Curating flood extent data and leveraging citizen science for benchmarking machine learning solutions

Shubhankar Gahlot[1], Muthukumaran Ramasubramanian[1], Iksha Gurung[2], Ronny Hansch[3], Andrew Molthan[4], and Manil Maskey[5]

[1]University of Alabama in Huntsville
[2]Unknown
[3]TU Berlin
[4]NASA Marshall Space Flight Center
[5]NASA

November 24, 2022

Abstract

We present a labeled machine learning (ML) training dataset derived from Sentinel 1 C-band synthetic aperture radar (SAR) data for flood events. In this paper, we detail the steps to collect, pre-process, label, curate, and catalog the training dataset. Development of benchmark ML models and usage of the training datasets for a data science competition are also presented.

Curating flood extent data and leveraging citizen science for benchmarking machine learning solutions

Shubhankar Gahlot[1], Muthukumaran Ramasubramanian[1], Iksha Gurung[1], Ronny Hänsch[2], Andrew L. Molthan[3], Manil Maskey[3]

[1]The University of Alabama Huntsville
[2]German Aerospace Center
[3]National Aeronautics and Space Administration

Corresponding author: Shubhankar Gahlot, `shubhankar.gahlot@uah.edu`

Abstract

We present a labeled machine learning (ML) training dataset derived from Sentinel 1 C-band synthetic aperture radar (SAR) data for flood events. In this paper, we detail the steps to collect, pre-process, label, curate, and catalog the training dataset. Development of benchmark ML models and usage of the training datasets for a data science competition are also presented.

1 Plain Language Summary

We discuss a machine learning (ML) training dataset designed for detecting flood extent in open waters from the cross-polarized and co-polarized returns of free and open Sentinel-1C-band synthetic aperture radar (SAR). We demonstrate the need for curating and providing this dataset, its detailed data structure, and the data processing procedures involved in generating the dataset. We also discuss how we leveraged citizen science to accelerate ML research for detecting flood extents.

2 Introduction

Floods are major natural disasters and contribute to widespread property damage, loss of agricultural productivity, loss of lives, displacement of those affected, and long-term socioeconomic consequences (Dawson et al., 2009; Boros & Nagy, 2014; Long et al., 2014; Inambao, 2013). Knowing the spatial extent of floods is crucial for federal agencies, local authorities, and nonprofits in providing emergency procedures and disaster relief. Flooding is caused by 1) persistent, above-normal rainfall (Alias et al., 2016), 2) flash flooding from severe thunderstorms (Boardman et al., 1996), 3) coastal flooding during high tides and strong onshore flow (Spicer et al., 2019), 4) storm surge and river backup during landfalling tropical cyclones, and 5) inland heavy rains from dissipating storms (Ullman et al., 2019). These aforementioned hazardous conditions render monitoring flood events in-situ difficult.

Remote sensing has been used extensively in the community to monitor these events (Sanyal & Lu, 2004; Schumann et al., 2009; Jain et al., 2005; Klemas, 2015). The temporal and spatial availability of remote sensing data provided by recent governmental and commercial satellites enable the community to make large scale analysis of flood events with greater detail than ever before. For example, synthetic aperture radar (SAR) (Curlander & McDonough, 1991) imagery has been used extensively for quantification and delineation of flood extents (Long et al., 2014; Matgen et al., 2011). Machine learning (ML) has also been leveraged for stochastically mapping the flood extents using SAR imagery (Benoudjit & Guida, 2019). Advancements in remote sensing coupled with the scalability of ML paradigms can vastly improve the volume and velocity of flood extent mapping.

However, finding an optimal machine learning solution is an exhaustive process in itself. Citizen science has been used extensively to find the best solution for problems in both scientific and commercial sectors (Beaumont et al., 2014; Borne & Team, 2011). As part of incorporating citizen science and involving the broader science community to find the best solution, we created a human-curated, ML ready flood extent dataset that can be used by data scientists from all discipline and sectors. Furthermore, we designed and hosted a competition on flood detection using the aforementioned dataset to estimate the flood extent based on satellite imagery. The competition is showcased by the International Conference on Emerging Techniques in Computational Intelligence (ICETCI), 2021.

The details of the curated flood dataset are explained in the following section.

