भारत सरकार
अंतरिक्ष विभाग
अंतरिक्ष उपयोग केन्द्र
आंबावाडी विस्तार डाक घर,
अहमदाबाद-380 015. (भारत)
दूरभाष : +91-79-26913050, 26913060
वेबसाईट : www.sac.isro.gov.in/www.sac.gov.in

Government of India
Department of Space
SPACE APPLICATIONS CENTRE
Ambawadi Vistar P.O.
Ahmedabad - 380 015. (INDIA)
Telephone : +91-79-26913050, 26913060
website : www.sac.isro.gov.in/www.sac.gov.in

वैज्ञानिक अनुसंधान एवं प्रशिक्षण प्रभाग (एसआरटीडी)

Scientific Research and Training Division (SRTD)

आरटीएमजी/मीसा/सैक RTMG/MISA/SAC

CERTIFICATE

This is to certify that **Ms. Mansi Shah**, student of MSc (Mathematics) of Sardar Vallabhbhai Nation Institute of Technology, Surat, Gujarat has completed a four months (08 December-2023 to 08 April-2024) project on "**Predicting presence of Mesoscale Convective System and its Location using Brightness Temperature on INSAT 3D Satellite Data**" under the supervision of Dr. Bipasha Paul Shukla, Sci/Engr-SG, EPSA-AOSG-ASD, Space Applications Centre (ISRO). The research work was carried out through Scientific Research and Training Division (SRTD) of Space Applications Centre, Ahmedabad.

डॉ सर्वेश्वर व्यास / Dr. S.P. Vyas
प्रधान, वैज्ञानिक अनुसन्धान एवं प्रशिक्षण प्रभाग
Head, Scientific Research and Training Division
एसआरटीडी-आरटीएमजी-मीसा / SRTD-RTMG-MISA
अंतरिक्ष उपयोग केंद्र (इसरो)
Space Applications Centre (ISRO)
अंतरिक्ष विभाग / Department of Space
भारत सरकार / Govt. of India
अहमदाबाद / Ahmedabad - 380015.

Predicting presence of Mesoscale Convective System and its location using Brightness Temperature on INSAT 3D satellite data

Report submitted for completion of training

(Satellite Meteorology and OceAnography Research and Training)
Programme of Space Applications Centre

Submitted by
Mansi Shah
Int M.Sc. (Mathematics)

Under the guidance of

Dr Bipasha Paul Shukla
SCI/ENG – ASG/AOSG/EPSA
HTDD/AMHTDG/EPSA
Space Applications Centre, (ISRO) Ahmedabad

Institution
Sardar Vallabhbhai National Institute of Technology
Surat – 395007
Gujarat

SRTD-RTCG-MISA
Space Applications Centre (ISRO)
Ahmedabad, Gujarat

08 December 2023 to 08 April 2024

1. Introduction:

Mesoscale Convective Systems (MCSs) represent a significant meteorological phenomenon. MCS is a large cluster of thunderstorms that forms in the atmosphere. These big groups of storms can cover a large area and last for several hours. Their occurrence, primarily during warm seasons and under specific atmospheric conditions, poses substantial risks to communities, infrastructure, and agriculture. When they happen, they often bring heavy rain, strong winds, lightning, and sometimes even severe weather like hail or tornadoes.

MCSs originate from a complex interplay of atmospheric dynamics, including the rapid ascent of warm, moist air driven by strong updrafts, often referred to as "jump updrafts." These updrafts are fueled by convective available potential energy (CAPE), which represents the energy available for upward motion within the atmosphere. As the warm, moist air rises, it undergoes overturning, forming towering clouds and generating intense precipitation within the MCS. Simultaneously, downdrafts, spurred by the cooling effects of evaporation and precipitation, lead to the formation of "overturning downdrafts," driving cool, dense air toward the surface and creating a localized area of colder air known as a "cold pool." The interaction between the descending cold air and the surrounding warm, moist air generates a boundary known as a "gust front," marking the leading edge of the MCS. This gust front acts as a trigger for further storm development and organization within the system.

Understanding MCSs is crucial for weather forecasting because they can cause significant impacts. By using Machine learning models, scientists can analyze lots of weather data to better predict when and where MCSs might form and how they might behave. This helps improve weather forecasts, giving people more time to prepare and stay safe during severe weather events.

The project aims to develop an automated Python model to predict the presence of Mesoscale Convective System (MCS). Provided data which contains complete MCS tracking details in NetCDF files is analyzed to show that the variable Brightness Temperature and Cloud track number(MCS cloud track number mask) can be used in the modeling process of identifying MCS. A Convolutional Neural Network(CNN) model is trained using input columns as brightness temperature, date, month, year, hour, min and target column as MCS presence. The MCS presence column is calculated from cloud track number.

2. Methodology

Understanding the Mesoscale Convective System through mathematical models and analyzing change in MCS with respect to change in its variables.

↓

Data preprocessing and analysis of The Global MCS tracking dataset developed by Feng et al. (2021) from 2015 to 2020

↓

Training Convolutional Neural Network on Brightness temperature and Cloud Track number from The Global MCS tracking dataset to predict if MCS is present on a specific file or not.

↓

Analysis and Processing of INSAT 3D satellite TIR data over region of Interest to use in model to obtain predictions

↓

Validating Predictions on INSAT 3D satellite data through plots and Visualisation of same date Global MCS tracking dataset

↓

Performing predictions for 6 months of Daily INSAT 3D data from Jan 2016 to July 2016

