B.Tech. Project Part-2

Report on

Using Cloud Top Temperature for the Forecast of Cyclone track using Machine Learning Approach

Submitted By

Vrishabh Kenkre
19ME01009

Under the Supervision of

Dr. Srinivasa Ramanujam Kannan
Assistant Professor

School of Mechanical Sciences
Indian Institute of Technology Bhubaneswar

May 2023

Using Cloud Top Temperature for the Forecast of Cyclone track using Machine Learning Approach

Vrishabh Kenkre
19ME01009

Abstract:

Cyclones are one of the deadliest natural disasters that can cause immense destruction. Knowing about natural disasters like cyclones in advance helps in planning and preparations as they can be extremely dangerous. Numerous methods have been developed in the past to track cyclones and measure their severity after an eye has formed. The purpose of this report is to track the movement of cyclones prior to eye formation in order to quickly and effectively avert cyclone-related damage. This work aims to explore the possibility of forecasting the occurrence of cyclones and predicting their track at an advanced time span by studying “Cloud Top Temperature (CTT) in the Bay of Bengal and Eastern Coastal Plains region of India. For this research, the more frequently available TIR-1 data of the INSAT 3D satellite was used. We chose observations with a thermal infrared sensor because of its high spatial and temporal resolution. Data were collected regularly at six-hour intervals from T-20 to T-5 days, where "T" indicates the day of landfall. This analysis sheds light on cyclones Amphan, Bulbul and some BoB category deep depressions.

Keywords: Cyclone, Bay of Bengal, INSAT 3D, CTT, forecast, Deep Depression.

Contents **Page No.**

ABBREVIATIONS

ANNs	Artificial Neural Networks
BTD	Best Track Data
CNN	Convolutional neural network
ECMWF	European Centre for Medium-Range Weather Forecasts
ECP	Estimated central pressure
ERA5	ECMWF Reanalysis version 5
IBTrACS	International Best Track Archive for Climate Stewardship
IMD	Indian Meteorological Department
LPS	Low-Pressure System
LSTM	Long Short-Term Memory Network
MSSWS	Maximum Sustained Surface Wind Speed
NI	North Indian
NOAA	National Oceanic and Atmospheric Administration
RMC	Regional Meteorological Centre
RNN	Recurrent Neural Network
SI	South Indian
SST	Sea Surface Temperature
TCs	Tropical Cyclones
WMO	World Meteorological Organization

1. INTRODUCTION

1.1 Theoretical Background

1.1.1 What are cyclones?

Cyclones are large, rotating storms that occur due to winds blowing around a central area of low atmospheric pressure. These storms are known as hurricanes or typhoons in the Northern Hemisphere and as cyclones in the Southern Hemisphere. The winds in these tropical storms blow in a counter-clockwise direction in the Northern Hemisphere and in a clockwise direction in the Southern Hemisphere. A cyclone is any storm that rotates around a low-pressure centre, with a closed low-level atmospheric circulation, strong winds, and a spiral pattern of thunderstorms that produce heavy rain. Cyclones are characterized by internal spiral winds that rotate around a low air pressure zone. Tropical cyclones and subtropical cyclones, which are warm-core cyclones, are large-scale weather systems, while mesocyclones and tornadoes are smaller-scale systems. Cyclones have also been observed on other planets, such as Mars and Neptune. Cyclones form as a result of latent heat, which is fuelled by significant storm activity.

1.1.2 When and where do cyclones occur?

Cyclones begin in tropical sea regions during the summer and tend to move initially westward, then eastward towards higher latitudes.

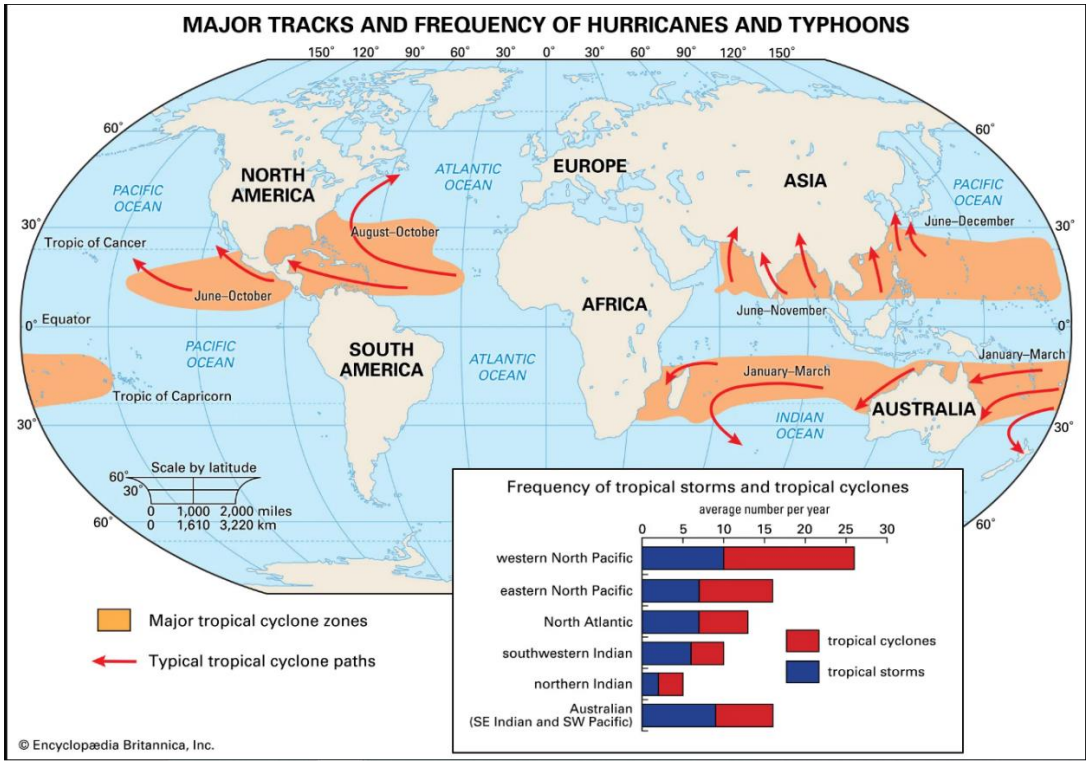


Fig 1:- Major Tracks and frequency of Hurricanes and Typhoons [2]

1.1.3 How do cyclones occur?

Cyclones typically form over warm seas near the equator. The sun heats the air, causing it to rise rapidly and creating areas of low pressure. This rising warm air carries a large amount of moisture, which condenses into massive thunderclouds. The surrounding air rushes in to fill the void left behind by the rising air, but due to the earth's rotation on its axis, the air is forced to bend inward and spiral upward. The swirling winds continue to intensify, forming a large circle that can be anywhere from 500-1000 km in diameter. At the centre of the storm is a calm, cloudless area known as the eye. In the eye of the cyclone, there is no rain and the winds are relatively light.

The development of cycle of tropical cyclones can be divided into three stages:

• Formation and initial development

• Full maturity

• Decay

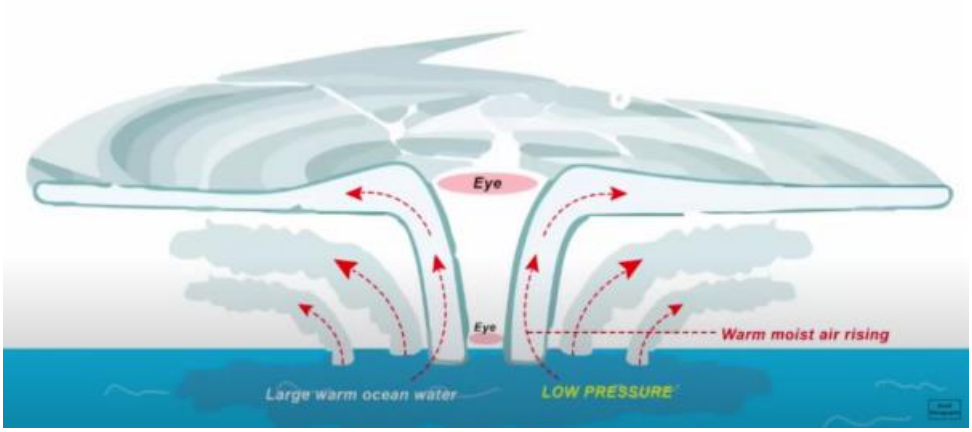


Fig 2:- Mechanism of Tropical cyclone formation [3]

1.1.3.1 Formation and initial development

• Low atmospheric pressure, which creates an area of reduced air density that encourages rising air to form the storm system.

• The Coriolis effect, which causes the rotating winds to form a circular pattern around the low-pressure center of the storm. The Coriolis effect is a result of the Earth's rotation, which causes air to deflect to the right in the Northern Hemisphere and to the left in the Southern Hemisphere. This effect helps to drive the rotation of the storm and strengthens it over time.

•Atmospheric instability refers to the condition where the atmosphere is unstable and tends to promote upward motion of air. This can result in the formation of massive vertical cumulus

cloud convection over the ocean, where the rising air cools and condenses to form clouds. This process can lead to the development of storms and cyclones as the clouds continue to grow and produce heavy precipitation. In addition, atmospheric instability can also induce strong updrafts and downdrafts, which can contribute to the intensification of storm systems by increasing the exchange of heat and moisture between the ocean and atmosphere. Overall, atmospheric instability is a critical factor in the formation and development of cyclonic storms over the ocean.

Together, these four conditions provide the necessary ingredients for the formation and intensification of a cyclonic storm.

1. The Tropical North Atlantic Ocean, including the Lesser Antilles and Caribbean, East of 70° W between July and October, North of the West Indies in June to October, Western Caribbean from June to early November, and Gulf of Mexico from June to November.
2. The Western North Pacific Ocean, including the Philippines, from May to November, although they can occur in all months.
3. The North Pacific, off the west coast of Central America, from June to October.
4. The Bay of Bengal and the Arabian Sea, from June to November to October.
5. The South Pacific Ocean, West of 140° W, from December to April.
6. The South Indian Ocean, from December to April, including the Northwest coast of Australia from November to April and west of 90° W from November to May. Tropical cyclones form over oceans worldwide except for the South Atlantic Ocean and Southeast Pacific.
7. A location of at least 4-5 latitude degrees from the equator allows the influence of the forces due to the earth's rotation to take effect for wind circulation around low-pressure centers.

1.1.3.2 Full Maturity

When the ocean and atmosphere conditions remain favourable, a cyclone may continue to strengthen and move towards higher latitudes. As it does so, the cloud system takes on a more circular shape and a well-defined eye form, which is a sign that the cyclone is at its most intense and hazardous stage. Roughly half of cyclones in this state go on to reach their peak intensity.

1.1.3.3 Decay

As a tropical cyclone moves over land, it begins to weaken due to the lack of warm, moist air from the ocean to sustain it. The central pressure begins to rise, and the eye becomes less defined or may disappear entirely. The high winds gradually weaken, and the storm system loses its circular shape. The severe weather associated with the cyclone, such as heavy rain and strong winds, may also decrease in intensity as the storm moves inland. However, the storm can still cause significant damage and flooding in areas along its path. Heavy rainfall may continue for several days, leading to flooding and landslides in some regions.

1.1.4 Classification of cyclones

Cyclones are categorized based on their wind velocity, but the classification differs depending on the region. In the US, they are classified into five categories based on the Saffir-Simpson scale (SS scale), which measures wind speed. The table below presents this classification.

Table 1:- Categorization of Cyclone based on Saffir-Simpson scale (Wind Velocity)

Category	Wind Speed in mph	Damage at Landfall	Storm surge in feet
1	74-95	Minimal	4 to 5
2	96-110	Moderate	6 to 8
3	11-130	Extensive	9 to 12
4	131-155	Extreme	13 to 18
5	>155	Catastrophic	19+

1.1.4 Hazards associated with cyclones

A cyclone poses three major threats that can lead to destruction. The first is storm surge, which is an abnormal rise in sea level caused by the cyclone that floods low-lying coastal areas, destroys infrastructure, erodes farmland and reduces soil fertility. The second threat is strong wind, which is powerful enough to topple buildings, uproot trees, and cause loss of life and property. The third danger is flooding, caused by heavy and prolonged rains, which can

waterlog low-lying areas, contaminate drinking water sources, and cause outbreaks of diseases like diarrhoea. Even after the cyclone has passed, flooding may still block transport routes, and the water may be contaminated with dead animals or spoiled food, posing a risk of infection to people.

1.1.5 Cloud top temperature

Baseline Cloud Top Temperature provides an estimation of the temperature of the top of a cloud in degrees Celsius. This estimation is considered to be more reliable than the brightness temperatures of individual channels, which may be influenced by gas absorption. This product is useful for tracking changes in the height of clouds during convection. It is also possible to set temperature thresholds for specific events using this product.

1.1.6 Information Entropy

Information theory defines entropy as a measure of the average level of uncertainty or surprise that is inherent to the potential outcomes of a random variable. Specifically, if we have a discrete random variable X that can take on values from an alphabet, and is distributed according to a certain probability distribution, then the entropy of X can be calculated as a function of that distribution.

$$p(x) = P[X = x] \quad (1)$$

$$\boldsymbol{H(x) = -\sum p(x) * log_2(p(x))} \quad (2)$$

where Σ denotes the sum over the variable's possible values. The concept of entropy in information theory refers to the average level of uncertainty or surprise inherent in a random variable's possible outcomes. This is calculated as the sum of the probability of each possible value multiplied by its logarithm to a chosen base. Different bases are used for different applications, with base 2 giving units of bits, base e giving natural units of nats, and base 10 giving units of dits, bans, or hartleys.

An alternative definition of entropy is the expected value of self-information, which is a function that increases as the probability of an event decreases. When an event is highly probable, its surprisal or self-information is low, but when its probability is close to zero, the surprisal is high. This relationship is captured by the function introduced by Claude Shannon in his seminal 1948 paper "A Mathematical Theory of Communication", and known as Shannon entropy. When p(E) is close to 1, the surprisal of the event is low, but if p(E) is close to 0, the surprisal of the event is high. This relationship is described by the function

$$log_2(1/p(E)) \quad (3)$$

OR

$$I(E) = - log_2(p(E)) \quad (4)$$

1.1.7 Thermodynamic Entropy

Entropy is one of the most important concepts in physics and in information theory. Informally, entropy is a measure of the amount of disorder in a physical, or a biological, system. The higher the entropy of a system, the less information we have about the system.

1.1.8 Pseudo-Adiabatic process

(Also called irreversible moist-adiabatic process.) A moist-adiabatic process in which the liquid water that condenses is assumed to be removed as soon as it is formed, by idealized instantaneous precipitation. The pseudo adiabatic process is only defined for expansion since a parcel that is compressed after such expansion will follow the dry-adiabatic lapse rate. A process similar to pseudo adiabatic descent can occur, however, if drizzle is evaporated into a relatively slow downdraft.

A set of approximate equations for pseudo adiabatic thermodynamics are developed. They are derived by neglecting the entropy of water vapour, and then compensating for this error by using a constant (but relatively large) value for the latent heat of vaporization. The subsequent formulations for entropy and equivalent potential temperature have comparable errors to previous formulations, but their simple form makes them attractive for use in theoretical studies. It is also shown that if the latent heat of vaporization is replaced with a constant value, then an optimal value should be chosen to minimize error; a value of 2.555 × 106 J kg−1 is found in tests herein.

The primary purpose of this formula is to demonstrate that a reasonably accurate formulation for θe that is applicable to precipitating clouds can be derived with a small number of reasonable assumptions.

$$s^r = (c_p + c_l r_t)\ln T - R\ln p_d + \frac{L_v r_v}{T} - R_v r_v \ln(\mathcal{H}). \quad (5)$$

wherein symbols are defined as follows: T is absolute temperature; pd is the partial pressure of dry air; R and Rv are respectively the gas constants for dry air and water vapor; cp is the specific

heat at constant pressure for dry air; cl is the specific heat of liquid water; rv and rt are respectively the mixing ratios of water vapor and total water; H ≡ e/es is relative humidity, wherein e is the partial pressure of water vapor, and es is the value of e at saturated equilibrium at the same temperature; and Lv is the latent heat of vaporization, which is a function of temperature according to Kirchhoff's Law, dLv/dT = cpv − cl , wherein cpv is the specific heat of water vapor at constant pressure. Total moist entropy (hereinafter referred to as reversible moist entropy, sr.

Surface entropy flux (wi)

$$wi = \frac{C_d * u_m * r_m}{\delta r}(s_b^0 - s_b^i) + C_k * u_m * (s_{SST}^0 - s_b^0) + \frac{C_d}{T_s} {u_m}^3 \quad (6)$$

C_k *and* C_d-Enthalpy and drag coefficients u_m is the maximum wind speed and r_m is mean radius. s_b^0, s_b^i cloud layer pseudo-adiabatic entropy in the inner region (r<r2) and outer regions (r>r2). $T_s$is the surface temperature and s_{SST}^0 is PAE at surface sea temperature.

$$\begin{aligned} C_{\mathrm{D}_{10\ \mathrm{m}}} &= [-0.053 u_{10\ \mathrm{m}} + 1.03] \times 10^{-3}, && \text{for } 0 \le u_{10\ \mathrm{m}} < 2.5\ \mathrm{m\ s^{-1}}, \\ C_{\mathrm{D}_{10\ \mathrm{m}}} &= [0.042 u_{10\ \mathrm{m}} + 0.8] \times 10^{-3}, && \text{for } 2.5 \le u_{10\ \mathrm{m}} < 33\ \mathrm{m\ s^{-1}}, \\ C_{\mathrm{D}_{10\ \mathrm{m}}} &= [2.18] \times 10^{-3}, && \text{for } 33 \le u_{10\ \mathrm{m}} < 40\ \mathrm{m\ s^{-1}}. \end{aligned} \quad (7)$$

1.1.9 Air-Sea thermodynamic Disequilibrium

Air-sea thermodynamic disequilibrium is calculated as $s_0^* - s_{600}$ where s^* is pseudo-adiabatic entropy of saturation at sea surface temperature and s_{600} is pseudo-adiabatic entropy at 600hPa. More the disequilibrium, more is the chances for genesis of cyclone.

