Introduction
Out of all the issues that California faces on a yearly basis, wildfires have been at the top of the damage list for decades. Wildfires not only cause deaths, they displace people from homes and jobs, leaving them hopeless. While measures to reduce the forming of wildfires exist, they simply aren't effective enough to solve the issue. California's 2018 wildfires alone cost the US economy $148.5 billion (Corey 18.)

Background or Literature Test

Hypothesis / Engineering Goal / Frameworks

Developing a Machine Learning Equipped Drone for Early Detection of Wildfires in Forests around the Bay Area

Materials, Methods, Procedures

DATA
Charts
Models
Photos
Graphs
Drawings

Results / Findings

Conclusions / Further Research

References / Acknowledgements

Developing a Mach
Equipped Dron
Detection of Wildfi
around the B

chine Learning
one for Early
lfires in Forests
Bay Area

Picture of us flying the drone
in Bay Area forests

Introduction

In 2018 alone, the economic toll of California's wildfires reached a staggering $148.5 billion (Corry 20), underscoring the persistent and severe impact of these incidents on both lives and livelihoods. Wildfires, a recurring challenge for the state, not only claim lives but also displace individuals, leaving them deprived of homes and employment opportunities. The destruction extends to buildings and cherished possessions, contributing to a profound sense of loss. Moreover, the toll on forests, integral to environmental balance, is equally substantial. While existing measures, such as fire hazard warnings, have proven beneficial by raising awareness and enabling strategic resource allocation by emergency response teams, inherent limitations persist. Manual collection of current fire hazard data by experts, though valuable, lacks the speed and efficiency demanded by the urgency of wildfire threats. In response, our machine learning model, trained with a diverse dataset, automates this process, operating almost in real-time through a drone. This technological innovation offers a more efficient and comprehensive solution, bridging the gap between expert knowledge and public accessibility. Our project introduces a user-friendly website, requiring only three simple inputs—temperature, humidity, and a satellite image of the area. This interface empowers users to assess wildfire risk in their specific location promptly. In addition, our website has multiple features, such as a map of respective zones around the bay area and the wildfire risk in each area. By increasing access to fire probability insights, our initiative aims to equip individuals with the knowledge needed to make informed decisions about their safety and the well-being of their communities. In essence, our project transcends the limitations of manual data collection and prediction, offering a transformative approach to wildfire risk assessment that is swift, automated, and accessible to all.

Background

Numerous models to detect the hazard risk of fires have been created in the past. For example, (Kondylatos et al., 22) used machine learning and deep learning to detect if an area was prone to wildfires in the mediterranean region. They used past data to train their model, and compared their outputs with the 2020-2021 Greek wildfire season to verify their results. In addition, The United States Department of Homeland Security also created a National Wildfire Risk Index Map. However, this map uses data from past wildfires, not a machine learning model.

Like our project, drones are also being used for visual images to detect wildfires. One example of this is SmokeD. "The SmokeD firefighting drones are drones equipped with cameras with sensitive optical sensors capable of spotting the first signs of fire and smoke from afar." (SmokeD). Similarly, (Sathishkumar et al. 23) used machine learning to detect wildfires. They mention Convolutional Neural Networks (CNN), a type of Artificial Intelligence, that "have been shown to outperform state-of-the-art methods in image classification and other computer vision tasks." However, these two examples differ from our project because they only detect fires once they have started, while ours aims to do it prior.

Finally, (Rubi et al. 23) developed a model using machine learning that predicted the behavior and spread of a fire once it had already started in Brazil. They explain how their model uses data from monitoring stations, satellites, and many other features. Some examples include distances to water sources, geography, vegetation, and urbanization index. Using multiple parameters like these, their model could accurately predict where and how the fire traveled.

Engineering Goal

Our engineering goal is to develop a specialized wildfire prediction model for the Bay Area's forests. We utilize temperature and relative humidity data, along with aerial pictures, as input data for our machine learning model. By utilizing easily accessible drone-captured imagery, we aim to provide access to wildfire risk assessment tools, making the model user-friendly for individuals with basic drone usage skills. This approach enables a diverse user base, including local authorities and community members to monitor and assess the likelihood of wildfires in the bay area in an instant. After collecting data using the drone, the image, temperature, and humidity should be manually inputted in the model on our website, and the likelihood of a wildfire is instantly displayed to show the chances of a wildfire in the area. In addition, our website will contain a map of zones around the bay, and the risk of fire in each area. Our project does not detect a wildfire after it has started, rather it identifies the probability of one occurring in a specific location in the future. Hopefully, at the end of our project, our model will be able to predict fire risk with above a 90% accuracy rate.

Results

Model (*see figure 1 and 12*)
In validating our results, we utilized the National Wildfire Risk Index Map provided by the US Department of Homeland Security. The conclusion of our analysis revealed an impressive 91-92% accuracy for our model (*see figure* 4), showcasing its robust reliability. Notably, as we expanded our dataset with additional data points, the model demonstrated, as anticipated, an improvement in accuracy.

Wildfire Risk Index over various Temperature and Humidity Values (*see figure* 5)
We collected 445 data points, each of which containing an image of the vegetation, and temperature & humidity values. We inputted these data points into our model, which gave us the risk index of each point. The plot shows the correlation of risk index and temperature/humidity values. Overall, areas with high temperature and low humidity tend to have higher risk indexes. We can also see that the risk index of areas with similar temperature/humidity values can still vary, which suggests that the vegetation image captured by the drone also plays an important role in determining the final risk index.

Bay Area Map (*see figure* 9)
We collected 94 data points consisting of a combination of drone and satellite images, and plotted the risk index on a map of the Bay Area. Overall, the wildfire risk of the Bay Area at this time seems to be moderate. This is likely due to the current season, as wildfires occur more commonly around August. Our data is also quite consistent with the USDHS's map, with forests in the northern Bay Area having a relatively low risk index compared to the southern Bay Area.

