Disaster Damage Categorization Applying Satellite Images and Machine Learning Algorithm

Farinaz Sabz Ali Pour
Old Dominion University

Adrian Gheorghe
Old Dominion University

Original Publication Citation
Pour, F. S. A., & Gheorghe, A. (2020) Disaster damage categorization applying satellite images and machine learning algorithm. *IIE Annual Conference. Proceedings,* (2020), 1-6.

Proceedings of the 2020 IISE Annual Conference
L. Cromarty, R. Shirwaiker, P. Wang, eds.

Disaster Damage Categorization Applying Satellite Images and Machine Learning Algorithm

Farinaz Sabz Ali Pour & Dr. Adrian Gheorghe
Engineering Management & Systems Engineering Department
Old Dominion University
Norfolk, VA, USA

Abstract

Special information has a significant role in disaster management. Land cover mapping can detect short- and long-term changes and monitor the vulnerable habitats. It is an effective evaluation to be included in the disaster management system to protect the conservation areas. The critical visual and statistical information presented to the decision-makers can help in mitigation or adaption before crossing a threshold. This paper aims to contribute in the academic and the practice aspects by offering a potential solution to enhance the disaster data source effectiveness. The key research question that the authors try to answer in this paper is how to apply remote sensor data in decision-making process for disaster preparedness and response. To achieve this goal, the satellite imagery is used as a data source for the decision-makers. Since the real time satellite imagery are a kind of meta data, big data helps in reading the information. Machine learning algorithms is applied to classify and analyze the data.

Keywords

Decision-making, Disaster Management, Big Data, Machine Learning, Satellite Imagery

1. Introduction

The frequency and impact of disasters have significantly increased in the past decades. One of the most critical challenges of disaster management is the scarcity of accurate data [1]. To empower societies against the high frequency and intensity of disasters, timely collecting, and communicating the real-time disaster information to all stakeholders is essential. This power derives from knowledge, which is the output of information generated from the data [1]. The lack of direct communication from the affected area leads to incomplete or incorrect information. The disaster-relief system should maximize efficiency and utilize all the available resources to minimize the impacts. Various level of disaster vulnerability limits the capabilities of decision-makers to respond in disaster situations [1]. Due to the chaotic and disorganized situation of disasters, the generation of big data can be a perfect fit for augmenting decisions [2]. Having more valuable data sources can reduce uncertainty in decision-making. Advance systems have enabled multi-directional communication as a new method of interfacing [3].

2. Literature Review

Disaster management is the process of developing and implementing policies that are concerned with all four phases of mitigation, preparedness, response, and recovery for a disaster [4]. The sentiment of the affected population during and after a catastrophe specifies the success of the disaster response and recovery process [3]. Recently, multiple studies have focused on communication technology in different phases of disaster management. The importance of accurate, relevant, and real-time geo-information, which can be captured through the internet, remote sensing, etc. was mentioned by Kamal [5]. However, the knowledge generation and dissemination through these sources are still limited to issues such as mobility and coverage [1].

In the last decade, massive numbers of data have been produced by satellite with several launched sensors [6]. Sakurai and Murayama [7] illustrate how information technologies can be applied in each stage of disaster management and the role of the information system. A variety of classification methods have been developed for land mapping utilizing remote sensing data in the literature. Thanh Noi and Kappas [8] conclude that Support Vector Machine (SVM), Random Forest (RF), and Key Nearest Neighbor (KNN) are the best classification methods for Landsat imagery classification.

This study proposes a framework on using real-time data that is driven through satellite imagery and remote sensing as one of the critical inputs to disaster management processes. The metadata can be analyzed via machine learning algorithm to support the disaster management team in making the proper decision at the tense short time of a catastrophe.

2.1. Disaster Categorization

Making the right decision is very important and challenging in a disaster response phase. Lack of direct communication from the affected areas can lead to an improper understanding of the disaster condition. Disaster situations are very chaotic and disorganized. Hence, applying computational intelligence approaches can offer effective handling of the environment by providing the right information, analyzing them, and taking the best course of action.

