CyINSAT: Cyclone Dataset from Indian National Satellite for Forecasting

Deap Daru, Aditya Thaker, Akshath Mahajan, Adithya Sanyal, Meera Narvekar, and Debajyoti Mukhopadhyay

Abstract Tropical cyclones in India are annual natural disasters that have taken a toll of 1 lakh casualties in total to date and established the North Indian Ocean (NIO) as the deadliest basin historically. Moreover, it causes property damage and can lead to soil erosion, having adverse effects wherever it makes landfall. Effective forecasting models are needed to minimise the destructive aftermath of such a hazardous phenomenon. The first step in doing so is a consolidated dataset that is publicly available for research to take place. This study aims to combine satellite image data with cyclone track record data, which is currently not publicly available in a usable format. Satellite images containing four channels-two infrared, mid-wave infrared, and water vapour-from the Indian National Satellite System (INSAT) are obtained from the Meteorological and Oceanographic Satellite Data Archival Centre. Additionally, the track record data of cyclones with wind speed, pressure, category of the cyclone on the Indian Meteorological Department scale, and the latitude and longitude of the storm eye are retrieved from the Regional Specialised Meteorological Centre, New Delhi. The cyclone research community can use this dataset to develop a holistic forecasting model that can predict the track and intensity of any depression forming in the NIO basin, subsequently improving disaster management systems.

D. Daru (✉) · A. Thaker · A. Mahajan · A. Sanyal · M. Narvekar
Dwarkadas Jivanlal Sanghvi College of Engineering, Mumbai, India
e-mail: deapdaru@gmail.com

A. Thaker
e-mail: adityathaker28@gmail.com

A. Mahajan
e-mail: akshathmahajan13@gmail.com

A. Sanyal
e-mail: adithyasanyal@gmail.com

M. Narvekar
e-mail: meera.narvekar@djsce.ac.in

D. Mukhopadhyay
WIDiCoReL Research Lab, Mumbai, India
e-mail: debajyoti.mukhopadhyay@gmail.com

S. Roy et al. (eds.), *Innovations in Computational Intelligence and Computer Vision*,
Lecture Notes in Networks and Systems 680,
https://doi.org/10.1007/978-981-99-2602-2_24

Two applications of this dataset are explored later in this study: forecasting wind speed using time series forecasting models and predicting the cyclone's eye track.

Keywords Tropical cyclones · North Indian basin · Satellite images · Historical data · Forecasting

1 Introduction

A tropical cyclone (TC) is a multi-hazardous natural disaster often accompanied by heavy rains (can be more than 30 cm in 24 h), storm surges (at landfall), and gale winds (exceeding speeds of 34 knots). It forms in low-pressure areas over the North Indian Ocean (NIO) basin, progressing into an intense whirl in the troposphere with extreme winds that circulate in the anti-clockwise direction. The difference in pressure influences the intensity or wind speed of a cyclone from the centre to the outside of the cyclone. About 5–6 TCs occur over the NIO annually, having an average lifespan of 4–5 days [1]. These cyclonic disturbances are predominant during the pre-monsoon season, from March to May, and the post-monsoon season, from October to December. The NIO basin comprises the Arabian Sea sub-basin and the Bay of Bengal sub-basin, where the TCs develop in the ratio of 1:4 [2].

Although merely 7% of TCs form over the NIO basin, it has climatologically established itself as the deadliest basin [1]. These annual occurrences bring immense misfortune to the place where they make landfall, leading to loss of life, property damage, and devastation of the coastal areas. The storm's eye produces wind with enough torque to twist man-made structures, roofs, and even vegetation, leaving trees uprooted. This leaves power poles and communication towers vulnerable, compromising local telephone, cellular, and long-distance service. It also takes its toll on natural resources due to its high wind speeds and rainfall, harming marine ecosystems over the oceans and accelerating soil erosion over land. Cyclone Amphan of the 2020 season took a death toll of 98 in West Bengal and was the costliest cyclone of the NIO basin [2].