Conclusion

In conclusion, the impact of this cutting-edge fire-detecting drone resonates across various aspects of global safety, helping the wildfire alert process by eliminating dependence on traditional news sources or weather updates and letting individuals easily assess the risk in their immediate surroundings. In the real world, people can now use our machine to determine the likelihood of wildfires around their area and wherever they want to test for wildfires. Looking ahead, our commitment to continuous improvement leads us to envision a massive leap in fire detection technology. Our primary objective is to enhance the system's responsiveness, striving to establish the world's first instantaneous fire detection capability. So far, our drone images can be transmitted to our back-end servers in real time, but the temperature and humidity data still needs to be transferred from the mobile device to our back-end servers. This will completely eliminate the need to manually input the data onto our website, and instead, a real time risk index will be shown while the drone is being flown. In addition, we can work to make the model more accurate in the future. It is already at 91-92% accuracy, and with more training data and model configuration, we plan to make this value rise even higher. We also plan on adding more features to the model in addition to drone image, temperature, and humidity. This may include elevation, wind speed, and fuel loading. By optimizing efficiency, speed, and accuracy, our vision is to create a seamlessly integrated system that sets a new standard for rapid and reliable fire risk assessment, which ultimately contributes to enhanced safety measures worldwide.

Acknowledgements

A special appreciation to Mr. Lester Leung for his support and guidance, without which nothing would have been possible. His expertise and encouragement have been really important in shaping our success in building this machine learning model and drone.

We would also like to thank our parents for taking the time to drive us around forests in the Bay Area so we could collect real world data.

Overall, we are deeply grateful for the valuable contributions of everyone involved, making this project a reality.

Methods

Data Collection

The dataset, comprising of approximately 5750 entries for both fire and non-fire data points, was gathered through a systematic process. Every single datapoint had a satellite image of the vegetation, the highest temperature of the day, and the humidity present at the time.. The data collection process was guided by the US Homeland Security map of wildfire risk. The tools we used for data collection were the Google Earth Pro app (satellite image), wikipedia (fire information), and timeanddate.com / wunderground.com (temperature/humidity data). We also manually flew the drone later to collect some data that could be used in the testing process (*see figure* 10). The drone costed $150, and the the sensor was $50. We attached a SensorPush sensor on the drone using rubber bands that collects the temperature and humidity data. The drone collects a bird's-eye view image of the vegetation and sends it directly to our back-end servers. After developing the model, we decided to build a website that allows the user to input data, and results in a fire risk probability for the user's location. As an additional feature, we added a map of the bay area with fire risk assessment data for specific locations (*see figure* 9).

5750

5750

2875

2875 2875

Machine Learning Model *(see figure 1)*

To create the machine learning model, we used Python's Tensorflow library, which offers a set of tools for building and training models. We created two separate model structures for the data, which consisted of temperature and humidity values, and image data, which is a satellite image of the vegetation. For the numeric data, we used a Multi-layer Perception (MLP) model, because this type of model is great at recognizing patterns in small amounts of data. For the image data, we used a Convolutional Neural Network (CNN) model, which is a common way to analyze images. The outputs of these two models are then concatenated together and processed with a linear activation layer to produce the final result, which is the risk index of the data.

Ophelia Go
is an app fit for drone control and aerial photography. We used/plan to use Ophelia Go to capture the aerial photos required for our machine learning model to predict the likelihood of a fire.

Sensorpush
is an app dedicated to give users environmental data. We used this device, and paired it with the Ophelia Go app to receive the temperature and relative humidity data.

