**Evaluating the impact of climate change on energy consumption with machine learning**

My proposal aims at evaluating the effects of climate -- and climate change -- on energy consumption profiles. Thus, the proposed methodology is aimed at investigating the correlation between climate data and energy consumption profiles referred to the same locations and time instants. The correlation matrix will be calculated with past data of hourly energy consumption profiles and atmospheric states. The highest correlations between the energy consumption profiles and climate variables indicate the climate variables which make the highest impacts on the energy demand. The analysis can be used to identify relevant trends relating climate variables with energy consumption and can be extended to different locations and time instants. In this context, I would like to receive some feedback on the choice of an adequate methodology and possible additional considerations in the proposed application.

**Machine learning applications for radar mosaics image analysis to understand climatic traits of convection initiation**

Hide abstract

The occurrence and strength of hazardous weather have been increasing in the changing climate. Convective systems are the major cause of hazardous weather, and these systems essentially affect the subsequent occurrence of organized convection and heavy rainfall. This project aims to investigate the climatological characteristics of convection initiation (CI) and associated interactions between CI and the large-scale circulations with the help of long-term observational data. It potentially indicates the regional climate change and provides feedback to the global climate . One of the major applications of machine learning algorithms is to process large amount of radar mosaic images to understand the initiation location and development of deep convection. It has been shown that the decision-tree-based methods and associated algorithms would be helpful to process raw images from radar for analysis. I look forward to feedback and suggestions regarding image processing and computer vision.

|  |  |
| --- | --- |
|  |  |
|  | **Adapting damage localization frameworks in the face of climate change adversity through probabilistic learning methods**  Hide abstract  Damage localization in structures such as bridges, buildings, airplanes is a popular research area. Common strategy is to use acoustic waves for localization. Correlation of a wave propagation model with experimental data is obtained to localize damage. Yet, in the presence of unusual environmental variations such as temperature swings, excess precipitation such methods provide inaccurate results (wave velocity has a temperature dependence). In the face of climate change adversity, we are experiencing such unusual climate patterns (temperature swings / anomalous rain seasons etc) at an alarming rate. These unusual patterns further exacerbate the complexity of damage localization. Indeed, it is the need of the time to investigate damage localization methods robust to environmental variations (hastened by climate change). Two points need to be considered in the context of robustness of localization algorithms. Firstly, the experimental data is different from the physical wave model simulation because of the above mentioned environmental variations. Deep learning methods can potentially leverage data with environmental uncertainties to learn robust mappings from wave data to damage location. Secondly, typically localization techniques give a point estimate. We contend that a point estimate is ineffective when dealing with a highly uncertain environment. We propose to represent the damage location estimate as a probability distribution. This can effectively represent the environmental uncertainty. We propose a deep neural network based framework for this problem. Currently, the project to me, seems like a mitigation strategy for a second-order effect of climate change (primary submission area chosen as buildings / cities). I would appreciate mentor-ship and feedback on how to frame the problem of damage localization in uncertain environment to be more relevant to climate change. If and how I can give more direct evidence of the criticality of this issue ? |

**A Sociotechnical Approach to Environmental Justice**

Hide abstract

Environmental hazards are not evenly distributed, in part due to zoning and planning of hazardous treatment, storage, and disposal facilities (TSDs) and toxic chemical releases into the air which are accounted for by the Environmental Protection Agency’s (EPA) Toxic Release Inventories (TRIs). Due to environmental racism, low-income people of color carry more environmental burden. Environmental racism and environmental sexism dictate that women of color face a double burden. Environmental justice (EJ) asserts that environmental hazards and benefits should be fairly distributed. EJ has been studied using spatial analysis and regression. Studies on spatial disparities in TRIs and TSDs based on race/ethnicity and socioeconomic status (SES) in U.S. cities, such as Charleston, SC, San Joaquin Valley, CA, and West Oakland, CA showed that there are more TRIs and TSDs in non-White and low SES areas in those cities. In Atlanta, 52.3% of the population is Black as of 2018. In 2016, Atlanta exhibited the highest rate of income inequality among major cities in the U.S. In August 2019, a smog alert was issued for Atlanta by the EPA due to air quality conditions being unhealthy for sensitive groups, such as people with lung disease or asthma. Consequently, Atlanta presents an opportunity for novel high-impact research. However, there is a lack of studies on EJ in Atlanta (and in general) that take advantage of machine learning analytics. Even more so, none of the previous studies account for gender. I would love to receive feedback on the machine learning methods I would like to use and whether or not they are sufficient. I am also interested in the validation of ideas and thinking through the data science process of this project.