3 Overview of the Dataset

3.1 Data Collection and Labeling

SAR imagery for various flood events were acquired from the European Space Agency (ESA) Sentinel-1A and Sentinel-1B missions, offering C-band, dual polarized (co-pol VV and cross-pol VH) imagery for a number of flood events of interest for the following regions within the United States and globally.

- Bangladesh (7150 sq. km.)
- Florence (7197 sq. km.)
- Nebraska (1741 sq. km.)
- North Alabama (13789 sq. km.)
- Red River North (6746 sq. km.)

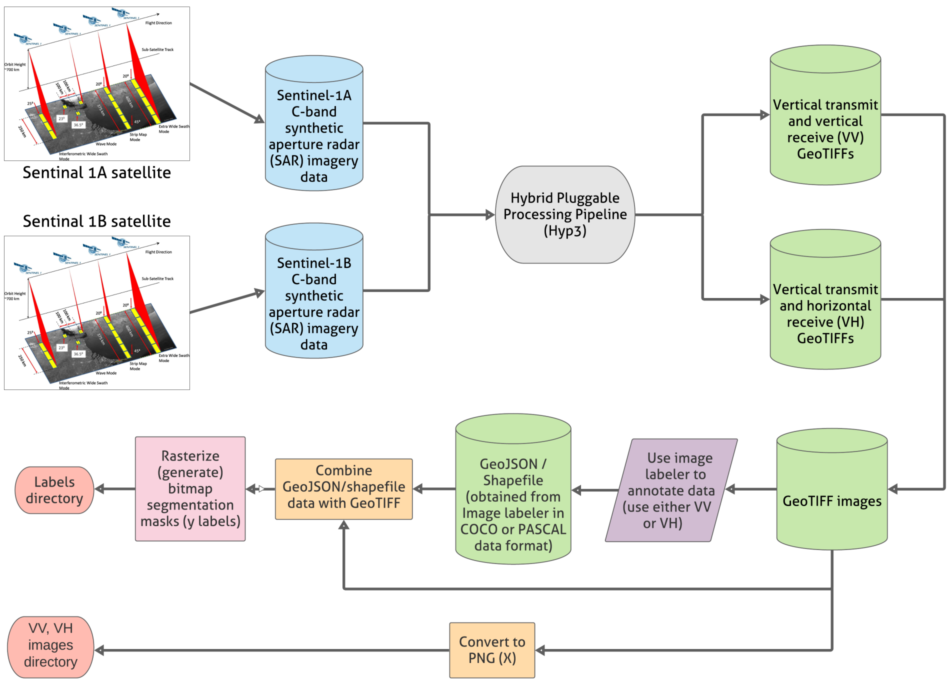

Figure 1. Data pre-processing & generation workflow

Images were processed to a radiometric and terrain-corrected (RTC) image of the radar amplitude, then converted to a grayscale image for visual analysis using the Hybrid Pluggable Processing Pipeline or "HyP3" system which takes the Sentinel archive and creates a set of processes to get to a consistent method of generating the VV / VH amplitude or power imagery. Here, emphasis was on the labeling of open water areas where specular reflection of the radar signal off of the relatively still, flat open water surface results in reduced backscatter, low amplitude, and an overall darkened appearance within the image. In normal conditions, ponds, lakes, and rivers will appear dark and usually include crisp edges where water adjoins the nearby vegetation and topography. Following heavy rains and flooding, additional dark features occur and often include expanded, flooding growth of dark regions along the normal water areas or standing water in fields or other topographic features where ponding of water is likely (Liang & Liu, 2020; Horritt et al., 2001). Emphasis was made on the labeling of these features for generating the training dataset.

Flood domain experts reviewed the data before it was provided to Earth science students for labeling. Imagery for the various flood events were made available in the ImageLabeler tool (*ImageLabeler*, 2021) developed by the NASA-IMPACT team. Multiple dates of post-event scenes were generated as grayscale imagery enhanced to focus on the contrast of dark, open water features for visual identification. Detailed polygons were drawn for suspected water areas and vetted through discussion with other analysts and project team members with additional SAR imagery expertise. Areas that were "dark" in the SAR images and might not have been water bodies were particularly challenging to examine. Alternate data sources were used to make sure that they were permanent water bodies. These polygons represent the open water class as expert labels and were used to classify open water pixels relative to vegetation and other classes in the image.