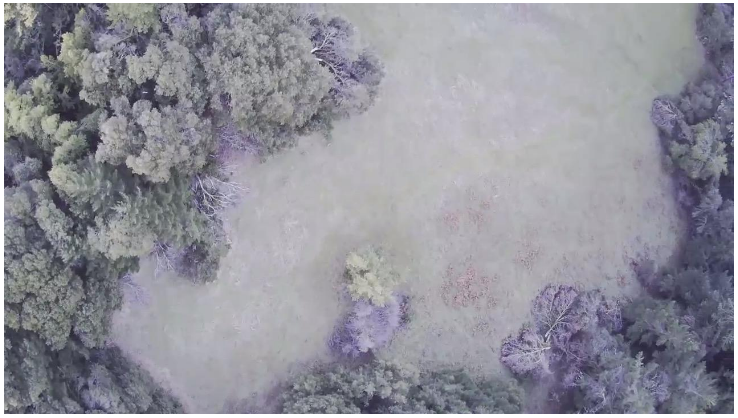
Real world data

Figure 10

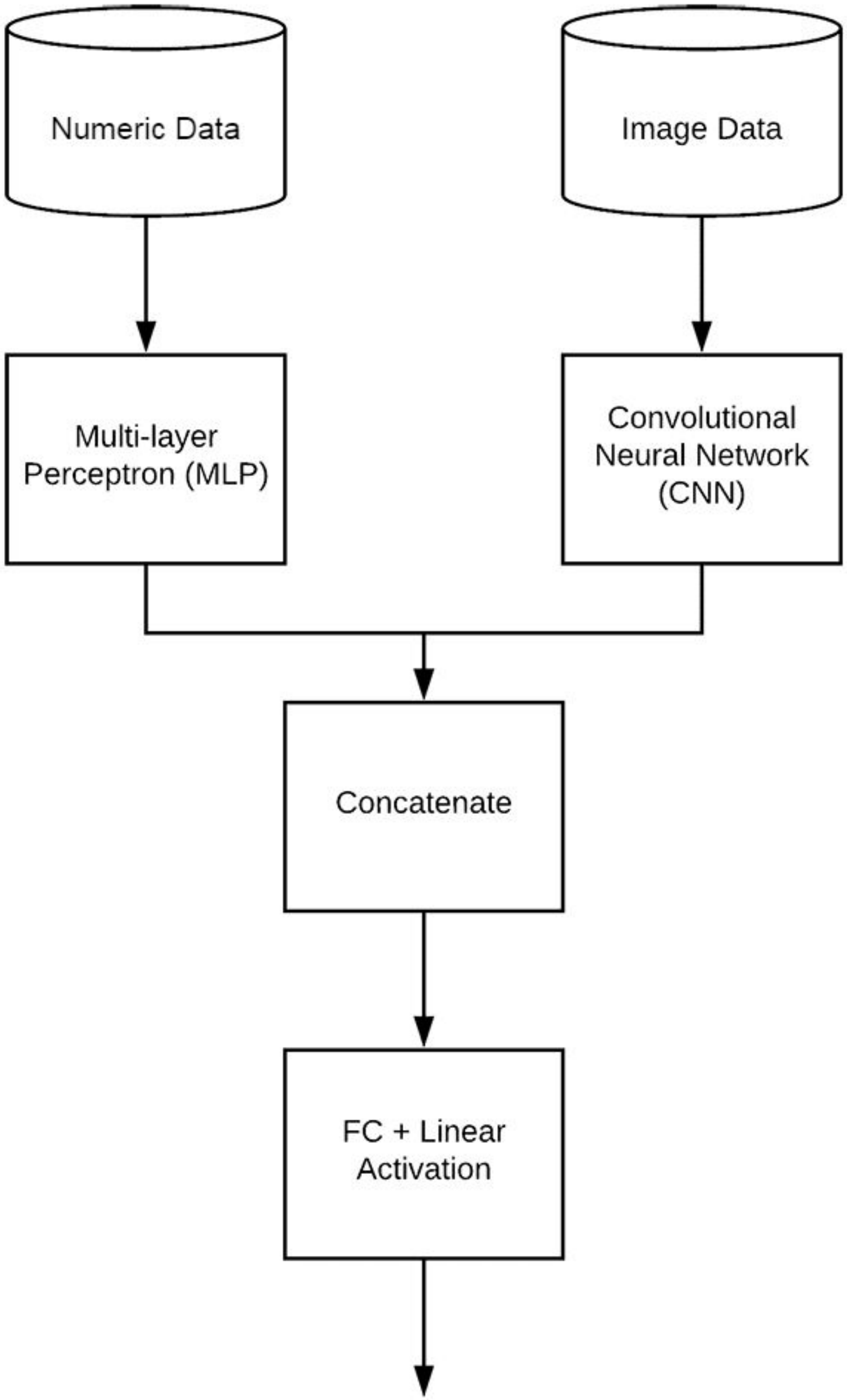
62 degrees Fahrenheit, 63% humidity
Risk Index: 46.502

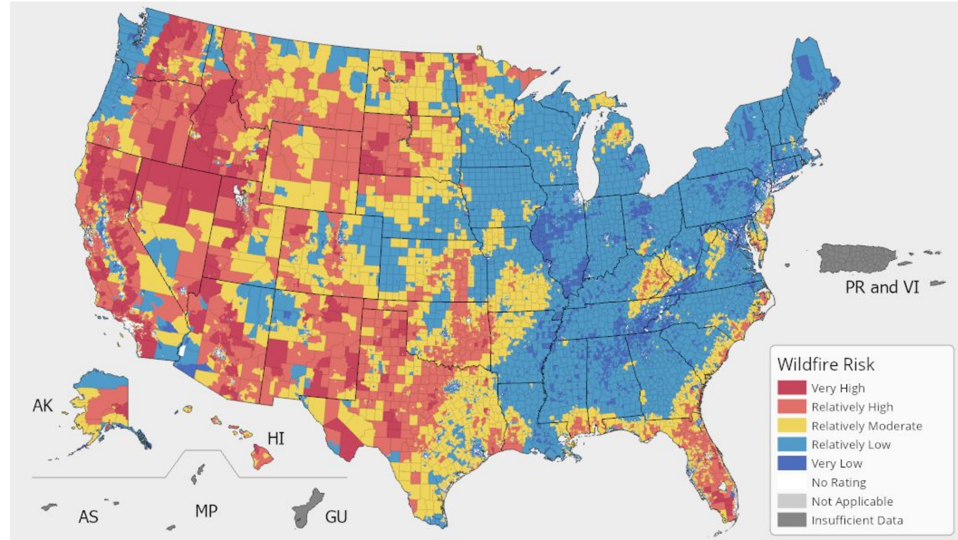
50 degrees Fahrenheit, 70% humidity
Risk Index: 24.849

59 degrees Fahrenheit, 72% humidity
Risk Index: 52.117

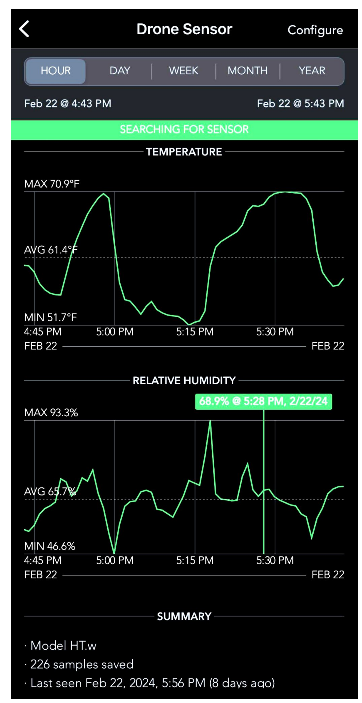


51 degrees Fahrenheit, 72% humidity
Risk Index: 44.715





We used the National Wildfire Risk Index Map by the US Department of Homeland Security to validate the accuracy of our model. (https://hazards.fema.gov/nri/map)



Temperature and humidity data from the SensorPush app.

```python
def create_cnn(width, height, depth, filters=(16, 64, 256), regress=False):
    # initialize the input shape and channel dimension, assuming
    # TensorFlow/channels-last ordering
    inputShape = (height, width, depth)
    chanDim = -1
    # define the model input
    inputs = Input(shape=inputShape)
    # loop over the number of filters
    for (i, f) in enumerate(filters):
        # if this is the first CONV layer then set the input
        # appropriately
        if i == 0:
            x = inputs
        # CONV => RELU => BN => POOL
        x = Conv2D(f, (3, 3), padding="same")(x)
        x = Activation("relu")(x)
        x = BatchNormalization(axis=chanDim)(x)
        x = MaxPooling2D(pool_size=(2, 2))(x)
        # flatten the volume, then FC => RELU => BN => DROPOUT
        x = Flatten()(x)
        x = Dense(16)(x)
        x = Activation("relu")(x)
        x = BatchNormalization(axis=chanDim)(x)
        x = Dropout(0.5)(x)
        # apply another FC layer, this one to match the number of nodes
        # coming out of the MLP
        x = Dense(4)(x)
        x = Activation("relu")(x)
        # check to see if the regression node should be added
        if regress:
            x = Dense(1, activation="linear")(x)
        # construct the CNN
        model = Model(inputs, x)
        # return the CNN
        return model
```