The only way to tackle and reduce the damage caused by TCs is to have a strong disaster prevention and mitigation system that can monitor and predict the intensity and path of the cyclone well in advance to deploy emergency services to aid the areas under threat. The Indian Meteorological Department (IMD) identifies two major inputs required to establish such a system: (1) synoptic analysis data that covers essential parameters necessary to assess the intensity of the cyclone and (2) satellite images from IMD containing infrared (IR) and water vapour (WV) channels [1]. The IMD has developed a credible cyclone warning system that utilises the above data for their numerical weather prediction (NWP) models to forecast and disseminate warnings to the public. Nevertheless, there is a lack of Deep Neural Networks implemented to predict the possibility of a cyclone and its lifetime and intensity [3]. This potentially will help improve the capacity of forecasting systems, consequently assisting the government in guaranteeing early warnings and advisories to be issued to the public.

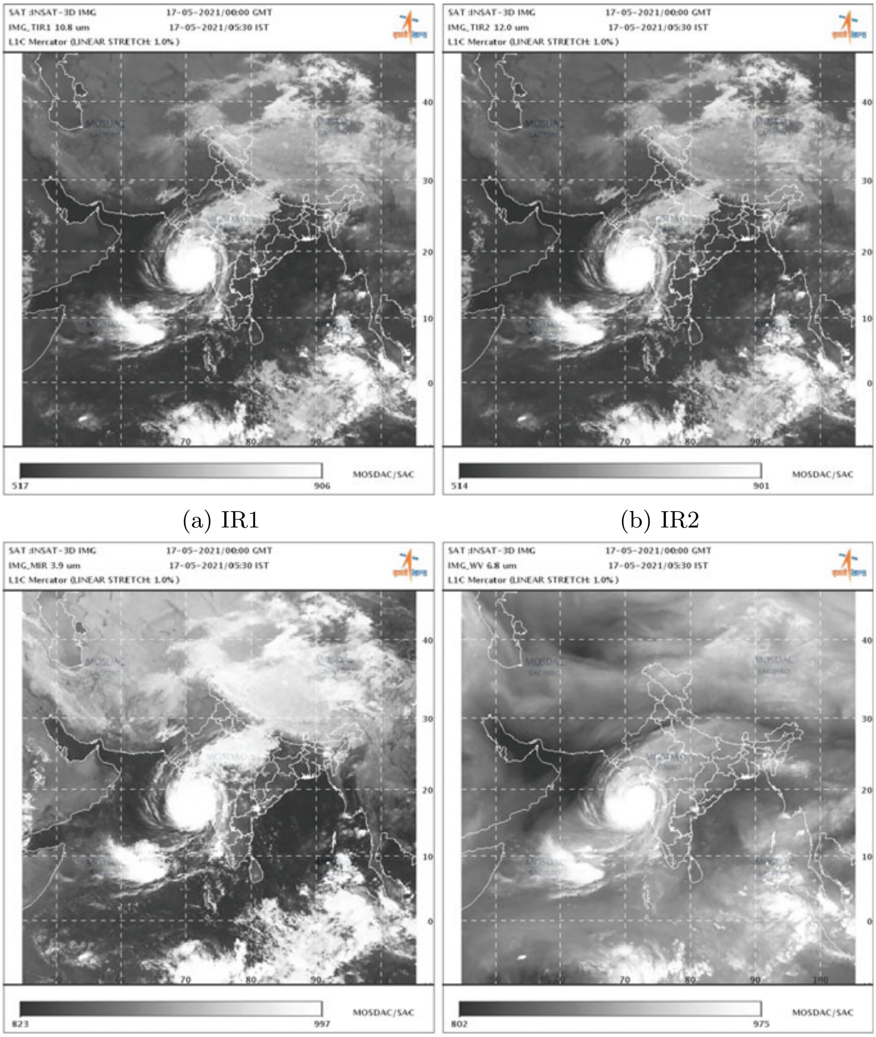
With the motivation to contribute to the betterment of cyclone prediction and forecasting, the following paper covers intricate details of the dataset in Sect. 2 with its properties in Sect. 2.1, constructions elaborated in Sect. 2.2 and previous datasets discussed in Sect. 2.3. Further, the paper explores a few of the applications of the dataset in the real-world setting using advanced Machine Learning (ML) and Deep Learning (DL) in Sect. 3 (Fig. 1).

