AI-based Wildfire Prevention System

Project ID : JR-CS-004

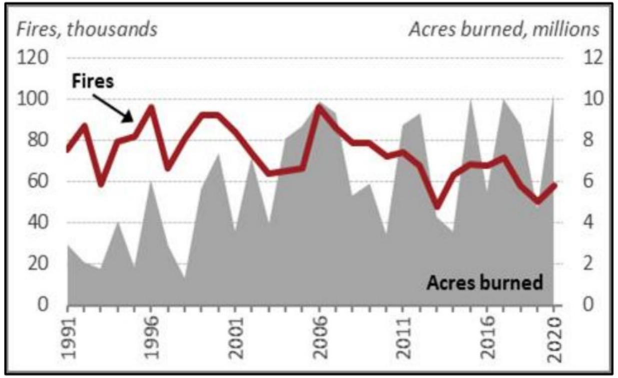
Source: Worldwide Wildlife Fund

Research Question: Can an AI-based wildfire prevention system be created which can effectively predict hotspots based on real-time satellite data and deploy drones to spray fire retardant for wildfire prevention?

Introduction to the Problem

- In the last few years, severe fires have besieged nations around the world, including United States, Canada, Australia, Greece, Spain, Portugal, Chile, Russia, China, Siberia, South Korea, Israel and Brazil.
- The length of the global fire season has increased on average by 19% (WWF).
- Wildfires pose a serious threat to the world's environment as forests account for more than 31% of the world's land surface and contribute to the continuity of ecological balance (EPI).
- The major effects of wildfires include loss of ecosystems and biodiversity, forest degradation, air pollution, soil degradation, economic losses, destruction of watersheds, and other impacts to human well-being and health.
- **In 2020, there were around 58,250 wildfires and more than 10.3 million acres burned in US (USNIFC). The costs for 2020 wildfires in Western US alone totaled between $130-$150 billion (AccuWeather).**
- The 2019–2020 bushfire season in Australia caused the burning of over 46 million acres of forest while destroying over 10,000 structures, resulting in over $100 billion in damages.
- "Right now, most resources are going to fire suppression, but it is clearly shown that investing more before the fire happens is much more efficient," research from Swansea University.

Annual Wildfires and Acres Burned (1991-2020) (US only)



Source: National Interagency Fire Center

There is a need for an AI-based wildfire prevention system which utilizes real-time satellite data to effectively predict hotspots and deploys drones to spray fire retardant to prevent wildfires.

Current Solutions and Design Criteria

Current Solution	Description	Disadvantages
Fire Urgency Estimator in Geosynchronous Orbit	• Fire detection based on satellite images • Uses drones to track the fire's progress • Alerts and dispatches air tanker and ground firefighters for control	• Does not prevent fire from occurring
San Diego Gas & Electric iPredict	• Uses AI based predictive model to increase the accuracy of weather forecast and predicts powerline caused wildfires	• Prevents only powerline caused wildfires
Smart Wildfire Sensor Device	• A device that consists of a network of sensors and uses AI to analyze biomass moisture content and amount of the fuel to predict wildfires in an area	• Device has to be manually installed in the areas for inspection • Cost increases as the number of devices to be deployed across the world will be huge • Only uses biomass as a factor for prediction

<u>Design Criteria:</u>

The system should be able to

- Use real-time satellite data of vegetation, temperature, precipitation, wind, soil moisture and land cover as model input
- Automatically identify hotspots (potential locations of where fires are more prone to start) using AI
- Attain higher than 90% accuracy in hotspot identification
- Eliminate the hotspots where fire retardant is prohibited to be sprayed
- Automatically notify a drone in the area to fly to the coordinates outputted by the classifier

Framework

<u>Machine Learning</u>: An intelligent system that learns and trains itself through data. This project uses supervised learning where the system is given sample training datasets with labels. Once familiarized, it is given real-time data which it labels on its own.

<u>Random Forest Classifier:</u> A classification model that uses a series of decision trees. Decision trees are a series of simple sequential decisions that branch to a specific end result. The random forest model uses multiple decision trees and chooses the answer that has appeared the most times to get the most accurate result.