3]

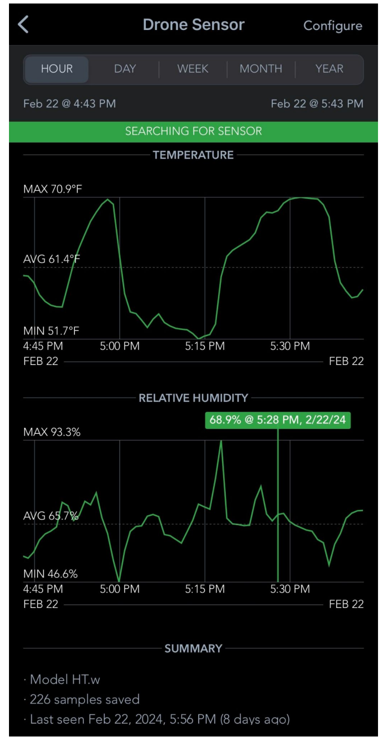
```python
def create_mlp(shape, regress=False):
    # define our MLP network
    model = Sequential()
    model.add(Dense(8, input_dim=shape, activation="relu"))
    model.add(Dense(4, activation="relu"))
    # check to see if the regression node should be added
    if regress:
        model.add(Dense(1, activation="linear"))
    # return our model
    return model
```

[82]

```python
# create the MLP and CNN models
mlp = create_mlp(numlist.shape[1], regress=False)
cnn = create_cnn(256, 256, 3, regress=False)
# create the input to our final set of layers as the *output* of both
# the MLP and CNN
combinedInput = concatenate([mlp.output, cnn.output])
# our final FC layer head will have two dense layers, the final one
# being our regression head
x = Dense(4, activation="relu")(combinedInput)
x = Dense(1, activation="linear")(x)
# our final model will accept categorical/numerical data on the MLP
# input and images on the CNN input, outputting a single value (the
# predicted price of the house)
model = Model(inputs=[mlp.input, cnn.input], outputs=x)
```

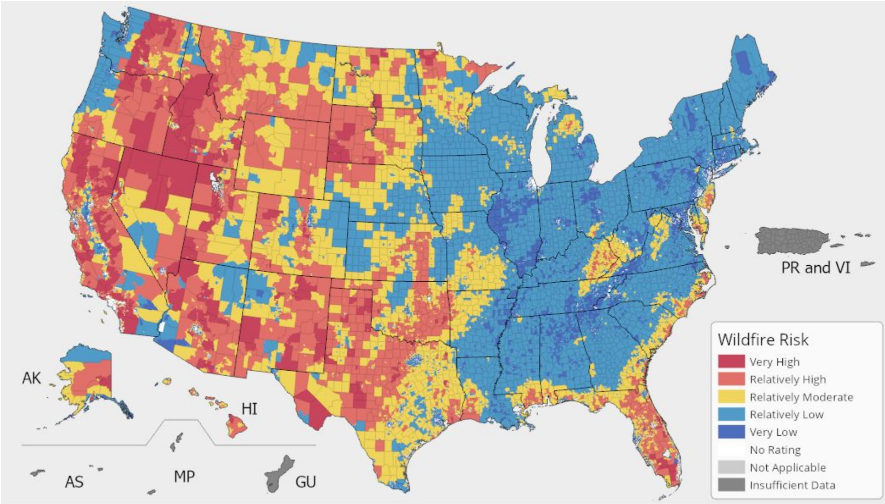
[112]

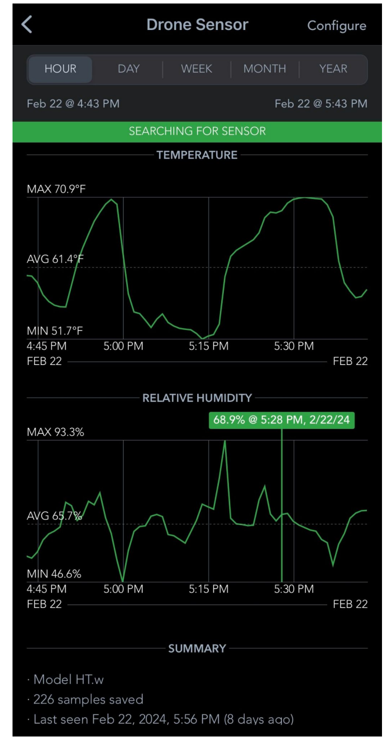
Python code for the Tensorflow Model Structure



Drone Sensor
Configure
HOUR
DAY
WEEK
MONTH
YEAR
Feb 22 @ 4:43 PM
Feb 22 @ 5:43 PM
SEARCHING FOR SENSOR
TEMPERATURE
MAX 70.9°F
AVG 61.4°F
MIN 51.7°F
4:45 PM
5:00 PM
5:15 PM
5:30 PM
FEB 22
FEB 22
RELATIVE HUMIDITY
68.9% @ 5:28 PM, 2/22/24
MAX 93.3%
AVG 65.7%
MIN 46.6%
4:45 PM
5:00 PM
5:15 PM
5:30 PM
FEB 22
FEB 22
SUMMARY
· Model HT.w
· 226 samples saved
· Last seen Feb 22, 2024, 5:56 PM (8 days ago)