1.1.10 Brightness temperature

Brightness temperature is the temperature at which a black body in thermal equilibrium with its surroundings would have to be in order to duplicate the observed intensity of a grey body object at a frequency nu. TIR-BT (Thermal infrared Brightness temperature) of 230-240 K, TIR-BT < 210 K, and TIRBT < 235 K are indicating the presence of convective clouds, deep convective cloud and high cloud respectively. TIR-BT < 208 K is a deep convective precipitating cloud. For our analysis we take three thresholds of TIR-BT namely deep cold clouds (TIR-BT < 210 K), cold clouds (210 K< TIR-BT < 240 K) and warm clouds (TIR-BT > 240 K).

1.2 Cyclone forecasting

1.2.1 What is Tropical cyclone forecasting?

Tropical cyclone forecasting is the practice of predicting where a tropical cyclone and its impacts are likely to be at a specific time in the future. It involves various components such as track, intensity, rainfall, storm surge, tornado, and seasonal forecasting. While track prediction accuracy has improved, intensity prediction accuracy has not seen much progress in recent years. Seasonal forecasting began in the 1980s in the Atlantic basin and later expanded to other regions.

In the early 1900s, identifying tropical cyclones relied on changes in weather conditions, sea level changes, and reports from affected areas, which left little time for early warnings, resulting in high death tolls. However, advancements in surveillance networks and technology, particularly the use of meteorological satellites in the 1960s, led to improved early detection and tracking of tropical cyclones.

1.2.2 Tracking and Forecasting

1.2.2.1 Synoptic scale meteorology

An extratropical cyclone is a synoptic-scale low-pressure weather system that has neither tropical nor polar characteristics, associated with fronts and horizontal gradients of temperature and dew point, otherwise known as "baroclinic zones".

This paragraph describes what an extratropical cyclone is and how it differs from tropical and polar cyclones. It is a low-pressure system associated with temperature and dew point gradients known as "baroclinic zones," occurring in the mid-latitudes and often referred to as "mid-latitude cyclones" or "depressions" by forecasters and the public. Extratropical cyclones control the weather over much of the Earth along with anticyclones. While they typically form along temperature and dew point gradients, they can become barotropic late in their life cycle. Under certain conditions, an extratropical cyclone can develop into a subtropical storm and then into a tropical cyclone.

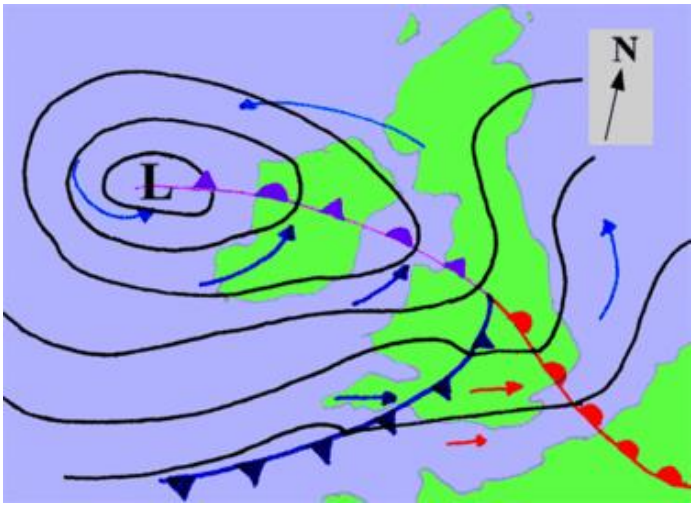


Fig 3:- Baroclinic Zones of a Hurricane [4]

1.2.2.2 One-two-three rule (1-2-3 rule)

The 1-2-3 rule, also known as the mariner's 1-2-3 rule or danger zone, is a guideline used by mariners to monitor and forecast severe storms, specifically hurricanes and tropical storms. It has two parts. The first part, the 34 knot rule, warns mariners to avoid a dangerous area. The second part, the 1-2-3 rule, is based on rounded long-term National Hurricane Center / Tropical Prediction Center (NHC/TPC) forecast errors of 100-200-300 nautical miles per 24-48-72 hours, respectively. These figures were close to the ten-year average for the period 1982-1991. However, due to improved accuracy, forecast errors have decreased to almost 50-100-150 nautical miles. To create the "danger zone," the forecast path is expanded by a radius corresponding to hundreds of miles plus forecast wind field radii of 34 knots.

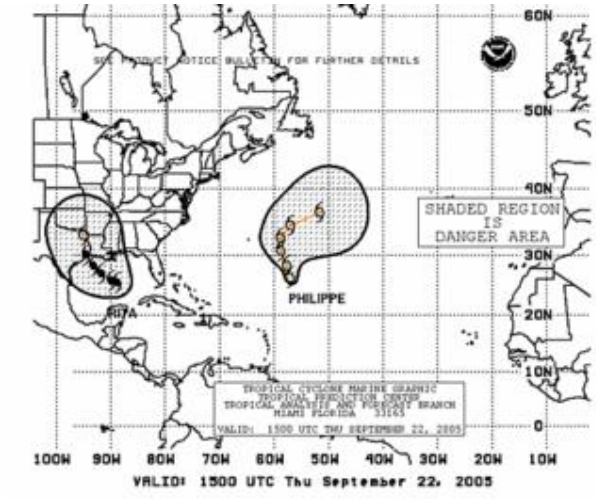


Fig 4:- Forecast of tropical cyclone using Mariner's 1-2-3 rule [5]

1.2.2.3 Use of satellites and aircraft

Several countries operate geostationary satellites that remain fixed in position above Earth and continuously capture images in visible and infrared wavelengths. Infrared images are crucial for tracking tropical cyclones as they reveal cloud top temperatures and can detect loosely and deeply organized convection associated with easterly waves and tropical cyclones, respectively. By observing specific cloud patterns, satellite images can also estimate the intensity of a storm. While satellite imagery provides general information on the location and intensity of tropical cyclones, detailed information on their strength and structure can only be obtained from aircraft, which is necessary for providing accurate warnings.

1.2.2.4 Landfall forecasts

Various national weather services designated as Regional Special Meteorological Centers (RSMCs) by the World Meteorological Organization (WMO) track the tropical cyclones that form in the world's ocean basins. To determine the current location and intensity of the storm, forecasters use various observation data from satellites and aircraft. They combine this information with computer forecasting models to predict the path and intensity of future storms. There are three main types of computer models available for this purpose. The first type is the simplest, which uses a statistical relationship based on the typical path of the storm in the region and the assumption that the storm will continue to move in the same direction. The second type is the statistical-dynamic model, which predicts large-scale circulation by solving equations that describe changes in atmospheric pressure, wind, and humidity. This model then uses statistical relationships that predict storm tracks based on large-scale conditions to forecast future storm conditions. The third type is the purely dynamic forecasting model, which solves equations describing changes in large-scale circulation and tropical cyclones. This model demonstrates the interaction of tropical cyclones with the environment but requires the use of powerful computers and a detailed description of the structure of tropical cyclones and their environment. Although current computer models are good at predicting the path of tropical cyclones, they cannot be fully trusted to predict changes in intensity beyond 24 hours in advance.

Once the forecasters detect that a tropical storm is potentially going to hit a land area, they issue a warning for that particular region. The observers use available observations and computer model output to give a 72-hour estimated track and peak wind speeds forecast, known as the "best track" forecast. By using the strike probability forecast, which determines the possibility of a tropical cyclone hitting a specific area during a particular time, local authorities

can trigger warning and evacuation plans. As the storm gets closer, a tropical cyclone watch is released for threatened regions. Every hour, evacuation procedures can begin in the vulnerable areas. Within 24 hours of a storm's arrival, a tropical cyclone warning is released if the tropical cyclone conditions are expected to occur. It is recommended to evacuate to the regions that may not be affected by storm surge and high water after a warning is released.

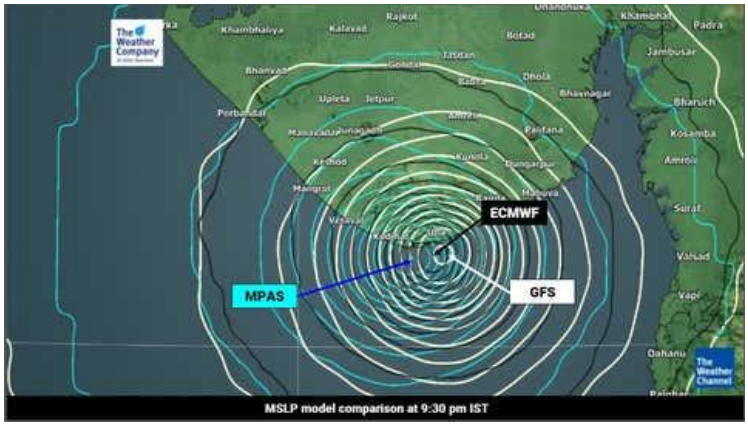


Fig 5:- Cyclone Tauktae landfall imapact [6]

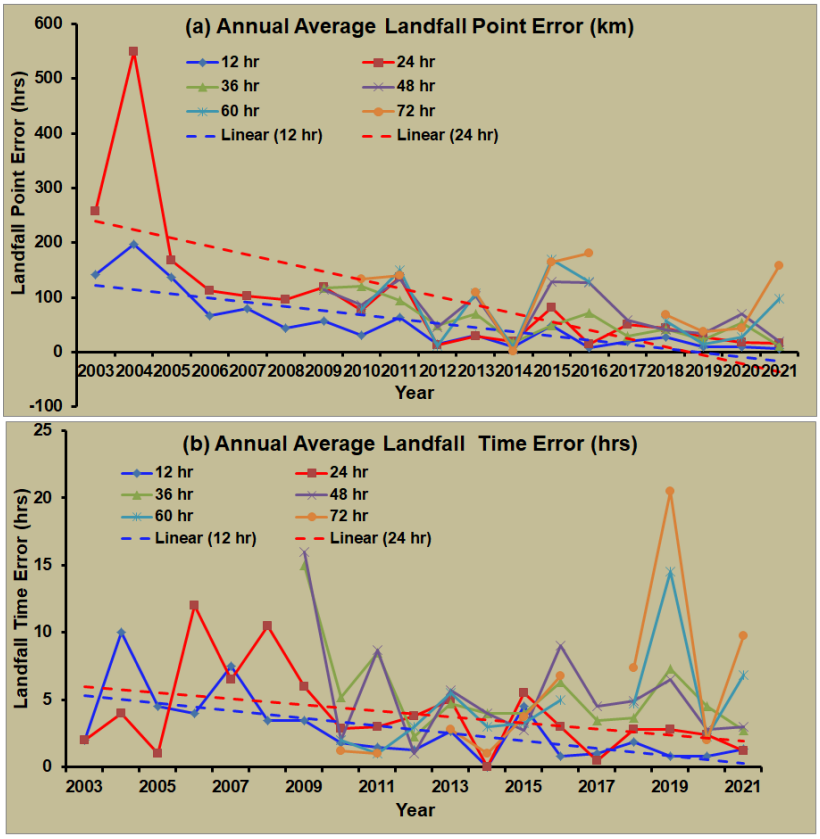


Fig 6:- Average landfall prediction error of RMSC [7]

1.2.2.5 Mathematical model for cyclone intensity

Around 1988, Dr. Kerry Emanuel developed a mathematical model known as the maximum potential intensity (MPI), which calculates the highest possible intensity that a tropical cyclone can reach based on sea surface temperature and atmospheric profiles from the most recent global model run. Using this equation, maps can be generated that illustrate the maximum achievable intensity based on the atmospheric thermodynamics at the time of the latest model run, either at 0000 or 1200 UTC. However, the MPI model does not consider the effects of vertical wind shear. MPI is computed using the following formula:

$$V = A + B \times e^{C(T-T_0)} \quad (8)$$

The equation for computing the maximum potential velocity in meters per second (V) takes into account the sea surface temperature (T) under the center of the tropical cyclone, with T_0 representing a reference temperature of 30°C, and constants A, B, and C that are curve-fitted. When the values of A, B, and C are set at 28.2, 55.8, and 0.1813 respectively, the resulting graph corresponds to the 99th percentile of empirical data on tropical cyclone intensity.

1.2 History of Development

Currently, the forecasting of tropical cyclones involves the use of multiple guidance models to predict both track and intensity. Dynamic models are based on the physical equations governing atmospheric motions and incorporate various physical models such as convective schemes, cloud microphysics, and ocean and land surface models, among others. These models are computationally intensive, and therefore, current practice involves using older model runs that are modified to be considered early methods. Consensus-driven methods or ensemble methods are also used to combine different dynamic models. On the other hand, statistical models rely on historical relationships between storm behavior and other parameters. However, these models are based on simple regressions and few statistical elements. By using state-of-the-art machine learning methods and incorporating large atmospheric data into statistical models, we can improve the accuracy of predictions and reduce computation time. Regional specialist weather centers such as the US Official NHC Forecast (OFCL) use consensus-driven or ensemble methods to provide real-time forecasts. The Global Ensemble Forecast System uses up to 20 different dynamic models.

1.3 Applications

Forecasting the future location and intensity of tropical cyclones is considered the most important function of tropical cyclone warning centres. One reason for the importance is the potential damage and loss of life that can occur when an intense tropical cyclone passes through. Early and accurate warnings do not necessarily eliminate the risk of damage or loss or life, but the effects can be greatly reduced.

1.4 Advantages and disadvantages

The impact of tropical cyclones is extensive and includes severe rain, high winds, storm surges near coastal areas, and tornadoes. The devastation caused by a cyclone is mainly dependent on its location, size, and intensity. A tropical cyclone can alter the landscape near coastal areas by destroying the forest canopy, reshaping sand dunes, and causing erosion along the coast. Even inland, the heavy rainfall can result in landslides in mountainous areas. The effects of a tropical cyclone can be observed for some time by monitoring the concentration of oxygen-18 isotope in caves. After the cyclone has passed, the destruction continues with fallen trees blocking roads, slowing down rescue efforts, and hindering repairs to power, communication, and water infrastructure. The standing water caused by the storm can also lead to the spread of diseases. The positive effects of tropical cyclones include bringing rain to arid regions and moving heat from the tropics towards the poles. In the sea, ships take advantage of the characteristics of weaker western cyclones. With nearly 2 million people worldwide having died due to tropical cyclones, accurate forecasting is critical to prevent potential damage and loss of life.

2. REVIEW OF THE TOPIC

2.1 Literature Review

Based on the article - Forecast of Tropical Cyclone Occurrences based on Fuzzy Logic Algorithm.

2.1.1 Why forecasting cyclones and implementation of fuzzy logic?

This research study proposes a novel approach to tropical cyclone (TC) forecasting by examining and predicting the occurrences of TCs rather than focusing on wind track and radius predictions or warnings. The study utilizes fuzzy logic (FL) models to predict TC occurrence based on five primary parameters of TC genesis, including low-level relative vorticity, upper-tropospheric horizontal wind, sea surface temperature, equivalent potential temperature, and specific humidity. To develop the FL model, trapezoidal and triangular fuzzy membership functions were used for the input and output variables, and fuzzy rules were derived from TC genesis parameter data with a daily calculation period spanning from 1989 to 2018. The FL model was created using the amount of TC genesis parameters when the lowest, middle, and upper tertile reconstruction cyclone occurred as a threshold. The results showed that the model satisfactorily simulated TC occurrence with an accuracy of 0.75, indicating that FL methods can be used effectively for forecasting TC occurrence. This research provides a promising avenue for improving TC forecasting and can contribute to better preparedness and prevention efforts in vulnerable regions.

2.1.2 Fuzzy logic algorithm approach

The Mamdani method is a widely used approach in FL algorithm for control and forecasting applications. It follows a clear and distinct set of procedures including fuzzification, logical decision-making, and defuzzification. On the other hand, the Takagi-Sugeno method lacks a clear defuzzification method, making it less commonly used in FL applications. In Mamdani's method, each IF-THEN rule generates a fuzzy set for the output variable, which requires a defuzzification step to obtain a crisp value for the output variable. Therefore, this study utilizes the Mamdani method to construct FL rules for predicting tropical cyclone occurrences based on five primary parameters of TC genesis.

2.1.3 TC forecasting model

Before determining the threshold of the fuzzy membership function, the paper focuses on the investigation of delayed input data to predict the occurrence of TC and TD. We used the cross-correlation method to evaluate the relationship between the lag time of each input variable and the occurrences of both TC and TD. The lag time with the best biserial correlation with TC and TD occurrences was chosen as the predictor. We further categorized the predictor threshold based on the 33rd percentile of TC and TD incidence for the period 1991-2010 and divided it into three classes (low, medium and high). After the fuzzy rule base is identified and the defuzzification algorithm is selected, the forecast accuracy is tested using a different set of historical data (training data) than that used to obtain the rule base. If unsatisfactory, the number of fuzzy membership functions and/or the shape of the fuzzy membership functions can be changed and a new FL rule base is obtained. The iterative process of rule base design, defuzzification algorithm selection, and system performance testing can be repeated several times with different numbers of fuzzy membership functions and/or different fuzzy membership shapes. The contingency table, also known as the pivot table or crosstab, is a method for evaluating the accuracy of a predictive model. The table consists of a matrix of rows and columns, with each row representing the actual classification of the data and each column representing the predicted classification. The diagonal cells of the matrix represent the cases where the predicted classification matches the actual classification, while the off-diagonal cells represent the cases where the prediction is incorrect.

Categorical statistics can be derived from the contingency table, including accuracy, precision, recall, and F1 score. Accuracy measures the proportion of correct predictions overall, while precision measures the proportion of true positives among all positive predictions. Recall measures the proportion of true positives among all actual positives, and the F1 score is a combination of precision and recall that provides a single measure of the model's performance.

In this study, the contingency table will be used to evaluate the performance of the FL model in predicting TC occurrence. The results will be analyzed using categorical statistics to determine the accuracy, precision, recall, and F1 score of the model.

Table 1. Contingency Table Form

Forecast	Observed		Total
	Yes	No	
Yes	Hits	False alarms	Forecast yes
No	Misses	Correct negatives	Forecast no
Total	Observed yes	Observed no	Total

The four combinations of forecasts (yes or no) and observations (yes or no), called the joint distribution, are hits (event forecast to occur, and did occur), misses (event forecast not to occur, but did occur), false alarm (event forecast to occur, but did not occur) and correct negative (event forecast not to occur, and did not occur).

$$Accuracy\ (AC) = \frac{Hits + Correct\ negatives}{Total} \tag{9}$$

Accuracy (AC) is a commonly used performance measurement algorithm in machine learning and classification tasks. It measures the overall correctness of the model's predictions and is calculated as the sum of the correct classifications divided by the total number of classifications. In the context of this study, the accuracy is determined using the contingency table and represents the proportion of the correct number of TC and TD predictions.

