Notice: the proposal content was published to the NASA Space Apps challenge website.

ML-Based Marine Spill Detector

We propose an oil spill detection web application that periodically reads NASA satellite data and generates a global geographic heatmap. The purpose of the heatmap is to indicate the occurrence chance of the spillage incidents using regular deep learning methods. Our method mainly consists of combining open source datasets for the NASA Space Apps to help natural disasters decision makers plan and allocate resources efficiently. Our vision is to have the contamination accumulation in the food chain de-accelerated, and the mission of the public user interface (SaudiSpaceShuttle.com) is to provide the first responders community with an additional deep model as well as to increase the public awareness about water pollution.

When it comes to water pollution, what goes around really comes around, since water pollution can put people at risk of diseases. One of the most pressing sources of water pollution is the oil spill problems. Even though the challenge of monitoring sea pollution has been addressed by many organizations, the stakes are too high, and additional resources could make a difference to solve the accumulating contamination in the food chain. There exist specialized monitoring organizations that use machine learning to reduce the impact of the oil spill. For instance, the National Oceanic and Atmospheric Administration, or NOAA for short, trains machine learning models to predict the trajectory of the oil spill spread. However, some spillages become irreversible within a couple of days after an oil tanker accident, and the negative effects from some types of oil could last for decades.

Therefore, the problem statement is about providing an additional oil spill model to double check the safety of the marine resources. However, our machine learning model uses existing open data from governmental organizations. The approach we adopt to addressing the water pollution is providing heat maps that can be easily read by emergency responders and relevant decision makers. For the end user, the workflow is as friendly as entering the area of interest to the Saudi Space Shuttle web app and receiving the requested information or having it posted to the public. Nevertheless, the underlying modeling is moderately sophisticated. The everyday data acquisition, cleaning, clustering, and mapping are part of the processing method.

In the ML-Based Marine Spill Detector, several pieces of open data are utilized, and two datasets are joined for extracting the main features for the marine oil-spill, chemical, and other incidents. The regular satellite imagery of the Earth API was prone to cloud blocking. Accordingly, the Synthetic Aperture Radar (SAR), from European Space Agency (ESA), which has a long-lasting partnership with NASA, was practical since it is less variant to weather conditions, and so it has multiple applications, one of which is oil pollution detection. In 2023, a new SAR addition is expected to be added by NASA's joint NASA ISRO SAR Mission (NISAR) program. In any case, the ability of reading the satellite imagery and radar readings isn't adequate without the historical data of the marine pollution incidents. For that reason, the ESA dataset was joined with the Raw Incident Data, which is open and is accessible via the IncidentNews website of National Oceanic and Atmospheric Administration (NOAA). In 2016. NASA launched an advanced weather satellite for NOAA. Nevertheless, the ESA data, along with several other datasets, is accessible through Google's Earth Engine API. It needs filtering libraries, available in Python and JavaScript, to prepare the data for downloads, and so does the Earth API.

We joined the NOAA dataset to ESA data to produce the imagery data, which becomes open to the public when it is published to the website, for our deep learning model. The website is www.SaudiSpaceShuttle.com. Our model also trains on the pre-processed ROBORDER data from MultiMoDal Data Fusion and Analytics Group. The ROBORDER data is accessible upon requesting the research group. Please find the list of the utilized data below.

The utilized data:

1. NASA, Earth API, [URL] https://api.nasa.gov
2. ESA, Google Earth Engine API, [URL] https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S1_GRD
3. NOAA, Raw Incident Data, [URL] https://incidentnews.noaa.gov/raw/index
4. "MultiMoDal Data Fusion and Analytics Group," ROBORDER Dataset, [1,2], [URL] https://m4d.iti.gr/oil-spill-detection-dataset

Both Python and JavaScript are open-source softwares, and so are the libraries, like Earth Engine, Geopandas/Pandas, Cupy/Numpy, and Tensorflow/Keras, and the operating system as well, which is Linux. We are using Google Cloud Platform to host and accelerate the processing, since some tasks require scaling up the computational capabilities. Our time is the main asset. The Saudi Space Shuttle (SSSh) app is a non-profit project, and its minimal viable service is displaying the oil detection chances of occurrences on a heatmap over a map object, by scanning specific areas of interest that are specified by the end-users through email or a form on the website.

In summary, the SSSh app gathers different types of data that can be joined geographically per the timestamp of the events, and its model is a binary classifier that trains to detect whether an incident is present or not in each of the satellite radar images using deep learning. The main tools were outlined. A Q/A table is included.

AhmadZuhair
The Saudi Space Shuttle Team

Table Q/A

	Question	Answer
1	What is the framing question of your analysis, or the purpose of the model/system you plan to build?	The problem statement is about adding a new model to the existing research body. The purpose of the model is to alert emergency responders about environmental incidents as the relevant satellite readings become available.
2	Who benefits from exploring this question or building this model/system?	It can lead to prosperity for the well-being of human beings since it addresses the human food chain and marine pollution.
3	What dataset(s) do you plan to use, and how will you obtain the data?	The data is SAR Sentinel-1, and it is obtainable via the Google Earth Engine APIs.
4	What is an individual sample/unit of analysis in this project? What characteristics/features do you expect to work with?	They are Satellite images. The samples have three band features: VV, HV, and angle.
5	If modeling, what will you predict as your target?	The output is either Oil-Spill or No-Oil-Spill.
6	How do you intend to meet the tools requirement of the project?	They're mostly free.
7	Are you planning in advance to need or use additional tools beyond those required?	Yes, we might scale up the tools should it be needed.
8	What would a minimum viable product (MVP) look like for this project?	It is a website with a JavaScript heatmap that reads the data from the server, and the data is updated throughout the day.

References: [1]. Krestenitis, M., Orfanidis, G., Ioannidis, K., Avgerinakis, K., Vrochidis, S., & Kompatsiaris, I. (2019). Oil Spill Identification from Satellite Images Using Deep Neural Networks. Remote Sensing, 11(15), 1762. [2]. Krestenitis, M., Orfanidis, G., Ioannidis, K., Avgerinakis, K., Vrochidis, S., & Kompatsiaris, I. (2019, January). Early Identification of Oil Spills in Satellite Images Using Deep CNNs. In International Conference on Multimedia Modeling (pp. 424-435). Springer, Cham.