2 Cyclone Dataset from Indian National Satellite

The Cyclone Dataset from Indian National Satellite (CyINSAT) is a usable and consolidated dataset for forecasting TCs curated through an amalgamation of data from the Meteorological and Oceanographic Satellite Data Archival Centre (MOSDAC) and the Regional Specialised Meteorological Centre, New Delhi (RSMC). This dataset will be extended publicly (through Kaggle) for the research community strictly (abiding by the terms and conditions of MOSDAC) to build state-of-the-art models for forecasting TCs over the NIO basin and advance the Indian disaster management systems. A better monitoring system would result in better warning dissemination and alert officials to make and take a plan of action to avoid human and economic loss.

The first systematic cyclonic studies can be dated back to the middle of the 19th century, when Henry Piddington, the then President of Marine Courts, used the meteorological logs of vessels navigating the seas of the Bay of Bengal [2]. Henry coined the term ‘Cyclone’ for the disturbances over the NIO basin because of their coil-like shape. RSMC—New Delhi, works under the Cyclone Warning Division of IMD and is responsible for the prediction and monitoring of any cyclonic disturbances over the NIO. It issues advisories for the World Meteorological Organisation (WMO) and Economic and Social Cooperation for Asia and the Pacific (ESCAP) Panel members for timely alerts. RSMC is also responsible for the collection and archival of data concerning these cyclonic disturbances and maintains the Best Track data for all the TCs. This Best Track data is revised and updated every January/February for the previous year’s record and is dispatched to the WMO/ESCAP. IMD categorises TC into five classes based on the sustained wind speed averaged over 3-minute for these disturbances, as mentioned in Table 1. There are two other cyclonic disturbances: Depression (17–27 kt) and Deep Depression (28–33 kt), but these are unnamed disturbances of the lowest classification that do not qualify as cyclones according to the RSMC and are therefore not considered for the sake of the paper and dataset creation.

IMD, under the Forecast Demonstration Project (FDP), has emphasised the need to improve current forecasting systems and satellite and radar technologies to aid experts in interpreting cyclonic activity and preparing bulletins efficiently and accurately. MOSDAC, which is a branch of the Indian Space Research Organisation


SAT :INSAT-3D IMG 17-05-2021/00:00 GMT
IMG_TIR1 10.8 um 17-05-2021/05:30 IST
L1C Mercator (LINEAR STRETCH: 1.0%)
517
906
MOSDAC/SAC
SAT :INSAT-3D IMG 17-05-2021/00:00 GMT
IMG_TIR2 12.0 um 17-05-2021/05:30 IST
L1C Mercator (LINEAR STRETCH: 1.0%)
514
901
MOSDAC/SAC
SAT :INSAT-3D IMG 17-05-2021/00:00 GMT
IMG_MIR 3.9 um 17-05-2021/05:30 IST
L1C Mercator (LINEAR STRETCH: 1.0%)
823
997
MOSDAC/SAC
SAT :INSAT-3D IMG 17-05-2021/00:00 GMT
IMG_WV 6.8 um 17-05-2021/05:30 IST
L1C Mercator (LINEAR STRETCH: 1.0%)
802
975
MOSDAC/SAC
40
30
20
10
0
70
80
90
100

(a) IR1

(b) IR2

(c) MIR

(d) WV

Fig. 1 INSAT 3D imager satellite images from MOSDAC (capturing Tauktae at its peak)