2.2. Satellite Imagery Technology

In all phases of a disaster, there are two types of input data one generated by users such as Twitter, Facebook, and the one sensor-generated data such as satellite images, drones [3]. A satellite-based communication system can be used as advanced forecasting, monitoring, and issuing early warnings system for a natural hazard and determine whether it has the potential of becoming a disaster [5]. The high-resolution data that are readily available to the public is a valuable source in monitoring the earth and the environment. Based on Navalgund, Jayaraman [9], "Remote sensing refers to the science of identification of earth surface features and estimation of their geo-bio-physical properties using electromagnetic radiation as a medium of interaction" [9]. Remote sensing is a tool for the identification of hazardous areas, monitoring the changes on earth on a real-time basis, and providing early warning to many potential disasters. Incorporating space technology as an input to natural disaster mitigation and monitoring is essential to reduce hazards [5]. Remote sensing tools can support all aspects of disaster management.

Along with GIS, they are essential and efficient in preparedness, communication, and training [5]. Satellite imagery technology mostly follow the mission of offering high spatial, temporal, and spectral resolution imagery. The critical mission is to support decision-makers for monitoring environment security and offer opportunities for the scientific community [6]. Satellite images are pixel wised data. Raster images consist of satellite images, Lidar data, and georeferenced maps. Raster includes a matrix that holds information about the location, such as elevation, temperature, and vegetation.

2.3. Machine Learning

Machine learning algorithms have been used in multiple research to develop a new object-based image analysis approach for mapping the critical environment using the satellite data. In remote sensing, a per-pixel classifier is applied as the classification algorithms. The techniques can be based on either identified computationally with minimum user input (unsupervised) or through used-defined training pixels (supervised) [6]. The durability and capability in the classification performance of machine learning algorithms made them an integral part of remote sensing studies. Several studies have applied machine learning algorithms in this regard. Whyte, Ferentinos [6] employed SVM and RF algorithms to evaluate the wetlands for the land cover classification. Pirotti, Sunar [10] made a comparison among nine machine learning algorithms to check the accuracy and speed in land-cover classification in a satellite dataset, and the results illustrate that RF method appeared to have the highest values.

In RF implementation, the number of trees and the number of features in each split are required [8]. The fascinating factors about RF for digital image analysis include unexcelled in accuracy among current algorithms. It can be run efficiently for large data sets and can handle a vast amount of input variables. The generated forests can be stored for future applications [10].

2.4. Big Data

Big data that has made a noticeable position in different industries were created from Twitter [3]. Disasters are big, messy, and devastating. However, big data analytics offers solutions to handle the input data effectively in all phases of a disaster [3]. Social networks are increasingly becoming helpful in emergency communications. Big data has recently been used to mine valuable insights from social media and mobile networks. The emergency request can be mined from the pool of big data to provide the timely support for the affected area. A big data driven approach for disaster response through sentiment analysis was proposed by Ragini, Anand [3]. The first incident that big data was used to help the affected area was the Haiti earthquake in 2010. This was when the digital humanitarian was first introduced as the process of using techniques to produce crisis maps [3]. Big data analytics is a critical resource for satellite image processing that can examine the data implementation challenges. The advent of new computing architectures and advances in Geographic Information System (GIS) software provides leveraging massive opportunities from volumes of imagery [11]. Real-time data from different devices, sensors, and social media can be displayed and analyzed using the Internet of Things (IoT). Big data offers an understanding of the hidden patterns and trends in massive data to help in making the right decision at the right time.

3. Methodology

This study proposes a framework driven by big data for disasters assessment and monitoring. This framework aims to improve the understanding of the situation in making more precise real-time decisions. In this study, data from satellite imagery and remote sensing is collected and categorized according to the damages and needs of the affected areas. A machine learning algorithm is used to classify the categorize data as part of the decision tree analysis in real-time scenarios. The general image processing includes collecting the image data, transform the data, classify the image, map them using the GIS database, and apply spatial analysis. All of these steps would lead to a better decision making approach [9].

3.1 Data collection

One of the great sources to collect data in case of a disaster is satellite images and remote sensing. Remote sensing data can provide a) a synoptic view, repetitive coverage with calibrated sensors which can detect changes b) different resolutions observations, and c) better alternative for natural resources management [9]. The near-real-time data can assist in monitoring the hazardous events, where current land-use layers can be analyzed through GIS to identify hazard zonation maps, the affected settlements, apply damage assessment, and relief planning.

3.2 Data categorization and classification

The collected data will be categorized using machine learning algorithms. Based on the objectives of the study, data is divided into training sample data and test sample data to determine the characteristics and noises in the data. Kussul, Lavreniuk [12] proposed four levels to classify satellite imagery as preprocessing, supervise classification, postprocessing, and geospatial analysis.

