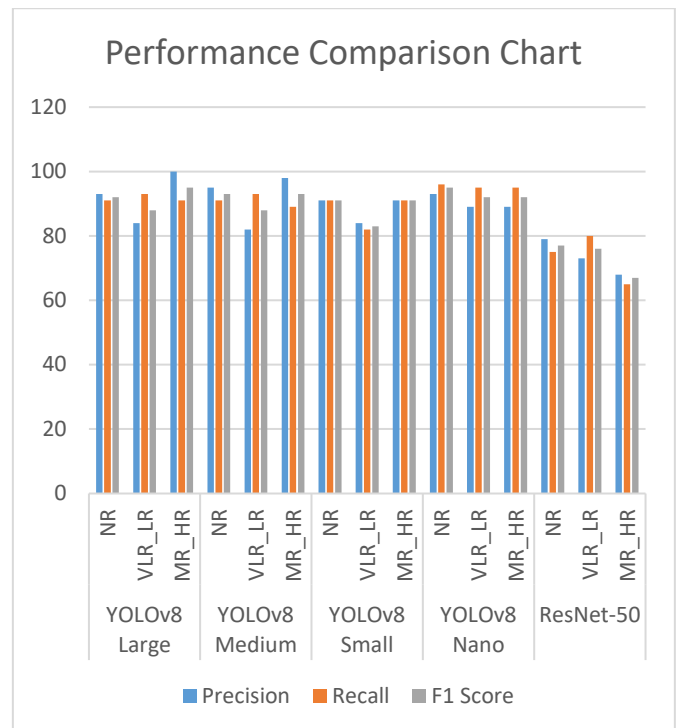


Figure 7. Performance comparison chart for various rainfall category prediction models

5. CONCLUSION

The results clearly indicate that all the YOLOv8 models' average accuracy values are above 90%, with the highest accuracy value of 95% for YOLOv8large. The inferences drawn from the performance comparison chart for each rainfall category are:

NR Category

- YOLOv8 Medium achieved the highest Precision (95%), indicating it has the best ability to avoid false positives in predicting "No Rain."
- YOLOv8 Nano has the highest Recall (96%), implying that it has captured almost all actual "No Rain" instances.
- The F1 Score, which balances Precision and Recall, is highest for YOLOv8 Nano (95%). This makes YOLOv8n the best model for predicting "No Rain".
- ResNet-50 has an F1 Score of only 77%, indicating that its ability to accurately predict No Rain is poor compared to the YOLOv8 variants.

VLR-LR Category

- YOLOv8 Nano shows maximum Precision (89%) and Recall (95%) values, resulting in the highest F1 Score (92%). This means YOLOv8 Nano is the most effective in identifying light rainfall events and minimizes false positives.
- YOLOv8 Large also exhibits better Recall (93%) value, making it capable to detect most light rain cases, but it has a slightly lower F1 Score (88%) due to lower Precision value.
- ResNet-50 has an F1 Score of 76%, indicating a higher rate of misclassification compared to YOLOv8 models

Moderate Rain to Heavy Rain (MR_HR) Category

- YOLOv8 Large achieves the ideal Precision (100%) value for this category, implying that it made no false positive predictions for "Moderate to Heavy Rain."
- Recall is slightly lower (91%), indicating that it has misclassified some true "Moderate to Heavy Rain" instances.
- YOLOv8 Large and Nano have identical F1 Scores (95%), indicating that they excel in detecting and classifying heavy rainfall events.
- ResNet-50 has an F1 Score of only 67%, indicating it frequently misclassifies heavy rainfall cases, likely leading to missed alerts for such important weather events.

Based on the above observations, the following conclusions can be drawn:

1). YOLOv8 models significantly outperform ResNet-50 in predicting rainfall categories across the board. This demonstrates the effectiveness of the YOLOv8 architecture in extracting relevant features from satellite images for weather classification tasks.

2). YOLOv8 Nano emerges as the best overall with the best F1 Score, making it the most reliable model for accurate predictions of rain intensity.