(ISRO), disseminates the satellite data collected from missions in near real-time through the Space Applications Centre, Ahmedabad (SAC). The Indian National Satellite (INSAT) 3D was launched on 26th July 2013 to receive images of the earth's surface of meteorological importance and offer satellite-aided search and rescue services. INSAT 3D has a six-channel Imager and a nineteen-channel Sounder to capture surface activity, including oceanic activity, vertical profiles of temperature and humidity in the atmosphere. The multi-spectral Imager can capture images in six

Table 1 IMD scale for TC intensity [1]

Category	Range of sustained winds (3-min average)
Cyclonic storm (CS)	34–47 kt
Severe cyclonic storm (SCS)	48–63 kt
Very severe cyclonic storm (VSCS)	64–89 kt
Extremely severe cyclonic storm (ESCS)	90–119 kt
Super cyclonic storm (SuCS)	$\geq$ 120 kt

different wavelengths, namely, visible (VIS), shortwave infrared (SWIR), mid-wave infrared (MIR), water vapour (WV), and two bands in the thermal infrared (IR1 and IR2) regions [4].

2.1 Properties

CyINSAT covers the TCs taken place from the 2014 NIO cyclone season to the present containing 39 cyclones in total. As labels it has 2 csv files namely *details.csv* and *paramters.csv*. The former contains data of all TCs in the columns: name, IMD-Scale, maxWindSpeedIMD (kmph), minPressure (hPa), startDate, peakDate, endDate, startToPeakDuration, and startToEndDuration. The latter contains data of all TCs in time series format in the columns: Folder, FileIR1, FileIR2, FileMIR, FileWV, Counter, Time (in GMT), WindSpeed (kt), Pressure (hPa), Latitude and Longitude. The column Folder and FileXXX give the path to access the file containing the image for the specific channel at the exact (approximation may be used, explained in Sect. 2.2) Time (at an interval of three hours).

Scale Containing 111,268 satellite images in total, the dataset takes up 12.3 GB of space on disk. Exhaustive statistics of the dataset is mentioned in Table 2.

Multimodal CyINSAT combines data of different formats, images, and numerical data. They provide varied information for ML and DL models to learn and deliver improved results.

Temporal Data contains the timestamp for each image and the respective maximum sustained wind speed, centre's pressure and latitude and longitude at that timestamp. This enables time series forecasting training to be conducted on the created dataset.

2.2 Construction

With high precedence given to the accuracy of the dataset, a pipeline is set up to generate a multimodal and temporal dataset rich enough to be researched upon.

Table 2 Statistics of CyINSAT

Total TCs	39
TCs from years	2014–2022
No. of INSAT images	111,268
Mean duration of TCs	5.49 days
Median duration of TCs	5 days
Mean category of TCs	2.44
Median category of TCs	2
Mean wind speed of TCs	129.36 kmph
Minimum from wind speeds of TCs	65 kmph
Maximum from wind speeds of TCs	240 kmph
Mean pressure of TCs	975.08 hPa
Minimum from pressures of TCs	920 hPa
Maximum from pressures of TCs	1000 hPa

Table 3 Utilised INSAT 3D imager specifications [5]

Spectral band	Range of wavelength (μm)	Ground resolution (km)
MIR	3.80–4.00	4
WV	6.50–7.10	8
IR1	10.3–11.3	4
IR2	11.5–12.5	4

Images Foremost, the INSAT 3D images were ordered from the MOSDAC portal covering the duration of the TCs (above the category of CS on the IMD scale) that took place from 2014 to 2022 since the data for the satellite is available after its launch in 2013 [6]. The Imager six-channel Mercator projection over the Asian Sector (3DIMG_L1C_ASIA_MER) images, which are taken at an interval of half-hour, was ordered, out of which only IR1, IR2, MIR, and WV channels, Table 3 which provides the specifications of these channels, were retained for the dataset for their importance in analysing TCs intensity and duration. This is done because VIS and SWIR cannot capture cloud cover at night since only the visible wavelengths are perceived by it. Furthermore, MIR aids in the capture of low clouds and fog during the nighttime, IR1 and IR2 include sea surface temperature with higher accuracy, and WV estimates the presence of humidity in the upper atmosphere [5]. When ordering satellite image data from the MOSDAC portal, five-day padding is included before and after the TC span to account for the varying duration of TCs over the years, which range from 2 to 10 days.