2.1.3 Summary

This study introduced a new approach to predicting the occurrence of tropical cyclones (TC) using the unsupervised machine learning FL algorithm. The authors developed FL rules and chose a defuzzification algorithm based on the 33rd percentile of historical TC and TD data from 1991-2010. The FL model accurately simulated TC occurrences in the study area, with a satisfactory level of error. The accuracy score of 0.75 indicates good agreement with observed TC and TD events. To facilitate user interaction with the TC genesis forecast, the authors used the MATLAB® commercial FL Toolbox to create a Graphical User Interface (GUI).

2.2 Major Focus Areas/ Objectives

- Fetching the data required for running analysis and converting the Cloud top Temperatures readings from satellite data into an Image-Time series.
- To find/predict a pattern in data, i.e.- the pattern of how tall clouds, Ice clouds spatially or temporally change- Increase or decrease in number, spread out or converge, increase or decrease in curvature, etc.
- Once we recognize the pattern, we need to check if the same pattern is followed in other cyclones as well as the depressions that did not result in the cyclone.
- Study of the neural network approach is to be done which will extract its features. Check if similar features are present in every cyclone formation.
- Make a Machine learning model that will predict the occurrence of cyclones 5-7 days before its occurrence based on the data fed with sufficient accuracy

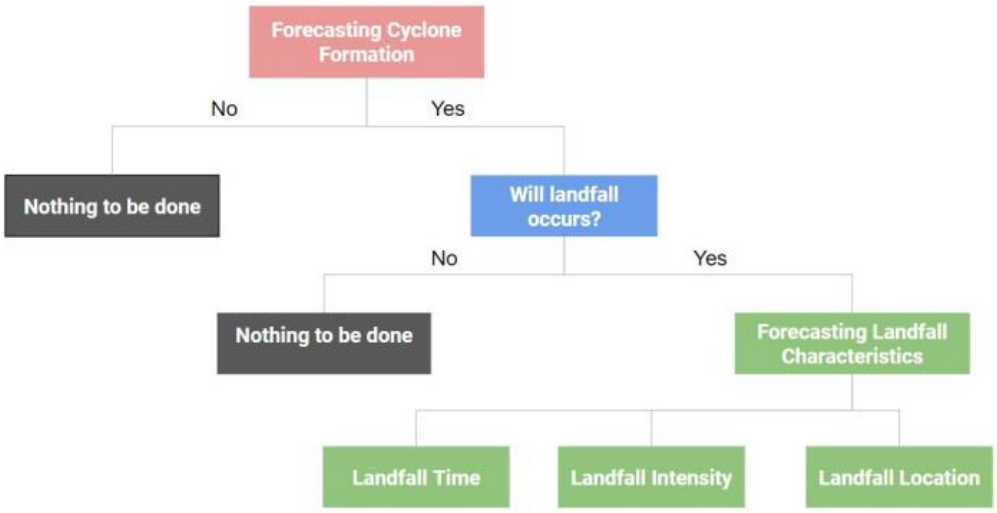


Fig 7:- Forecasting methods flowchart

2.3 Methodology

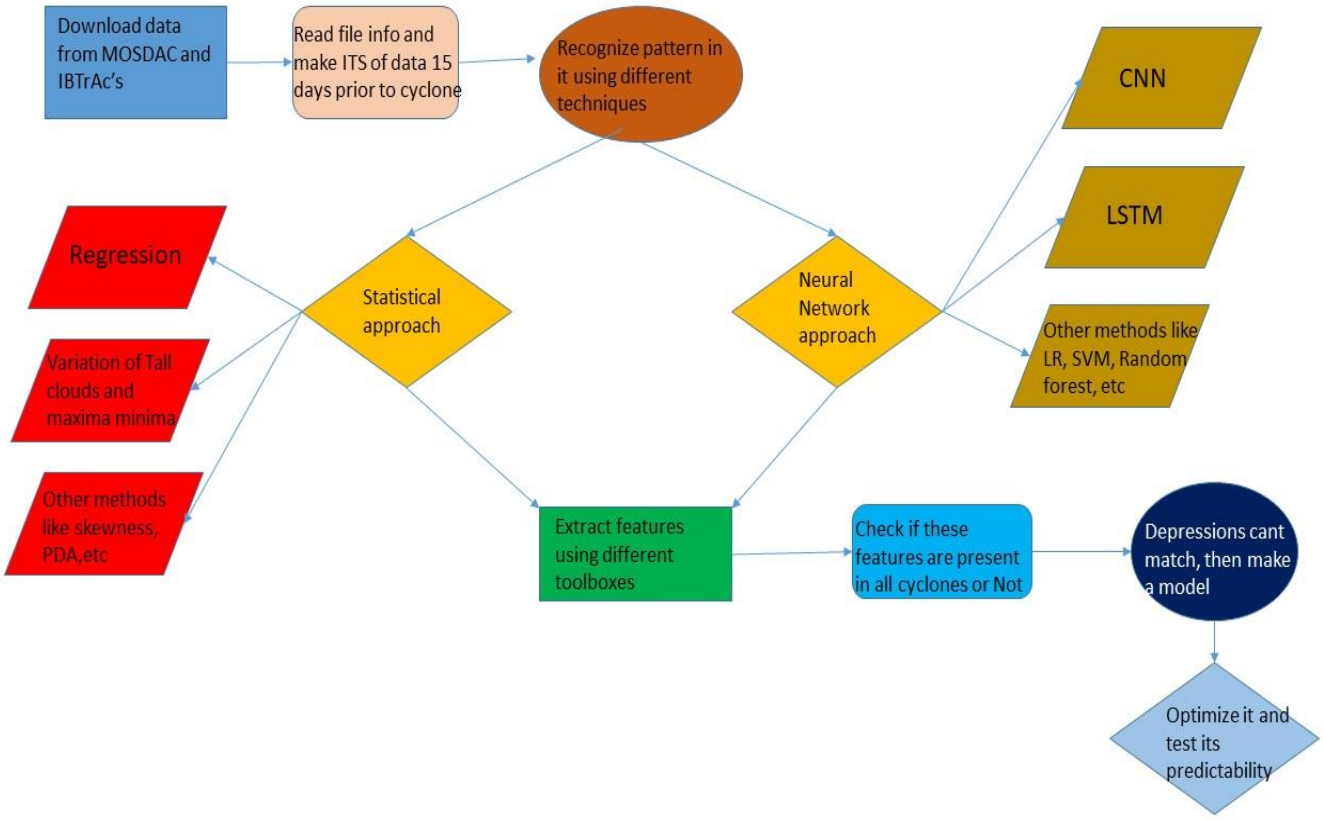


Fig 8:- Methodology of Predicting Cyclone track

3. EXPERIMENTAL DETAILS/ MODEL DESCRIPTION

To achieve the above objectives, the work plan followed is-

First, we fetch the data from MOSDAC (Meteorological and Oceanographic Satellite Data Archival Centre). MOSDAC is a data repository for the missions of the Indian Space Research Organisation and the Government of India, dealing with meteorology, oceanography, and tropical water cycles. Data acquired from missions is disseminated in near real-time from Space Applications Centre, Ahmedabad through the MOSDAC website.

We take data on Cloud top temperature over the Bay of Bengal and the eastern coast of India over the period of T-20 days to T-5 days, where T denotes the day of the landfall of a cyclone. As we have images in latitude and longitude format but the data matrix in X-Y coordinates we use meracters projection to take the Asia-Mer region and then crop the data to focus on eastern coastal plains and the Bay of Bengal area. So, we crop the data to make a matrix over this zone.

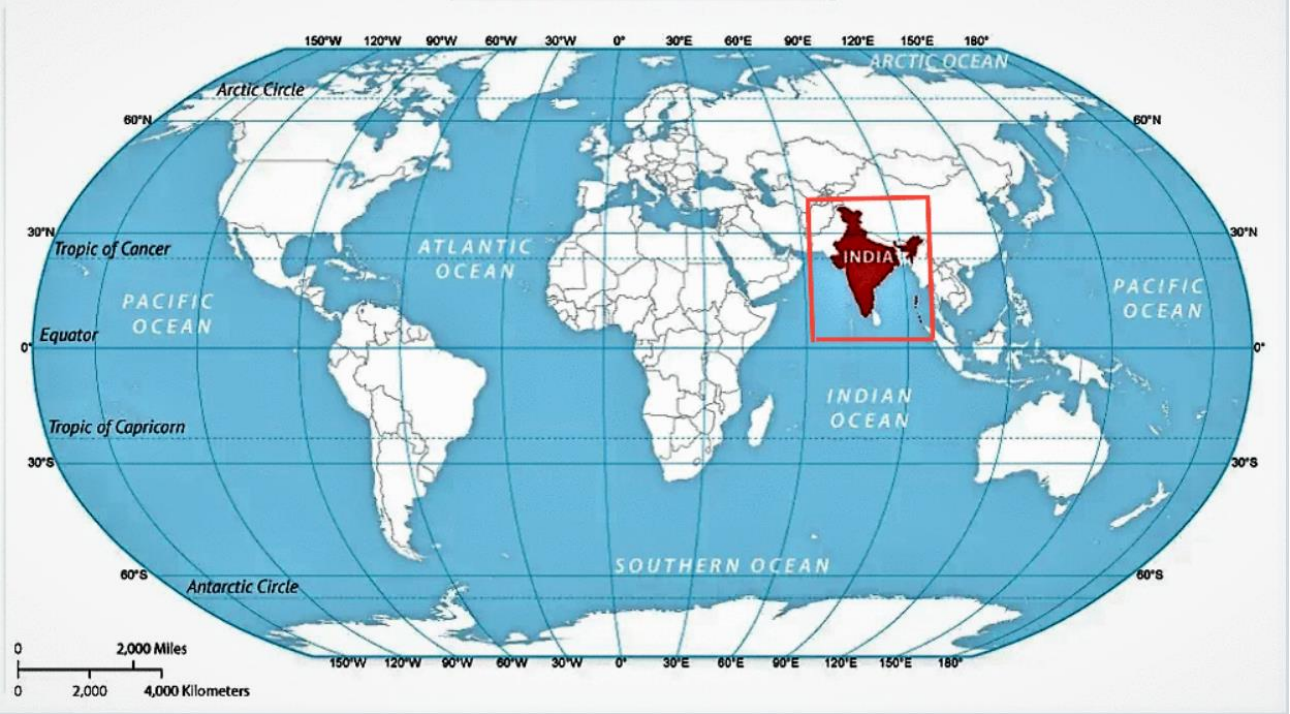


Fig 9-: Location of India's Latitude and Longitude

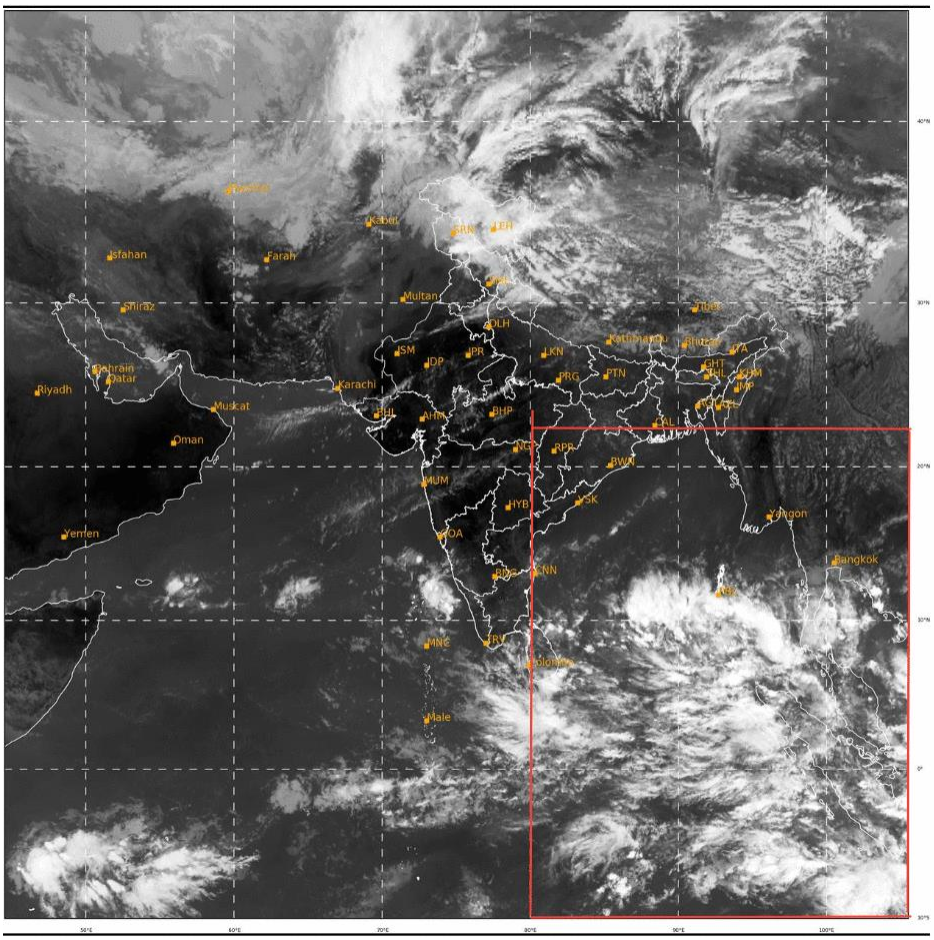


Fig 10:- Located area for analysis (25N to 10S and 80E to 110E) taken from [8]

For converting this data, we have over a span of 15 days into an image time series matrix (ITS), we use space-filling curves (Z-curve mapping) to first reduce the dimensionality of the data and then make a 2-D matrix with spatial variation in the Y domain and temporal variation in the X domain.

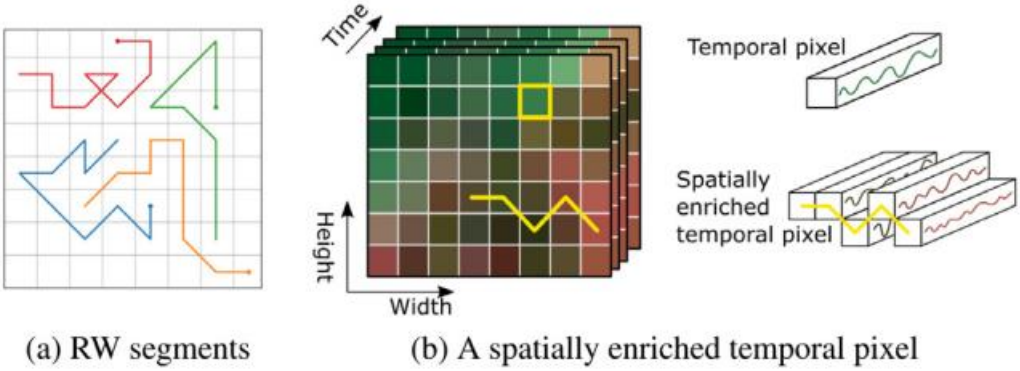


Fig 11:- (a) Random walk method segments and (b) spatially enriched temporal pixel

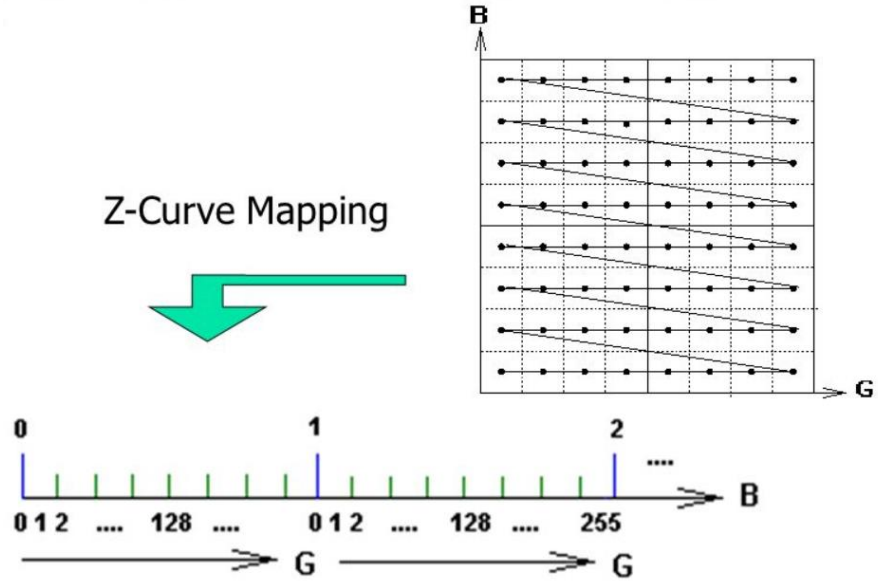


Fig 12 :- Z-curve mapping

To achieve our next objective of finding a pattern in the data, we perform Clustering through machine learning toolbox in MATLAB. In clustering, you want a neural network to group data by similarity, mimic the output data and copy its features. Weight here represents the coordinates of neurons. We will take distance as a weight so that the self-organizing map can mimic the pattern followed by the input data set. The SOM algorithm essentially computes how close or far and how concentrated are tall clouds using distance as its weight.

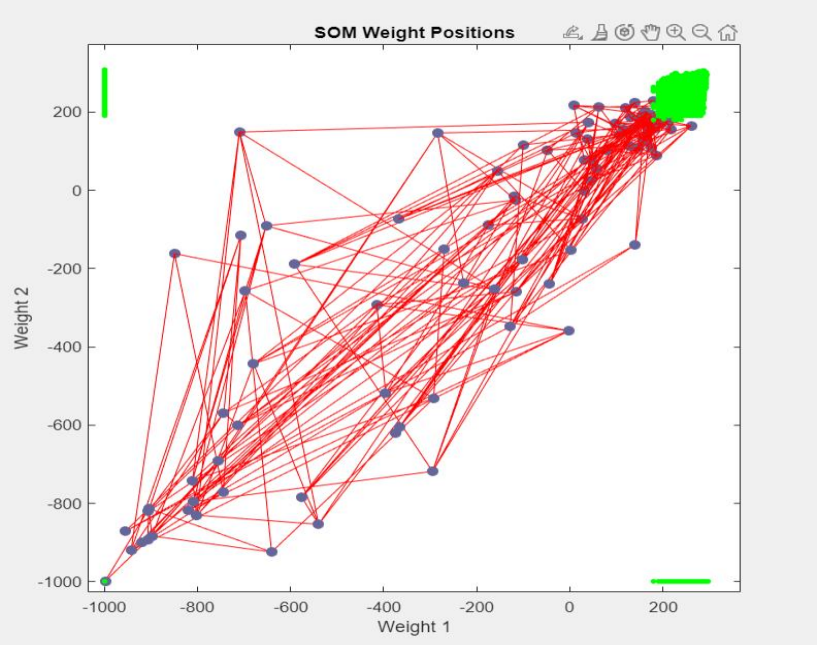


Fig 13:- Self Organizing Map with weights as distance and No. of clouds.