A machine learning algorithm is required to map the extracted features for decision-making. There are various algorithms. Based on Whyte, Ferentinos [6], RF and SVMs are the most popular machine learning algorithms used in remote sensing studies. For the proposed framework, first, the preprocessing phase is applied to deal with the missing values in the imagery and attribute them. Then to explore the spectral and spatial features of the imagery, RF classifier is used.

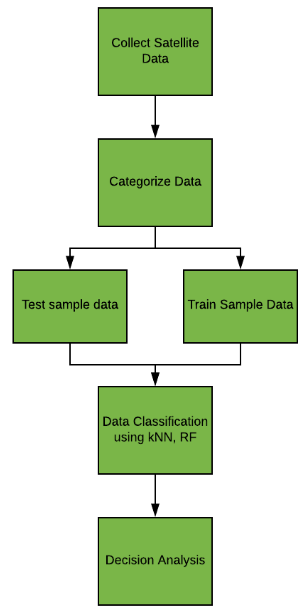

Figure 1: Methodology

3.3 Experimental results and analysis

The results of the machine learning algorithm are evaluated by calculating precision, recall, and F-measure that are the primary metrics of classification system performance measurement [3]. The method is implemented and tested using Apache Spark Big Data framework and Python programming language. The decision tree-based classification has two levels of training and classification. The training level builds a training model to predict the input as positive or negative on the classification level. The class under investigation is considered positive, and the remaining classes are negative [3]. Figure 1 depicts the methodology steps for the proposed framework. The results can be applied as a valuable input in disaster management. The results provide a visualization, estimation of the event expansion, and the potential consequences. This information can help the decision-makers to have a better understanding of the disaster with more accurate real-time data to make the optimum decision.

4. Case Study

The combination of fuel, weather, and ignition can lead to fires that can be disastrous by affecting human resources [13]. Fire is a crucial factor in the ecology of the natural environment and is a potential hazard to human life and property [14]. There can be significant physical, economic, and social consequences from a natural disaster. Lack of having a proactive management approach can lead to ecosystem health and biodiversity erosion [15]. In case of a fire, satellite data can be used to map and monitor the fire-affected areas, damage assessment, hazard zoning, and post-fire survey of configuration and protection. Integrating information technology into the management approach can help in the determination of potential fire risk areas. Since the last decade, the use of satellite remote sensing offers an investigation of the complexity of the fires. Landsat and MODIS are the most commonly used sensors [16]. Landsat satellite images are freely available from the USGS website. Landsat series provides temporal recode of multispectral data of the earth's surface. It can help in natural resources management. The metadata indicates the data process, location, and structure. Remote sensing data that is collected from satellite permits gathering information about the landscape over time. Landsat series offers the longest temporal record of moderate resolution multispectral data of the earth's surface. The data is collected using a platform installed on a satellite in space that orbits the earth. The United State Geological Survey (USGS) offers an archive of the Landsat record for almost 50 years. All the USGS Landsat data are available free via the internet [17]. The level 2 data includes atmospherically, topographically corrected, georecitified surface reflectance. Level 3 data contains operational metrics such as the normalized burn ratio (NBR) and the normalized difference vegetation index (NDVI) [16].

A dataset is retrieved from USGS that contains Landsat 8 Operational land imager and Thermal Infrared Sensor from April 2013 till present to illustrate the application of satellite remote sensing on determining the burnt area in united states in this timeline. To show how this type of data can be helpful, the cold spring wildfire occurred on July 2016 in Colorado is selected as a case for this study. The fire spread rapidly, and 528 acres were burnt. The cause of the fire was attributed to a campfire that was not extinguished correctly. It was estimated that almost 2000 people were affected by the fire [18].