Labels Next from the RSMC, Best Track data provides three-hourly data on the parameters of the cyclone, like wind speed, category of the cyclonic disturbance, central pressure, drop in pressure, and the latitude and longitude of the cyclone eye. This data is available from 1982 to 2022, where the previous years' parameters are verified the following year in January or February. To combine the above data with this, we have to sync the timestamps for each data. Fortunately, both agencies already do this by having the time in GMT. A few duplicate images are present in the image data that have been deleted. A few of the timestamps in image data are not rounded to the nearest half-hour that has been rounded to map the Best Track data to the exact timestamp image (roughly rounded by 1–2 min). One major issue is that a few images for some timestamps are entirely missing (not retrieved from MOSDAC). To tackle the mapping of parameters for which the exact timestamp image does not exist, the data for that has been assigned to the nearest available timestamp image assigned algorithmically. About 3.2% of such data exists, making 96.8% of data align perfectly. This creates the *paramters.csv*, which is formatted to enable forecasting models to be trained on it. Several tests are run on this data to ensure the high accuracy of the produced dataset.

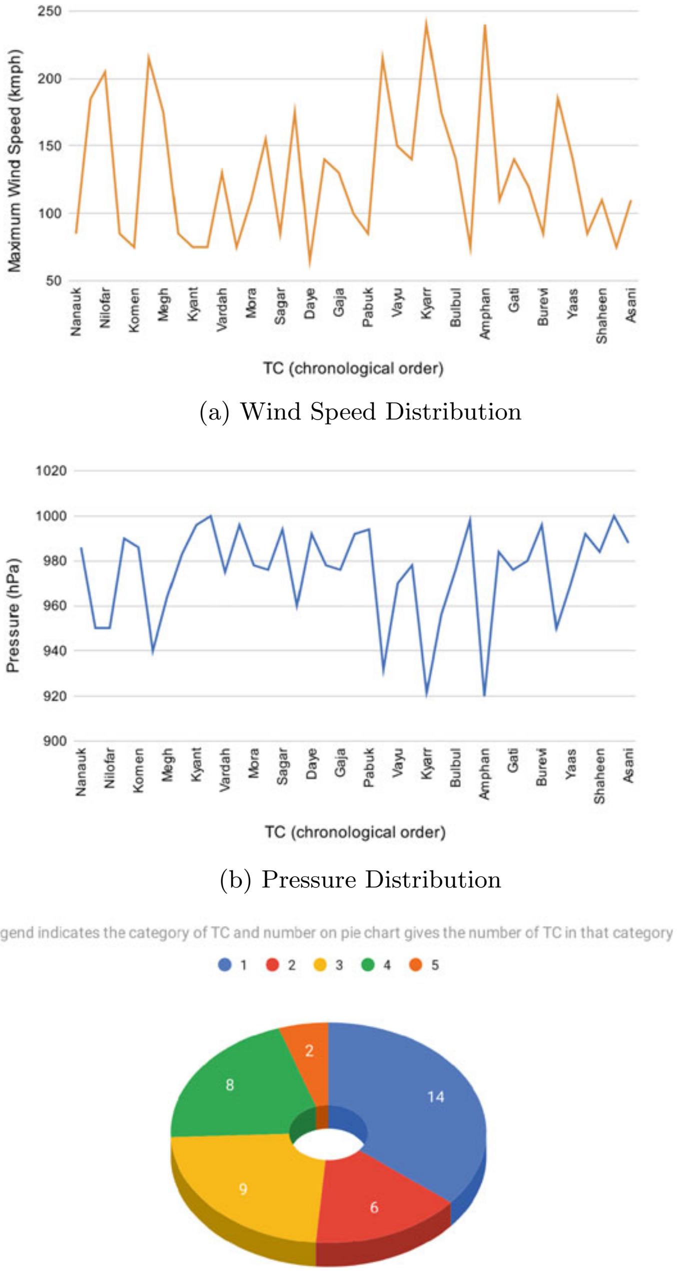
This dataset pipeline shall be employed to update the dataset for future TCs that take place over NIO, with the plan of updating the dataset half-yearly.