**Investigating the Ground-level Ozone Formation using Machine Learning**

Hide abstract

Understanding the tropospheric ozone formation mechanism is essential since it has a significant influence on the ecological community. Furthermore, recent studies indicate that tropospheric ozone might increase due to climate change. However, traditional theoretical studies of ozone formation mechanisms are insufficient to accurately simulate the ozone in climate models. Therefore, this proposal aims to apply different machine learning algorithms, such as logistic regression, gradient boosting regression, and long short-term memory network, on investigating ozone formation mechanisms. Compared with traditional theoretical studies, up-to-date machine learning algorithms are beneficial to identify the unveiled correlation between ozone and other variables in countless data. It will be supportive to receive feedback about how to improve the prediction with different methodologies. Moreover, we are glad to have advice for utilizing the latest algorithms to boost the prediction performance.

**Federated Learning for Climate Change**

Hide abstract

As recent privacy-regulation laws like HIPAA and GDPR have become more stringent, there is a need to democratize Machine Learning(ML) models to promote wider usage. Federated Learning is a technique where the model is sent to the clients in the protocol and processed locally and the corresponding model updates are sent to the global model at the server, hence preserving the local sensitive data of the user. In this paper, we discuss the different applications of Federated Learning with respect to Climate Change and how it can help us understand Climate Change by impacting data-gathering and data-processing at a massive scale in this field, without compromising on the privacy-sensitive information present in the devices of the users.

**A Machine Learning Approach to Address Methane Emissions from Oil and Gas Operations**

Hide abstract

Methane emissions mitigation in the oil and gas sector is an important component of climate policy across North America. One of the common features of methane regulations is the implementation of periodic leak detection and repair (LDAR) surveys. These surveys are typically conducted by a two-person crew visiting each site with a handheld sensor to detect methane leaks. However, this is time-consuming and expensive, particularly for companies that manage thousands of production sites. Furthermore, recent research has shown that emissions follow a skewed distribution where a small number of sites (‘super emitters) are responsible for a majority of total methane emissions. Thus, finding these super-emitters in a timely manner will optimize both time and financial investments in emissions reduction programs. Here, we propose to develop a machine learning model to predict high-emitting sites that can be prioritized for LDAR surveys. Such prioritization can reduce the cost of surveys and increase emissions reductions compared to a random survey approach currently used in the industry. Prior efforts have failed because of limited data availability. By taking advantage of recent measurements across North America, we will assemble the largest complete data-set on methane emissions to predict the occurrence of super-emitters. Through this mentorship program, we look forward to discussions on the role of machine learning in addressing emissions reduction policies. Specifically, we are looking for 1) feedback on research design, including defining potential attributes for predictive models and 2) suggestions on improving model performance.

|  |  |
| --- | --- |
|  |  |

**Short-term PV output prediction using convolutional neural network: learning from an imbalanced sky images dataset via sampling approaches and data augmentation**

Hide abstract

As the largest contributor to global GHG emissions, the current energy supply sector, dominated by high-GHG fossil fuel conversion technologies, requires a fundamental reformation. Integrating renewable resources, such as wind and solar power, into the current electricity grids is one of the major pathways towards a clean and sustainable energy system. Photovoltaics (PV), the primary way of utilizing solar energy, has been growing rapidly in the last decade and is expected to continue growing in the near future. However, the biggest challenge for large-scale deployment of PV is the strong fluctuation in its power generation, which may seriously affect the stability of the power grids. A reliable and accurate PV output prediction is thus critical to help reduce the uncertainty and mitigate the stability risks of the grids. In this study, we empirically investigate using different sampling approaches and different types of image data augmentation techniques to handle an imbalanced sky images dataset. The baseline architecture is a specialized CNN developed by our group in previous work and the dataset we used is an in-house sky images dataset with about 1 year of data. The following feedback from the mentor would be beneficial: (1) feedback on machine learning methodology and comments on the experiments we have done; (2) insights on improving the results interpretability from a machine learning perspective; (3) polish the results presentations by providing insights on tabulation and visualization.