Fig 1 - Methodology

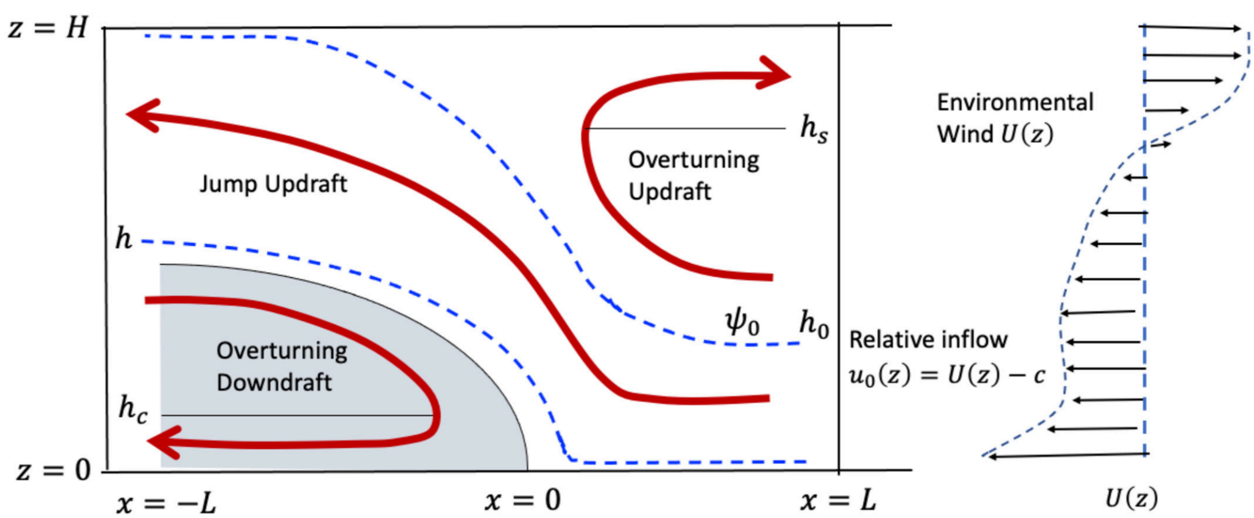
The platform which is used in the project is Jupyter Notebook and Google Colab. The language used for programming is Python.

2. Method:

2.1 Modeling MCS:

A prominent type of MCS is Squall Lines. A Squall line is a narrow, elongated band of thunderstorms characterized by a line or series of closely spaced cells. This weather phenomenon often forms along or ahead of a cold front, where warm and moist air is lifted and forced to rise rapidly over a relatively short distance.
It consists of three major sections: Jump updraft, Overturning updraft and Overturning downdraft.

A Jump updraft refers to a sudden and vigorous upward movement of air within a thunderstorm, often associated with strong updrafts. An Overturning downdraft occurs when a descending column of air within a thunderstorm becomes disrupted or tilted, leading to a change in its vertical motion. An Overturning updraft refers to a disruption or tilting of the upward-moving column of air within a thunderstorm.



Source : [An Analytical Model of Two-Dimensional Mesoscale Circulation and Associated Properties Across Squall Lines - Zhang - 2022 - AGU Advances - Wiley Online Library]

Fig 2 - MCS components

The storm relative speed is given as below where $U(z_0)$ is an environmental wind profile and c is the propagation speed.

$$u_0(z_0) = U(z_0) - c$$

2.2 Effect of changes in storm relative speed on MCS components

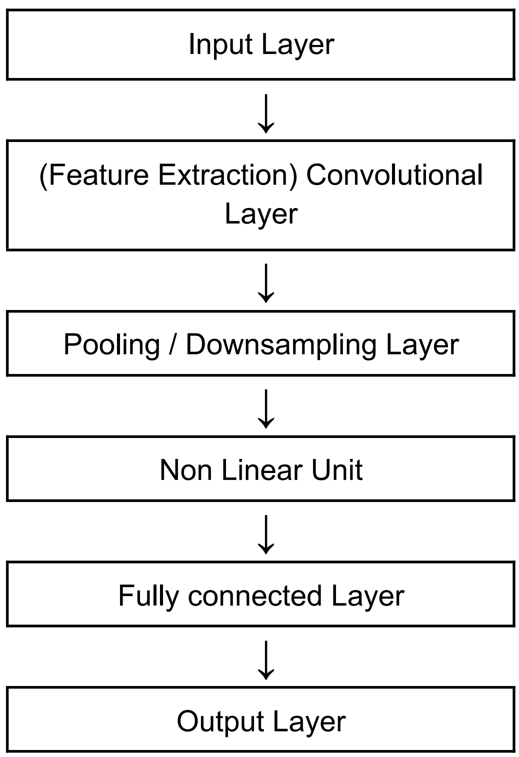
If we change the storm relative speed to other types of function for example Exponential then equation becomes $u(z_0) = ae^{z_0} - c$ and the overall effect would be that. The base state velocity would now increase exponentially with height i.e More Complex Wind profile. Stronger vertical shear would affect Updraft and Downdraft strength Potential formation of low level jets and changes in Squall line structure.

If we change storm relative speed to power function then equation becomes $u(z_0) = az_0^b - c$. The inflow velocity now varies with height according to a power law, potentially leading to different vertical wind profiles and influencing squall lines structure.

2.3 Training Model

Convolutional Neural Network(CNN) Model

Convolutional Neural Networks (CNNs) are a class of deep learning algorithms widely used for image recognition and computer vision tasks. CNNs operate by applying convolutional filters to input images, extracting features such as edges and textures. These filters slide over the image, performing mathematical operations that capture patterns and structures. Activation functions like ReLU introduce non-linearity, enabling the network to learn complex relationships between features. Pooling layers downsample the feature maps, reducing computational complexity while preserving important information. Fully connected layers further process the features, allowing the network to learn high-level representations and make predictions. During training, CNNs adjust their parameters through backpropagation to minimize the difference between predicted and actual outputs. This iterative process enables the network to learn hierarchical representations of input data, automatically extracting relevant features for the task at hand. By leveraging these hierarchical representations, CNNs excel in tasks such as image classification, object detection, and image segmentation.



Source - [Texture classification using convolutional neural network optimized with whale optimization algorithm | Discover Applied Sciences]

Fig 3 - Flowchart describing CNN

The basic Convolutional Neural Network (CNN) model consists of the following key components: convolutional layers, activation functions, pooling layers, fully connected layers, and a final output layer. Here's a summary with necessary equations:

Convolution Operation:
The convolution operation is the core of CNNs. It involves sliding a filter (kernel) over the input image and computing element-wise multiplications followed by a summation to produce a feature map.

$$\mathrm{Conv}(i,j) = \sum_{m=0}^{F_h-1} \sum_{n=0}^{F_w-1} \mathrm{Input}(i+m, j+n) \times \mathrm{Kernel}(m,n) + b$$

where Conv(i, j) is the value of the feature map at position (i, j), Input(i+m, j+n) is the pixel value of the input image at position (i+m, j+n), Kernel(m, n) is the filter value at position (m, n), b is the bias term, and F_h and F_w are the height and width of the filter respectively.