<u>Why Random Forest?</u>: Random Forest algorithms make decisions based off of multiple decision trees, giving the most accurate result. To get the most accurate results, parameters can be tuned such as n-estimators (the number of decision trees in the classifier) and max-depth (the number of sequential branches on the decision tree).

<u>Underfitting:</u> The model performs poorly on the training data. This is because the model is unable to capture the relationship between the input examples (often called X) and the target values (often called Y).

<u>Overfitting:</u> The model performs well on the training data but does not perform well on the evaluation data. This is because the model is memorizing the data it has seen and is unable to generalize to unseen examples.

<u>Precision:</u> shows out of the number of times a model predicted 0 or 1, how often was it correct?

<u>Recall:</u> shows out of the total actual 0 or 1, how many of them were correctly predicted as 0 or 1?

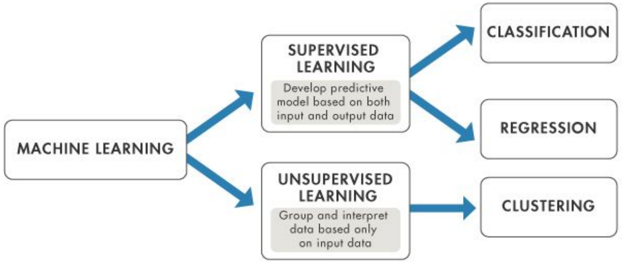


Image Source: MathWorks

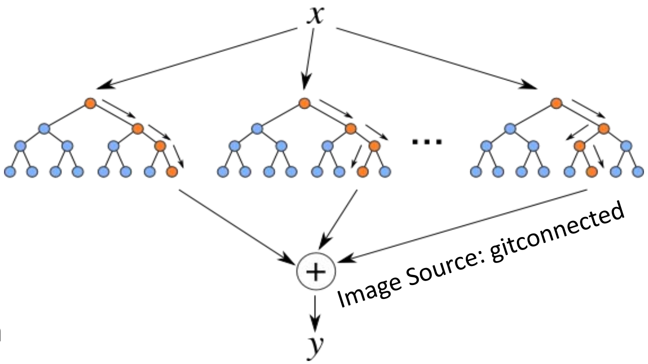


Image Source: gitconnected

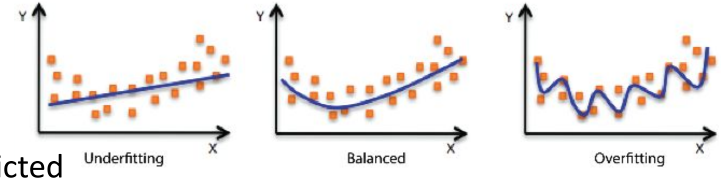
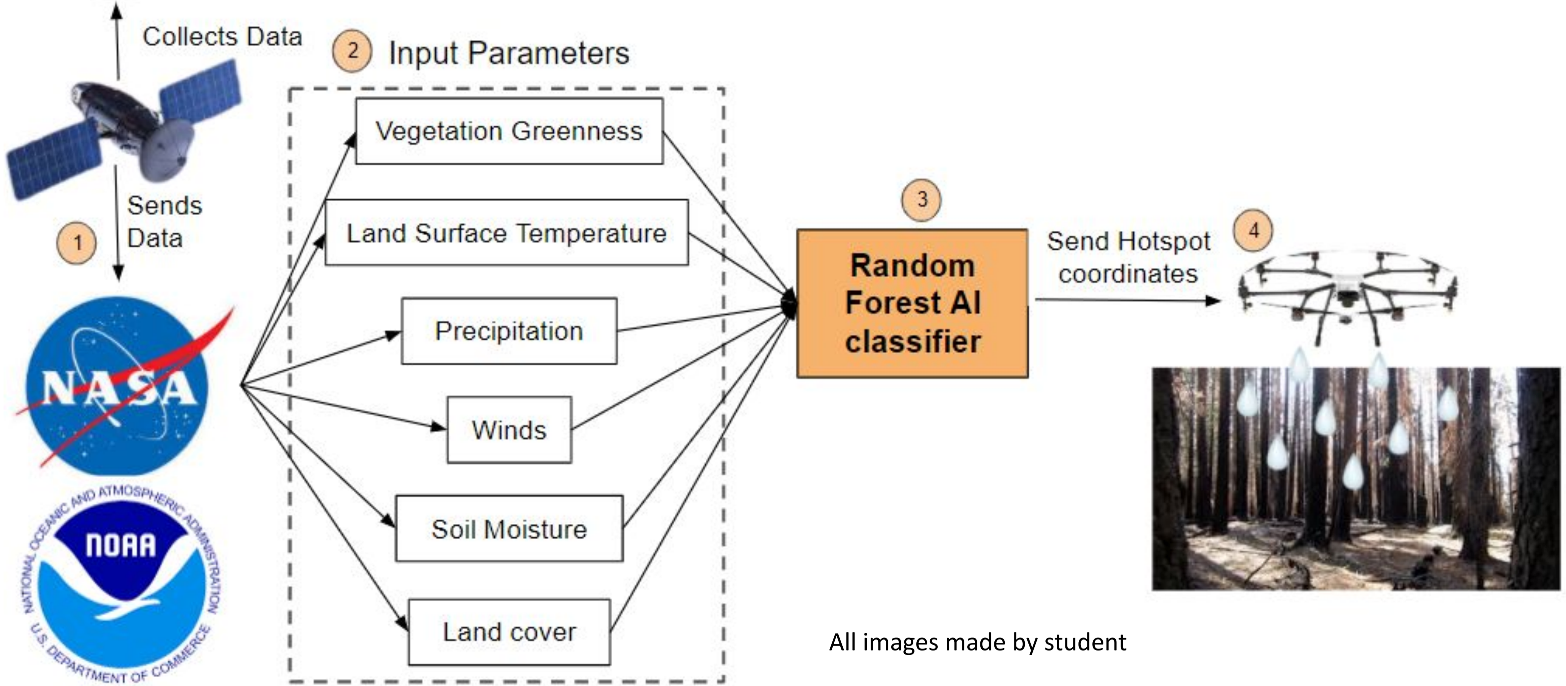