3.2 Data Preprocessing

Following the data collection, the imagery is then preprocessed and converted to 0-255 grayscale images using various GIS libraries. A total of 54 labeled GeoTiff files are converted into grayscale images before subsetting them into 256×256 tiles (scenes) by eliminating overlaps and omitting the tiles outside of the valid SAR boundary. In addition to the flood data, World Water Bodies GeoTiff data from UCLA Geoportal (*UCLA Geo-portal world water bodies*, 2021) is also preprocessed into water body labels and provided for the respective regions which should improve model training (see 1). Next, the data is divided into train (Nebraska, North Alabama, Bangladesh), validation (Florence), and test set (Red River North region). They are divided at geographically to make sure that the train-validation and test distributions are not similar:

- train
- validation
- test

Nebraska, north Alabama, Bangladesh and Florence regions are used for the train and validation set whereas the Red River North region is used for the test set. Each region directory contained the following sub-directories with the corresponding image types (normalized / contrast enhanced polarization amplitudes described above):

- VV (polarization amplitude) (Fig. 2 top left)
- VH (polarization amplitude) (Fig. 2 top right)
- Water body label (Fig. 2 right)
- Flood label (Fig. 2 left)

3.3 Scientific utility of the flood dataset

ML is being extensively used as part of many scientific workflows to solve many problems. Since ML models depend on the quality and quantity of the data they are trained on, it is necessary to provide such data for their training and continuous improvement.

Most current practical examples of ML are applications of supervised learning (*Supervised Learning: Model Popularity from Past to Present*, 2018). Supervised learning is used when labeled data is available and the preferred target variables are known (Liu & Wu, 2012). Training data is used to help a system learn relationships of given inputs to a given output—for example, to recognize objects in an image or to transcribe human speech. More recently, advanced supervised learning algorithms have shown to outperform existing benchmarks in many applications. However, these advances can be traced back to the availability of large scale training datasets. For example, ImageNet (Krizhevsky et al., 2012) for image classification tasks or the Spoken Wikipedia Corpora for speech recognition tasks, etc. If it wasn't for these datasets we wouldn't have had the cutting edge com-

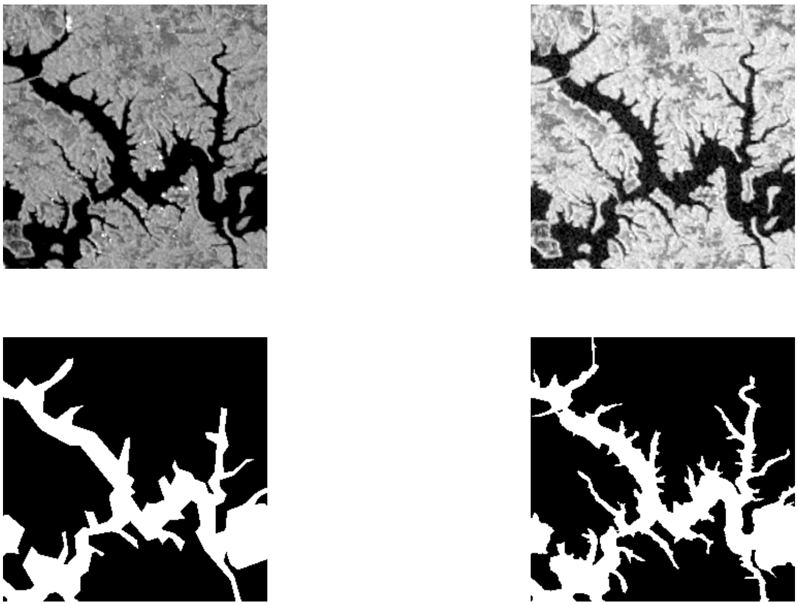

Figure 2. SAR image tiles with flood areas in northern Alabama. Left to right and top to bottom : VV, VH, Flood label, Water body label

puter vision and speech recognition that we have today. These datasets provided the clean and curated data that ML models are trained and tested on.

Thus, by creating this dataset and making it open source (Gahlot et al., 2021) we want to lower the barrier to use machine learning for flood extent detection.

4 Baseline models

Two image segmentation models were trained on the flood data for generating model baseline. The models trained were:

1. FPN (Feature Pyramid Network) (Ronneberger et al., 2015) with a ResNet50 backbone. An FPN is a feature extractor that takes a single-scale image of an arbitrary size as input and outputs proportionally sized feature maps at multiple levels, in a fully convolutional fashion.
2. UNet (Lin et al., 2017) with a ResNet50 backbone. U-Net is an architecture for semantic segmentation. It consists of a contracting path and an expansive path. The contracting path follows the typical architecture of a convolutional network.