Data-

Each Regional Meteorological Centre (RMC) across the world maintains the historical dataset of TCs that is known as Best Track Data (BTD). In this research work, we have used three publicly available datasets: IMD BTD, International Best Track Archive for Climate Stewardship (IBTrACS) and ERA5 reanalysis data. The BTD maintained by IMD contains the record of latitude, longitude, MSSWS, grade, estimated central pressure (ECP), and other variables from 1990 onwards. Now, we will briefly describe the dataset - IBTrACS.

IBTrACS

IBTrACS project maintained by World Meteorological Organization (WMO) is a comprehensive dataset of global TCs which collects and merges the TCs data from various RMCs and creates a unified, freely available BTD. The most recent version 4 contains the three hourly BTD across all ocean basins from 1980 onwards and is updated daily. It covers all the TCs originating in the region 70°N to 70°S and 180°W to 180°E.

The dataset contains information about ocean basins, reporting regional agencies, wind speed, pressure, grade, track (latitude and longitude), distance to land, speed, direction, and many other variables. The genesis location of all the TCs in six ocean basins, North Indian (NI), South Indian (SI), North Atlantic (NA), West Pacific (WP), East Pacific

(EP), and South Pacific (SP), is shown in the figure below.

The IBTrACS dataset is available in network Common Data Form (NetCDF) and comma-separated values (CSV) file format. For this work, we have used version 4 of IBTrACS dataset available in CSV file format**.**

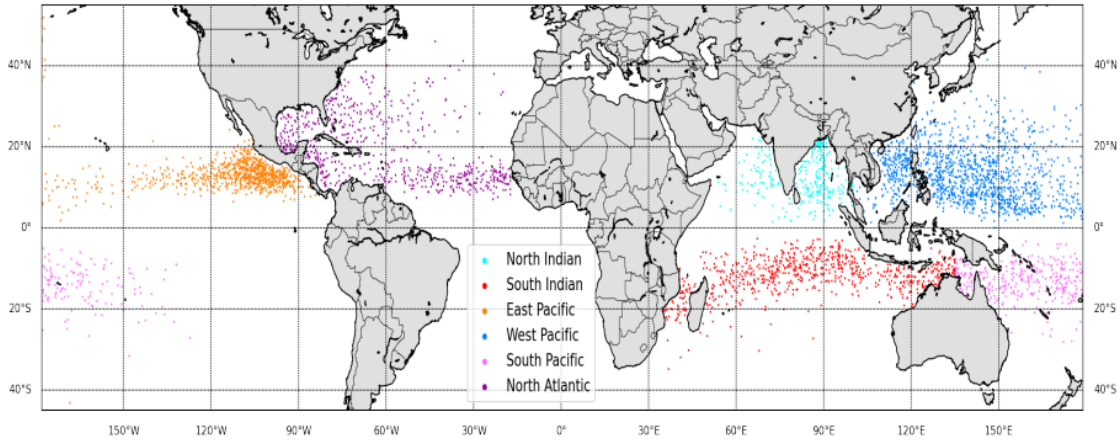


Fig 14:- Genesis Points of Cyclones all over the World present in IBTrAc's Dataset

4. RESULTS AND DISCUSSION

The Clustering toolbox was utilized to conduct a batch unsupervised learning self-organizing map (SOM) analysis to determine the intensity of tall clouds. SOM is an artificial neural network trained via unsupervised learning to generate a two-dimensional representation of the input space, referred to as a map, which facilitates dimensionality reduction.

During training, data points compete to be represented within the network. The network is first initialized with weights, and training data is randomly selected, and the distance between nodes (weights) and sample vectors is calculated. The node with the shortest distance is designated as the BMU (Best Matching Unit), and the neighborhood of the BMU is determined. The weights of the neighboring nodes are then updated, and this process is repeated for each training data point until the network is fully trained.

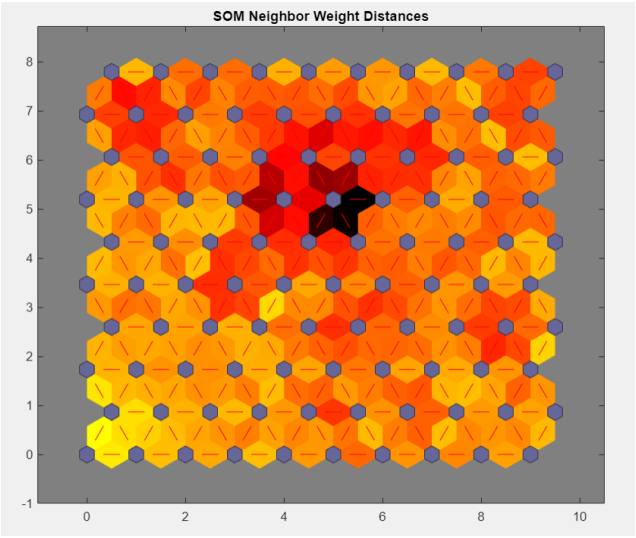


Fig 15:- SOM of Neighbour weight distances.

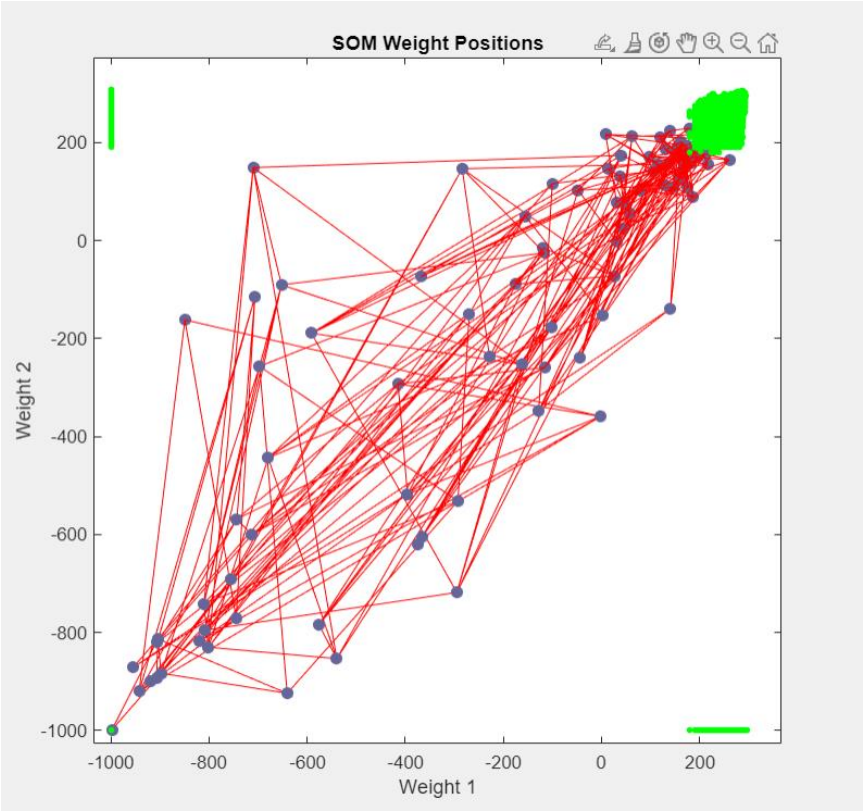


Fig 16:- SOM weight positions.

As we had taken the pixels of cloud images as data points, nearby pixels of logical 1's together would have been a single cloud (as we preserved the spatial relationship by chronologically placing them in a column). Then to see the pattern, we plotted no. of tall clouds v/s days for

Cyclone Bulbul. As we had taken all the tall clouds in space. So, we targeted only the Bay of Bengal area.

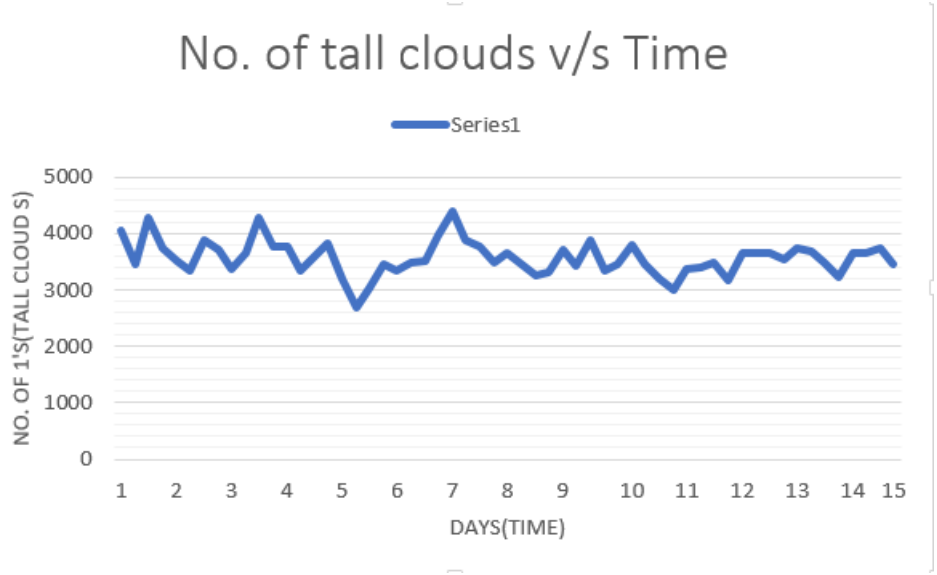


Fig 17:- Plot of tall clouds v/s days considering the whole map of India

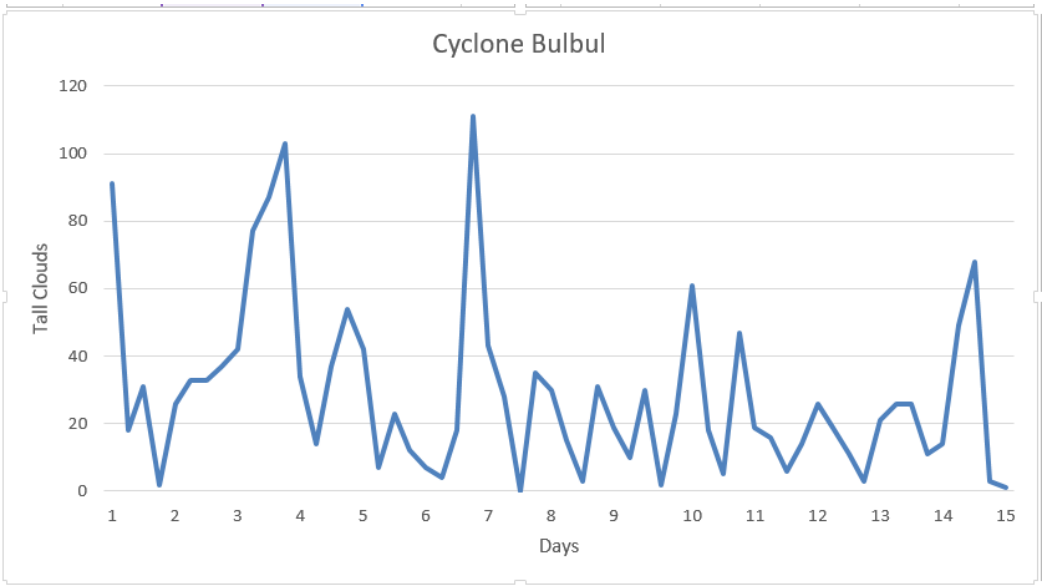


Fig 18:- Plot of tall clouds v/s days considering only the Bay of Bengal part.

Here, we found out that even if the pattern is dissimilar but there are large variations in number of tall clouds before the occurrence of a cyclone which is characterized by some maxima on the plot.

Then we performed regression analysis for tall and medium clouds for cyclone Bulbul and Amphan. We performed Logarithmic, Linear, Polynomial, and Sinusoidal regression on both cyclones for 3 types of clouds.

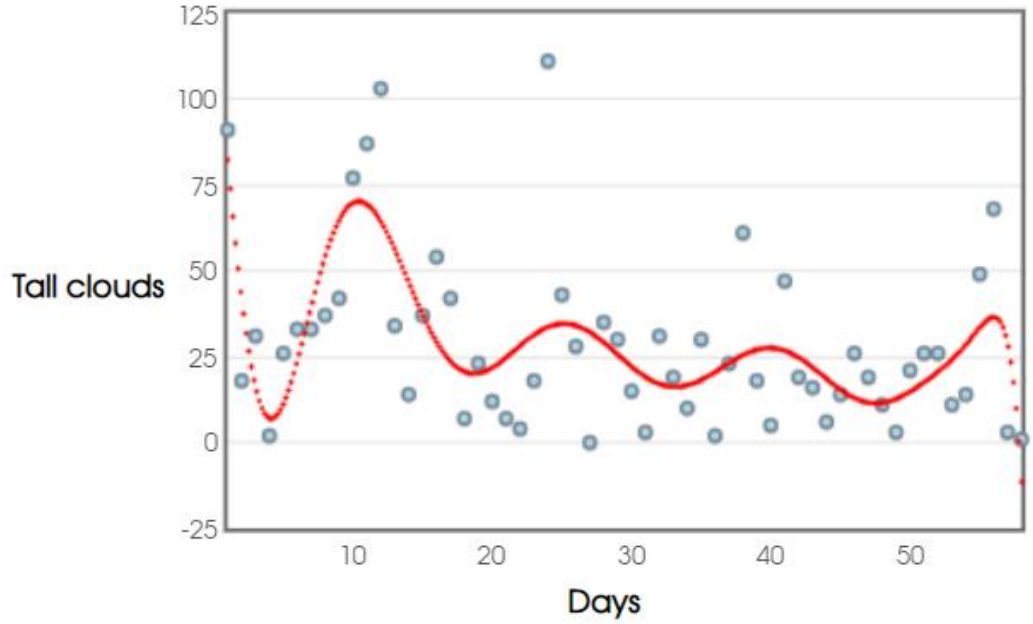


Fig 19:- Variation of Tall clouds before Cyclone Bulbul

Regression Polynomial: $\mathbf{TALL\ CLOUDS} = 0.0000 \cdot \mathbf{DAYS}^{14} + 0.0000 \cdot \mathbf{DAYS}^{13} + 0.0000 \cdot \mathbf{DAYS}^{12} + 0.0000 \cdot \mathbf{DAYS}^{11} + 0.0000 \cdot \mathbf{DAYS}^{10} + 0.0000 \cdot \mathbf{DAYS}^{9} + 0.0000 \cdot \mathbf{DAYS}^{8} + 0.0001 \cdot \mathbf{DAYS}^{7} - 0.0032 \cdot \mathbf{DAYS}^{6} + 0.0747 \cdot \mathbf{DAYS}^{5} - 0.9537 \cdot \mathbf{DAYS}^{4} + 0.0273 \cdot \mathbf{DAYS}^{3} - 10.9407 \cdot \mathbf{DAYS}^{2} - 37.3648 \cdot \mathbf{DAYS} + 125.4143$
R-squared: $r^2 = 0.4274$
Adjusted R-squared: $r^2_{adj} = 0.2582$
Residual Standard Error: 22.3029 on 43 degrees of freedom

Coefficient	Estimate	Standard Error	*t*-statistic	*p*-value
β_0	125.4143	78.8277	1.591	0.1189
β_1	-37.3648	96.9	-0.3856	0.7017
β_2	-10.9407	41.0806	-0.2663	0.7913
β_3	0.0273	8.4861	0.7103	0.4814
β_4	-0.9537	0.9921	-0.9614	0.3417
β_5	0.0747	0.0707	1.0577	0.2961
β_6	-0.0032	0.0031	-1.0178	0.3145
β_7	0.0001	0.0001	0.8362	0.4077
β_8	0	NaN	NaN	0
β_9	0	NaN	NaN	0
β_{10}	0	NaN	NaN	0
β_{11}	0	0	0.4185	0.6777
β_{12}	0	0	-0.7093	0.482
β_{13}	0	0	0.8704	0.3889
β_{14}	0	0	-0.9638	0.3405

Analysis of Variance Table

Source	df	SS	MS	*F*-statistic	*p*-value
Regression	14	16488.6183	1177.7585	2.3677	0.0154
Residual Error	43	21380.1244	497.4215		
Total	57	37353.3793	655.3224		

Fig 20:- Polynomial equation for Cyclone bulbuls' tall clouds and values of its coefficients.

