Wildfire Detection in the U.S.

By: Karan Bhalla, Nyah Strickland, Tanisha Damle, Yusra Abdelrahman, Yu Wati Nyi

Background Information

- **Uncontrolled fire** that burns in the **wildland vegetation**
- **Global feature** occuring in **every month of the year**
- Most of the **area burned** by wildfires occurs in **grasslands and savannas**
- **Humans** are responsible for starting over **90% of wildland fires**, and **lightning** is responsible for almost all of the **remaining ignitions**
- Impact humans either **directly** through **loss of life** and **destruction to communities** or **indirectly through smoke exposure**
- **Billions of dollars spent** every year on **fire management activities**
- **Wildfire management** includes **emergency response, ecosystem management, land-use planning, and climate adaptation, etc.**
- Wildland fire is a **complex process. Behavior and occurrence** impacted by several interrelated factors, including **ignition source, fuel composition, weather, and topography**

Evolution

- Initially focused on predicting wildfire spread, then pivoted to wildfire detection
- Read articles to deepen understanding of which machine learning model is best for image classification problems
- Major challenges
  - Creating directories and 2 classes (fire/nonfire)
  - Testing different models to determine best accuracy rate

Research Question

How can machine learning techniques be leveraged to classify outdoor images as fire or non fire to enhance wildfire monitoring systems for improved safety and alertness?

Dataset

Source: FIRE Dataset (Kaggle)
Licensed by: CC0 Public Domain

- The dataset was created by a team during the NASA Space Apps Challenge in 2018, the goal was using the dataset to develop a model that can recognize the images with fire.
- Data was collected to train a model to distinguish between the images that contain fire (fire images) and regular images (non-fire images), so the whole problem was binary classification.

Dataset

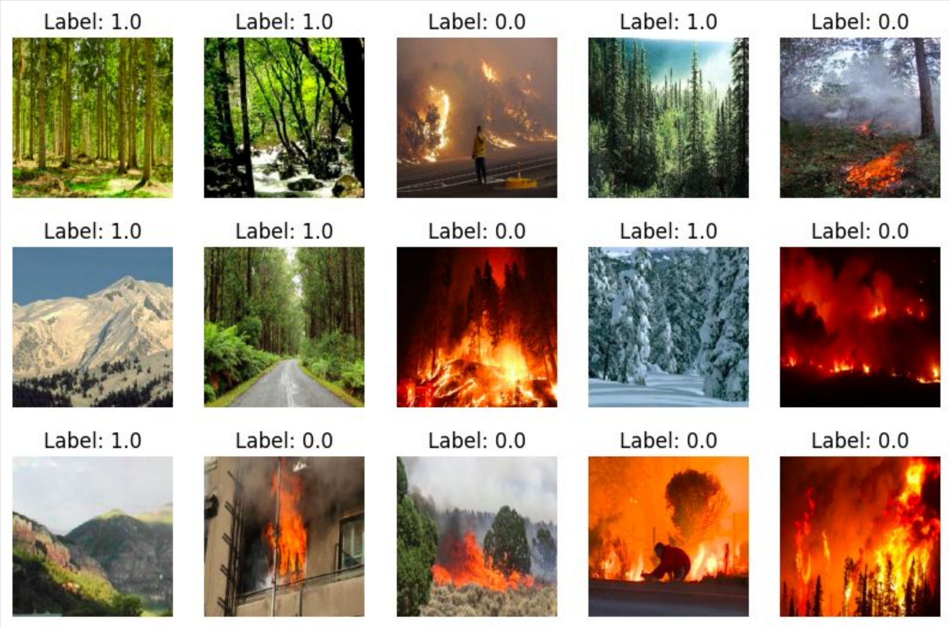
Source: Wildfire Detection Image Data (Kaggle)
Licensed by: Database: Open Database, Contents

- Forest fire data that can be used for Machine Learning and Deep Learning Process
- This data was used because our first dataset had an imbalance between fire vs bonfire images, so the non-fire images from this dataset were randomly added to our dataset

Data Preprocessing

- Image Generator (rescaling/normalization)
  - Adjusting the range of the pixel values in an image to a standard scale, typically between 0 and 1
  - This normalization of pixel values improved the performance of our model
- To improve the accuracy of the model, we did not use any other data preprocessing measures (using many other preprocessing techniques made accuracy go down)

Data Visualization



- Five images per generator are printed at random with the labels to distinguish between fire vs non-fire images

Machine Learning Algorithms

We used supervised learning and a computer vision technique.

- **Convolutional Neural Networks**
  - ResNet50
  - Inception ResNet V2
  - Dropout layer (dropping out some inputs to avoid overfitting)
  - Dense Layer (sigmoid activation function, outputs probability)
- **Inputs and outputs of the model**
  - Input: Image
  - Output: classification of fire or not fire
- **Infrastructure**
  - Google colab
  - Tensorflow
  - Numpy
  - Matplotlib

Inception-ResNet-V2

- Convolutional neural network based on a combination of Inception and ResNet architectures
- This function returns a Keras image classification model pre-trained on ImageNet.

What is ImageNet?

- A dataset organized based on WordNet. WordNet has more than 100,000 words/phrases with 80,000+ nouns
- In ImageNet, there are on average 1000 images to illustrate each word/phrase.