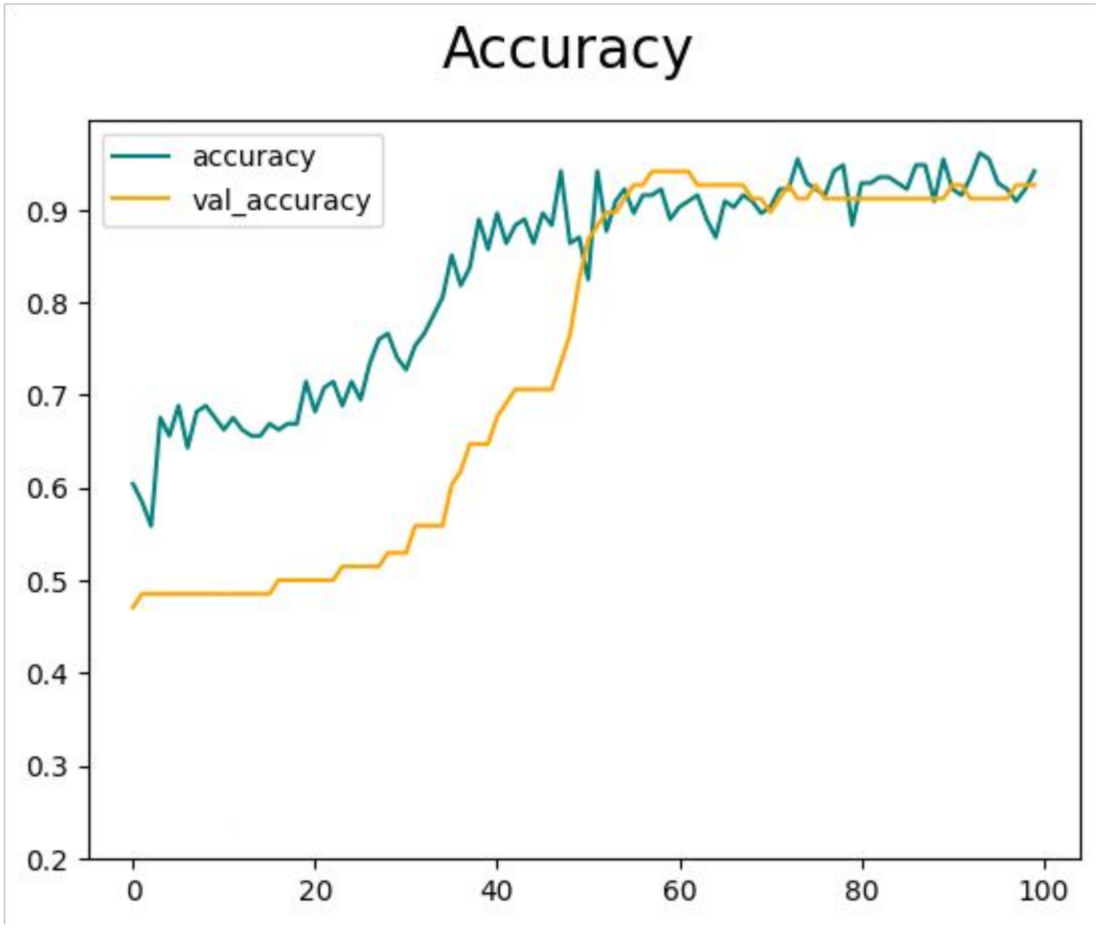
Figure 2





We used the National Wildfire Risk Index Map by the US Department of Homeland Security to validate the accuracy of our model. (https://hazards.fema.gov/nri/map)

Figure 3



<u>accuracy</u>:

- Accuracy with training data
- Steady growth
  - Logistic growth
  - Growth slowed at 50th epoch
- Ups and downs/choppy
  - Data is hard to learn

<u>val_accuracy</u>:

- Accuracy with validation data
  - new data the model hasn't seen before
- Logistic growth
  - Model couldn't find much of a pattern in the first 15 epochs
  - After a pattern was found, accuracy quickly improved
- Relatively smooth

We trained the model with around 5750 pieces of data in total (2875 fires, 2875 nonfires)
The model trained for 100 epochs with batch sizes of 32. We found that over 100 epochs wouldn't give us the best result as the model tends to memorize the training data instead of finding patterns, which is why the validation accuracy would go down. Anything less than 100 epochs wouldn't let the model learn fully.
In the end, our model had an accuracy of around 91% ~ 92%