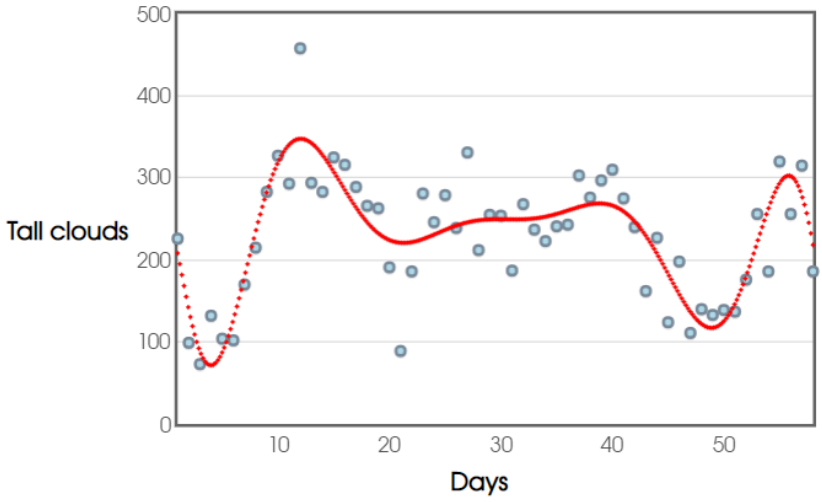


Fig 21:- Variation of Middle height clouds before Cyclone Bulbul

Regression Polynomial: $\text{TALL CLOUDS} = 0.0000 \cdot \text{DAYS}^{12} + 0.0000 \cdot \text{DAYS}^{12} + 0.0000 \cdot \text{DAYS}^{11} + 0.0000 \cdot \text{DAYS}^{10} + 0.0000 \cdot \text{DAYS}^{9} + 0.0000 \cdot \text{DAYS}^{8} + 0.0002 \cdot \text{DAYS}^{7} - 0.0087 \cdot \text{DAYS}^{6} + 0.1950 \cdot \text{DAYS}^{5} - 2.5907 \cdot \text{DAYS}^{4} + 18.4183 \cdot \text{DAYS}^{3} - 51.6749 \cdot \text{DAYS}^{2} - 9.5763 \cdot \text{DAYS} + 253.4364$

R-squared: $r^2 = 0.7132$

Adjusted R-squared: $r^2_{\text{adj}} = 0.6367$

Residual Standard Error: 46.9539 on 44 degrees of freedom

Coefficient	Estimate	Standard Error	t-statistic	p-value
β_0	253.4364	158.4299	1.5997	0.1168
β_1	-9.5763	192.151	-0.0498	0.9605
β_2	-51.6749	81.3303	-0.6354	0.5285
β_3	18.4183	16.9731	1.0851	0.2838
β_4	-2.5907	2.0237	-1.2802	0.2072
β_5	0.195	0.1475	1.3222	0.1929
β_6	-0.0087	0.0066	-1.3093	0.1972
β_7	0.0002	0.0002	1.4357	0.1582
β_8	0	NaN	NaN	0
β_9	0	NaN	NaN	0
β_{10}	0	NaN	NaN	0
β_{11}	0	0	-0.9877	0.3287
β_{12}	0	0	0.4485	0.656
β_{13}	0	0	-0.2864	0.776

Analysis of Variance Table

Source	df	SS	MS	F-statistic	p-value
Regression	13	242585.7702	18660.4439	8.464	0
Residual Error	44	97005.5634	2204.6719		
Total	57	338229.8793	5933.8575		

Fig 22:- Polynomial equation for Cyclone Bulbuls' middle height clouds and values of its coefficients.

Similarly, we have results for Cyclone Amphan.

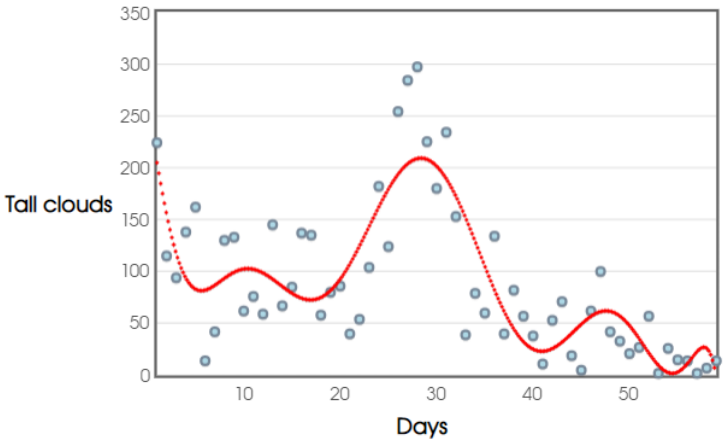


Fig 23:- Variation of Tall clouds before Cyclone Amphan

Regression Polynomial: $\text{TALL CLOUDS} = 0.0000 \cdot \text{DAYS}^{10} + 0.0000 \cdot \text{DAYS}^{9} + 0.0000 \cdot \text{DAYS}^{8} + 0.0001 \cdot \text{DAYS}^{7} - 0.0021 \cdot \text{DAYS}^{6} + 0.0504 \cdot \text{DAYS}^{5} - 0.6862 \cdot \text{DAYS}^{4} + 4.7391 \cdot \text{DAYS}^{3} - 9.5304 \cdot \text{DAYS}^{2} - 45.125 \cdot \text{DAYS} + 255.2257$

R-squared: $r^2 = 0.6665$

Adjusted R-squared: $r^2_{\text{adj}} = 0.6053$

Residual Standard Error: 46.599 on 48 degrees of freedom

Coefficient	Estimate	Standard Error	t-statistic	p-value
β_0	255.2257	120.8125	2.1126	0.0399
β_1	-45.125	122.0879	-0.3696	0.7133
β_2	-9.5304	41.9525	-0.2272	0.8213
β_3	4.7391	6.9494	0.6819	0.4986
β_4	-0.6862	0.6472	-1.0602	0.2944
β_5	0.0504	0.0366	1.3757	0.1753
β_6	-0.0021	0.0013	-1.6308	0.1095
β_7	0.0001	0	1.8298	0.0735
β_8	0	0	-1.9797	0.0535
β_9	0	0	2.0888	0.0421
β_{10}	0	0	-2.1648	0.0354

Analysis of Variance Table

Source	df	SS	MS	F-statistic	p-value
Regression	10	208206.5798	20820.658	9.5883	0
Residual Error	48	104230.5169	2171.4691		
Total	58	312576.6441	5389.2525		

Fig 24:- Polynomial equation for Cyclone Amphan tall clouds and values of its coefficients.

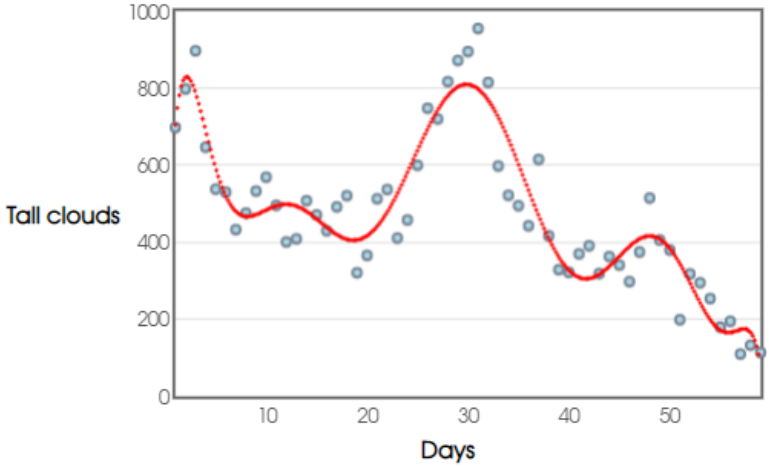


Fig 25:- Variation of Middle height clouds before Cyclone Amphan

Regression Polynomial: $\text{TALL CLOUDS} = 0.0000 \cdot \text{DAYS}^{10} + 0.0000 \cdot \text{DAYS}^{9} + 0.0000 \cdot \text{DAYS}^{8} + 0.0003 \cdot \text{DAYS}^{7} - 0.0120 \cdot \text{DAYS}^{6} + 0.3257 \cdot \text{DAYS}^{5} - 5.5023 \cdot \text{DAYS}^{4} + 55.3329 \cdot \text{DAYS}^{3} - 302.3690 \cdot \text{DAYS}^{2} + 716.3091 \cdot \text{DAYS} + 240.707$

R-squared: $r^2 = 0.8697$

Adjusted R-squared: $r^2_{\text{adj}} = 0.8458$

Residual Standard Error: 79.6117 on 48 degrees of freedom

Coefficient	Estimate	Standard Error	t-statistic	p-value
β_0	240.707	206.4012	1.1662	0.2493
β_1	716.3091	208.5801	3.4342	0.0012
β_2	-302.369	71.6734	-4.2187	0.0001
β_3	55.3329	11.8727	4.6605	0
β_4	-5.5023	1.1057	-4.9765	0
β_5	0.3257	0.0626	5.2039	0
β_6	-0.012	0.0022	-5.3501	0
β_7	0.0003	0.0001	5.4244	0
β_8	0	0	-5.4392	0
β_9	0	0	5.4077	0
β_{10}	0	0	-5.342	0

Analysis of Variance Table

Source	df	SS	MS	F-statistic	p-value
Regression	10	2026190.9799	202619.098	31.9688	0
Residual Error	48	304225.4164	6338.0295		
Total	58	2334632.7458	40252.2887		

Fig 26:- Polynomial equation for Cyclone Amphan middle height clouds and values of its coefficients.

Next, we take into account the Ice clouds. Ice clouds have temperatures between -20C and -110C. We observe the spatial variation of Ice clouds and how the count of tall clouds varies with the variation of ice clouds. Below is a MATLAB plot for 15 days' time.

Fig 27:- Cyclone bulbuls ice clouds variation every 6 hours for 15 days prior to landfall.

Fig 28:- Cyclone amphan’ s ice clouds variation every 6 hours for 15 days prior to landfall.

We can see that while the tall clouds are spatially curving, there is always an increase in the number of clouds.

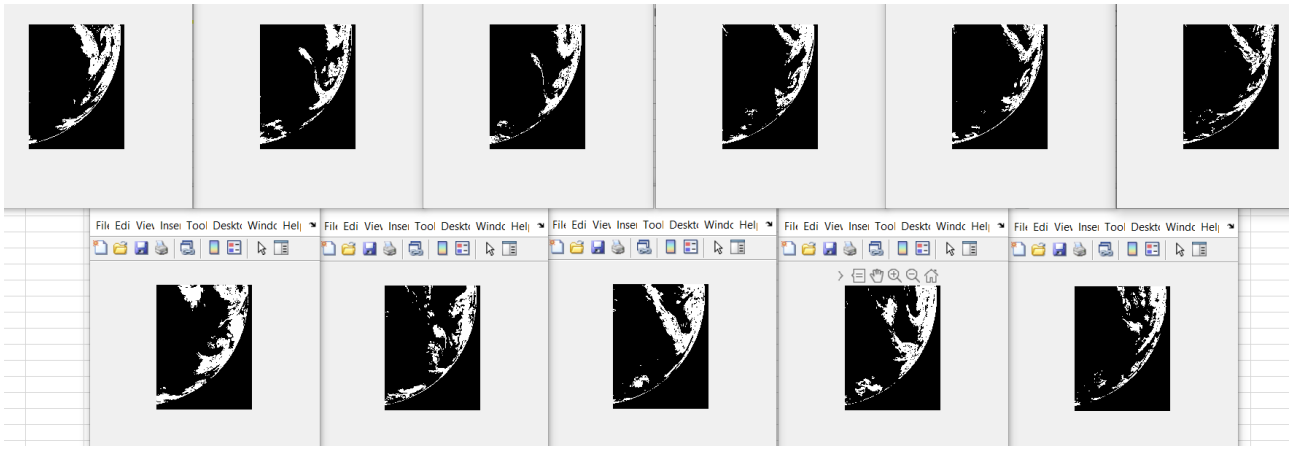


Fig 29:- Variation of Ice clouds of cyclone bulbul spatially.

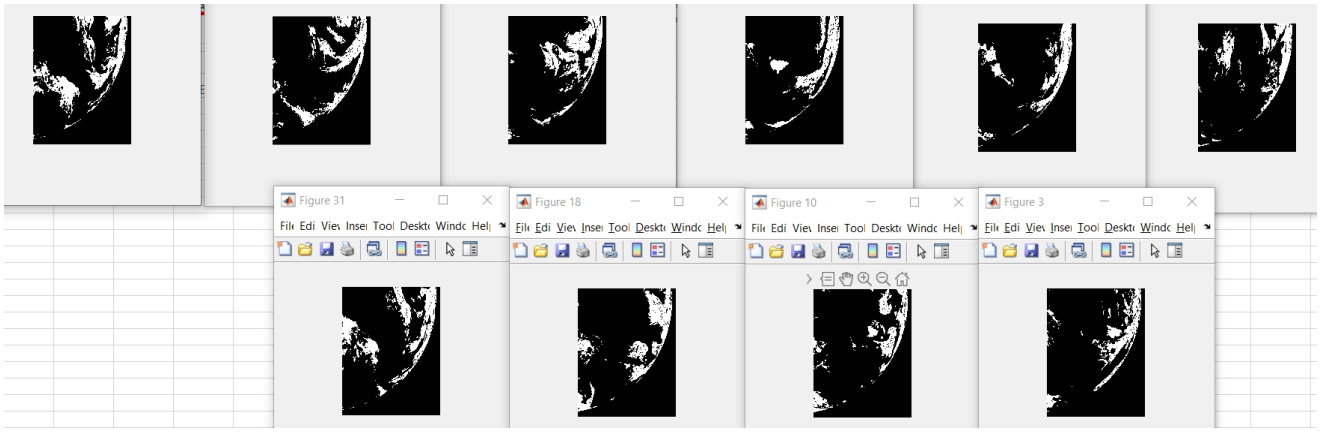


Fig 30 :- Variation of Ice clouds of cyclone Amphan spatially.

Next, we have found entropy i.e. - how much variation is present in a pixel as compared to its surroundings. Here we take each row of the image matrix and pass it to the entropy function where it converts this array into a normal distribution and finds each values entropy with respect to the distribution. We plot the image of cyclone bulbul, its entropy image and find that its entropy is greater than its surroundings and also the eye of the cyclone.

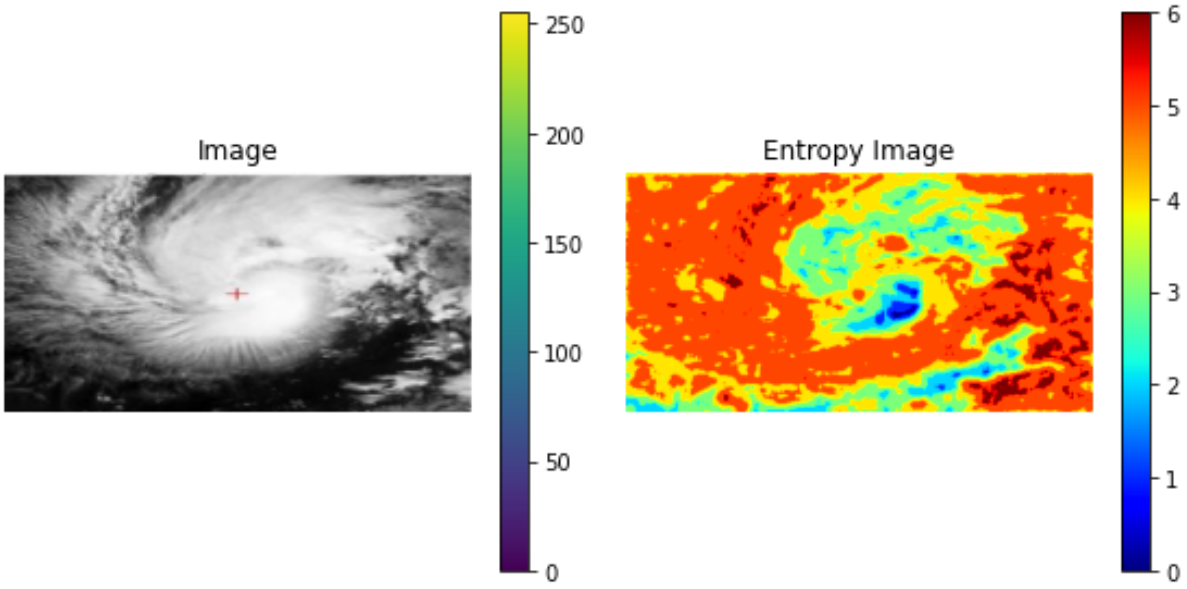


Fig 31 :- Black and White image and Entropy Image of Cyclone

Next, we average the entropy of each pixel by standardizing it. Standardization is done by dividing it with total no. of pixels. A time series of entropies of Cyclone Bulbul and Amphan was plot.

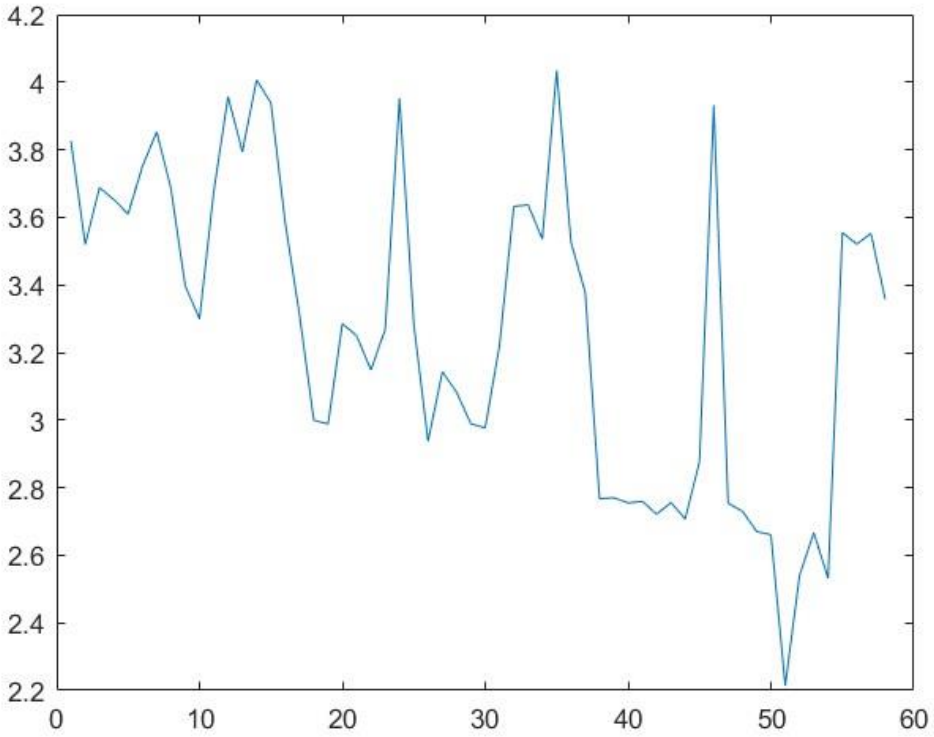


Fig 32:- Cyclone Bulbuls variation of Entropy

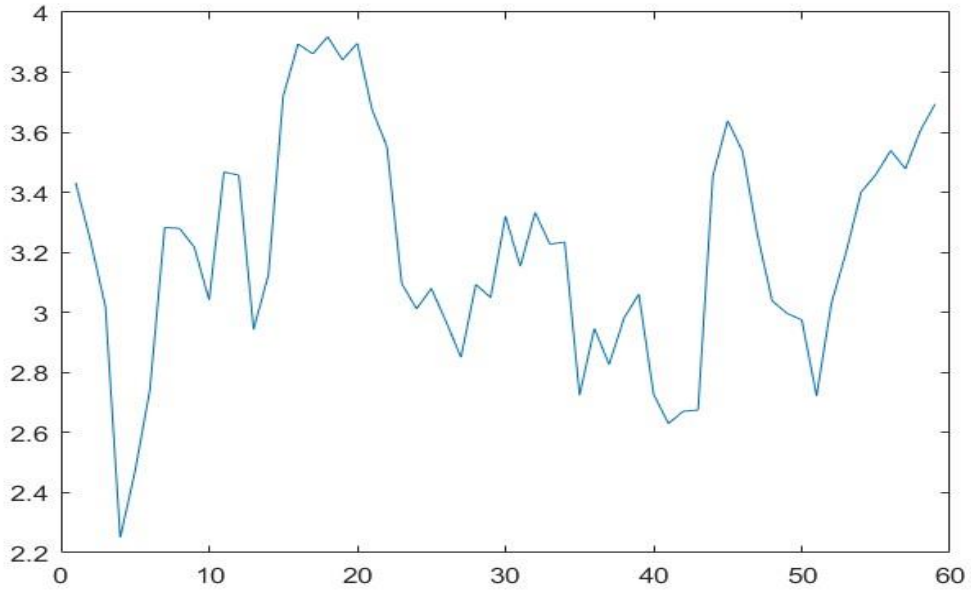


Fig 33:- Cyclone Amphan variation of Entropy

Further, we apply thermodynamic entropy called pseudo-adiabatic entropy and integrate it over area. We find the simplified form surface entropy flux. We take data of best track from IMD and JTWC and temperature value from MOSDAC, find the surface entropy flux which depends on maximum wind speed, radius of the area we take, and surface temperature, etc. We plot these values for Cyclone Vardah and study their variations.