Image Source: Amazon Machine Learning

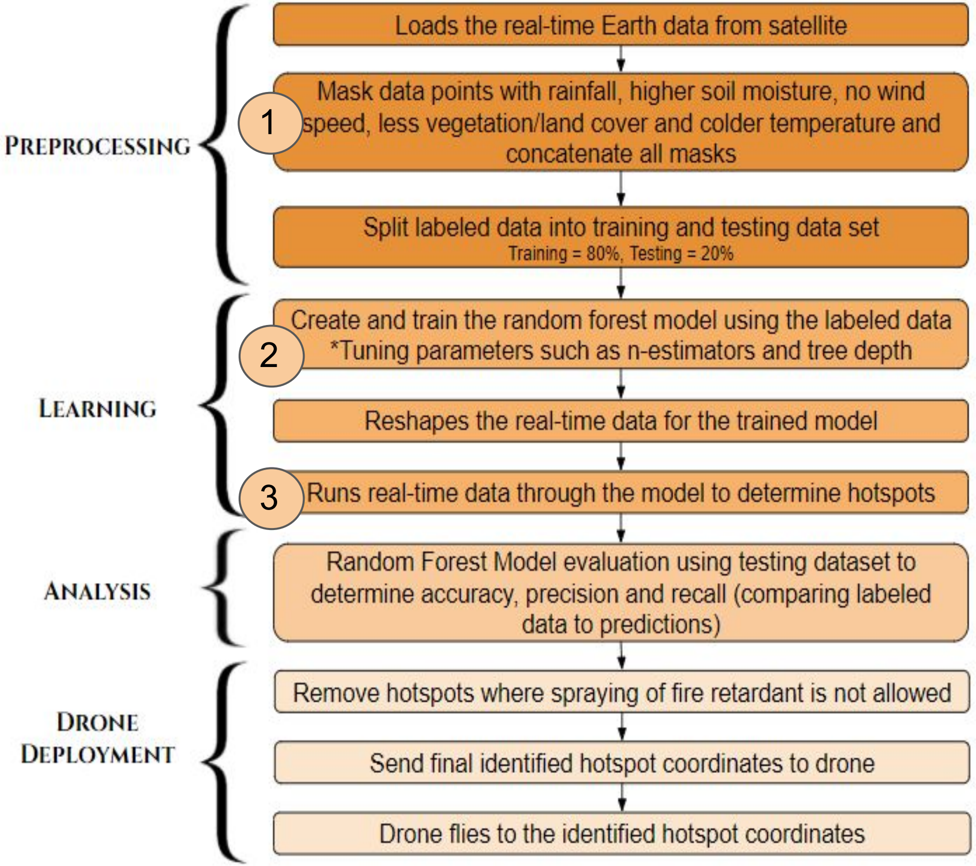
WPS Design



All images made by student

1. Loads real-time Earth data from NASA and NOAA
2. Takes multiple parameters as input to predict hotspots accurately
3. Uses a Random Forest AI classifier to determine whether a lat-lon coordinate is a hotspot or not
4. Deploys a drone to hotspot location

Build Procedures and Code



All images and graphs made by student

```
# load landcover image coordinates which doesn't contain: open water, perennial ice/snow,
# Developed land (open space, low intensity, medium intensity, high intensity), barren land
# (Rock, sand, clay), moss, woody wetlands and emergent herbaceous wetlands.

# Stiches all the images of the world into a map so it is easy for visualization
img = wf.ImageCollection.from_id('nlcd:land_cover').mosaic()
# Creating a mask for the data and points that should not be included in the classifier
unwanted_areas = (img == 11) | (img == 12) | (img == 31) | (img == 90) | (img == 21) |
(img == 22) | (img == 23) | (img == 24) | (img == 95) | (img == 74)
img = img.mask(unwanted_areas)
# Visualizing the dataset on the map
img.visualize('land_cover', scales=[[0, 95]], colormap='viridis')
```

1

```
# Load the Random Forest Classifier for finding hotspots
from sklearn.ensemble import RandomForestClassifier
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score