**Developing LSTM networks for drought impact assessment based on text-based data**

Hide abstract

This study employs the text-based drought reports in the Drought Impact Reporter (DIR) developed and maintained by the National Drought Mitigation Center (NDMC) as a surrogate data set of drought impacts. Using California as a case study, this work builds long short-term memory (LSTM) networks to estimate and predict drought impacts with various hydrometeorological and remote sensing predictors. The goal of this study is to quantitatively identify the relationship between the physical environment and drought socio-economic impacts by applying a well-trained LSTM network. I am looking for some professional suggestions and comments on how to train LSTM networks for the time-series analysis, and if possible, some recommendations of other machine learning or deep learning algorithms that might work better.

**Few-shot semantic segmentation for deforestation change detection**

Hide abstract

Deforestation is one of the primary concerns hampering control of greenhouse gases and this makes it a significant focus area for climate change studies. Besides, deforestation does also result in loss of biodiversity and this adversely affects the surrounding ecosystem. Remote sensing technologies provide a cost-effective approach to study the deforestation, especially in inaccessible regions. The existing methods for deforestation identification, such as the DETER and Terra Class project, involves manual operations. Therefore, few studies focus on an automatic end-to-end system to identify the potential regions of deforestation using standard deep learning architectures (Early Fusion CNN and Siamese CNN). However, these approaches require the availability of manually annotated images during training, which is a time consuming and error-prone task. Few-shot learning is a popular approach to overcome the lack of ground truth labels. Dense pixel labeling on new classes with few support examples is referred to as few-shot semantic segmentation. Despite the recent advances in few-shot learning, there is limited application of this approach for change detection of deforestation (Marc et al, CVPRW 2020). The proposed work evaluates the performance of two state-of-the-art few shot semantic segmentation algorithms (PFENet: Tian et al, 2020 PAMI, and PANet: Wang et al, ICCV 2019) for identification of forest regions. The two methods are evaluated on DeepGlobe land cover segmentation dataset. This performance evaluation would help identify challenges in the segmentation of forest using few shot segmentation algorithms, paving the way for the development of a more robust approach. In future, this work will be extended for identifying the regions of deforestation by comparing the segmentation map to two images acquired at a different time interval. An ideal mentor research interest would be few-shot learning, and I expect an in-depth discussion on change detection.

**Reliable demand response in buildings and neighborhoods with reinforcement learning**

Reinforcement learning have been successfully tested in-game environments where the sample cost is far lower than in demand response scenarios. The goal of this paper/proposal is developing an demand response solution for both, for individual consumers in a single building(HVAC, domestic appliances, EV chargers and water boilers) as well as for cooperation between a number of buildings in a single neighbourhood/microgrid. This would be implemented by developing and testing both, indiviudual and cooperative reinforcement learning agents. The goal is maximizing the use of available renewable electricity, while reducing the use of non-renewable electricity. The subgoals for achieving this can be outlined as reducing the overall consumption, reducing the peak demand, the averaged daily peak demand and reducing the ramping. I would very much appreciate mentor feedback on the topics: - application of reinforcement learning and other control theory methods in demand response - achieving safety and sample efficiency in those applications - what is state-of-the-art in the aforementioned topics - feedback on the proposed solution and possible improvements

**Oil spill detection, recovery and aquatic behaviour analysis using ML**

Hide abstract

In this project , i want to address the common problem of oil spill in oceans.The main problem with this once a oil spill if faster action and precautions aren't taken it will lead to catastrophic effects in terms of both human and aquatic life.since man is dependent on the aquatic life , given an oil spill , how to detect a oil spill , get the region of interest and possible hotspots for it to seeping off the ocean.Once such a map is present it will be easy to recover the oil and also it gives a way to analyze the aquatic life being affected in that region.By doing that we can keep track of what life got affected , what part of it enters in to food chain and how we can reduce the damage through ML.I would like my mentor to guide me on the unsupervised and semi supervised learning aspects of this problem with an additional help in how to write a proper paper.

res. California, particularly, has seen some of the worst wild res, that have caused loss of human life and large scale infrastructure destruction, such as residential and commercial buildings, roads and bridges. The top ten most destructive wild res in California have resulted in almost 40,000 structures destroyed and 169 lives lost (CADFP 2019). Wildres also contribute to emissions, and when compared to the overall emissions inventory in California for 2017, their impact is almost as large as the emissions due to the entire electricity sector of the state (CARB 2019). The WUI is where the urban dwellings and wildland vegetation meet and where wildre problems are most pronounced. WUI has grown in the US, both in terms of new houses (30.8 million to 43.4 million, 41% growth) and land area (581,000 to 770,000 km2)(Kramer et. al 2018). In recent years, an increase in houses in the WUI has resulted in increased risk to lives and homes (Randelf et. al 2018). While WUI is critical, research suggests that the characteristics of a small portion of the WUI, the home ignition zone (HIZ), largely deter mine potential for ignition from wildres. This paper presents a deep learning based data fusion approach to characterise building risk to wildfires in the WUI. The framework can be applied to remote sensing data for other extreme events thus providing a tangible framework for risk assessment due to extreme events.