Activation Function:
After convolution, an activation function is applied element-wise to introduce non-linearity. Common activation functions include ReLU (Rectified Linear Unit):

$$\mathrm{ReLU}(x) = \max(0, x)$$

Pooling Operation:
Pooling layers are used to reduce the spatial dimensions of the feature maps, thereby reducing computational complexity and controlling overfitting. Max pooling is a common pooling operation:

$$\mathrm{MaxPooling}(i,j) = \max_{m,n \in \text{pooling window}} \mathrm{Conv}(i+m, j+n)$$

Fully Connected Layer:
After several convolutional and pooling layers, the feature maps are flattened into a vector and fed into fully connected layers (also called dense layers). Each neuron in the fully connected layer is connected to every neuron in the previous layer.

$$\mathrm{FC}(i) = \sum_{j=1}^{N} \mathrm{PreviousLayer}(j) \times \mathrm{Weight}(j, i) + b$$

where FC(i) is the value of neuron i in the fully connected layer, PreviousLayer(j) is the output of neuron j in the previous layer, Weight(j, i) is the weight connecting neuron j to neuron i, and b is the bias term.

Output Layer:
The output layer typically consists of neurons equal to the number of classes in the classification task. For multi-class classification, softmax activation is often used:

$$\mathrm{Softmax}(x_i) = \frac{e^{x_i}}{\sum_{j=1}^{N} e^{x_j}}$$

where x_i is the input to neuron i in the output layer and N is the total number of neurons in the output layer.

These equations represent the basic operations and structure of a CNN model, though in practice, there might be variations and additional components based on the specific architecture and task requirements.

2.1 Description of the Study Area:

The study area of the project is over India at particular latitude and longitude i.e **lon_min = 60.05, lon_max = 95.05 , lat_min = 0.05 and lat_max = 15.05**.

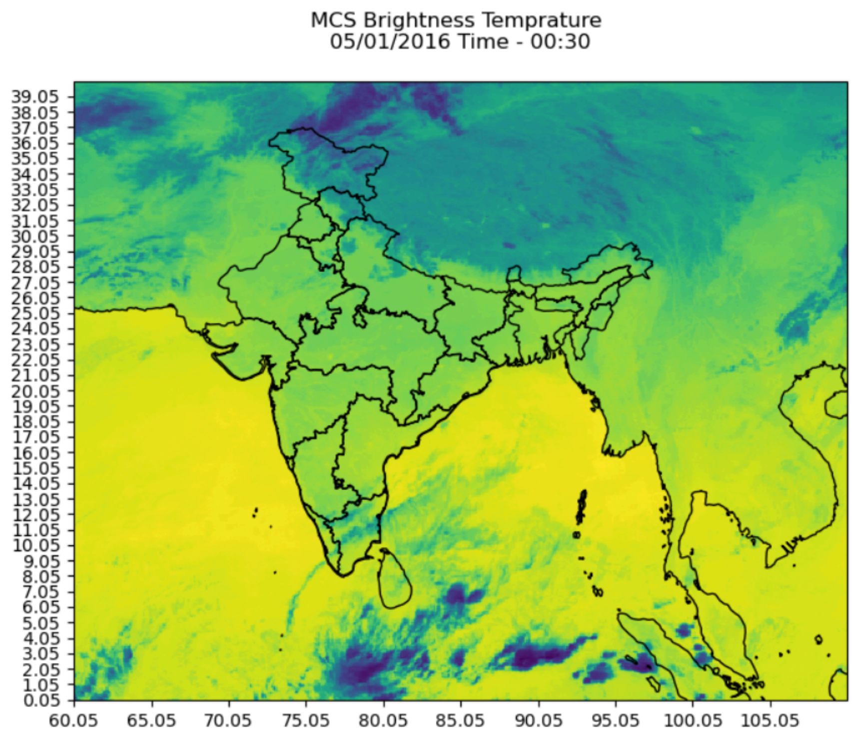


Fig 4 - BT over complete INDIA

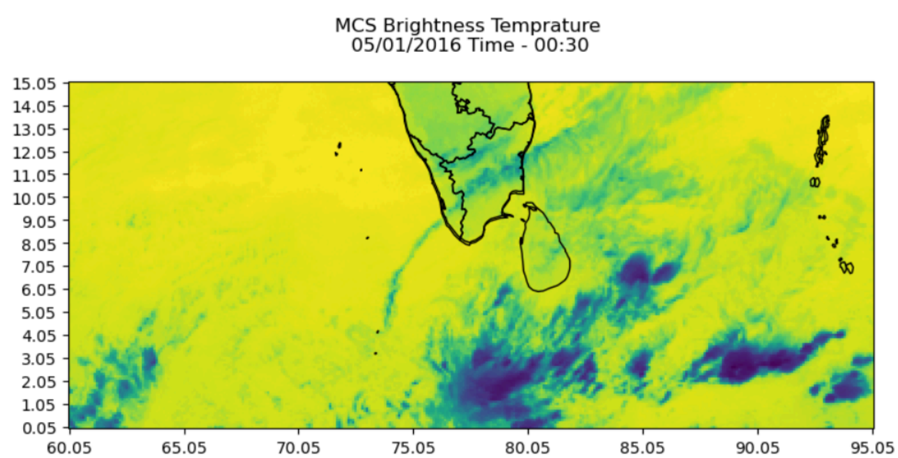


Fig 5 - BT over Required area

2.2 Data and Sources:

MCS_TRACKING DATA

The Global MCS tracking dataset developed by Feng et al. (2021). The current v2 of the dataset is produced by the Python FLEXible object TRacKeR (PyFLEXTRKR) algorithm (Feng et al., 2023). The period is from June 2000 to December 2020. The geographic coverage is 180W-180E, 60S-60N.