3). YOLOv8 Large excels in detecting Moderate to Heavy Rain with perfect Precision, meaning it is the best choice in avoiding false alarms for heavy rains.

4). ResNet-50 shows weaker performance in all categories, with its highest F1 Score being only 77% for the "No Rain" category.

There is a probability of misclassifying the MR-HR categories as VLR-LR and this poses significant challenges in effectively preparing the public for potential weather-related calamities. Such misclassifications can hinder timely responses and appropriate resource allocation during adverse weather events.

To enhance the model's accuracy, one effective approach would be to increase the volume of satellite imagery included in the dataset. Additionally, integrating more climatic parameters-such as humidity, temperature, and atmospheric pressure-could provide the model with valuable contextual information that enhances its predictive capabilities.

The authors' future research will focus on improving the model's accuracy through these strategies. By expanding the dataset and refining the input features, we aim to reduce misclassification rates and ultimately improve the reliability of rainfall predictions. This will not only enhance the model's performance but also ensure that communities are better prepared to respond to weather-related challenges.

To deploy this model, a Telegram-based application can be developed utilizing the proposed YOLOv8 model for rainfall prediction, allowing the public to access this bot at any time to view daily rainfall category forecasts.

ACKNOWLEDGMENT

The authors are grateful to Eumetsat and to the Indian Metrological Department (IMD) for allowing us to use their data.

REFERENCES

- [1] Kerala. (2025). Encyclopaedia Britannica. <https://www.britannica.com/place/Kerala>, accessed on Apr. 21, 2025.
- [2] Kerala at a Glance. (2025). Department Of Soil Survey & Soil Conservation. <https://www.keralasoils.gov.in/en/kerala-glance>.
- [3] Peel, M.C., Finlayson, B.L., McMahon, T.A. (2007). Updated world map of the Köppen-Geiger climate classification. Hydrology and Earth System Sciences, 11(5): 1633-1644. <https://doi.org/10.5194/hess-11-1633-2007>
- [4] Box, G.E.P., Jenkins, G.M., Reinsel, G.C., Ljung, G.M. (2015). Time Series Analysis: Forecasting and Control. John Wiley & Sons. <https://books.google.com/books?id=rNt5CgAAQBAJ>.
- [5] Kidson, J.W., Thompson, C.S. (1998). A comparison of statistical and model-Based downscaling techniques for estimating local climate variations. Journal of Climate, 11(4): 735-753. [https://doi.org/10.1175/1520-0442\(1998\)011%3C0735:ACOSAM%3E2.0.CO;2](https://doi.org/10.1175/1520-0442(1998)011%3C0735:ACOSAM%3E2.0.CO;2)
- [6] Shi, X., Chen, Z., Wang, H., Yeung, D.Y., Wong, W.K., Woo, W.C. (2015). Convolutional LSTM network: A machine learning approach for precipitation nowcasting. Advances in Neural Information Processing Systems.

