Research Proposal Form

Name: Rayan Yu	Research Topic: Machine learning applications to wildfires

Share your problem statement andyour research question and/or project goal.
Problem: Currently, uncontrolled forest fires devastate our natural surroundings due torising temperatures and climate change.
Project Goal: Spatiotemporally optimizing prescribed burns with machine learning to aid in controlling wildfires

Describe three-five key studies that have informed your understanding of the scholarly conversation about your topic.

- Study 1 citation
- Piyush Jain, Sean CP Coogan, Sriram Ganapathi Subramanian, Mark Crowley, Steve Taylor, and Mike D Flannigan. A review of machine learning applications in wildfire science and management. arXiv preprint arXiv:2003.00646, 2020
- Broad, encompassing overview on the research done on machine learning methods for wildfire-based problems; very recent research; categorizes six approximate domains for fire-related machine learning problems: 1) fuels characterization, fire detection, and mapping; 2) fire weather and climate change; 3) fire occurrence, susceptibility, and risk; 4) fire behavior prediction; 5) fire effects; and 6) fire management; analyzes data size, computational requirements, generalizability, and interpretability, potential advances in of wildfires management within machine learning applications; reviews over 300 papers in this issue; this will likely be my most useful source

- Study 2 citation
- Christopher D O'Connor, David E Calkin, and Matthew P Thompson. An empirical machine learning method for predicting potential fire control locations for pre-fire planning and operational fire management. International journal of wildland fire, 26(7):587–597, 2017
- Leverages slightly varying conditions like topography and fuels to quantify the effects of topography, fuel characteristics, road networks and fire suppression effort on the perimeter locations of large fires; they develop a prediction model for geolocational fire control taking into account fuel types, topographic features and natural and anthropogenic barriers; uses boosted logistic regression to a decently accurate margin in spatiotemporal terms; highly relevant, may look to open source from authors to build off of or test another model for

- Study 3 citation
- Anxhelo Agastra, Charley Fite, and Christopher Holmes. Evaluating a probability-based model for prescribed fire forecasting with machine learning. In 100th American Meteorological Society Annual Meeting. AMS, 2020.
- Highly relevant to my original ideas–dicusses finding optimal conditions for prescribed burns; quantifies a machine learning approach in XGBoost for temperature, RH, 10 m wind speed, boundary layer depth, Lavdas Atmospheric Dispersion Index, daily total precipitation, and the climatological probability of fire; works towards probability-based prediction for wildfires to be aided by prescribed burns

- Study 4 citation
- Sriram Ganapathi Subramanian and Mark Crowley. Using spatial reinforcement learning to build forest wildfire dynamics models from satellite images. Frontiers in ICT, 5:6, 2018.
- Discusses spatially spreading processes (SSPs) in depth, and machine learning applications to model and account for these intricacies; this source adopts a more complex reinforcement learning approach to "game" the model, with an agent policy rewarded by the correct identification of cells with/without fire as validated by satellite images; very extensive implementations and validation through five RL algorithms: value iteration, policy iteration, Q-learning, Monte Carlo Tree Search, and Asynchronous Advantage Actor-Critic (A3C).]

- ➔ Study 5 citation
- ➔ David Radke, Anna Hessler, and Dan Ellsworth. Firecast: Leveraging deep learning to predict wildfire spread. In IJCAI, pages 4575–4581, 2019.
- ➔ Contains extensive information on the intersection of geographic information systems and using AI to process those formats of data; implements a 2D convolutional neural network with simple parameters; evaluates historical fire perimeters from GeoMAC, a United States Geological Survey (USGS) database, for each of the training and testing fires; fairly preliminary work, limited by data amount and quality

Identify the gap addressed by your proposed research, and explain how the gap is situated into the scholarly conversation. Provide sources to justify the gap your proposed research is addressing.

The gap addressed by my proposed research is using machine learning to optimize for the control of wildfires in the area of prescribed burns. There exist studies to forecast the size of fires, and attempt to prevent them based on environmental factors, but there has yet to be relevant work in the particular area of spatiotemporally optimizing prescribed burns. The closest research I have found explores climate conditions and environmental factors, but do not investigate the geospatial level of prescribed burns.

- ➔ Piyush Jain, Sean CP Coogan, Sriram Ganapathi Subramanian, Mark Crowley, Steve Taylor, and Mike D Flannigan. A review of machine learning applications in wildfire science and management. arXiv preprint arXiv:2003.00646, 2020
- ➔ Anxhelo Agastra, Charley Fite, and Christopher Holmes. Evaluating a probability-based model for prescribed fire forecasting with machine learning. In 100th American Meteorological Society Annual Meeting. AMS, 2020.

Describe your chosen or developed research method and defend its alignment with your research question.

Machine learning is an effective tool to make accurate, computational predictions based on large sets of data. In geospatial and spatiotemporal applications, machine learning has proven accurate in its ability to learn off of temporal autocorrelation and geospatial patterns.

Identify additional approval processes (check all that apply)

- ❑ Human subjects (requires additional IRB review and approval if student wants to publish and/or publicly present)
- ❑ Animal subjects (requires additional review or approved by the school or district process)
- ❑ Harmful microorganisms (requires additional review o r approval by school or district process)
- ❑ Hazardous materials (requires additional review or approval by school or district process)
- ❑ No additional review or approval required

Describe the data or additional scholarly work that will be generated to answer your proposed research question or achieve your project goal.

- Data agglomeration and preprocessing from sources including the GeoMAC, a United States Geological Survey (USGS) database and other cited sources in my bibliography and related studies

Describe the way you will analyze the data or additional scholarly work generated by your method and justify its alignment with your research question or project goal.

- Will test several machine learning algorithms to fit the finished data set, at the current moment these include: XGBoost (based on previous success in related studies), random forests (industry-known for solid efficacy in most applications), and linear/logistic regressions (simple and effective in most cases).
- Will apply more complex methods like neural networks including convolutional and recurrent if other

methods don't appear to work effectively or if time permits
- Machine learning algorithms are one of the most effective computational methods to analyze and make predictions based on large datasets

List any equipment, resources, and permissions needed to collect data or information. Attach the initial drafts that apply to your proposal if engaged in human subject research informed consent forms, surveys, interview questions, questionnaires, or other data gathering forms, or letters/flyers that will be distributed to study subjects.
- NVIDIA RTX 2060 GPU (already acquired)
- Google cloud if extra computation is necessary

Describe the anticipated logistical and personnel challenges for your research project (to collect and analyze data or to pursue research methods appropriate to a paper that suppers a performance/exhibit/product).
- Data collection and preprocessing will likely surface as the most difficult portion of my project, particularly normalizing and aggregating across different formats of data and discovering how to weigh each aspect when not using a neural network
- Outside of data, designing architecture for more complex deep learning methods like recurrent neural networks may prove difficult as well due to the amount and mathematical complexity of parameters

Provide a brief timeline that outlines your process from now through project completion.
- December→work through data preprocessing
- January→finish data preprocessing + finalize list of algorithms to test
- February→test, optimize, and select most effective algorithms
- March→ if time permits, apply more complex analysis using neural networks, else begin paper edits and revisions from the rough notes taken through the process
- April→finalize research paper
- May→develop/finalize presentation

Discuss the anticipated value and/or broader implications of your research project.
- If successfully completed, my machine learning framework could offer a low-cost, accurate alternative to in-person analysis of high-risk wildfire environments
- By replacing in-person analysis, this framework could be implemented in high-risk zones to pre-determine areas most volatile and prescribe spatiotemporally ideal prescribed burns to minimize and prevent uncontrolled wildfires

Please share any other information that might be relevant to your proposal here.