For the training model, a daily dataset from 5 years - 2015 to 2020 is used, therefore a total of 2,191 NetCDF files. Each NetCDF file in Global MCS tracking dataset includes variables such as brightness temperature, cloud track number(MCS cloud track number mask), cloud type, precipitation, base time, latitude, longitude and other variables which are used in tracking of MCS.

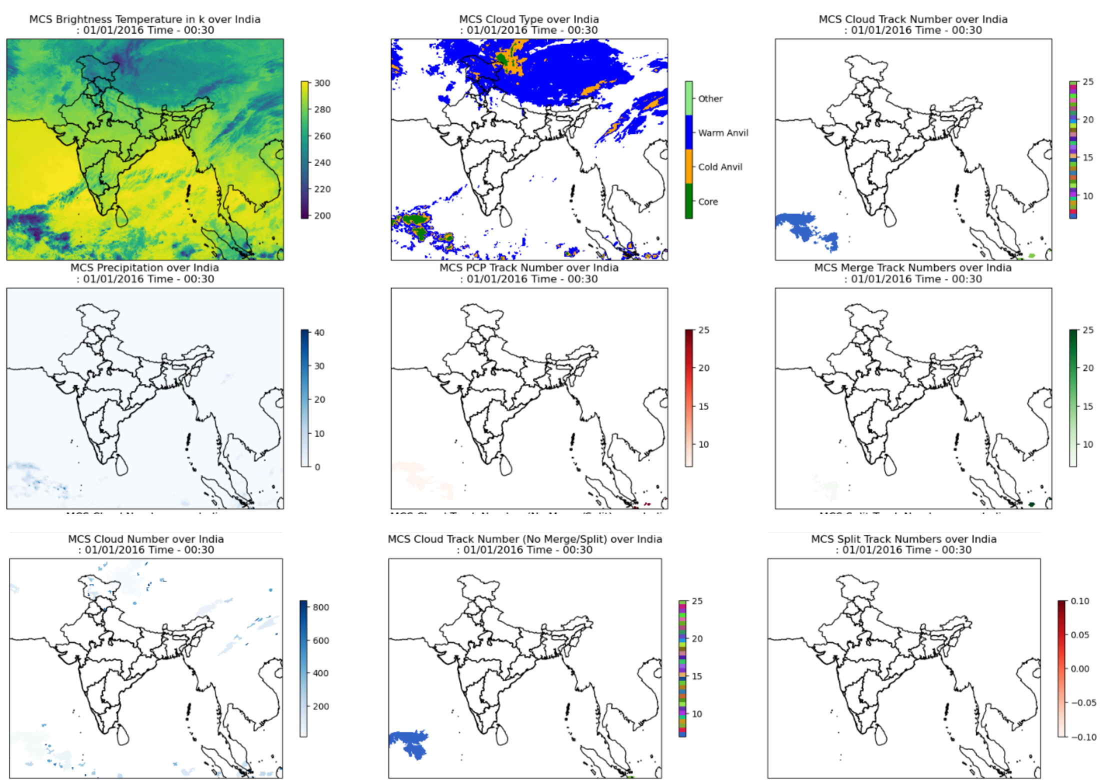


Fig 6 - Visualisation of All Variables in .nc file

INSAT-3D DATA

INSAT or the Indian National Satellite System is a series of multipurpose Geo-stationary satellites launched by ISRO to satisfy the telecommunications, broadcasting, meteorology, and search and rescue needs of India. Commissioned in 1983, INSAT is the largest domestic communication system in the Asia Pacific Region. The satellite is monitored and controlled by Master Control Facilities that exist in Hassan and Bhopal. INSAT-3D is a multipurpose geosynchronous spacecraft with main meteorological payloads (imager and sounder). The main objectives for this mission are to provide an operational, environmental and storm warning system to protect life and property. INSAT-3D is monitoring the earth's surface, oceanic observations and also provides data dissemination capabilities. It provides Broadcast Satellite Services (BSS) through two S-band transponders. The data acquisition and processing system is established at Space Applications Centre, Bopal Campus, Ahmedabad, India. The processing of INSAT-3D data takes place broadly in four steps.

1. Ground receiving system to receive data
2. Data Reception (DR) system to generate raw data (L0) files
3. Data Processing (DP) system to process L0 data and produce L1B data files (Calibrated and Geo located)
4. Product generation and Dissemination system

The INSAT 3D product named **3DIMG_L1C_SGP, Level1 IMAGER 6** with channel data of TIR1, TIR2, WV, VIS, SWIR, MIR Bands in Mercator projection L1C from 2016-01-01 2016-07-01, daily at time 00:30 a.m is used for generating predictions. Each file is in the form of a TIF file therefore 174 files. Each TIF file contains the IMG_TIR1 band with brightness temperature value over INDIA.