2.3 Related Datasets

The motivation for developing CyINSAT came from the sheer fact that only monomodal data exists for cyclones. Explored in Sect. 1, RSMC New Delhi has Best Tracks data publicly available that contains numerical parameters regarding TCs over NIO only. Similarly, RSMC Miami has developed Atlantic Hurricane Dataset (HURDAT, or recently updated to HURDAT2), covering the Atlantic basin and containing the best track data of TCs developed from 1851 to 2021 [7]. The International Best Track Archive for Climate Stewardship (IBTrACS), made as a result of the joint effort of WMO and the various RSMCs throughout the globe, is the most complete best track data of TCs [8]. These datasets still lack holistic information about the cyclonic developments since most data focuses on the centre of the cyclone's development (having information on its pressure, wind speed, etc.) [9]. Satellite imagery provides crucial information on how widespread the cyclonic depression is, its track with respect to its outer boundaries and the extent of land area that is vulnerable to impact and damages. With this aim, CyINSAT aims to fill in the gaps existing in the lack of publicly available multimodal data on TCs over the NIO.

3 Applications

The uses of the CyINSAT dataset are illustrated throughout this part. The precise and well-defined graphics acquired in the dataset are employed in the application described in the first subsection to forecast hurricane parameters. The second subheading looks at how we could exploit the photographs to forecast the track that a cyclone would follow (Fig. 2).



(a) Wind Speed Distribution

(b) Pressure Distribution

(c) Category Distribution

Fig. 2 Distribution of parameters in CyINSAT

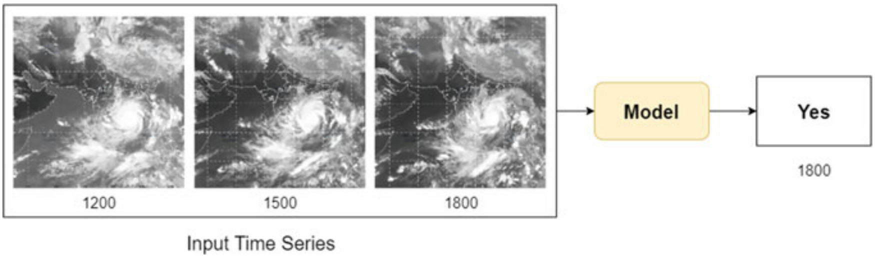


Fig. 3 Cyclone genesis classification

3.1 Forecasting Parameters

Parameters such as wind speed, air pressure, humidity, and temperature are critical environmental factors that determine the intensity and potential track of a tropical cyclone. However, these atmospheric variables seem to elude the general understanding due to their strong interdependence and hypersensitivity to any change in the environment. Forecasting these parameters has always been a high priority in not only predicting the course of a cyclone but also estimating the time it will take for the calamity to subside. Additionally, this provides deep insight into the locations in danger as well as the losses likely to be incurred. Hence, a robust system that can accurately predict the values of these variables is quintessential. Recent innovations in deep learning techniques pave the way for the development of such systems, by learning important information from data collected on cyclones. One such system can be created that learns the time series of cyclones along with these atmospheric parameters. They require data that includes images of cyclones over several days and the corresponding variables measured for every 3 h of every day.