**Realistic satellite imagery generation for climate migration impacts**

Hide abstract

Faced with climate change related environmental degradation, humans have two choices: stay and adapt or move. However, where these climate migrants will go, and what changes they will cause to existing settlements, are difficult questions. Here, we aim to integrate predictive models combining behavioral science and climate change with photorealistic image generation. We use predictions from the world bank internal migration model (2018) as input into a conditional generative adversarial network trained on urban satellite imagery from the Sentinel 2 dataset, in order to visualize the extent of climate change in terms of realistic settlement population changes. The flexibility of GAN gives us the possibility to control the extent and severity of climate migration and its land-use effects. Finally, we explore and discuss how such generated imagery can be further used in downstream remote sensing analytics. At this point, we require mentoring with how exactly to extract predictions from the world bank model.

|  |  |
| --- | --- |
|  | **Earth Science: Lightning and Severe Climate**  Hide abstract  Utillize GOES-R geostationary lightning mapper (GLM) data along with colocated advanced baseline imager (ABI) data to characterize and predict severe weather/climate events in the Contiguous United States. Lightning behavior before and after severe events will be analyzed to extract relevant features and characteristics which may inform the temporal evolution of events with cloud features. Temporal characteristics will be analyzed using recurrent neural network architectures along with traditional computer vision techniques. In terms of mentorship, the expansion of the project to look into the characterization of dry lightning events and the spread of wild/bushfires is of interest and may provide another subtopic to explore and develop. The proposal could use some tailoring in terms of specific directions or individual severe events. The goal would be to augment current state of art with machine learning techniques from computer vision to present and propose novel work with respect to a specific domain need. |

**Expert-in-the-loop Systems Towards Safety-critical Machine Learning Technology in Wildfire Intelligence**

Hide abstract

Wildfires are recurrent natural disasters, which have a brutal impact on the environment and natural ecosystems. These phenomena have drastic effects on communities on social and economic levels, since these can lead to the loss of lives and amount to extensive material damages. The recent escalation in frequency and severity of fire events has prompted an urgent need for improvements in the stages of preparedness, firefighting and aftermath of these events. To mitigate its effects, wildfire intelligence plays a pivotal role, e.g. to monitor the evolution of wildfires or for early detection in high risk areas such as wildland-urban-interface regions. Recent works have proposed deep learning solutions for fire detection tasks, however the quality and limited size of databases used often does not offer generalization guarantees for reliable deployments in real contexts. To address this issue, we propose the development of expert-in-the-loop systems that combine the automation advantages provided by machine learning with the introduction of relevant domain knowledge expertise. The main objective of this approach is the development of semi-automated software tools that can support data curation by wildfire experts, improving the quality of data sources with the ultimate objective of enabling the development of large-scale databases for wildfire-related problems. More importantly the expert-in-the-loop approach allows involving domain experts and end-users in the development procedure, thus improving the relevance of machine learning applications developed for wildfire intelligence in real contexts.

**A Risk-Cognizant Framework for Predicting Wildfire Location by Using Data-driven Machine Learning Techniques**

Hide abstract

Wildfires are considered as one of the deadliest crises, especially for Californians, who are very susceptible to wildfires due to the hot and dry weather conditions. Frequent wildfires in recent years pose a great threat to many aspects of our life, ranging from poor air quality and loss of habitation to the damage of lots of valuable assets in power system infrastructure. If an accurate prediction of wildfire occurrences in a particular location is available, we can then assess the risk followed by appropriate measures to minimize the potential loss. Due to the adverse effect of climate change, we need to consider different meteorological factors and vegetation condition to capture the pattern of this type of event. In this context, I propose a data driven machine learning (ML) based technique to accurately predict the fire hotspots effectively 3 days or a week ahead. I have already built a dataset from scratch, containing relevant weather factors and vegetation indices around a fire location from past event. From a mentor, I would like to get feedback on how to build a new or adaptive ML algorithm to classify different fire risk levels that will outperform the existing algorithms.