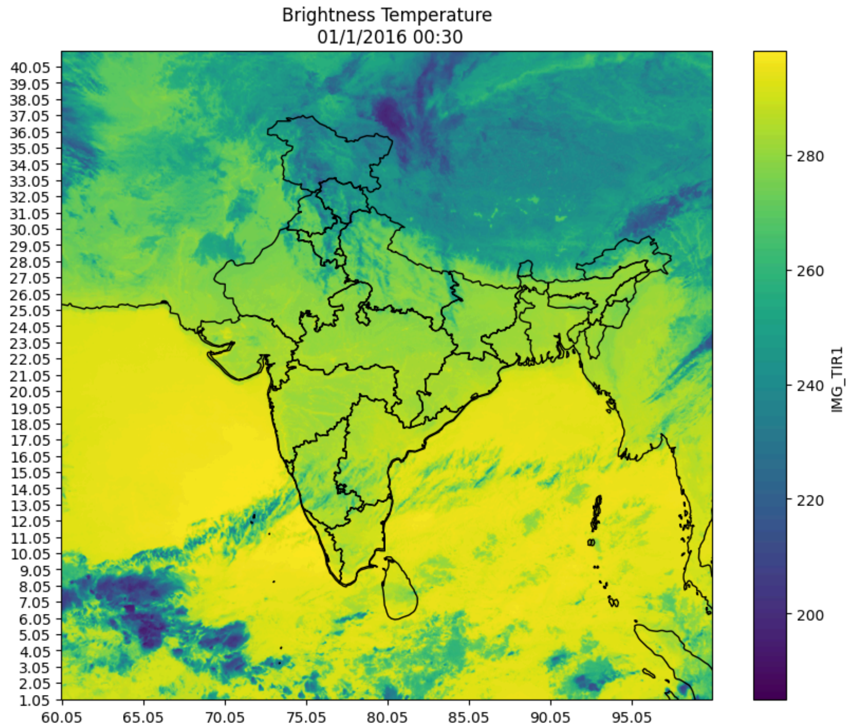


Fig 7 - TIR Band in INSAT 3D Tiff file

2.3 Data Processing and Analysis:

By understanding and analyzing each variable of data it has been noticed that cloud track number and brightness temperature are key factors in determining the presence of MCS. Using these two variables along with date, month, year and time the data frame has been prepared. With the help of cloud track number a new column introduced in the data frame which is MCS presence.Further the columns which are not much useful for the training process of the model are removed. The final data frame is shown in the below figure

	Year	Day	Month	Hour	Min	Brightness_Temp	MCS_Present
0	2015	01	01	00	30	[[[292.11508, 292.54602, 292.51083, 293.03833,...	1
1	2015	02	01	00	30	[[[291.7246, 292.05948, 292.27228, 292.40466, ...	1
2	2015	03	01	00	30	[[[283.20786, 280.60205, 280.8905, 284.9672, 2...	1
3	2015	04	01	00	30	[[[293.0, 292.91785, 292.9831, 293.0, 293.0, 2...	0
4	2015	05	01	00	30	[[[292.95917, 292.37076, 291.95767, 291.77173,...	1
...	...	...	...	...	...	...	...
2166	2020	27	12	00	30	[[[295.0, 295.0, 295.0, 295.0, 295.0, 295.0, 2...	1
2167	2020	28	12	00	30	[[[295.40466, 295.27228, 295.0169, 295.0, 295....	1
2168	2020	29	12	00	30	[[[296.0, 296.0, 296.0, 295.9567, 295.1557, 29...	1
2169	2020	30	12	00	30	[[[291.29767, 294.73572, 295.82108, 295.50784,...	1
2170	2020	31	12	00	30	[[[294.14984, 294.8783, 295.0, 295.0, 294.7585...	1

Table 1 - Training dataset in form of Data frame prepared from The Global MCS tracking dataset developed by Feng et al. (2021).

2.4 Model development and evaluation:

Reshaping the variables

Brightness temperature matrix is reshaped in the format required by the CNN model.

Feature Scaling

In order to ensure uniformity and convergence during model training, feature scaling is applied to the dataset using StandardScaler(). This process standardizes the distribution of features by removing the mean and scaling to unit variance.

Model Architecture

A neural network model is constructed using the Keras API with TensorFlow backend. The model architecture consists of three fully connected layers:

Input Layer: Dense layer with 32 units and ReLU activation function, corresponding to the number of features in the input data.

Hidden Layer: Dense layer with 64 units and ReLU activation function, providing nonlinear transformations to capture complex patterns in the data.

Output Layer: Dense layer with 1 unit and Sigmoid activation function, which outputs a probability score indicating the likelihood of the sample belonging to the positive class.

Model Compilation

The model is compiled using the Adam optimizer and binary cross-entropy loss function. Additionally, model performance will be evaluated based on accuracy metric during training.

Model Training

The model is trained on the training dataset for 10 epochs with a batch size of 32. Validation data is used to monitor model performance during training.

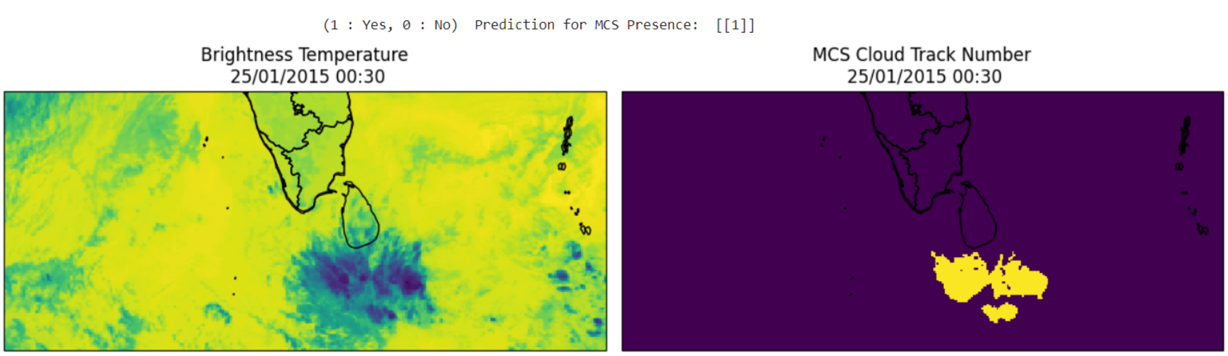
80 percent of data i.e 5 years from 2015 to 2020 are used to train the model and the year 2016 is used as a Test dataset.

Model Evaluation

After training, the model's performance is evaluated on the testing data. Predictions are made on the scaled test data to obtain probability scores for each sample. These probability scores are then thresholded at 0.5 to obtain binary predictions, indicating the predicted class labels. Finally, the accuracy of the model is calculated using the accuracy_score function. ***The Model provides a 90 % Accuracy on the test dataset.***

2.5 Validating the predictions from the model on 5 randomly selected files from the Global MCS tracking dataset through Plotting and visualizations.

