



Thesis title

Can Generative Artificial Intelligence be used to synthesise images of the rarest tropical cyclones?

Short description

Tropical cyclones (TCs), also known as hurricanes or typhoons, are among the most devastating extreme climate events: over the past 50 years, they killed over 700,000 people and caused \$1,400 billion in economic losses [1]. Unfortunately, TCs are still poorly understood phenomena, due to the highly complex physical processes that govern them. Furthermore, trustworthy records of TC go back only as far as the late 1970s (when satellite programs to monitor TCs started), meaning there is little data on TCs, especially for the most intense, rare ones.

Such lack of data inhibits the use of state-of-the-art deep learning models, which require enormous datasets to be trained. This is especially true for image-based analyses of TCs, which are typically done using large, data-hungry Convolutional Neural Networks (CNNs). Therefore, researchers studying TCs often use *data augmentation*: they apply transformations (e.g. flipping, rotating) to images of cyclones, treating the transformed images as new samples and thus expanding their dataset. However, the transformed images are only marginally helpful to train deep learning models because they are highly correlated with the ones from which they have been generated, meaning they add little information to the dataset.

With the rise of generative artificial intelligence models (gAI) like GANs, Stable Diffusion, and GPT, researchers have started using them as data augmentation tools, in place of or in addition to traditional data augmentation techniques. A crucial advantage of a sample generated by a gAI model is that it is not directly derived from any single sample from the starting dataset and therefore adds as much information as the original samples. Furthermore, the number of ways in which you can combine traditional data augmentation to generate new samples is limited; using gAI, one could generate infinite new samples.

A branch of TC analysis that has been particularly active in exploring data augmentation techniques is TC intensity estimation (TCIE): given a satellite image of a TC (e.g. Fig 1), the goal is to estimate—usually with a CNN—the corresponding wind speed, from which one can derive an estimate of potential damages [2]. Only traditional data augmentation techniques have been used for TCIE, but such techniques are unable to generate large enough numbers of images for the rarest, most intense TCs, leaving datasets highly imbalanced even after augmentation. In this thesis project, one or two students will assess whether gAI models can be applied in this setting, and whether they improve the performance of TCIE models compared to traditional data augmentation techniques. To achieve these goals, the student(s) will carry out the following activities:

- Literature review: review the state of the art for TCIE models (i.e., CNNs), data augmentation, and generative AI models.
- Data elaboration: retrieving GRIDSAT-B1 satellite images from the National Oceanic and Atmospheric Administration (NOAA) archive [3], pre-processing them to remove outliers and images with missing data, culminating in the formation of a dataset of images to be used for training deep learning models.

- Computational experiments:
 - Implementing (or adapting off-the-shelf) generative AI models to synthesise images of TCs, training them on the dataset acquired.
 - Using the generated samples as a form of data augmentation for TC intensity estimation (using an off-the-shelf CNN).
 - Implementing traditional data augmentation techniques as a benchmark against which to evaluate the proposed method.

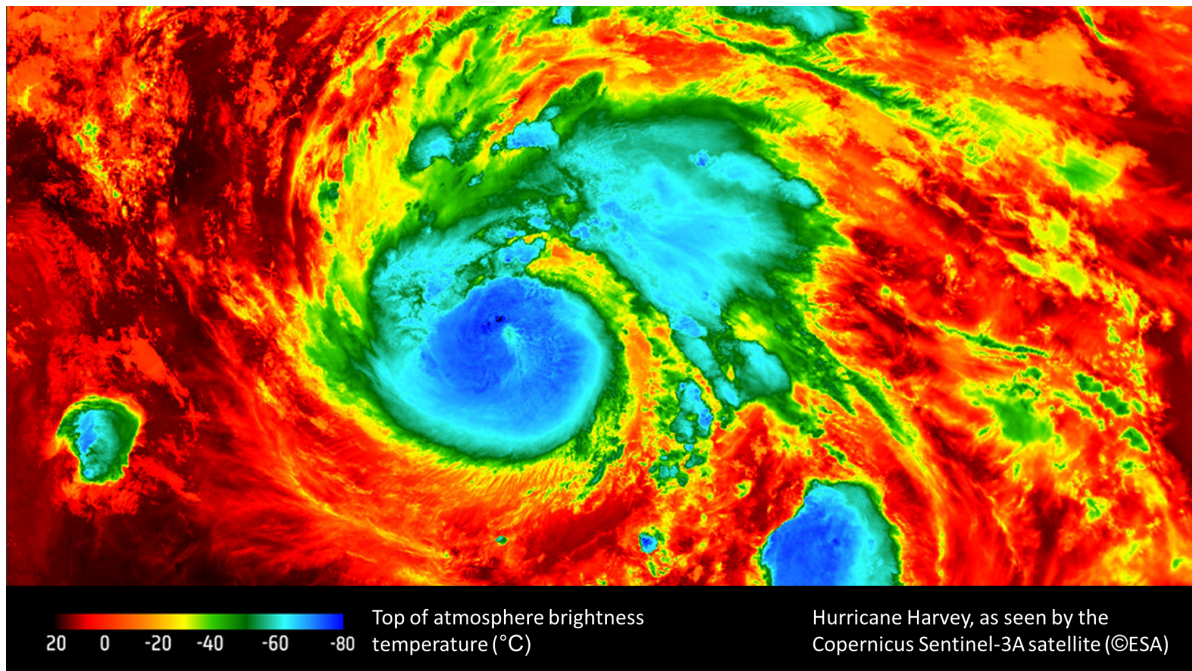


Fig 1: Example of the type of images used for Tropical Cyclone Intensity Estimation (TCIE). This image is a satellite view of Hurricane Harvey, which hit the U.S. in 2017 killing 100 people and causing \$125 billion damages.

Relevant courses and knowledge: Natural Resources Management

Number of Students: 1 or 2

Requisites: The student(s) **must** be comfortable with coding (Python will be used for the project). Knowledge of machine learning tools is a major plus.

References

- (1) <https://public.wmo.int/en/our-mandate/focus-areas/natural-hazards-and-disaster-risk-reduction/tropicalcyclones#:~:text=Over%20the%20past%2050%20years,million%20in%20damages%20every%20day>
- (2) <https://www.ncdc.noaa.gov/gridsat/>
- (3) Pradhan, R., Aygun, R. S., Maskey, M., Ramachandran, R., & Cecil, D. J. (2017). Tropical cyclone intensity estimation using a deep convolutional neural network. *IEEE Transactions on Image Processing*, 27(2), 692-702.