<u>Figure 4</u>

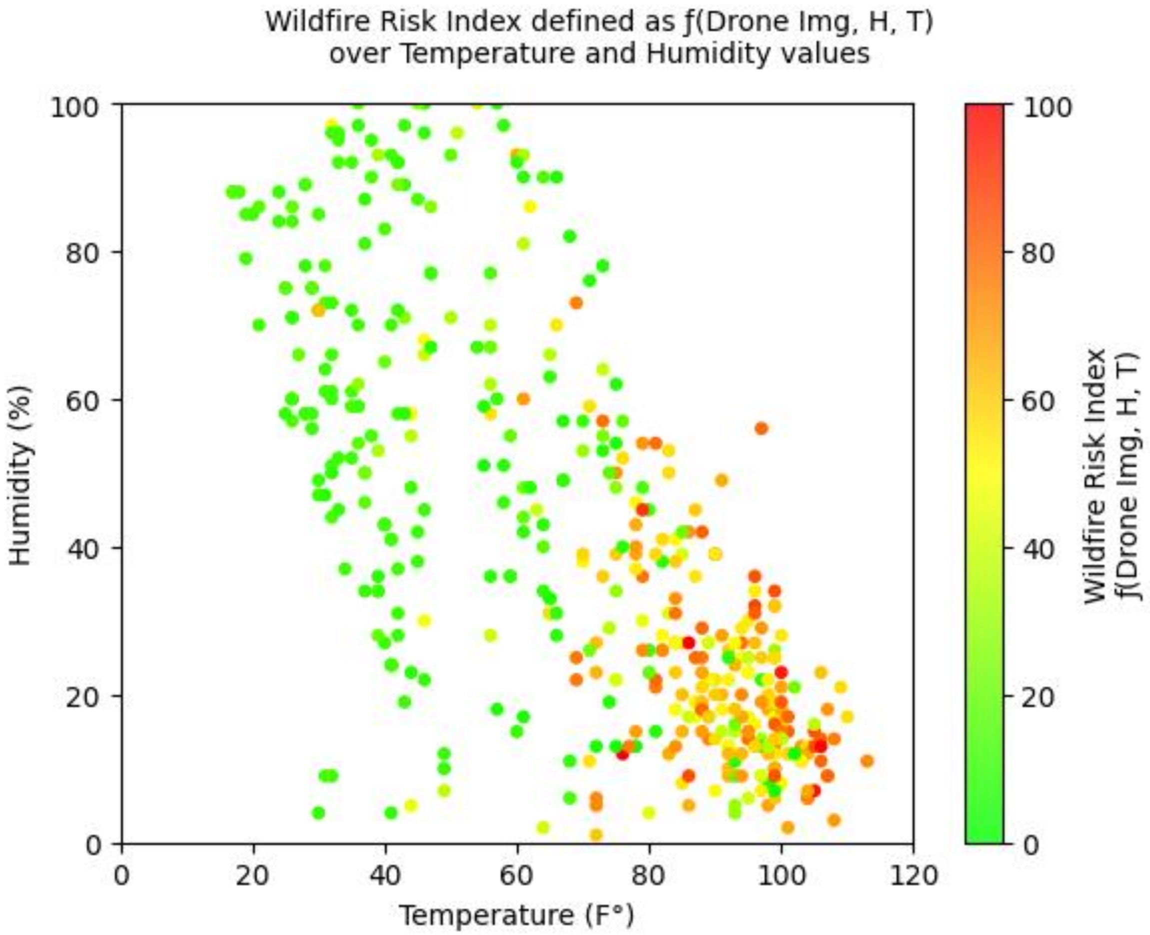
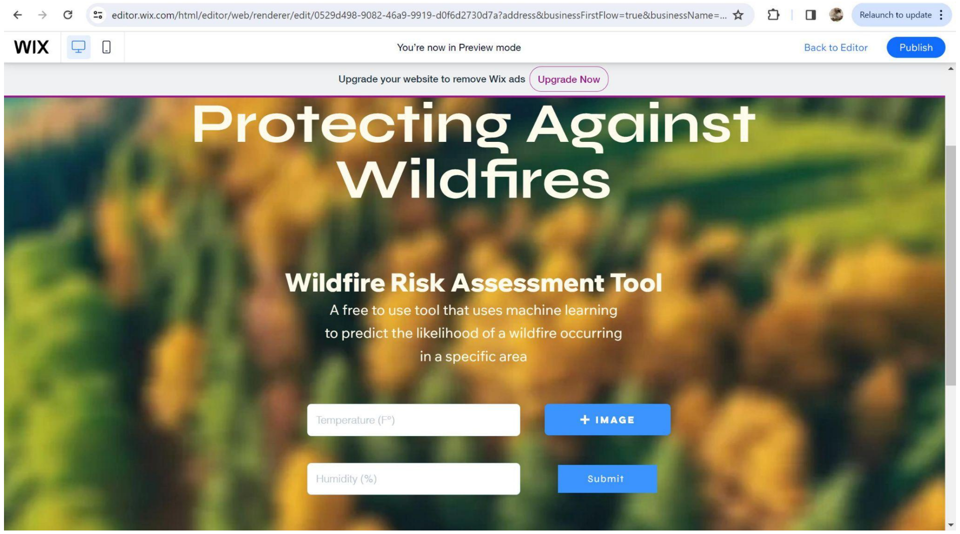
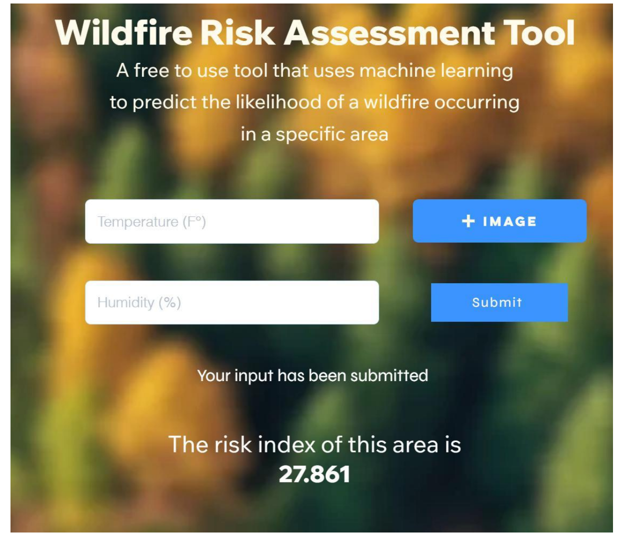
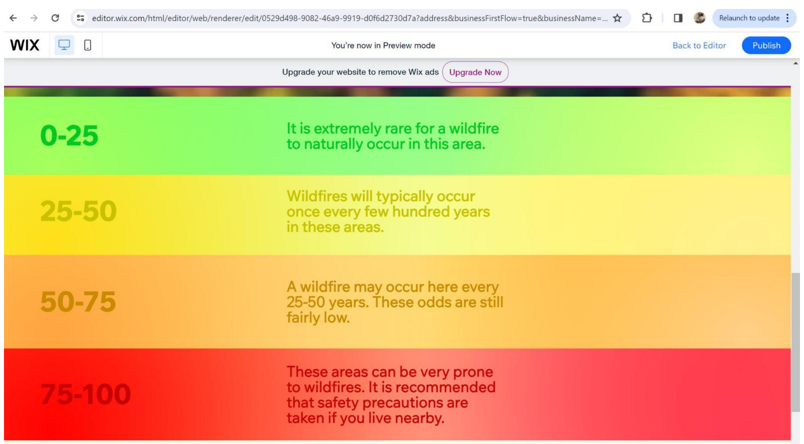