Both networks were trained for 100 epochs without any regularization by stacking VV, VH and Water body label images as 3-channels of the input tensor. The models were tested using the IOU scoring function also known as Jaccard Index given by Eq. 1.

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} \tag{1}$$

where A and B are the estimated and reference flood masks respectively.

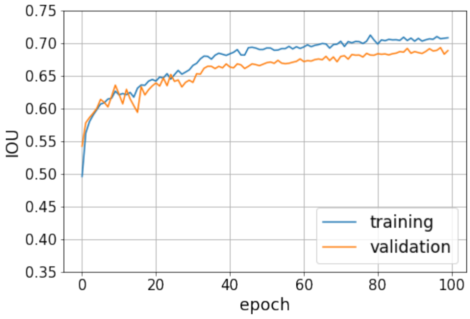

Figure 3. Train-validation scores for FPN

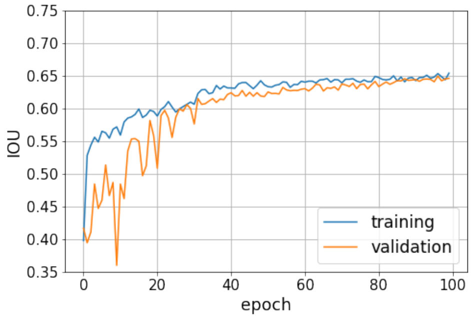

Figure 4. Train-validation scores for UNet

4.1 Results

Fig. 3 and Fig. 4 show the accuracy plots for Unet and Feature Pyramid Network respectively with ResNet50 backbone. The UNet training is noisy and relatively slower in comparison to the FPN training because for the same number of initial epochs FPN gives a higher accuracy. However, UNet closes in as the training progresses with each epoch.

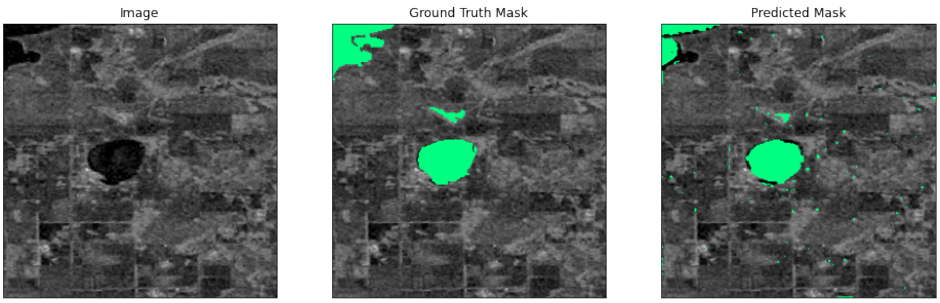

Figure 5. A sample UNet prediction

The IOU scores on the test set are 0.6021 and 0.6198 for FPN and UNet models, respectively. The difference in accuracies could be attributed to the initialization. The baseline models provide a starting point for more advanced deep learning techniques. Fig. 5 shows a sample from Unet predictions. The image on the left is the actual VV sample. The image in the center is the ground truth flood mask, and on the right is the predicted flood mask overlaid on top of the actual image.

5 Community Involvement

The dataset was made publicly available (Gahlot et al., 2021) for a machine learning competition which also helped in reaching out to the broader science community to solve the flood detection problem. The competition was organized in collaboration with the International Conference on Emerging Techniques in Computational Intelligence (ICTE) and Geoscience and Remote Sensing Society (GRSS). The competition commenced on April 15, 2021 and concluded on July 15, 2021. The competition received a total 137 participants and more than 200 submissions. The competition was hosted using the Codalab competition platform (*ETCI flood competition portal*, 2021) (*ETCI competition page*, 2021) which is an open source, community driven data science competition platform.

There were total of 309 submissions from 142 participants. three winners were chosen based on their IOU score. All winners used some version of the UNet deep neural network with different pre- and post-processing steps. The highest IOU score achieved was 0.7681 followed by 0.7654 and 0.7506. The competition was divided into 2 phases (Phase 1: Development and Phase 2: Test)(*ETCI competition page*, 2021). Phase 1 ran from April 15 - May 15 and phase 2 from May 16 - June 15. The competition timeline is shown in Fig. 6.