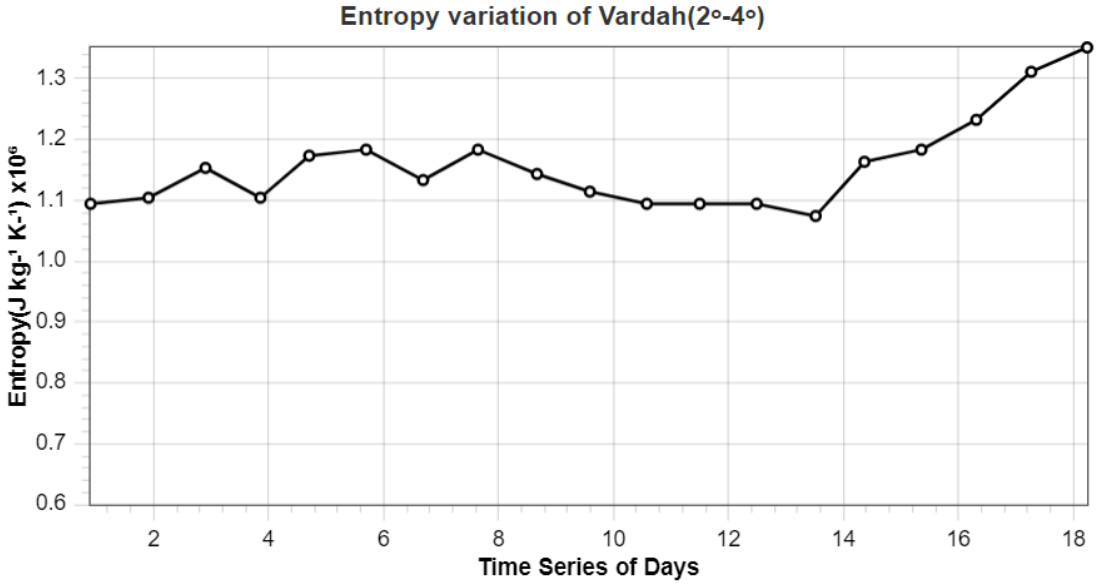


Fig 34:- Cyclone Vardah thermodynamic entropy variation

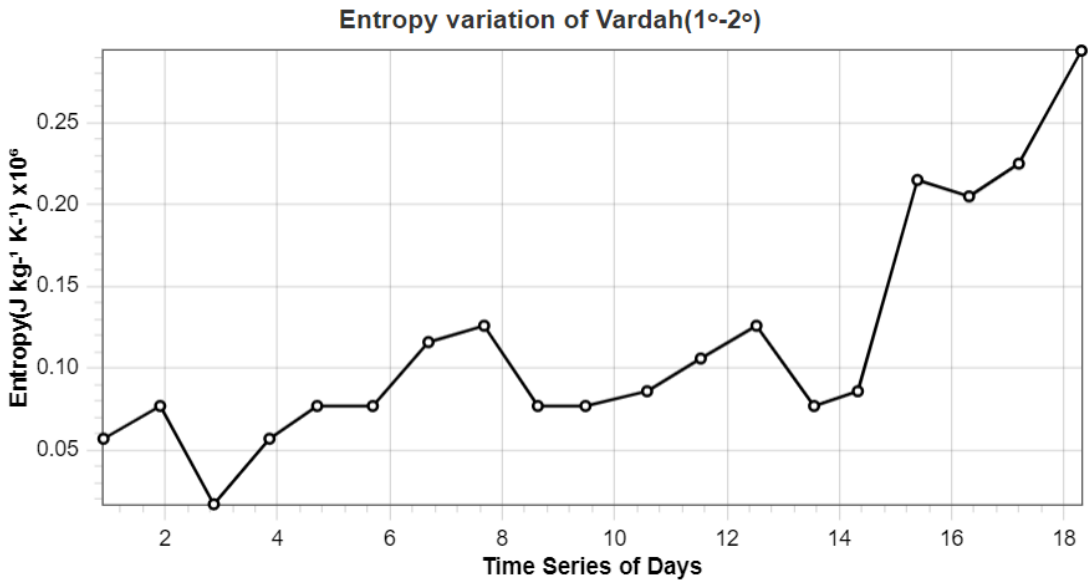


Fig 35:- Cyclone Vardah thermodynamic entropy variation

Air-sea thermodynamic disequilibrium is measured as the difference between saturation entropy at sea surface temperature and entropy at 600 hecta Pascal. More the disequilibrium, the more the difference in entropy for the chances of cyclone genesis occurring. We plotted the ASTD values over time for cyclone Vardah and studied its variations.

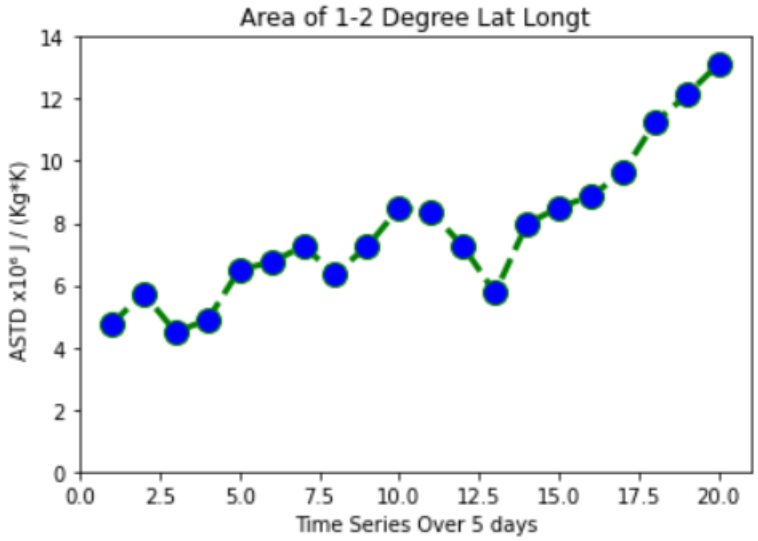


Fig 36:- ASTD of Cyclone Vardah for 1^o-2^o Lattitude

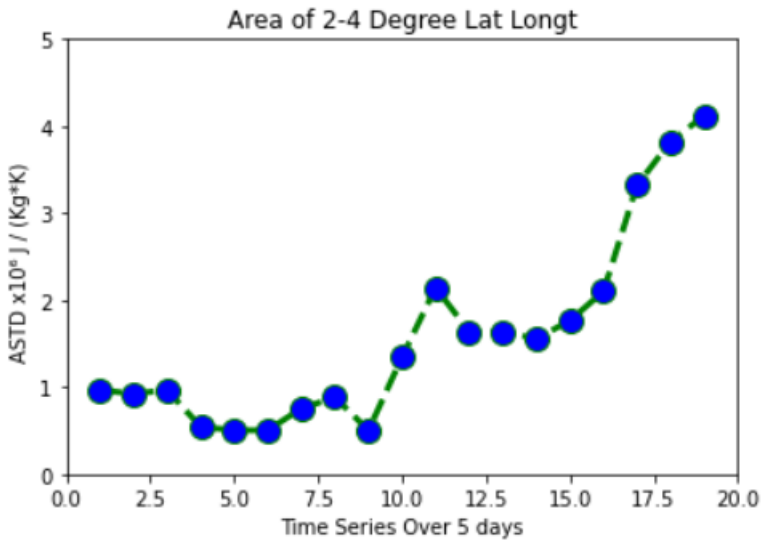


Fig 37:- ASTD of Cyclone Vardah for 2^o-4^o Lattitude

Brightness temperature data from MOSDAC was used and clouds were classified based on 3 temperature ranges. The deep convective cold clouds were taken as having BT<210K, cold clouds were taken as having 210K<BT<240K. We calculated brightness temperature images for hudhud cyclone and studied its diurnal variation (how it changes in a day).

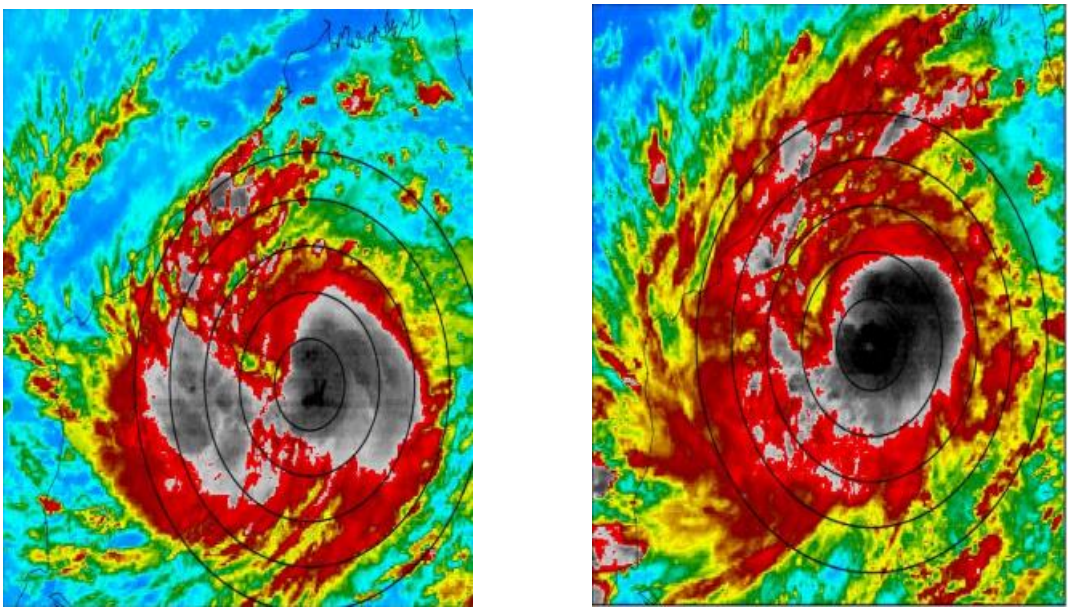


Fig 38:- TIR-BT of Cyclone Vardah at an interval of 6 hours.

Next, we plotted mean cloud cover area of deep convective cold clouds and cold clouds with time and found that they are in an out of phase relationship. Mean cloud cover area was calculated by counting the number of pixels having temperature in required range, then scaling it to actual area in km^2. Then dividing by the whole area for mean.

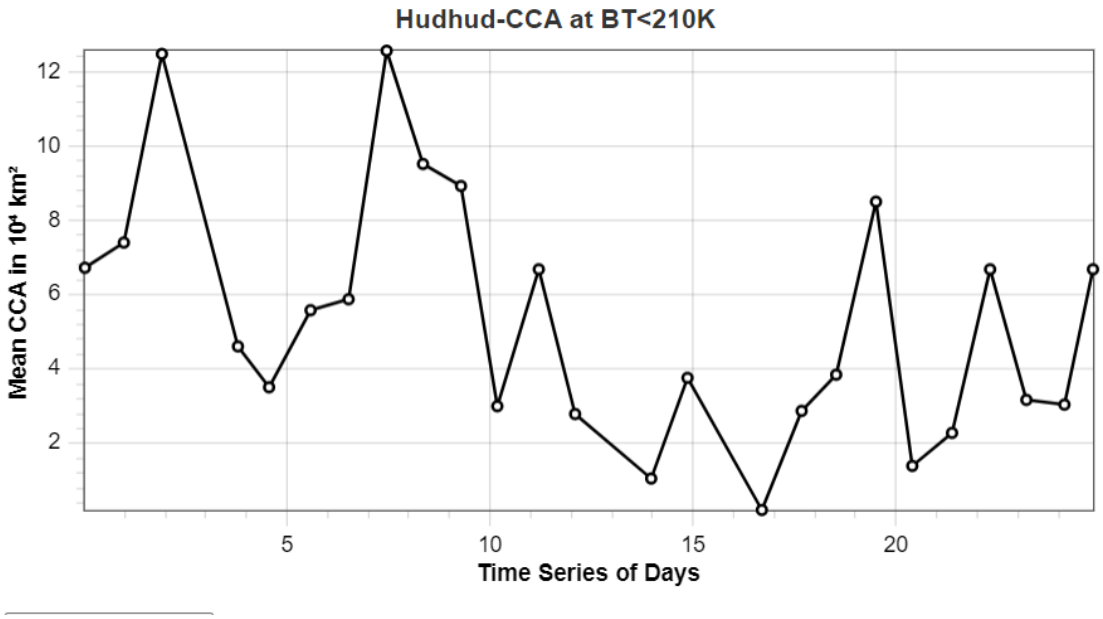


Fig 39:- Variation of Mean cloud cover area for Cyclone Hudhud having BT<210K

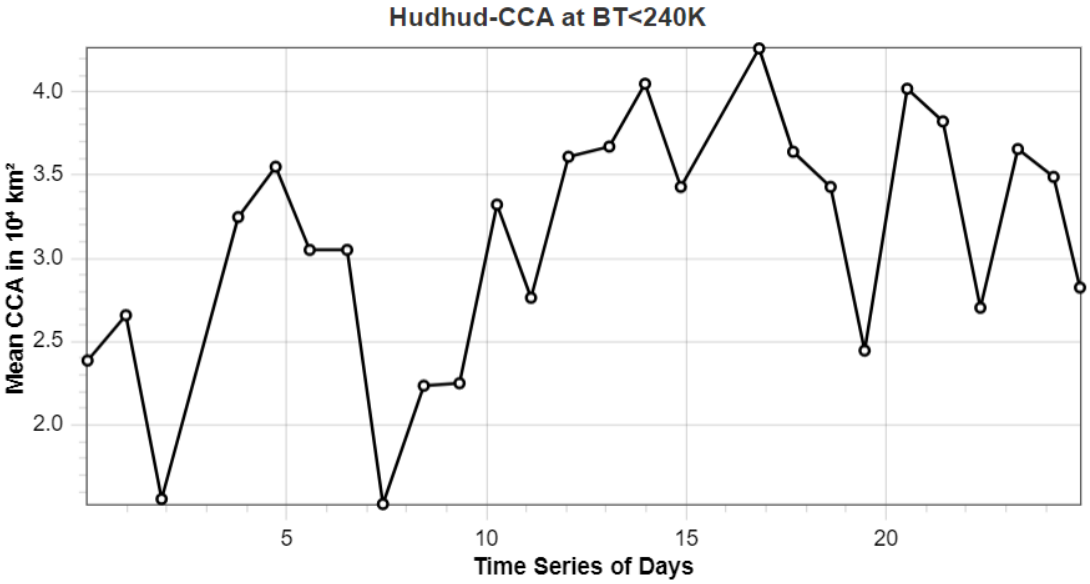


Fig 40:- Variation of Mean cloud cover area for Cyclone Hudhud having 210K<BT<240K