Figure 5

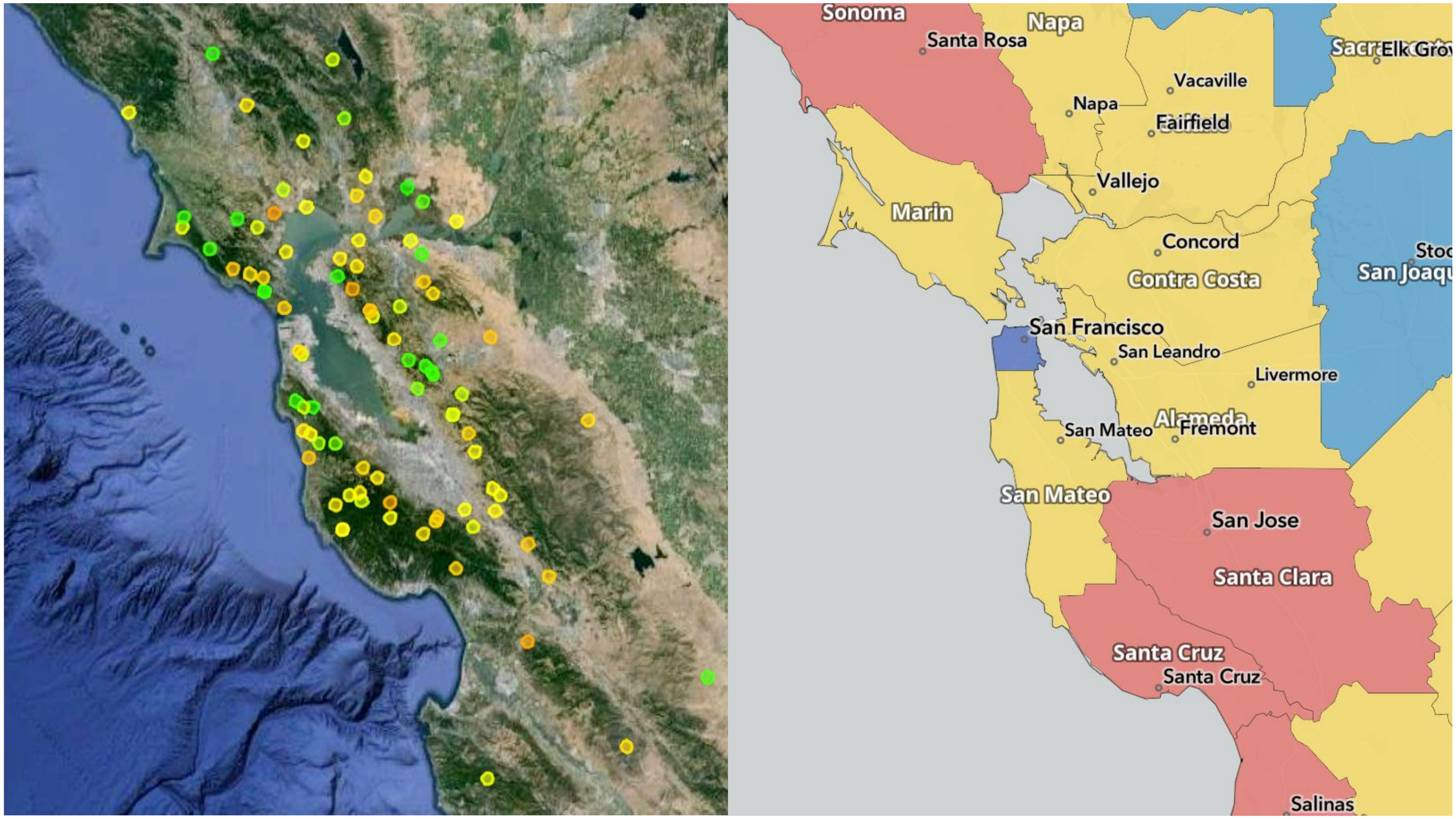
We collected 445 data points, each of which containing an image of the vegetation, and temperature & humidity values. We inputted these data points into our model, which gave us the risk index of each point. The plot shows the correlation of risk index and temperature/humidity values. Overall, areas with high temperature and low humidity tend to have higher risk indexes. We can also see that the risk index of areas with similar temperature/humidity values can still vary, which suggests that the vegetation image captured by the drone also plays an important role in determining the final risk index.