dateUpdated	sensor	PANCHROMATIC_SAMPLES	GROUND_CONTROL_P	LANDSAT_PRODUCT_ID	acquisitionDate	upperRightCornerLongitude	PROCESSING_SOFTWARE_VER	GRID_CELL_SIZE_REFLECTIVE	lowerRightCornerLongitude	lowerRightCornerLatitude	sceneCenterLongitude
1/19/2020	OLI_TIRS	15181	4	LC08_L1TP_095054_20200120_20200120_01_RT	1/20/2020	153.0924	LPGS_13.1.	30	152.714	7.62333	152.0587
1/19/2020	OLI_TIRS	15181	4	LC08_L1TP_095055_20200120_20200120_01_RT	1/20/2020	152.7776	LPGS_13.1.	30	152.4016	6.17787	151.7479
1/19/2020	OLI_TIRS	15201	4	LC08_L1TP_095061_20200120_20200120_01_RT	1/20/2020	150.9218	LPGS_13.1.	30	150.5537	-2.49835	149.901
1/19/2020	OLI_TIRS	15221	4	LC08_L1TP_095062_20200120_20200120_01_RT	1/20/2020	150.6156	LPGS_13.1.	30	150.2477	-3.94467	149.5937
1/19/2020	OLI_TIRS	15221	4	LC08_L1GT_095063_20200120_20200120_01_RT	1/20/2020	150.3095	LPGS_13.1.	30	149.9416	-5.39055	149.2861
1/19/2020	OLI_TIRS	15221	4	LC08_L1TP_095064_20200120_20200120_01_RT	1/20/2020	150.0034	LPGS_13.1.	30	149.6351	-6.83675	148.9775
1/19/2020	OLI_TIRS	15241	4	LC08_L1TP_095065_20200120_20200120_01_RT	1/20/2020	149.6969	LPGS_13.1.	30	149.3278	-8.2827	148.6679
1/19/2020	OLI_TIRS	15241	4	LC08_L1TP_095066_20200120_20200120_01_RT	1/20/2020	149.3898	LPGS_13.1.	30	149.0197	-9.72841	148.357
1/19/2020	OLI_TIRS	15241	4	LC08_L1TP_095067_20200120_20200120_01_RT	1/20/2020	149.0813	LPGS_13.1.	30	148.7099	-11.17367	148.0445

Figure 2: A sample view of the Landsat data sheet (Apr/2013 – Jan/2020)

Jupyter notebook, a python library that offers searching, retrieving, and downloading is used to access the imagery data for the case study. This data includes burning severity, changes before versus after the burn in vegetation, and differences between the burn site and areas around it that did not burn. This data can also be used to calculate quantitative metrics and monitor ecosystem recovery. Applying machine learning algorithm provides valuable information from the collected data. RF algorithms have high accuracy that finds the robust outliers and noise, depicts the relative importance of the input variables, and is rapid. The bootstrap aggregating algorithm is applied to generate sub-datasets from the training dataset. The datasets construct decision trees and are grouped to form the random forest [19]. The set of observations is located at each tree of the forest. Each tree vote for the class, and the classification that got the most votes would be selected by the forest.

The classified data can provide a plan for the area fire management, which can be represented as an asset for human settlement and economic value. For each asset, the risks have been assessed and have been assigned appropriate treatment strategies. The human settlement line predominately represents areas where people may be located, such as roads, walking trails, or a series of campsites. Economic lines predominately represent utility or agency infrastructure, such as pipeline or the dog fence. This data can be used as an assistant to priorities fire prevention and preparedness activities.

Table 1: Landsat Instrument Band settings

Band #	B1 – coastal aerosol	B2 - Blue	B3 – Green	B4 - Red	B5 – near-infrared (NIR)	Band 6 – SWIR 1	Band 7 – SWIR 2	Band 8 – Panchromatic	Band 9 - Cirrus
Wavelength range (nm)	430-450	450-510	530-590	640-670	850-880	1570-1650	2110-2290	500-680	1360-1380
Spatial Resolution (m)	30	30	30	30	30	30	30	15	30
Spectral Width (nm)	2	6	6	0.03	3	8	18	18	2

Interpretation of data derived from the satellite imagery assist in determining patterns of the fire geometry, the burn history, and weather condition. The results provide a realistic basis for fire forecasting procedures development and suitable control measures specification [14]. Each band in a Landsat scene is stored in a ".tif" file. So, first, the relevant bands are needed to be obtained and converted to NumPy array in Python. For this study, a secondary dataset from the cold spring wildfire is retrieved from Wasser, Palomino [20], to illustrate how beneficial can remote sensing data be for disaster management. The bands in the Landsat 8 sensor are described in table 1.

The Normalized Difference Vegetation Index (NDVI) is a vegetation index that is used to measure the greenness of the area. For the purpose of studying wildfires like the Cold Springs Fire, it allows visualize and quantify vegetation health before and after the fire.