# Spliting the dataset into testing and training data
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=0)
# Setting up the classifier and the data
clf = RandomForestClassifier(n_estimators=10, random_state=0).fit(X_train, y_train)
# Testing the model and evalutaing the mode
y_test_pred = clf.predict(X_test)
print(accuracy_score(y_test, y_test_pred))
```

2

```
# Running data through the model to get predictions
if arr.shape[0] > 0:
    predictions = clf.predict(arr)
    xx_wm_pos = xx_wm[predictions == 1]
    yy_wm_pos = yy_wm[predictions == 1]
    if xx_wm_pos.size > 0:
        lon, lat = transform(Proj(init='epsg:3857'), Proj(init='epsg:4326'), xx_wm_pos, yy_wm_pos)
        lat_lons = np.stack((lat, lon), axis=1)
        lat_lons = np.unique(np.round(lat_lons * 1e1) / 1e1, axis=0)
        print(lat_lons)
        np.savetxt('coordinates.txt', lat_lons)
    else:
        print("There are no hotpots found.")
else:
    print("All data masked.")
```

3

Model Parameters

The Model Parameters are used for both Preprocessing (Masking) and Learning (Classification).

Model Parameters

National Land Cover Dataset (NLCD) Land Cover

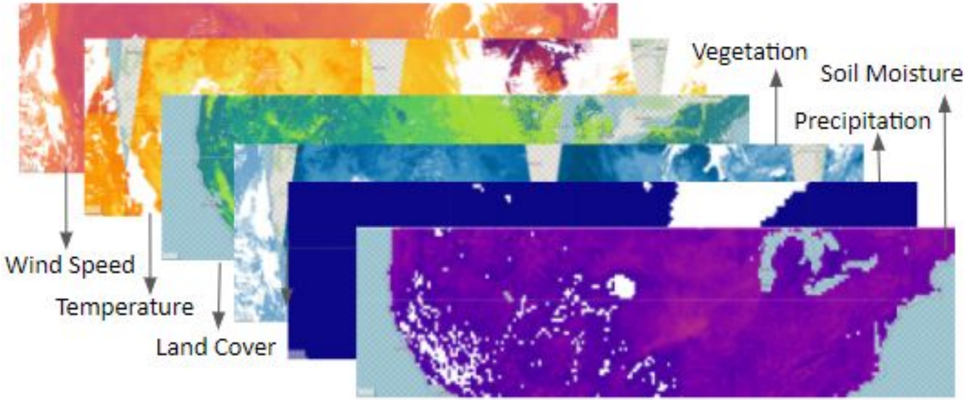
1. **Land Cover:** masks areas with open water, perennial ice/snow, developed land (open space, low intensity, medium intensity, high intensity), barren land (rock, sand, clay), moss, woody wetlands and emergent herbaceous wetlands