- [7] Unnikrishnan, P., Jothiprakash, V. (2020). Hybrid SSA-ARIMA-ANN model for forecasting daily rainfall. *Water Resources Management*, 34(11): 3609-3623. <https://doi.org/10.1007/s11269-020-02638-w>
- [8] Ridwan, W.M., Sapitang, M., Aziz, A., Kushiar, K.F., Ahmed, A.N., El-Shafie, A. (2021). Rainfall forecasting model using machine learning methods: Case study Terengganu, Malaysia. *Ain Shams Engineering Journal*, 12(2): 1651-1663. <https://doi.org/10.1016/j.asej.2020.09.011>
- [9] Boonyuen, K., Kaewprapha, P., Srivihok, P. (2018). Daily rainfall forecast model from satellite image using convolution neural network. In 2018 International Conference on Information Technology (InCIT), Khon Kaen, Thailand, pp. 1-7. <https://doi.org/10.23919/INCIT.2018.8584886>
- [10] Novitasari, D.C.R., Supatmanto, B.D., Rozi, M.F., Farida, Y., Setyowati, R.D., Junaidi, R., Arifin, A.Z., Fatoni, A.R. (2020). Rainfall prediction based on himawari-8 IR enhanced image using backpropagation. *Journal of Physics: Conference Series*, 1501(1): 012011. <https://doi.org/10.1088/1742-6596/1501/1/012011>
- [11] Simanjuntak, F., Jamaluddin, I., Lin, T.H., Siahaan, H.A.W., Chen, Y.N. (2022). Rainfall forecast using machine learning with high spatiotemporal satellite imagery every 10 minutes. *Remote Sensing*, 14(23): 5950. <https://doi.org/10.3390/rs14235950>
- [12] Ionescu, V.S., Czibula, G., Mihuleț, E. (2021). DeePS at: A deep learning model for prediction of satellite images for nowcasting purposes. *Procedia Computer Science*, 192: 622-631. <https://doi.org/10.1016/j.procs.2021.08.064>
- [13] Liyew, C.M., Melese, H.A. (2021). Machine learning techniques to predict daily rainfall amount. *Journal of Big Data*, 8: 1-11. <https://doi.org/10.1186/s40537-021-00545-4>
- [14] Manandhar, S., Dev, S., Lee, Y.H., Meng, Y.S., Winkler, S. (2019). A data-Driven approach for accurate rainfall prediction. *IEEE Transactions on Geoscience and Remote Sensing*, 57(11): 9323-9331. <https://doi.org/10.1109/TGRS.2019.2926110>
- [15] Praveena, R., Babu, T.G., Birunda, M., Sudha, G., Sukumar, P., Gnanasoundharam, J. (2023). Prediction of rainfall analysis using logistic regression and support vector machine. *Journal of Physics: Conference Series*, 2466(1): 012032. <https://doi.org/10.1088/1742-6596/2466/1/012032>
- [16] Jocher, G. (2023). Ultralytics YOLOv8. <https://github.com/ultralytics/ultralytics>.
- [17] Barde, V., Nageswararao, M.M., Mohanty, U.C., Panda, R.K., Ramadas, M. (2020). Characteristics of southwest summer monsoon rainfall events over East India. *Theoretical and Applied Climatology*, 141: 1511-1528. <https://doi.org/10.1007/s00704-020-03251-y>
- [18] Sohan, M., Sai Ram, T., Rami Reddy, C.V. (2024). A review on yolov8 and its advancements. In *International Conference on Data Intelligence and Cognitive Informatics*. Springer, Singapore, pp. 529-545. https://doi.org/10.1007/978-981-99-7962-2_39
- [19] Wang, C.Y., Liao, H.Y.M., Wu, Y.H., Chen, P.Y., Hsieh, J.W., Yeh, I.H. (2020). CSPNet: A new backbone that can enhance learning capability of CNN. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, Seattle, WA, USA, pp. 390-391. <https://doi.org/10.1109/CVPRW50498.2020.00203>
- [20] Yaseen, M. (2024). What is yolov9: An in-depth exploration of the internal features of the next-Generation object detector. *arXiv Preprint arXiv: 2409.07813*. <https://doi.org/10.48550/arXiv.2409.07813>
- [21] Lin, T.Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S. (2017). Feature pyramid networks for object detection. In *Proceedings of The IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, pp. 2117-2125. <https://doi.org/10.1109/CVPR.2017.106>
- [22] Terven, J., Córdova-Esparza, D.M., Romero-González, J.A. (2023). A comprehensive review of YOLO architectures in computer vision: From YOLOv1 to YOLOv8 and YOLO-NAS. *Machine Learning and Knowledge Extraction*, 5(4): 1680-1716. <https://doi.org/10.3390/make5040083>
- [23] He, K., Zhang, X., Ren, S., Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, pp. 770-778. <https://doi.org/10.1109/CVPR.2016.90>
- [24] Performance metrics in machine learning: Complete guide (2023). <http://neptune.ai/blog/performance-metrics-in-machine-learning-complete-guide>.