5. Results and Discussion

Making decisions about mitigating the detrimental impacts that fire can have on assets and values is very critical [15]. Having a structured risk management approach yields an appropriate framework to articulate actions to mitigate and manage the wildfires effectively [15]. One of the strategies for treating fire risk is readiness as part of consequence reduction, which can include early detection, having accurate information, and communication technologies.

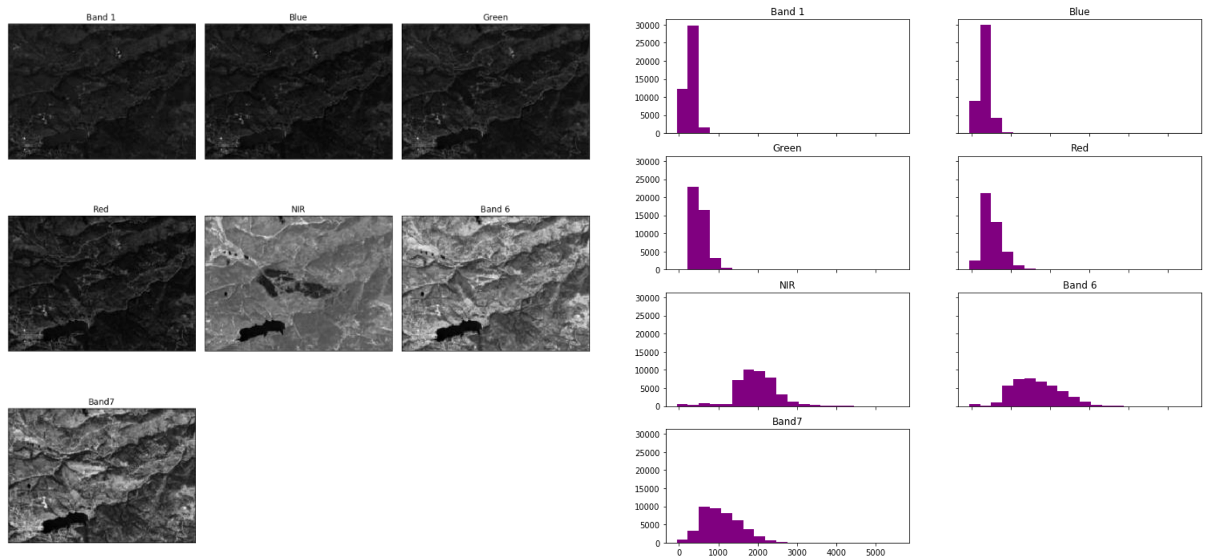

Figure 3: Cold Spring Fire Landsat 8 Bands and Histogram Retrieved from Wasser [20]

This study aims to indicate the continuing necessity of data availability for decision-making activities throughout the disaster response and recovery phases. This approach helps decision-makers in establishing fundamental strategies for information management in the dynamic disaster environment. This method enables effective damage detection and evaluation. A massive flow of information regarding the damage categories and scale are provided, which enhances the awareness and operation of the emergency team. The proposed framework offers better strategy development to the emergency responder for effective information management in the dynamic environment of disasters. This method tries to bridge the gap between the information scarcity from the affected area to indicate the required actions. This can help to identify the exact area that is in the most needs.

The proposed approach proves that satellite imagery can be used as a valuable data source for disaster needs classification and effective response and recovery, as well as an input for the preparedness phase for potential disasters. Raster functions implementation yields dynamic and well-suited distributed processing to envisage large datasets. A pixel histogram is created by the raster to show the distribution of values in the bands and is depicted in figure 2.

6. Conclusion

Disaster management relies on accurate, relevant, and on-time geoinformation. The advanced information and communication technology, such as remote sensing, can be useful in planning and implementing hazards reduction programs.

The lack of standard crisis data set and difficulty in collecting the disaster-related data makes the accurate evaluation of the victims very challenging. Wildfire can get beyond the capacity to contain it and pose a high risk to people, buildings, infrastructure, and the environment. The risk level is dependent on several dynamic variables, including the severity of the fire, the affected assets, number of days, etc. Thus, decision-makers should analyze all the risks and potential consequences as part of disaster management. The high uncertainty and nature of a disaster make more complexity in the process. Decision-makers need to have real-time data and the trend of the disaster expansion to have a better understanding of the disaster situation. The integration of satellite data and machine learning algorithms provides more accuracy in the classification of disaster data. This approach can lead to constructing an ontology for