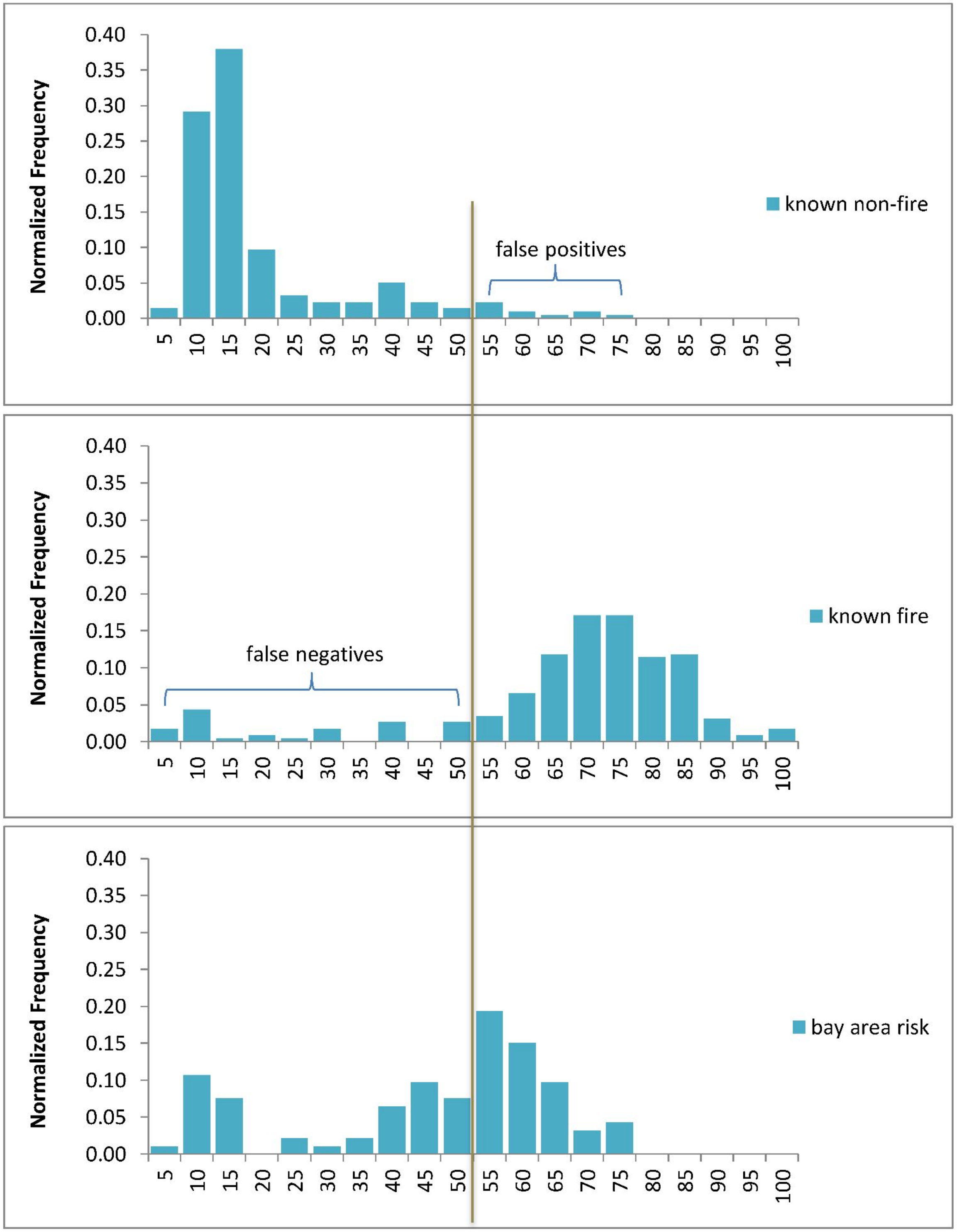


We also created a website using a platform called Wix, to allow people around the world to gain access to our tool. Simply input the temperature and humidity values, as well as upload a drone image of the vegetation, and a risk index score ranging from 0-100 will be shown to evaluate the wildfire risk. Scores from 75-100 means the area may be prone to wildfires.



We collected 94 data points consisting of a combination of drone and satellite images, and plotted the risk index on a map of the Bay Area. Overall, the wildfire risk of the Bay Area at this time seems to be moderate. This is likely due to the current season, as wildfires occur more commonly around August. Our data is also quite consistent with the USDHS's map, with forests in the northern Bay Area having a relatively low risk index compared to the southern Bay Area.

Figure 9



f(Drone Img, T, H) based Wildfire Risk Index

Figure 1

Figure 2

Figure 3

Figure 5

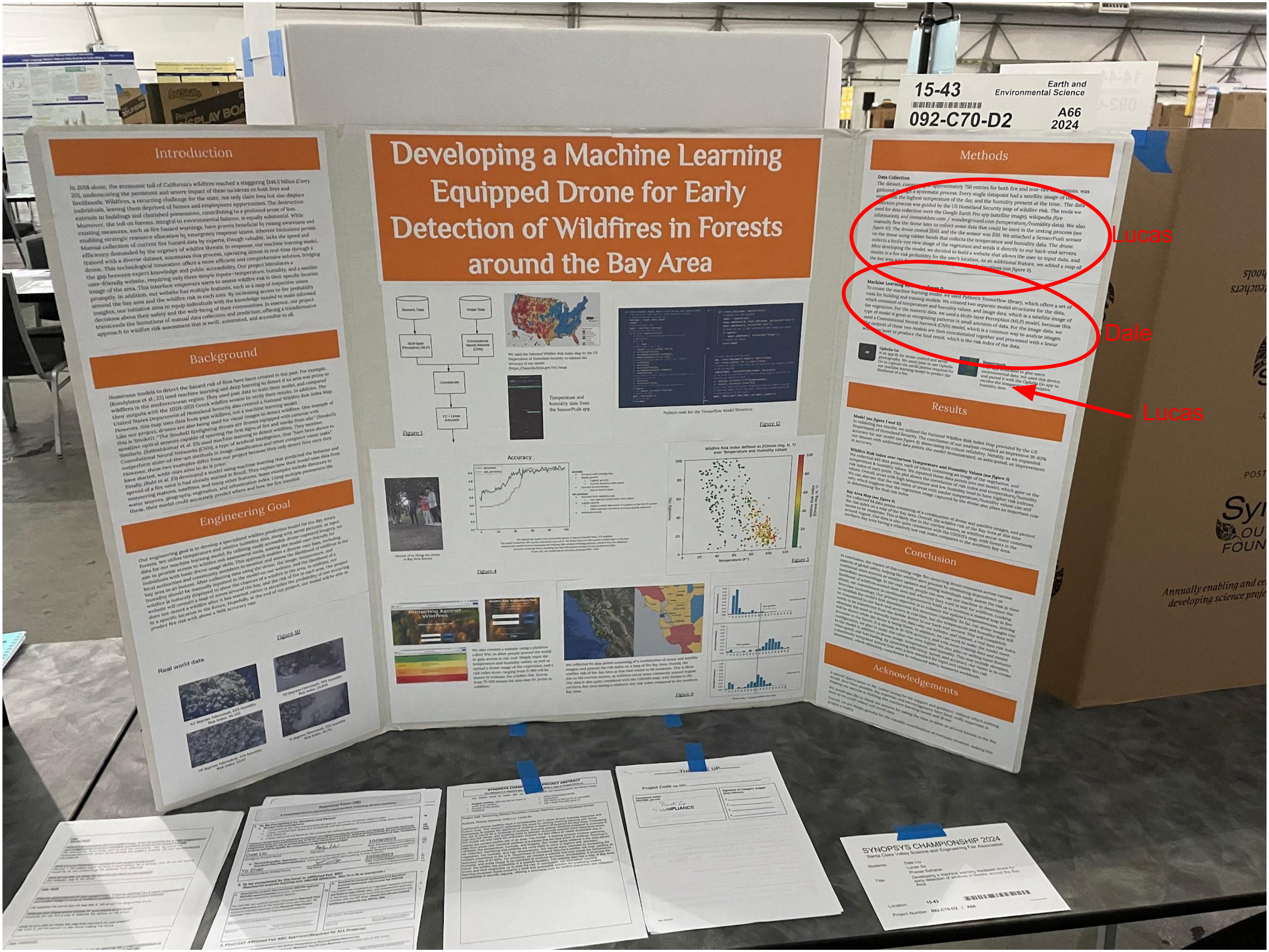
Figure 9

Figure 4

Figure 10

Figure 12

Figure 11


15-43
Earth and Environmental Science
092-C70-D2
A66
2024
Introduction
Developing a Machine Learning Equipped Drone for Early Detection of Wildfires in Forests around the Bay Area
Methods
Lucas
Dale
Lucas
Background
Engineering Goal
Results
Conclusion
Acknowledgements
Accuracy
Real world data
SYNOPSYS CHAMPIONSHIP 2024
Santa Clara Valley Science and Engineering Fair Association