CyINSAT dataset is capable of providing time series data for predicting these markers. The availability of temporal satellite images containing a myriad of channels. This provides deeper insight into image features over traditional RGB images, allowing the model to learn more meaningful representations. Among the several forecasting methods, [10] available are genesis and intensity predictions that can be performed leveraging the CyINSAT dataset.

The preliminary goal of genesis forecasting is to determine precisely whether or not a storm will originate. The following phases that make up the genesis of a cyclone are the transition from a tropical perturbation to a tropical depression and the subsequent growth from a tropical depression into a cyclone. The beginning of a cyclone time series can be used at the smaller granularity of time steps in order to learn genesis prediction. As depicted in Fig. 3, the model takes the input of a time series of cyclone images and a trained model produces a label indicating if a cyclone is being formed. On the other hand, the intensity of a cyclone is measured with the help of maximum wind speeds or minimum sea level pressure at the eye of the cyclone. A model can be trained to correlate the maximum wind speeds and cyclone image to learn a mapping of cyclone image features and the maximum intensity it

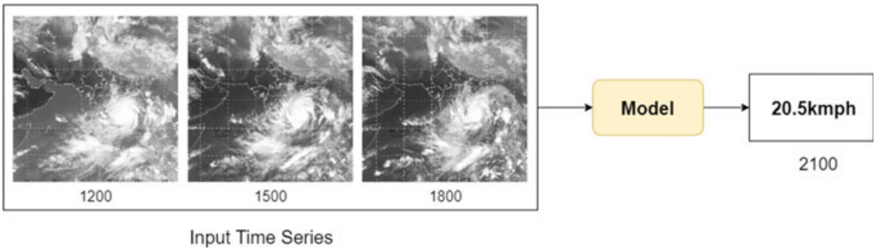


Fig. 4 Maximum intensity prediction

can attain. This is useful to determine the strength that a cyclone will achieve through its journey using the CyINSAT dataset's temporal half-hourly aggregation of these cyclone visuals and the associated wind speeds. As shown in Fig. 4, the model takes the input of a time series of cyclone images and a trained model produces a value for a given parameter, here wind speed.

3.2 *Cyclone Track Forecasting*

Measuring atmospheric variables isn't the most precise marker of the cyclone's course, a system that can depict the future position of the storm's eye can provide deeper insight to rescue and evacuation teams effectively minimising the adverse effects. Track prediction aims to produce an image of how the cyclone would evolve in the next time steps, by observing patterns in images across time series.

A time series of satellite photos of the cyclonic storms that transpired throughout the North Indian Ocean provided by CyINSAT can be used along with Generative Adversarial Networks (GANs) as proposed in [11] can be used to predict the course of any tropical hurricane in that region. It leverages the cyclone photos from a temporal stream to produce visuals that predict the cyclone's whereabouts at the next time step. This network uses satellite imagery and a designated cyclone centre as its input, along with its latitude and longitude data; without full knowledge of the wind velocity and direction in order to anticipate the positioning of the cyclone's eye. The marked red square is used to assess the precision of the vortex centre, as the model learns to look for it in a given image by tweaking feature maps. This is done because usually, networks have difficulty in learning a mapping between Cartesian coordinates and coordinates in the pixel-based representations [12]. Whilst the model is training, it utilises a colour filter on the produced pictures as well as a convolution with the scale of the red square upon that filtered image to detect the red quadrilateral in the frames. The position of the projected cyclone core is revealed by the pixel with the greatest value in the convoluted picture. After determining the location of the designated quadrilateral's pixel position, geospatial reference is used to translate those coordinates into the matching latitudinal and longitudinal location.