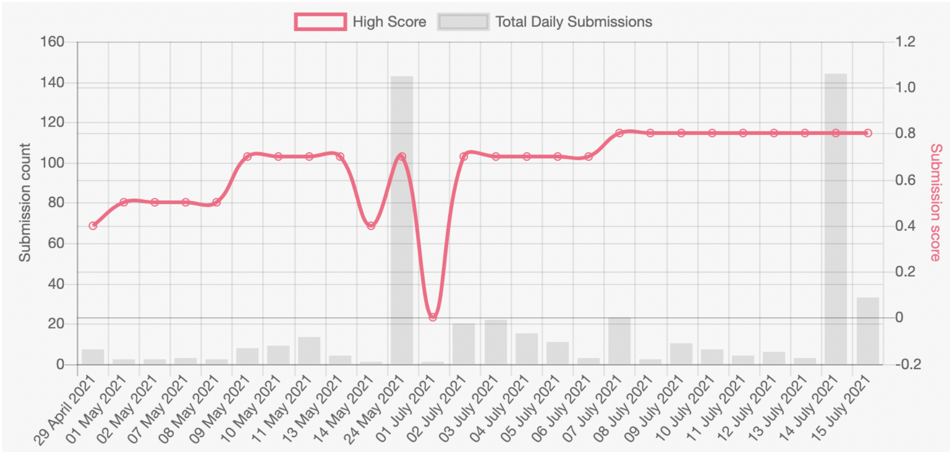

Figure 6. Best IOU scores and number of submissions throughout the competition

6 Conclusion

Machine learning has become an important part of the workflows for solving many scientific problems which requires large amounts of clean data. The saying "garbage in garbage out" hasn't been more true in any other domain than it has been in ML (Geiger et al., 2020). This paper documents the process of generating and curating a high quality flood data to help lower the barrier of entry for flood extent detection using machine learning, and how citizen science can be leveraged to involve the broader scientific community and use its collective efforts to crowd source better machine learning solutions.

Acknowledgments

We would like to thank the NASA-IMPACT, IEEE GRSS, University of Alabama in Huntsville, NASA Disaster Team at MSFC for supporting training data generation and International Conference on Emerging Techniques in Computational Intelligence (ICETCI) for hosting the competition. Special thanks to Dr. Ronny Hänsch for his support throughout the competition and chairing this event at ICETCI and the students: Jacob Robinson, Kiahna Mollette, Kaitlyn Wheeler, Stefanie Mehlich and Zachary Helton; and research staff Ankur Shah and Ronan M. Lucey for their help curating dataset.

Dataset for this research is available in this in-text data citation reference: (Gahlot et al., 2021) [CC-BY-4.0]. Such dataset must be findable and accessible from `https://registry.mlhub.earth/10.34911/rdnt.ebk43x`

References

Alias, N. E., Mohamad, H., Chin, W. Y., & Yusop, Z. (2016). Rainfall analysis of the kelantan big yellow flood 2014. *Jurnal Teknologi*, *78*(9-4).

Beaumont, C. N., Goodman, A. A., Kendrew, S., Williams, J. P., & Simpson, R. (2014). The milky way project: leveraging citizen science and machine learning to detect interstellar bubbles. *The Astrophysical Journal Supplement Series*, *214* (1), 3.

Benoudjit, A., & Guida, R. (2019). A novel fully automated mapping of the flood extent on sar images using a supervised classifier. *Remote Sensing*, *11* (7), 779.

Boardman, J., Burt, T., Evans, R., Slattery, M., & Shuttleworth, H. (1996). Soil erosion and flooding as a result of a summer thunderstorm in oxfordshire and berkshire, may 1993. *Applied Geography*, *16* (1), 21–34.

Borne, K., & Team, Z. (2011). The zooniverse: A framework for knowledge discovery from citizen science data. In *Agu fall meeting abstracts* (Vol. 2011, pp. ED23C–0650).

Boros, L., & Nagy, G. (2014). The long-term socioeconomic consequences of the tisza flood of 2001 in szabolcs-szatmár-bereg county, hungary. *Belvedere Meridionale*, *26* (4), 122–130.

Curlander, J. C., & McDonough, R. N. (1991). *Synthetic aperture radar* (Vol. 11). Wiley, New York.