NCEP GFS Weather (NOAA)

2. **Wind Speed** (gust) [m/s]: 100 < wind speed <-100
3. **Precipitation** (rate) [kg/(m^2 s)]: rainfall > 0
4. **Soil Moisture** (Volumetric Soil Moisture Content, 0 - 0.1 m depth below land layer level): soil moisture < 2000

MODIS Aqua/Terra Satellite (NASA)

5. **Temperature** (2 datasets of surface temperature (Band 20 (3660-3840 nm) and Band 31 (10780-11284 nm)): temperature > 0.5
6. **Vegetation** (2 datasets Near Infrared and red combined for Normalized Difference Vegetation Index) [ndvi = (nir - red) / (nir + red)]: ndvi < 0



The process of masking is eliminating areas where fires will never be able to start such as water bodies, rainfall, etc. In the above images, the areas being masked are white.

Test Procedures

Training Dataset Collection (one-time manual process for supervised learning) (Fig 1)

1. Identify and label the hotspots (1) based on the 6 model input parameters
2. Identify and label non-hotspots (0) which are coordinates different from above
3. Complete steps 1 and 2 until there is a minimum of 350 data points

Tuning of parameters: n-estimators and max-depth for maximum accuracy of classifier

1. Modify n-estimators from 1-15 and store the test and train accuracy
2. Modify max_depth from 1-15 and store the test and train accuracy
3. Determine the values for the highest model accuracy

Evaluating the Random Forest AI Classifier (Fig 2)

1. Run the classifier with the real-time data to determine hotspots
2. Evaluate model based on the following metrics:
   a. Accuracy
   b. Precision
   c. Recall

Deploying of Drone (Fig 3)

1. Identified hotspots are sent to drone using python code
2. Drone flies to location

Hotspot Identification Process

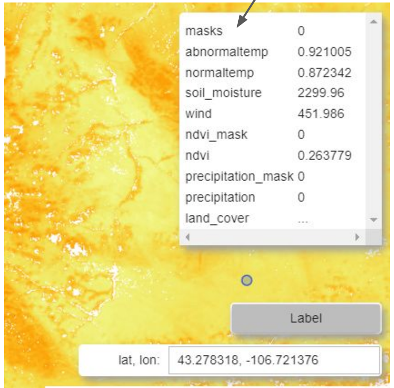
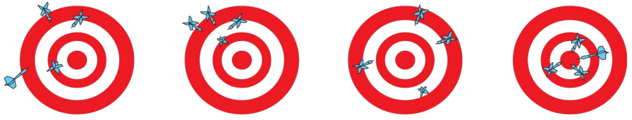


Image made by student Fig 1



Accuracy ✗ Precision ✗ | Accuracy ✗ Precision ✓ | Accuracy ✓ Precision ✗ | Accuracy ✓ Precision ✓