5 Random days are picked from the dataset and used as input for the model. The plots of Brightness temp, cloud track number are generated to validate the predicted results



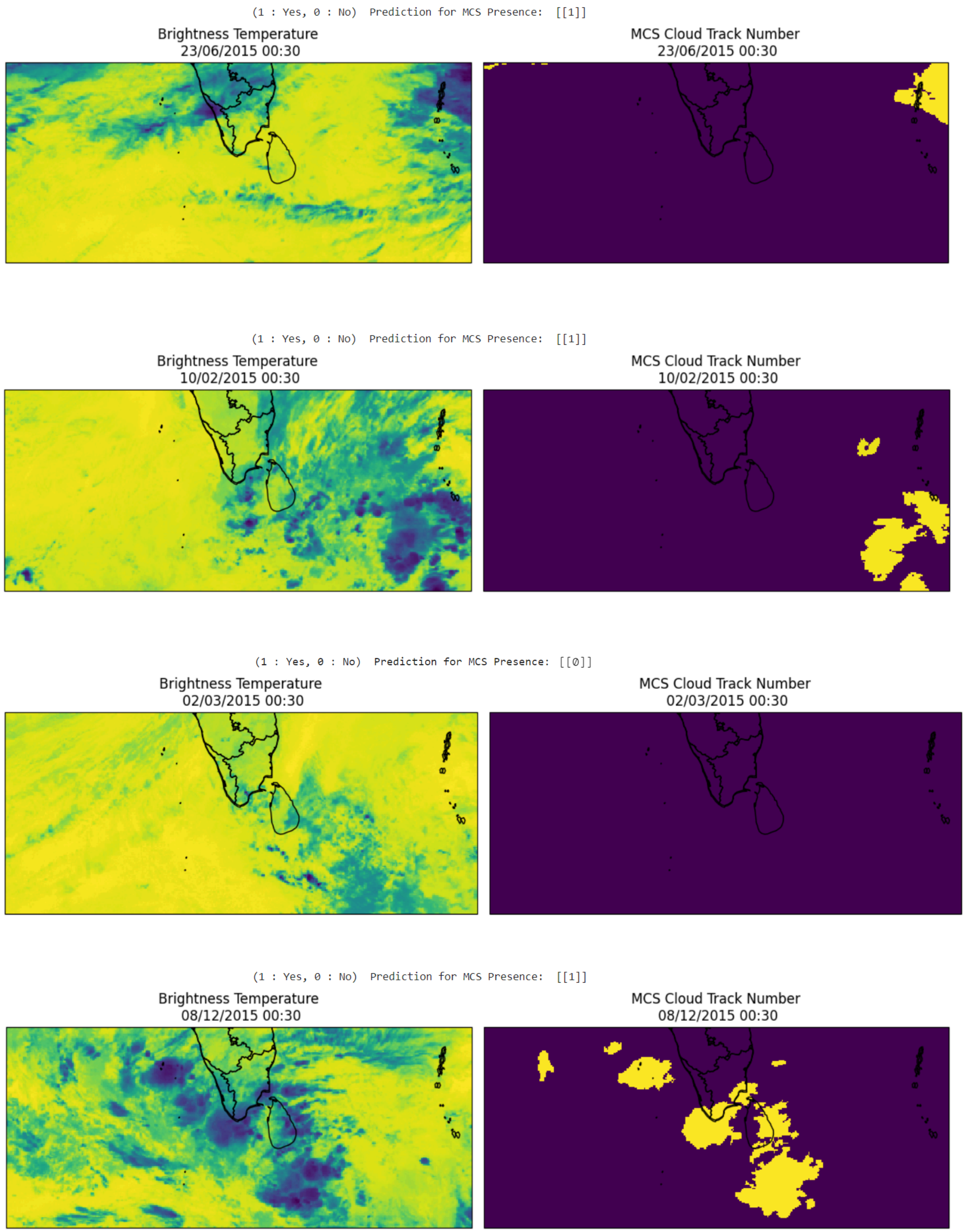


Fig - 8 Validating the predictions on custom inputs using plots

3. Preprocessing INSAT 3D data :

The Insat 3D data is in .h5 format which contains Brightness temperature in the form of Thermal Infrared 1 in variable IMG_TIR1. The matrix is cropped into the shape of required **lon_min = 60.05, lon_max = 95.05 , lat_min = 0.05 and lat_max = 15.05**

The shape of IMG_TIR1 is (1, 3207, 3062). After cropping it to the selected region and making it as 2D the shape of the file would become (-906:, :2808). To use the model for this data the reshaping would be needed. The shape which is required for the model is (151,351). Dividing the value (-453, 1053) with (3,3) will get (151,351) which is the required shape for the model. This process is downsampling.

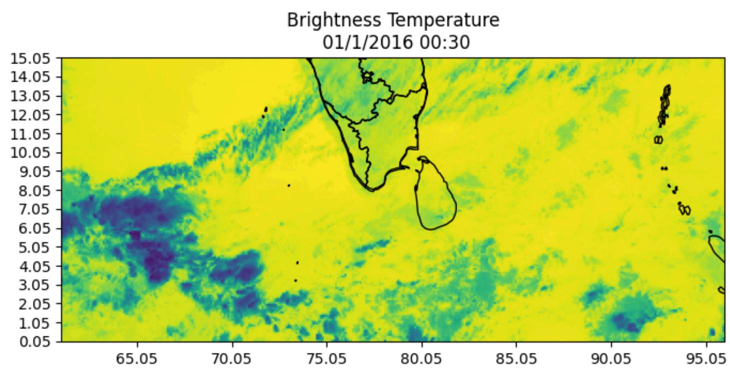


Fig 9 - Cropped to required Lat Lon

4. Results and Discussion:

4.1 Predictions on INSAT 3D satellite data.

The prepared CNN model is used to predict MCS on preprocessed INSAT 3D data.

	Year	Day	Month	Hour	Min	Brightness_Temp	MCS_Present	MCS_Presence_Pred
0	2016	01	1	00	30	[[[290.4385, 291.23294, 292.2755, 292.62637, 2...	1	1
1	2016	02	1	00	30	[[[285.12247, 286.28287, 288.46045, 291.0901, ...	1	1
2	2016	03	1	00	30	[[[293.2004, 293.52454, 293.8342, 293.65076, 2...	0	0
3	2016	04	1	00	30	[[[292.78372, 293.45877, 293.76105, 293.41476,...	1	1
4	2016	05	1	00	30	[[[294.02673, 293.97467, 293.91614, 294.1771, ...	1	1
...	...	...	...	...	...	...	...	...
173	2016	26	6	00	30	[[[266.8912, 273.44885, 279.57877, 280.1216, 2...	1	1
174	2016	27	6	00	30	[[[285.6744, 286.79126, 280.8583, 281.06522, 2...	1	1
175	2016	28	6	00	30	[[[288.44205, 289.39212, 286.63373, 278.33517,...	1	1
176	2016	29	6	00	30	[[[258.27982, 260.4826, 257.57382, 254.60864, ...	1	1
177	2016	30	6	00	30	[[[245.81519, 248.44516, 250.60356, 266.70102,...	1	1

Table 2 - INSAT six month data from 2016 January to June at 24 hour intervals i.e 178 files are used for predictions and added in MSC_presence_Pred column. The predictions are validated using the same date Global MCS tracking dataset.