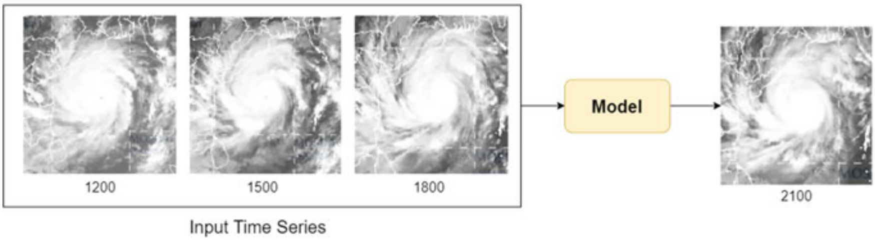


Fig. 5 Cyclone track prediction

The generator model is then trained adversarially to produce cyclone images that look similar to the ones in the next time step. This encourages the discriminator to discredit the generator model's output as fake for the slightest change in dissimilarity score values. Hence, a system to approximate future storm images can be built using CyINSAT. As shown in Fig. 5, the model takes an input of a time series of images cropped to only include the cyclone with its eye at the centre and the GAN model produces a cyclone image with the next approximate position of the eye.

4 Conclusion

The immediate plan to grow the CyINSAT dataset includes the inclusion of INSAT-3A images ordered from MOSDAC that covers 2009 to 2016, which shall be used to add all TCs from 2010 to 2014 to CyINSAT [13]. The CyINSAT dataset currently provides cyclone images for forecasting tasks including forecasting tracks as well as forecasting parameters.

However, these tasks might not be enough to accurately analyse these tropical storms. Another really important task is semantic segmentation, which is the task of identifying and colouring separate objects in a given image. This task facilitates analyses of cyclone shapes which can be correlated to their damage and change in course, essentially providing a mapping between different shapes and their potential lethality. Another interesting problem is predicting wind speed and other parameters at different regions of the cyclone, this could provide deeper insight into exactly what type of conditions cause these storms and if some remedial measures can be taken early on in the genesis period to avoid damage to human life and property.

References

1. IMD (2021) Cyclone warning in India standard operation procedure, Mar 2021
2. IMD (2020) Super cyclonic storm, 'amphan' over southeast Bay of Bengal: a preliminary report
3. WMO (2021) Tropical cyclone operational plan for the Bay of Bengal and the Arabian Sea

4. Choudhary RK, Siddharth Srivastav AP, Kaushik NK (2022) Incorporation of version information in INSAT-3D and INSAT-3DR imager and sounder data products
5. IMD (2014) Insat-3d data products catalog
6. MOSDAC. https://doi.org/10.19038/SAC/10/3DIMG_L1C_ASIA_MER, https://mosdac.gov.in
7. Landsea CW, Franklin JL (2013) Atlantic hurricane database uncertainty and presentation of a new database format. Mon Weather Rev 141(10):3576–3592
8. Knapp KR, Kruk MC, Levinson DH, Diamond HJ, Neumann CJ (2010) The international best track archive for climate stewardship (IBTRACS) unifying tropical cyclone data. Bull Am Meteorol Soci 91(3):363–376
9. Knapp KR, Diamond HJ, Kossin JP, Kruk MC, Schreck C et al (2018) International best track archive for climate stewardship (IBTRACS) project, version 4. NOAA National Centers for Environmental Information
10. Chen R, Zhang W, Wang X (2020) Machine learning in tropical cyclone forecast modeling: a review. Atmosphere 11(7). https://doi.org/10.3390/atmos11070676, https://www.mdpi.com/2073-4433/11/7/676
11. Rüttgers M, Lee S, Jeon S, You D (2019) Prediction of a typhoon track using a generative adversarial network and satellite images. Sci Rep 9(1):1–15
12. Liu R, Lehman J, Molino P, Petroski Such F, Frank E, Sergeev A, Yosinski J (2018) An intriguing failing of convolutional neural networks and the CoordConv solution. In: Advances in neural information processing systems, vol 31
13. MOSDAC. https://doi.org/10.19038/SAC/10/3A-VHR-L1, https://mosdac.gov.in