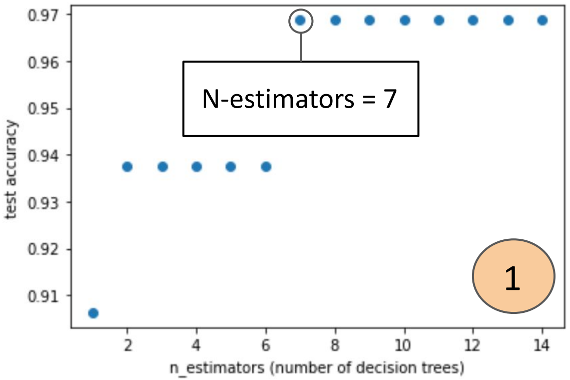
Source: Exploring our fluid Earth Fig 2

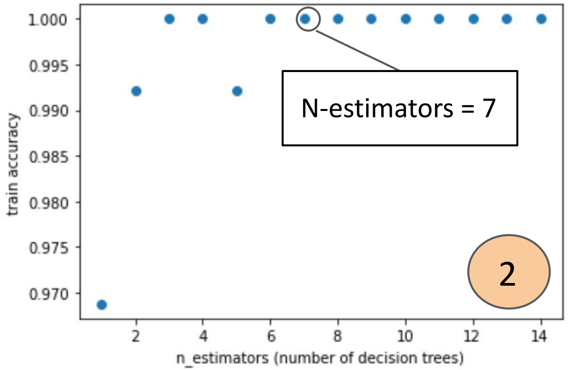
Drone Deployed

Image taken by student Fig 3

Findings (Analysis)

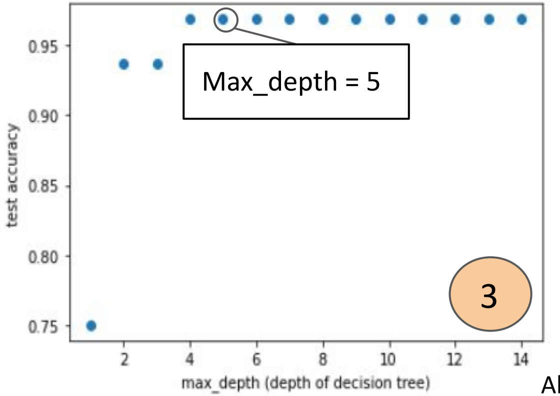
For identifying hotspots of where wildfires could start using the Random Forest Classifier, following are the results of tuning the parameters of n-estimators (number of decision trees) and max_depth (maximum tree depth) for highest testing accuracy

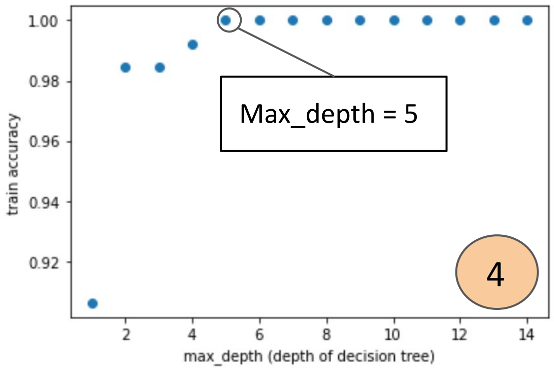




- When the number of n-estimators < 7, the model is **underfitting** the data.
- When the number of n-estimators > 7, the model is **overfitting** the data.

From graphs 1 and 2, it was analyzed that for the highest model accuracy, the number of n-estimators should be 7.



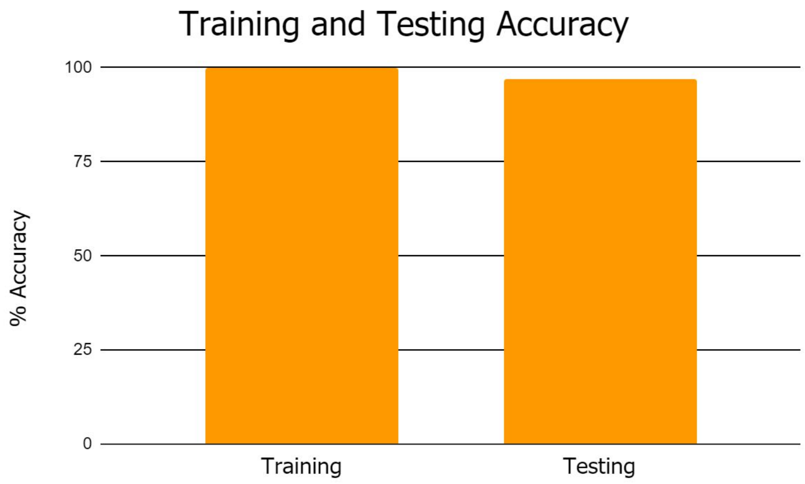


All graphs made by student

- When the max_depth < 5, the model is **underfitting** the data.
- When the max_depth > 5, the model is **overfitting** the data.