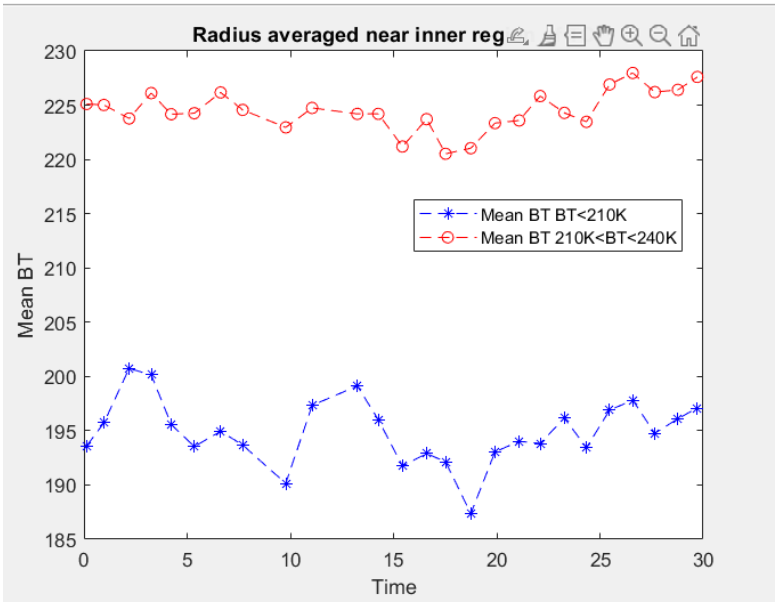


Fig 41:- Variation of Radius averaged BT in Inner region of Cyclone

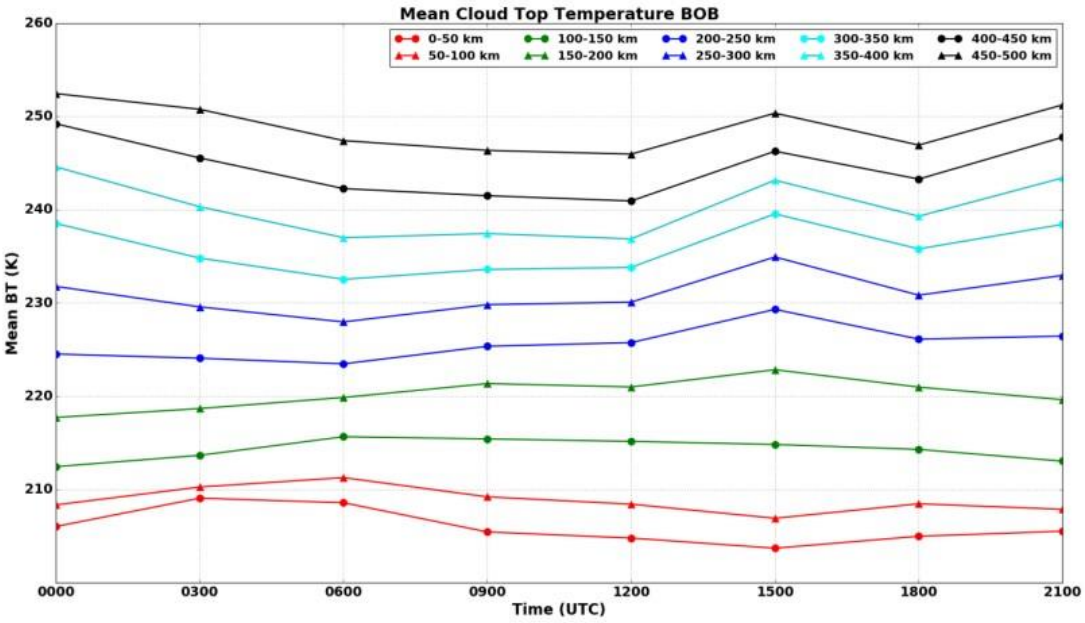
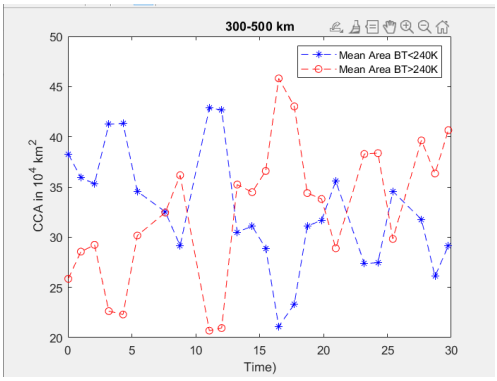
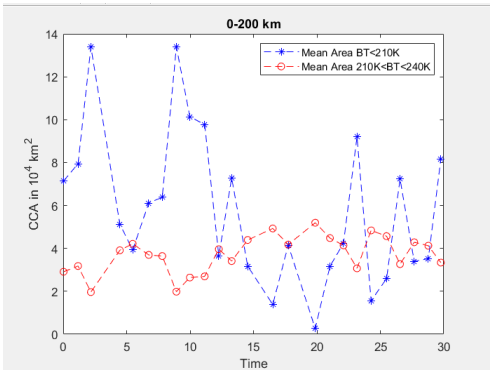
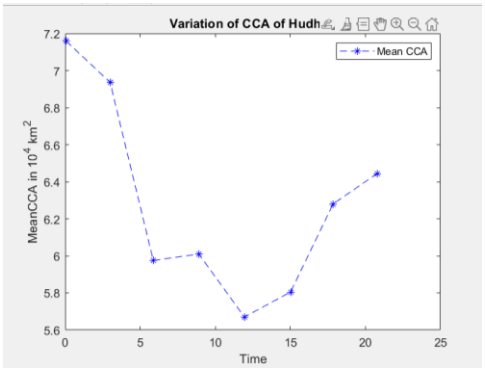
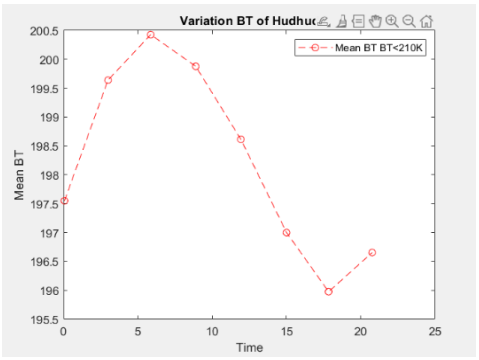


Fig 42:- The mean diurnal variation of CTT at 0-50, 50-100, 100-150, 150200, 200-250, 250-300, 300-350, 350-400, 400-450 and 450-500 km annular region from the TC centre









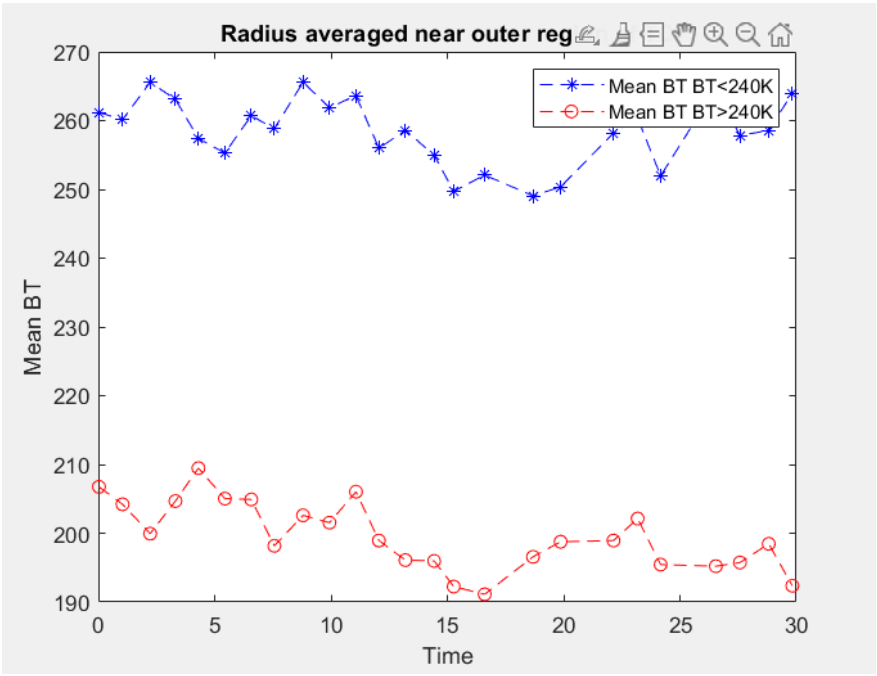


Fig 43:- Variation of CCA and Mean BT in Cyclone vardah (a) 300-500 km (b) 0-200 km

And in Cyclone Hudhud (c) Mean CCA (d)Mean BT (e) Radius Averaged in Outer region of cyclone

Variation of cloud cover area and cloud top temperature of specific cyclone

Plots above show the changes in the areal extent of clouds and TIR-BT for Cyclone Hudhud throughout its life cycles. The CCA plots indicate that the diurnal variation of deep cold clouds and cold clouds in the inner region of the cyclones are opposite in phase and magnitude. Similar trends are observed for the outer region of cold clouds and warm clouds. The diurnal variation of CTT for deep cold clouds and cold clouds are also out of phase, but there is a 3-hour lag in the inner region. The same trend is observed in the outer region for cold clouds (<240 K) and warm clouds (>240 K). It is noteworthy that CCA and CTT are out of phase. When the area covered by cold clouds (<240 K) decreases, the area covered by warm clouds (>240 K) increases and vice versa. The two leading out-of-phase conditions, CCA and CTT, result in two peaks for the day for the average temperature of the outer region (300-500 km).

Diurnal variation of mean cloud cover area and mean cloud top temperature

The analysis of the diurnal variation of the mean Cloud Cover Area (CCA) and Cloud Top Temperature (CTT) for all the considered cyclones during 24 hours. They divided the analysis between the inner and outer regions of the cyclones and for both deep cold clouds and cold clouds.

For the BoB (Bay of Bengal) and ARB (Arabian Sea) cyclones which formed during 2008-2012 using Kalpana-1 data, the authors observed that the area covered by deep cold clouds reaches a maximum during the morning (0600-0800 IST) and then decreases after sunrise and becomes a minimum during the evening (1800-2100 IST) in both the inner and outer regions.

For cold clouds, the maximum CCA is found in the afternoon (1200-1400 IST) for both the inner and outer regions. The CTT of deep cold clouds in both the inner and outer regions are maximum during the afternoon (1200-1500 IST) but the minimum occurs during midnight (2400-0600 IST). In the case of cold clouds, the CTT is found to be maximum during early morning hours in both the inner and outer region of the cyclones.

Similar trends and results are found when the INSAT-3D data is used in the calculation of CCA and CTT for the cyclones (2013-2016).

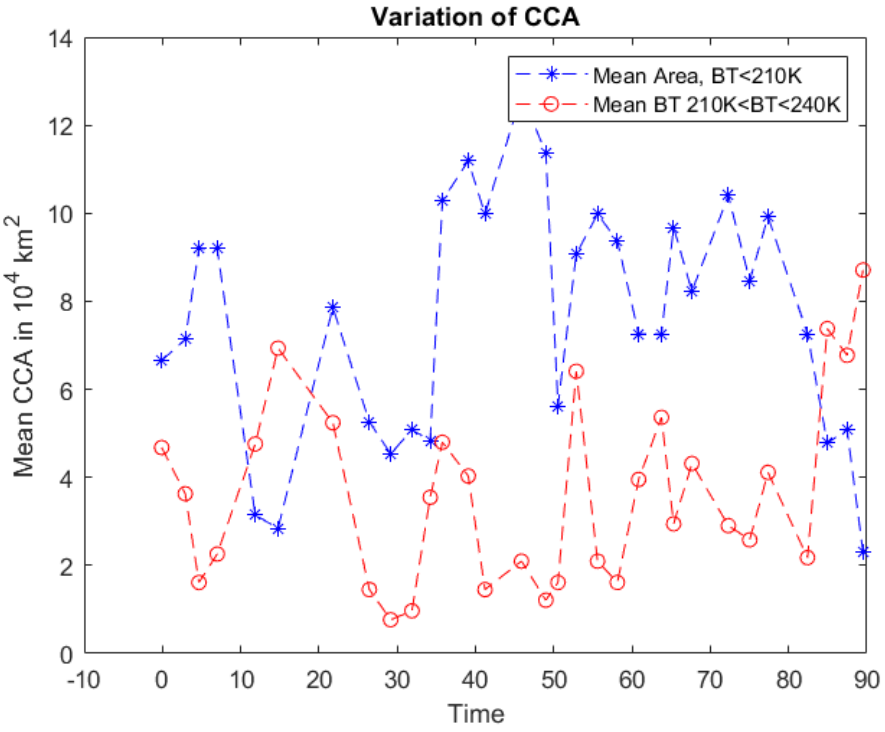


Fig 44:- Variation of CCA for cyclone Hudhud

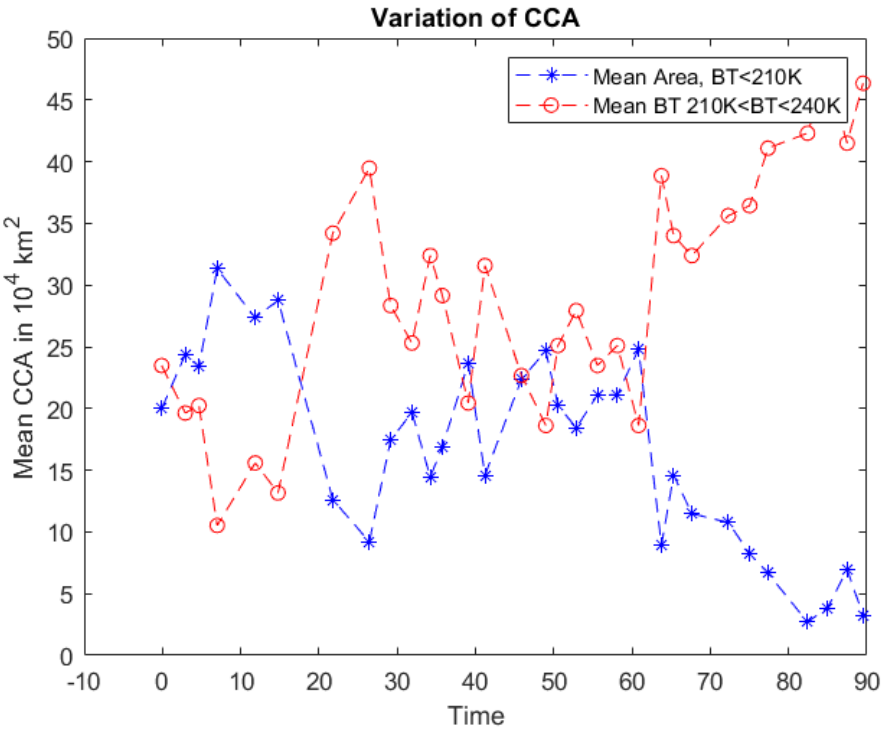


Fig 45:- Variation of CCA for cyclone Vardah

In this research, two types of data were used: Atlantic hurricane data from the NOAA database and BTD of tropical cyclones in the NI ocean from the Regional Specialized Meteorological Centre, New Delhi. The Atlantic Ocean data contained information on hurricanes from 1920 to 2012, with recordings provided at 6-hour intervals. The dataset consisted of a total of 16,394 recordings, with each recording containing details about the hurricane's year, time, name (if applicable), latitude, longitude, ECP, and MSSWS. The figure below shows the trajectories of all the hurricanes. After removing invalid and missing data, there were a total of 789 hurricanes left, with an average of 21 recordings per hurricane. The largest cyclone had 96 recordings. To analyze the relative change of hurricane track progression, two additional features, distance and direction of change, were generated between two consecutive recordings of a hurricane. The study utilized ECP, MSSWS, distance, and direction features.

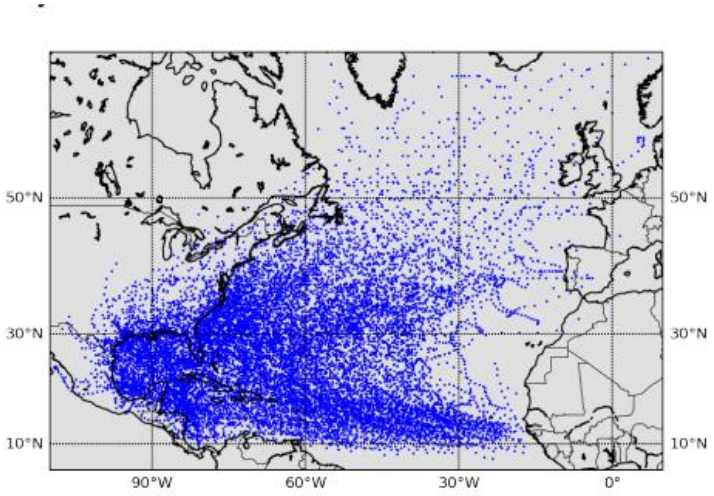


Figure 5.1: Atlantic hurricanes' track.

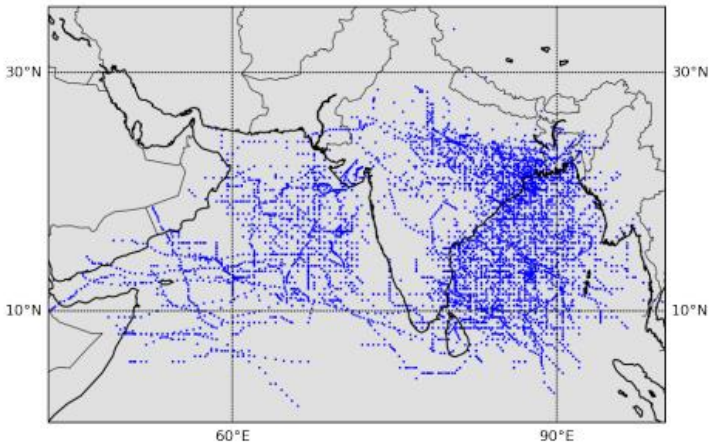


Figure 5.2: North Indian ocean cyclones' track.

The dataset used in this study includes information about 341 cyclones that occurred in the North Indian ocean from 1990 to 2017. The data is provided at 3-hour intervals and contains a total of 7,662 recordings. Each recording contains details such as the basin of origin, latitude, longitude, ECP, MSSWS, and the date and time of occurrence of the cyclone.

GridID

To forecast the path of a cyclone, it is necessary to predict its future latitude and longitude. Rather than directly estimating these values, a grid-based approach is used, as previously done in a study by Alemany et al. (2018). The region of latitude and longitude is divided into square grid blocks, with a fixed size of x degrees along both dimensions. The latitude and longitude ranges in the data are denoted as [amin, amax] and [bmin, bmax], respectively. Then, there will be a total of

$$\frac{(a_{max}-a_{min})(b_{max}-b_{min})}{x^2} \quad (10)$$

grid blocks. We assign a GridID to each grid block. For the block in ith row (starting from below) and jth column, we assign GridID

$$\frac{(b_{max}-b_{min})}{x}(j-1)+i \quad (11)$$

Each vertex in the grid has coordinates of the form (amin + ix, bmin + jx). Now, given any general latitude and longitude coordinate (a, b) in the region of the grid, we will have a = amin + ix + α, b = bmin + jx + β for some i, j ∈ N and $0 \leq \alpha, \beta < x$. Then

$$i = \frac{(a-a_{min})}{x} \quad (12)$$

$$j = \frac{(b-b_{min})}{x} \quad (13)$$

Therefore, we can define a GridID function for given latitude a and longitude b as

$$GridId(a,b) = \left[\frac{b_{max}-b_{min}}{x}\right]\left(\frac{b-b_{min}}{x}-1\right)+\left[\frac{a-a_{min}}{x}\right] \quad (14)$$

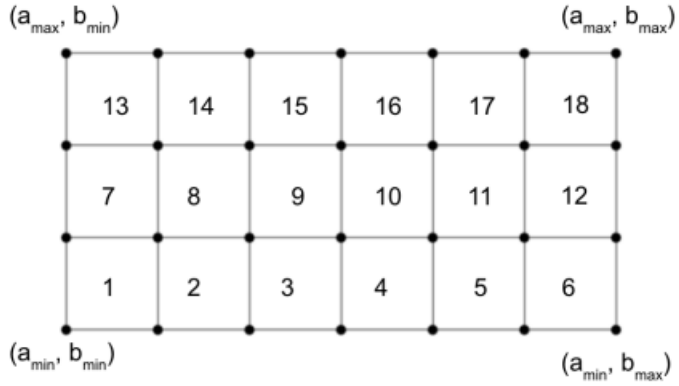


Fig 46:-Example of 6x3 grid

which will assign (a, b) the GridID for the grid block in which it lies. To predict the path of tropical cyclones, we use a grid-based approach by dividing the latitude and longitude ranges into square blocks of fixed size (x◦). For the Atlantic hurricane data, the latitude range is (7, 66) and the longitude range is (−110, 14), resulting in 7316 grid blocks of 1◦ × 1◦ size. For the Indian Ocean cyclone data, the latitude range is (1, 41) and the longitude range is (42, 103), resulting in 2440 grid blocks of 1◦ × 1◦ size. The 1◦ × 1◦ grid blocks have an average distance of 50 kilometers from the center of the block, making it acceptable to predict the grid block location (GridID) instead of directly predicting the latitude and longitude coordinates. Suppose we have S data points available at a time interval of D hours for a tropical cyclone, and we want to predict the GridID for the next T2 number of time points (at an interval of D hours) using T1 number of data points ((T1 − 1) × D hours of data).We generate $S - T1 - T2 + 1$ data points, where a single data point is a sequence of T1 vectors of the form:

$$(MSSWS(t), ECP(t), distance(t), direction(t), GridID(t)), k \leq t \leq T1 + k - 1$$

and output is a sequence of GridIDs (scaled between 0 and 1) for the next T2 time points, that is:

$$(GridID(T1 + k), GridID(T1 + k + 1), GridID(T1 + k + 2), \cdots, GridID(T1 + k + T2 - 1))$$

where k varies from 1 to $S - T1 - T2 + 1$. The collection of all such data points for all the cyclones form the dataset.