Dawson, R. J., Dickson, M. E., Nicholls, R. J., Hall, J. W., Walkden, M. J., Stansby, P. K., . . . others (2009). Integrated analysis of risks of coastal flooding and cliff erosion under scenarios of long term change. *Climatic Change*, *95* (1), 249–288.

Etci competition page. (2021, Apr). Retrieved from https://nasa-impact.github.io/etci2021/

Etci flood competition portal. (2021, Apr). Retrieved from https://competitions.codalab.org/competitions/30440

Gahlot, S., Gurung, I., Molthan, A., & Maskey, M. (2021). *Flood extent data for machine learning.* Radiant MLHub. Retrieved from https://registry.mlhub.earth/10.34911/rdnt.ebk43x doi: 10.34911/RDNT.EBK43X

Geiger, R. S., Yu, K., Yang, Y., Dai, M., Qiu, J., Tang, R., & Huang, J. (2020). Garbage in, garbage out? do machine learning application papers in social computing report where human-labeled training data comes from? In *Proceedings of the 2020 conference on fairness, accountability, and transparency* (pp. 325–336).

Horritt, M. S., Mason, D. C., & Luckman, A. J. (2001). Flood boundary delineation from synthetic aperture radar imagery using a statistical active contour model. *International Journal of Remote Sensing*, *22* (13), 2489-2507. Retrieved from https://doi.org/10.1080/01431160116902 doi: 10.1080/01431160116902

Imagelabeler. (2021). Retrieved from https://impact.earthdata.nasa.gov/labeler/

Inambao, C. (2013). Namibia: Caprivi floods reach historic mark. *Relief Web, Accessed*, *28*.

Jain, S. K., Singh, R., Jain, M., & Lohani, A. (2005). Delineation of flood-prone areas using remote sensing techniques. *Water resources management*, *19* (4), 333–347.

Klemas, V. (2015). Remote sensing of floods and flood-prone areas: an overview. *Journal of Coastal Research*, *31* (4), 1005–1013.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, *25*, 1097–1105.

Liang, J., & Liu, D. (2020). A local thresholding approach to flood water delineation using sentinel-1 sar imagery. *ISPRS journal of photogrammetry and remote sensing*, *159*, 53–62.

Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). *Feature pyramid networks for object detection.*

Liu, Q., & Wu, Y. (2012). Supervised learning. In N. M. Seel (Ed.), *Encyclopedia of the sciences of learning* (pp. 3243–3245). Boston, MA: Springer US. Retrieved from https://doi.org/10.1007/978-1-4419-1428-6_451 doi: 10.1007/978-1-4419-1428-6_451

Long, S., Fatoyinbo, T. E., & Policelli, F. (2014). Flood extent mapping for namibia using change detection and thresholding with sar. *Environmental Research Letters*, *9*(3), 035002.

Matgen, P., Hostache, R., Schumann, G., Pfister, L., Hoffmann, L., & Savenije, H. (2011). Towards an automated sar-based flood monitoring system: Lessons learned from two case studies. *Physics and Chemistry of the Earth, Parts A/B/C*, *36*(7-8), 241–252.

Ronneberger, O., Fischer, P., & Brox, T. (2015). *U-net: Convolutional networks for biomedical image segmentation.*

Sanyal, J., & Lu, X. X. (2004). Application of remote sensing in flood management with special reference to monsoon asia: a review. *Natural Hazards*, *33*(2), 283–301.

Schumann, G., Bates, P. D., Horritt, M. S., Matgen, P., & Pappenberger, F. (2009). Progress in integration of remote sensing–derived flood extent and stage data and hydraulic models. *Reviews of Geophysics*, *47*(4).

Spicer, P., Huguenard, K., Ross, L., & Rickard, L. N. (2019). High-frequency tide-surge-river interaction in estuaries: Causes and implications for coastal flooding. *Journal of Geophysical Research: Oceans*, *124*(12), 9517–9530.

Supervised learning: Model popularity from past to present. (2018). Retrieved from https://www.kdnuggets.com/2018/12/supervised-learning-model-popularity-from-past-present.html

Ucla geo-portal world water bodies. (2021, Apr). Retrieved from https://apps.gis.ucla.edu/geodata/dataset/world_water_bodies

Ullman, D. S., Ginis, I., Huang, W., Nowakowski, C., Chen, X., & Stempel, P. (2019). Assessing the multiple impacts of extreme hurricanes in southern new england, usa. *Geosciences*, *9*(6), 265.