**Mangrove Ecosystem Detection using a Hybrid-Convolutional Neural Networks and Mixed-Resolution Imagery**

Hide abstract

Mangrove forests are one of the most productive ecosystems in the world, as well as a big contributor to carbon sequestration. However, they are currently under threat from anthropogenic activities and thus monitoring their health, extent, and productivity is vital to our ability to study, inform policy, and protect these important ecosystems. While advances in unmanned aerial vehicles (UAVs) technology has allowed us to gather the needed data in the form of high-resolution imagery of the mangrove canopy, the classification of this imagery cannot be done efficiently by mere visual inspection. As a solution, we propose the use of a hybrid neural network, which takes combines a Convolutional Neural Network (CNN) feature extractor with a Multilayer-Perceptron, to accurately distinguish between mangroves and non-mangroves utilizing both high resolution drone imagery and medium resolution satellite imagery. We present a study and comparison of our novel Hybrid CNN to algorithms algorithms previously applied on mangrove image classification to a data set we collected of dwarf mangroves from consumer UAVs in Baja California Sur, Mexico, where we show a 95% intersection over union (IOU) score for mangrove image classification. These results are meant as a real-world case study of these techniques, as well as illustrating how they are powerful tools to aid conservation efforts.

**Efficient Room Environment Control through Occupancy Prediction**

Hide abstract

Introduction We propose a new framework that learns the individual room’s usage patterns and makes prediction for space-specific future room usage. With this room-specific prediction result, we can turn on HVACs at optimal time that the room before it to be occupied (e.g., 20 minutes prior to the first time that the room to be occupied) and also we can turn off the HVACs at optimal time that the room will be emptied. Typically, temperatures of the room is controlled by HVACs on predefined scheduling (e.g., run only weekdays, working hours) and certain temperature setpoints. But this predefined scheduling may not be the optimal approach as actual usage of the building can be different from the predefined time schedule. Step 1) Site-wide Model Training Our first step approach is to train a site-wide (e.g., a campus) predictive ML model. This model uses large amounts of historical data points. As we have no limitation of computation capacity with this model, we can leverage large amounts of data. Step 2) Space-specific Model-adaptation and Prediction One challenge of the site-wide model is its parameter is trained to predict future variables of all spaces. To address this issue, we propose to train space-specific (e.g., each room) models with a newer, real-time site-specific data collected from the specific location. For the final prediction for each space, the space-specific model is used. Approach One important strategy to orchestrate the two different subcomponents is to leverage what is already implemented for distributed communication: Volttron. Once site-wide model is trained, it is shared to individual space-specific computing units (e.g., Raspberry) located at each room via Volttron. With the newly fetched site-wide model, each room controller adapts space-specific model. Once the adaptation is done, we use the updated model to make a prediction for space-specific variables for next certain time-window.

**Forest Cover Forecasting using Generative Modelling**

Hide abstract

Forest canopy cover is important to mitigate the impact of climate change. Yet, existing quantification of urban greenery is either manual and not scalable, or use traditional computer vision methods that are inaccurate. A Generative Adversarial Network is proposed on datasets used for self-driving cars to estimate urban greenery instead. Climate change is making it more and more difficult for forest to survive. The proposed method will use a Conditional GAN approach for forest cover forecasting

**Application of text mining to the analysis of climate-related disclosures**

Hide abstract

In this project we apply text-mining techniques to analyse the TCFD recommendations on climate-related disclosures of the 12 significant Spanish financial institutions using publicly available corporate reports from 2014 until 2019. In our analysis, applying expert judgement, first we create a taxonomy of concepts present in disclosures commonly associated to each of the four areas described in the TCFD recommendations. This taxonomy is then linked together by a set of rules in the form of queries of selected concepts. The queries are crafted so that they identify those excerpts most likely to relate to each of the TCFD’s eleven recommended disclosures. By applying these rules we estimate a TCFD compliance score for each of the four main areas for the period 2014-2019, using five types of reports: Annual Reports, ESG Reports, Corporate Governance Reports, Remuneration of Directors Reports and Pillar 3 Reports. We also describe some challenges in analysing climate-related disclosures in corporate reports. With some caveats, the results give an overview of the evolution of the level of climate-related financial disclosures present in the corporate reports of the Spanish banking sector. The results indicate that the amount of climate-related disclosures reported by the banking sector is growing each year. We also show that to get a clear picture of climate-related disclosures Annual Reports and ESG Reports are not enough since some disclosures are only present in other corporate reports.