Model

We used a stacked LSTM model to analyze the time-series data. The input size of the model is (T, 5), where T is the length of sequential data and 5 is the number of variables. The model comprises an input layer, two hidden layers, and an output layer, with the hidden layers of size 200 each. The authors utilized Mean Square Error (MSE) as the loss function, Adam optimizer function with a default learning rate of 0.01, and ReLU as the activation function. The proposed model produced the best results after testing various configurations.

The MSE and Mean Absolute Error (MAE) is defined as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y - \overline{y_i})^2 \quad (15)$$

$$RMSE = \sqrt{MSE} \quad (16)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \overline{y_i}| \quad (17)$$

where y_i is the actual value and $\overline{y_i}$ is the model predicted value. MAE (Mean Absolute Error) is another commonly used metric to evaluate the performance of a model. It measures the average absolute difference between the predicted values and the actual values. While MSE gives more weight to larger errors, MAE treats all errors equally.

So, even though MSE is the loss function used for training the model, it is possible to report the performance of the model using MAE. This can provide a better understanding of how well the model performs in terms of the actual prediction error, without being affected by the magnitude of the errors.

.

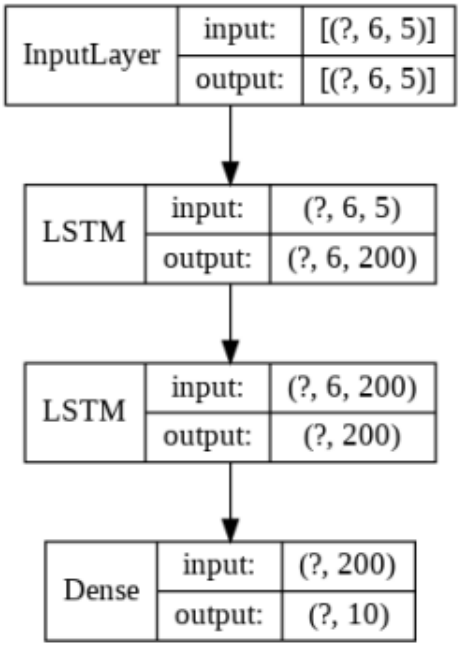


Fig 47:- Model structure for T1=6, T2=10

The proposed model is designed to take in T1 consecutive time points of cyclone data as the length of the training data point and predict the GridID for the next T2 successive time points. For example, if T1 is four and T2 is eight, then the model uses four consecutive time points of cyclone data to predict the GridID for the next eight successive time points.

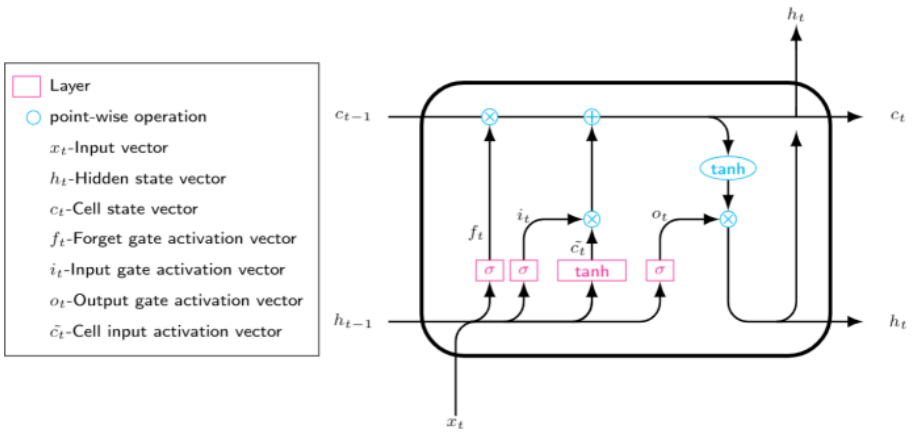


Fig 48:- LSTM Cell Inner working

For each combination of T1 and T2, the size of the training set along with the number of cyclones with at least T1 + T2 number of data points are reported. We have reported the MAE of 5-fold cross-validation. To further establish the efficacy of our model, we have reported the MAE for the two most significant cyclones -Vayu and Fani in the Indian ocean in recent times. All these named cyclones are not part of the training data.

Data Used			Mean Absolute Error		
Training Points T_1 (hours)	Predict Points T_2 (hours)	Training data size (no of cyclones)	5-fold MAE	Vayu	Fani
4 (12)	1 (3)	6428 (295)	0.0104	0.0092	0.0114
	2 (6)	6133 (290)	0.0127	0.0152	0.0112
	4 (12)	5560 (279)	0.0156	0.0118	0.0266
	6 (18)	5011 (247)	0.0198	0.0173	0.0224
	8 (24)	4524 (236)	0.0233	0.0211	0.0395
	10 (30)	4063 (217)	0.0264	0.0225	0.0428
	12 (36)	3637 (204)	0.0301	0.0281	0.0274
	16 (48)	2865 (174)	0.0334	0.0369	0.0348
	20 (60)	2214 (149)	0.0364	0.0412	0.0296
6 (18)	1 (3)	5843 (283)	0.0107	0.0102	0.0095
	4 (12)	5011 (247)	0.0159	0.0102	0.0185
	8 (24)	4063 (217)	0.0234	0.0304	0.0170
	12 (36)	3233 (187)	0.0316	0.0322	0.0271
	16 (48)	2527 (159)	0.0359	0.0336	0.0318
	20 (60)	1918 (136)	0.0386	0.0320	0.0412
	24 (72)	1421 (104)	0.0452	0.0362	0.0431
8 (24)	1 (3)	5281 (270)	0.0122	0.0068	0.0111
	8 (24)	3637 (204)	0.0260	0.0230	0.0184
	16 (48)	2214 (149)	0.0426	0.0282	0.0434
	24 (72)	1219 (92)	0.0469	0.0368	0.0493
12 (36)	1 (3)	4288 (225)	0.0137	0.0136	0.0163
	8 (24)	2865 (174)	0.0258	0.0174	0.0173
	16 (48)	1652 (122)	0.0441	0.0557	0.0321
	24 (72)	881 (69)	0.0527	0.0455	0.0600
	32 (96)	427 (40)	0.0540	0.0512	0.0558

Different combinations of T1 and T2 are tested to evaluate the performance of the model.

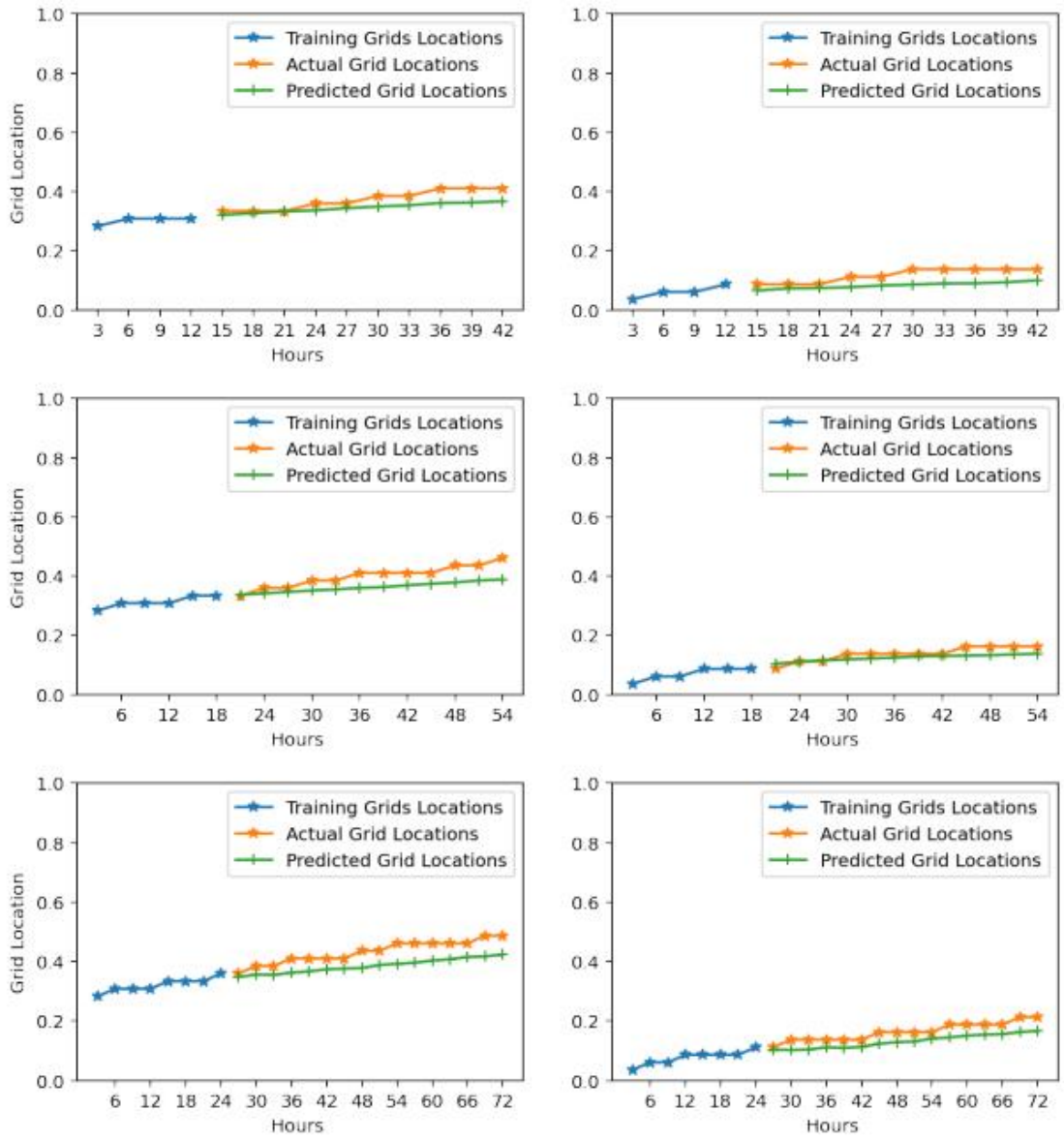


Fig 49 :- Grid predictions for Cyclones Vayu and Fani for T1 = 4,6,8 and T2 = 10,12,16

It is mentioned that the MAEs for the NI ocean cyclone data are small for all combinations of T1 and T2, indicating good performance of the proposed model. The model is also tested on recent cyclones such as Vayu and Fani, and it is reported that the model performed well on them. The actual and predicted grid locations for these cyclones are shown above, suggesting that the model was able to accurately predict their path.

5. CONCLUSION

Floods and storms are among the most devastating natural disasters in terms of their count, the number of affected people, and economic loss. TCs are the primary reason behind storms and floods in tropical and subtropical coastal regions. TCs bring various hazards in the form of flooding, storm surge, thunderstorms, lightning, and tornadoes. As per a report in the last 50 years, around 1942 disasters happened due to TCs that killed about 7,79,324 people and caused a financial loss of US$1407.6 billion. Equivalently, the disasters due to TCs cause, on average, 43 deaths and US$78 million in economic loss daily. Moreover, natural disasters affect the lower-middle group income countries the most, where it accounts for around 1.77% of their GDP, which is way higher than the IMF's threshold of 0.5% for a major economic disaster. All this makes the efforts towards mitigating the disaster caused by tropical cyclones central to sustainable and inclusive development.

An obvious way to mitigate and reduce the adverse effects of TCs is to have models that can forecast various development phases of TCs with high accuracy. Deep learning models have been successfully deployed to predict different weather-related phenomena, including TCs. Motivated by this, in this research work, we have developed various deep learning models using reanalysis and IBSTrACS data to forecast various essential phases of a TC. If we look at the different phases of a tropical cyclone, the first significant problem is to detect its formation well in advance. If a cyclone dies over the sea, then one need not worry, and it can be simply ignored. The next most critical and important phase of a cyclone is its landfall, which is the event when it hits the coastline and moves over to land. The location, intensity, and time of landfall of a cyclone determine the extent of disaster caused by it. An early information about whether a cyclone will make landfall or not, and if it is going to make landfall; where, at what speed and at what time will help the administration in mitigating the adverse effects of disaster. Therefore, we proposed a model that forecasts the track and time of landfall of a cyclone from the early phase of its development.

We proposed an LSTM based model is proposed for track prediction of a cyclone in terms of a GridID, which is obtained by applying the grid function on the ranges of latitudes and longitudes. Recently in (Alemany et al., 2018), a new track prediction approach is proposed in terms of GridID, in place of directly predicting latitudes and longitudes. In the comparison of (Alemany et al., 2018) where the model predicts GridID just one time-step ahead, we proposed a comparatively better model that predicts GridID several time steps ahead with better accuracy. The model performance is reported for Atlantic and NI ocean basins.

Predicting cyclone tracks well in advance can help in taking timely preventive measures to minimize human and property loss. One possible extension of this work could be to train an RNN model that can predict the actual location parameters, i.e., latitude and longitude, from GridIDs with a low conversion error. It may also be possible to experiment with different sizes of grids by changing the x value in the grid definition, which could help reduce the error margin. However, a challenge in reducing x would be to maintain the low MAE of GridIDs' prediction achieved by the proposed LSTM model.

6. FUTURE WORKS

The main contribution of this work is to answer the TC's related three prediction problems - TC formation, landfall event, and landfall's characteristics using deep learning models based on LSTM and CNN. A possible and exciting extension of the work will be in the following directions:

• One can explore using the reanalysis dataset to answer other TC-related forecast problems like storm surge, rainfall, continuous tracking of the path, and intensity.

• Reanalysis dataset ERA5 contains an extensive record of various atmospheric and oceanographic variables. Many studies have used the wind fields - u and v, geopotential z, and SST. It will be interesting to explore the usage of other variables like humidity, temperature, cloud cover, vorticity, and many others in answering above stated prediction problems.

• As per our knowledge, the proposed deep learning model that predicts landfall events and its characteristics is a promising attempt to answer these forecast problems. One can explore numerical or statistical methods to target these problems in the future.

• A promising research direction could be to propose ensemble models for these prediction tasks by combining numerical or statistical methods with deep learning techniques.

REFERENCES

1. https://whatsanswer.com/map/latitude-and-longitude-map-of-india-where-is-india/
2. https://www.google.com/url?sa=i&url=https%3A%2F%2Fkids.britannica.com%2Fstudents%2Fassembly%2Fview%2F115790&psig
3. https://www.mapsofworld.com/hurricane/mechanism-of-tropical-cyclone-formation.html
4. https://en.wikipedia.org/wiki/File:UK-Cyclone.gif
5. https://en.wikipedia.org/wiki/File:1-2-3_Rita_and_Philippe.png
6. https://weather.com/en-IN/india/news/news/2021-05-17-what-is-tauktae-landfall-impact
7. https://rsmcnewdelhi.imd.gov.in/landfall-forecast.php
8. https://mosdac.gov.in/
9. Pegahfar, N., Gharaylou, M. Entropy evolution characteristics during an intense tropical cyclone. *Meteorol Atmos Phys* **132**, 461–482 (2020).
10. arXiv:2212.06149v1 **[physics.ao-ph]**
11. Moradi Kordmahalleh, Mina & Gorji Sefidmazgi, Mohammad & Homaifar, Abdollah. (2016). A Sparse Recurrent Neural Network for Trajectory Prediction of Atlantic Hurricanes.957-964.10.1145/2908812.2908834.
12. Sharma, Vikas et al. "Diurnal Variation of Clouds Cover Area and Clouds Top Temperature over Tropical Cyclone in the North Indian Ocean Basins." (2019).
13. arXiv:2103.16108v1 **[cs.LG]** https://doi.org/10.48550/arXiv.2103.16108
14. Alemany, Sheila & Beltran, Jonathan & Perez, Adrian & Ganzfried, Sam. (2018). Predicting Hurricane Trajectories Using a Recurrent Neural Network. Proceedings of the AAAI Conference on Artificial Intelligence. 33. 10.1609/aaai.v33i01.3301468.
15. Abhishek, K., M. Singh, S. Ghosh, and A. Anand (2012). Weather forecasting model using artificial neural network. Procedia Technology, 4, 311 – 318. ISSN 2212- 0173. URL http://www.sciencedirect.com/science/article/pii/ S221201731200326X. 2nd International Conference on Computer, Communication, Control and Information Technology(C3IT-2012) on February 25 - 26, 2012.
16. Alemany, S., J. Beltran, A. Perez, and S. Ganzfried (2018). Predicting hurricane trajectories using a recurrent neural network. Proceedings of the AAAI Conference on Artificial Intelligence, URL https://ojs.aaai.org/index.php/AAAI/ article/view/3819.
17. Ali, M., P. Jagadeesh, I.-I. Lin, and J.-Y. Hsu (2012). A neural network approach to estimate tropical cyclone heat potential in the indian ocean. IEEE Geoscience and Remote Sensing Letters, 9, 1114–1117.
18. Bowman, K. P., and Fowler, M. D. (2015). The diurnal cycle of precipitation in tropical cyclones. Journal of Climate, 28(13), 5325-5334
19. Gray, W. M., and Jacobson Jr, R. W. (1977). Diurnal variation of deep cumulus convection. Monthly Weather Review, 105(9), 1171-1188