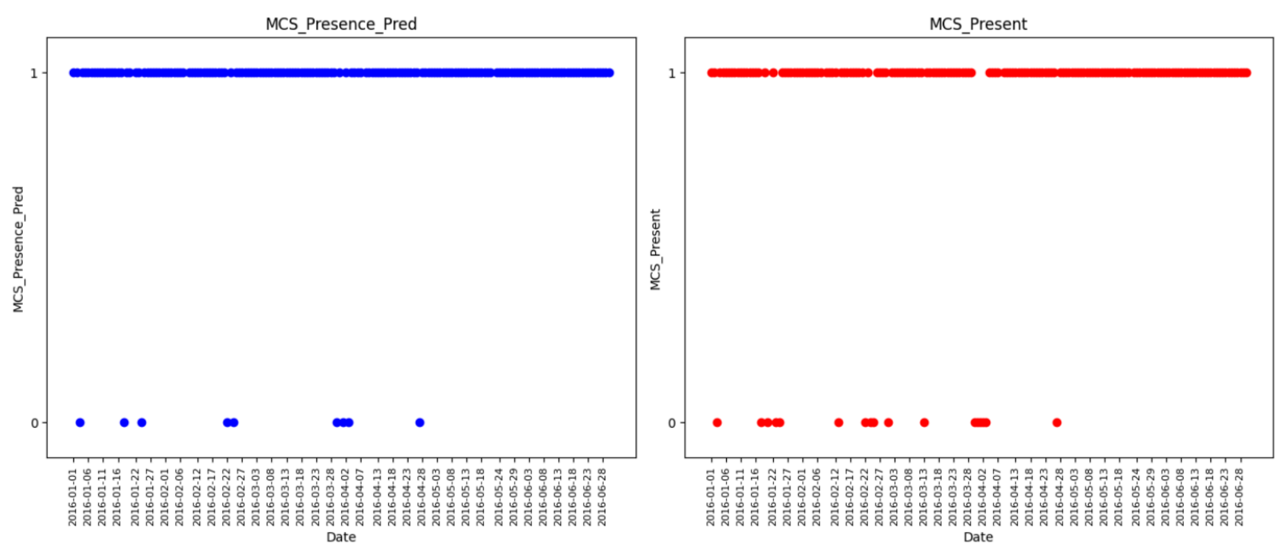
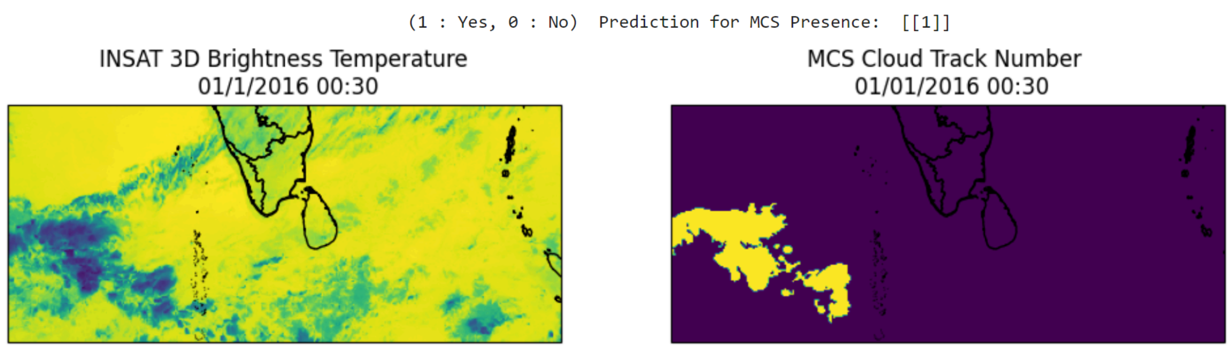


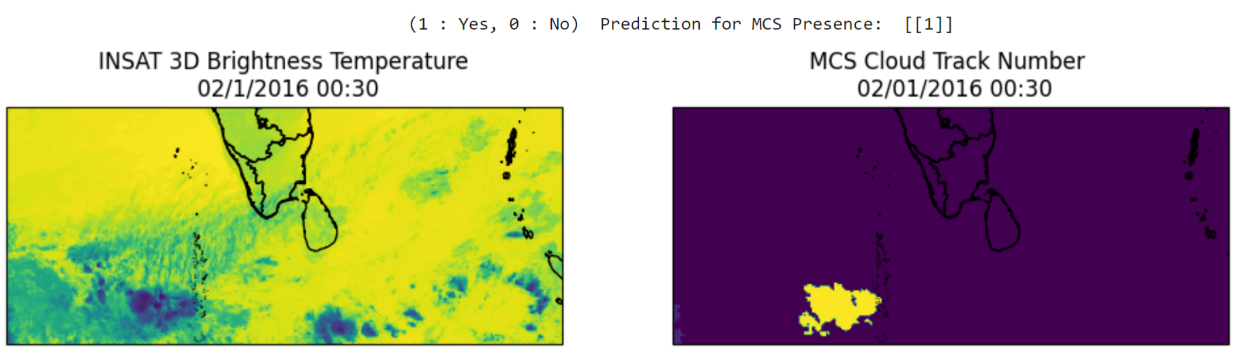
Fig 10 - Predictions of model plotted along with Actual data

Out of 178 files, 170 days predictions match and 8 days prediction does not match. **Accuracy - 95.50 on INSAT 3D satellite data.**

4.2 Validating the predictions from the model on INSAT 3D data through Plotting and visualizations.

After predictions, we validate the model's results with the same date .nc file data which has Cloud track Number to ensure results are correct.