From graphs 3 and 4, it was analyzed that for the highest model accuracy, the max_depth should be 7.

Findings (Analysis)



The WPS Random Forest (RF) Classifier was able to attain 96.8% testing accuracy in identifying the hotspots where wildfires can start.
(n-estimators = 7 and max depth = 5)

All images and graphs made by student

To validate that the WPS RF classifier was able to identify hotspots and not hotspots, it was analyzed using a confusion matrix and precision and recall.

Confusion Matrix

	Actually Hotspot (1)	**Actually Not Hotspot (0)**
Predicted Hotspot (1)	59.38% True positives (TP)	3.13% False positives (FP)
Predicted Not Hotspot (0)	0 False negatives (FN)	37.5% True negatives (TN)

The WPS RF classifier predicted minimal (3.13%) false positives (not a hotspot but identified as a hotspot), and no (0%) false negatives (hotspots being missed).

Precision and Recall

	Precision	**Recall**
1.0 (Hotspot)	95% (TP/FP+TP)	100% (TP/TP+FN)
0.0 (Not a hotspot)	100% (TN/FN+TN)	92% (TN/TN+FP)

The WPS RF classifier was 97.5% precise while identifying a hotspot and 96% able to recall its previous accurate decisions.

Conclusion

Wildfire Prevention System successfully meets the design criteria:

1. WPS uses real-time satellite data from NASA and NOAA of soil moisture, precipitation, vegetation, wind speed, land cover, and temperature to identify hotspots.
2. WPS attains 96.8% accuracy in identifying hotspots.
3. WPS automatically notifies drones to fly to the hotspot coordinates.

Comparison with current methods:

1. WPS predicts the hotspots using multiple parameters, while current methods use limited parameters.
2. WPS effectively deploys drones to hotspots while current methods only predict locations.
3. WPS is preventative compared to other methods being reactive.
4. WPS is not only cost-efficient but also fully automated reducing the chance of human error.

The power of the AI-based Wildfire Prevention System developed can be applied to any location in the world to identify hotspots, deploy drones to the hotspots and therefore prevent wildfires.

1. Help prevent Climate Change
2. Save precious lives
3. Preserve and protect ecosystems and biodiversity
4. Avert huge economic losses (e.g. in US alone, ~ 10.3 million acres from burning and ~ $130-$150 billion annually)

Future Research and Bibliography

Current Limitations

- A large amount of real-time data requires significant computer processing power
- Currently the WPS drone can fly to the location, however, it can neither carry nor spray the fire retardant

Future Research

- To use cloud computing or supercomputers to decrease the time taken to process huge amounts of real-time data from satellites
- Use an agricultural drone which can carry and spray the fire retardant on hotspots

Applications

- The power of the AI-based Wildfire Prevention System developed can be applied to any location in the world to identify hotspots, deploy drones to spray fire retardant and therefore, prevent wildfires. It can be used by fire departments to prevent wildfires from starting as well as mitigate existing wildfires from spreading.

Bibliography

1. “Types of Machine Learning: 3 Machine Learning Types You Must Know.” UpGrad Blog, 19 June 2020, www.upgrad.com/blog/types-of-machine-learning/.
2. Yiu, Tony. “Understanding Random Forest.” Medium, Towards Data Science, 14 Aug. 2019, towardsdatascience.com/understanding-random-forest-58381e0602d
3. “Facts + Statistics: Wildfires.” III, www.iii.org/fact-statistic/facts-statistics-wildfires.
4. “Wildfire Statistics for the US and Abroad: Policy Advice.” PolicyAdvice, 5 Feb. 2021, policyadvice.net/insurance/insights/wildfire-statistics/.
5. “Reviewing the Horrid Global 2020 Wildfire Season " Yale Climate Connections.” Yale Climate Connections, 6 Jan. 2021, yaleclimateconnections.org/2021/01/reviewing-the-horrid-global-2020-wildfire-season/.