disaster data, which can help in needs categorization during disasters and provide a disaster damage database for forecasting and making comparisons in the future possible disasters.
The major impact of the model is the real-time segregation of the images according to the needs of the affected population. The system output provides the emergency responders better insights into the inclination of the event.
Remote sensing satellite provides extensive geographical coverage, valuable information on the earth at an efficient cost.
This study aims to provide a semi-automated workflow that can be scaled. However, there are some limitations in satellite applications for response operations, such as low coverage in cloudy conditions. For future works, this study will use other instruments datasets to compare the efficiency in disaster management processes. Moreover, a more in-depth analysis can be implemented to evaluate the influence of satellite imagery classified information on disaster management.

References

1. Sukhwani, V. and R. Shaw, *Operationalizing crowdsourcing through mobile applications for disaster management in India.* Progress in Disaster Science, 2020. **5**: p. 100052.
2. Horita, F.E., et al., *Bridging the gap between decision-making and emerging big data sources: An application of a model-based framework to disaster management in Brazil.* Decision Support Systems, 2017. **97**: p. 12-22.
3. Ragini, J.R., P.M.R. Anand, and V. Bhaskar, *Big data analytics for disaster response and recovery through sentiment analysis.* International Journal of Information Management, 2018. **42**: p. 13-24.
4. Petak, W.J., *Emergency management: A challenge for public administration.* Public Administration Review, 1985. **45**: p. 3-7.
5. Kamal, M.A., *Role of information and communication technology in natural disaster management in India.* ICT in Disaster Management, 2015. **182**: p. 188.
6. Whyte, A., K.P. Ferentinos, and G.P. Petropoulos, *A new synergistic approach for monitoring wetlands using Sentinels -1 and 2 data with object-based machine learning algorithms.* Environmental Modelling & Software, 2018. **104**: p. 40-54.
7. Sakurai, M. and Y. Murayama, *Information technologies and disaster management – Benefits and issues.* Progress in Disaster Science, 2019. **2**: p. 100012.
8. Thanh Noi, P. and M. Kappas, *Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using Sentinel-2 imagery.* Sensors, 2018. **18**(1): p. 18.
9. Navalgund, R.R., V. Jayaraman, and P. Roy, *Remote sensing applications: An overview.* Current Science (00113891), 2007. **93**(12).
10. Pirotti, F., F. Sunar, and M. Piragnolo, *BENCHMARK OF MACHINE LEARNING METHODS FOR CLASSIFICATION OF A SENTINEL-2 IMAGE.* International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences, 2016. **41**.
11. Brown, C. and C. Harder, *The ArcGIS Imagery Book: New View, New Vision*. 2016: Esri Press.
12. Kussul, N., et al., *Deep learning classification of land cover and crop types using remote sensing data.* IEEE Geoscience and Remote Sensing Letters, 2017. **14**(5): p. 778-782.
13. Cheney, N.P., *Bushfire disasters in Australia, 1945–1975.* Australian Forestry, 1976. **39**(4): p. 245-268.
14. Milne, A.K., *The use of remote sensing in mapping and monitoring vegetational change associated with bushfire events in Eastern Australia.* Geocarto International, 1986. **1**(1): p. 25-32.
15. Ellis, S., P. Kanowski, and R. Whelan, *National inquiry on bushfire mitigation and management.* 2004.
16. Stavros, E.N., et al., *Unprecedented remote sensing data over King and Rim megafires in the Sierra Nevada Mountains of California.* Ecology, 2016. **97**(11): p. 3244-3244.
17. Loveland, T.R. and J.L. Dwyer, *Landsat: Building a strong future.* Remote Sensing of Environment, 2012. **122**: p. 22-29.
18. Paul, J., *With 5 homes destroyed, crews report good day battling Cold Springs fire*, in *The Denver Post*. 2016.
19. Vafaei, S., et al., *Improving accuracy estimation of forest aboveground biomass based on incorporation of ALOS-2 PALSAR-2 and sentinel-2A imagery and machine learning: A case study of the Hyrcanian forest area (Iran).* Remote Sensing, 2018. **10**(2): p. 172.
20. Wasser, L., J. Palomino, and C. Holdgraf. *earthlab/earth-analytics-python-course: earthlab/earth-analytics-python-course: Version-1.0.1*. 2019; Available from: https://zenodo.org/record/3523193#.XiYCCvlKiUk.