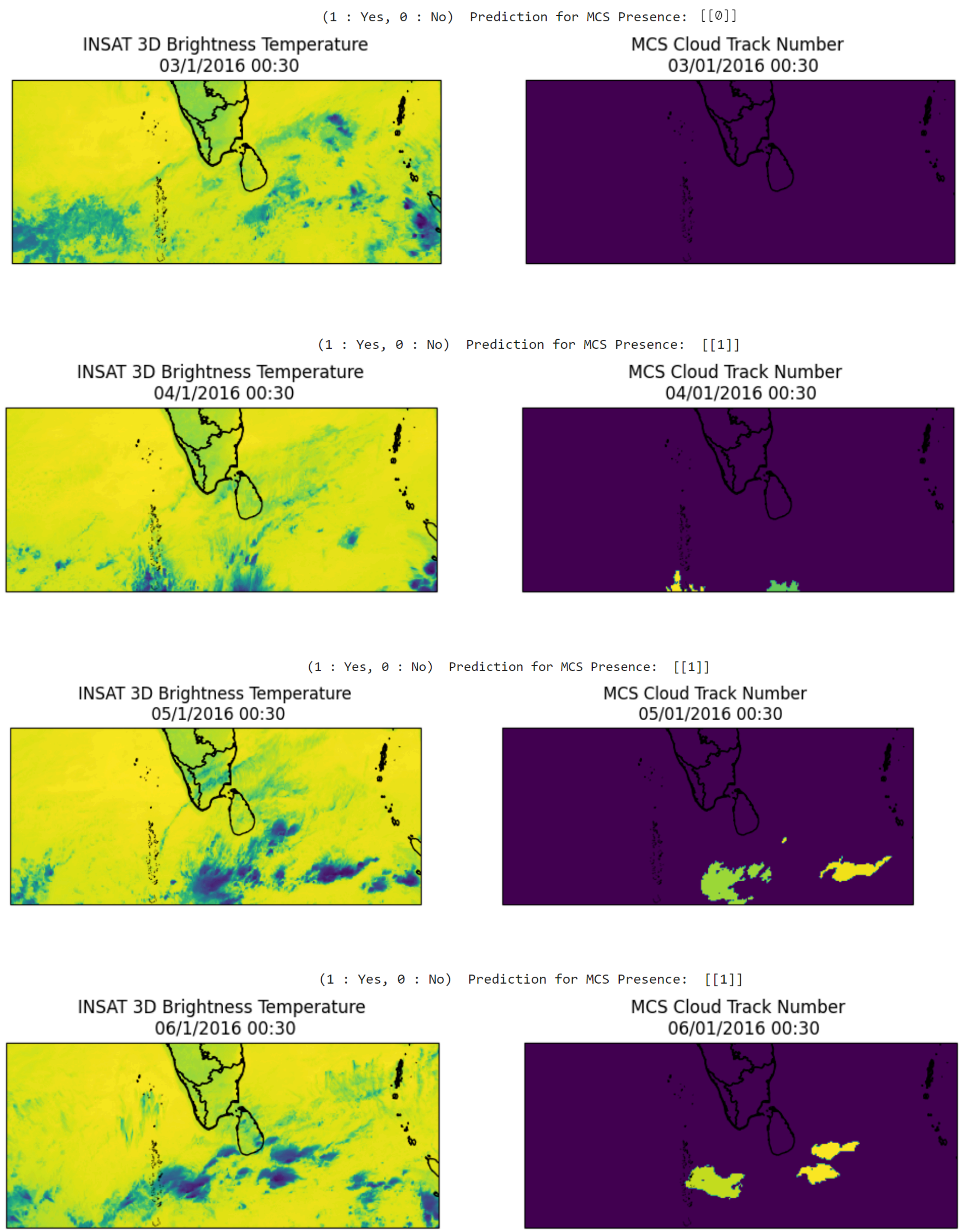


Fig 8 - Validation of Model result on INSAT 3D data using MCS plot generated by applying threshold on Brightness temperature.

Future Directions:

Computational resources were a limitation so in future following can be explored
- Training model on Hourly data which can increase accuracy of the model
- Further Implementation of a model which tracks the MCS path.

Reference

Zhang, Y., & Fu, Q. (2022). An Analytical Model of Two-Dimensional Mesoscale Circulation and Associated Properties Across Squall Lines. AGU Advances, 3(3), e2022AV000726. https://doi.org/10.1029/2022AV000726

Liu, H., Xie, P., & Li, J. (2020). A Global High-Resolution Mesoscale Convective System Database Using Satellite-Derived Cloud Tops, Surface Precipitation, and Tracking. Journal of Geophysical Research: Atmospheres, 125(18), e2020JD034202. https://doi.org/10.1029/2020JD034202

Kolios, S., Feidas, H., & Kotsopoulos, S. (2013). An automated nowcasting system of mesoscale convective systems for the Mediterranean basin using Meteosat imagery. Part I: System description. Meteorological Applications, 20(4), 425-434. https://doi.org/10.1002/met.1282

Venkatesan, R., Priyadharshini, P., & Paramasivam, P. (2019). Texture classification using convolutional neural networks optimized with whale optimization algorithms. SN Computer Science, 1(4), 1-9. https://doi.org/10.1007/s42452-019-0678-y

Cho, H., & Oh, H. (2000). Initialization of a fine-scale model for convective-system prediction: A case study. Quarterly Journal of the Royal Meteorological Society, 126(567), 963-982. https://doi.org/10.1002/qj.49712657004

Nesbitt, S. W., & Zipser, E. J. (2020). The formation, character and changing nature of mesoscale convective systems. npj Climate and Atmospheric Science, 3(1), 1-10. https://doi.org/10.1038/s43017-020-0057-7

Liu, H., Xie, P., & Li, J. (2020). A global high-resolution mesoscale convective system database using satellite-derived cloud tops, surface precipitation, and tracking. Journal of Geophysical Research: Atmospheres, 125(18), e2020JD034202. https://doi.org/10.1029/2020JD034202

Bryan, G. H., & Fritsch, J. M. (2010). On the predictability of mesoscale convective systems: Three-dimensional simulations. Monthly Weather Review, 138(3), 1095-1108. https://doi.org/10.1175/2009MWR2961.1