Model Training

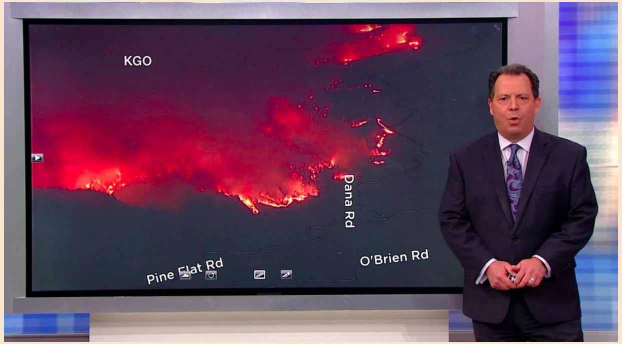
- **Training size** : 799 (to train the model)
- **Validation size:** 101 (to evaluate on unseen data)
- **Testing size:** 99 (final metrics to see how well model can predict)
- **Hyperparameters:**
    - **Image-Size:** (299, 299) pixels
    - **Batch-Size:** 32
    - **Learning Rate:** Adam classifier -> default
    - **Epochs:** 5
- No fine-tuning done (we froze the layers in the pretrained base)

Demo & GitHub Repository

GitHub: https://github.com/Aran203/wildfire_predict

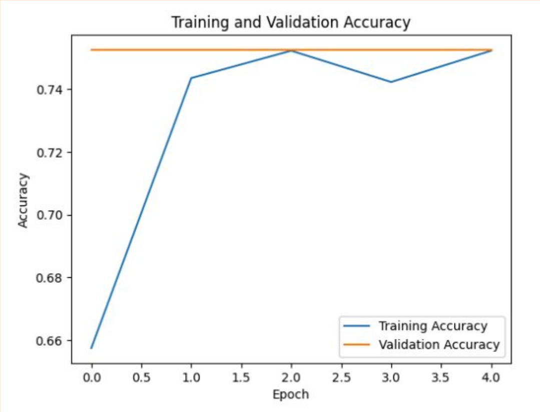
Adversarial Testing Images

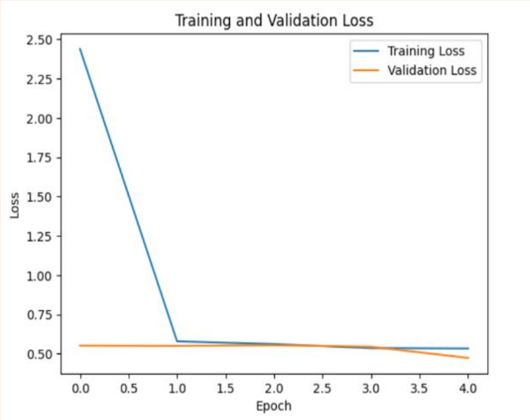
1. Wildfire on the tv
2. Stop sign
3. Outdoor Barbecue/grill
4. Campfire
5. Candle
6. Overhead/airplane view of wildfire
7. Volcano
8. Orange and red flowers
9. Fire on the stove
10. Sunset
11. orange/red leaves on ground



Evaluation

1. Limit false negatives (when the model says it is not a fire when in reality it is) and improve true positive rate
2. Focus on recall (sensitivity) since it is the true positive rate
3. ResNet50:
   a. Accuracy: 75%
   b. Recall: 1.00
4. InceptionResNetV2:
   a. Accuracy: 94%





Essential Question

The development processes of responsible AI/ML solutions require thoughtful choices, considering potential biases at each step, along with an awareness of societal implications across diverse demographics. Projects are oriented towards a specific ethical goal with clear principles to ground them from the outset. This approach facilitates the mitigation of adverse impacts while strengthening beneficial outcomes because it permits an informed understanding of these impacts and their origins. With that knowledge, one can take action to minimize negative effects and maximize positive ones.

Impact

Positive impact:

1. **Early Detection and Response** of Wildfires
2. **Improved Monitoring** of Vulnerable Areas
3. **Environmental Conservation** and lesser risk of human life in the process

Not so positive impact:

1. **Algorithmic Bias of the Model** due to the nature of the data it has been trained on
2. **Privacy Concerns** with respect to Data Collection
3. **Increased Dependency** on Technology will cause in concentration of such technologies in richer pockets

Sources of Bias

1. Sampling bias: Is the dataset collected from all over the world?
    a. This dataset was uploaded by another group, not NASA, so some information might be lost?
2. Representation bias: A small dataset can lead to certain type of images being less represented and not fully capturing the diversity of real-world fire, both wildfire and non-wildfire, image distribution
    a. The dataset only has outdoor pictures.
    b. The model will wrongly predict indoor fires (e.g. candle fire) as wildfires, as seen with adversarial testing.

Challenges

1. **GPU limitations** - not all of us were able to run the model
2. **Splitting the dataset** - first we were just able to split the names of files into different arrays, then we ended up changing the original dataset. Then we realized it was best to make copies of the dataset.
3. **First-Timers** - not sure if we implemented the machine learning algorithm perfectly, we were learning and building at the same time.

Future Work

- By **predicting the detection** of a wildfire and the locations it can spread to, there can be quick evacuation plans put in place to endanger less lives.
- Quicker **Resource Allocation** and **Dispatch of Firefighters**
- **Limit property damage** through early prediction/threat assessment
- Property damage is a **high risk for insurance companies** as it can lead to severe financial losses for them, especially in areas where a whole neighborhood was destroyed
- Can focus on **protecting individual lives** and **financial stability of insurance companies**.

Applications: Preparedness & Recovery, Wildfire Surveillance, Threat Assessment, Insurance Companies

Next Steps

a. Karan - lead a project about natural language processing in the coming semester
b. Yu - want to implement machine learning in web application project in the future.
c. Yusra - plan to apply knowledge gained towards making an extra credit AI project for the AI foundations course I am taking
d. Tanisha - continue exploring AI and creating new AI projects with this experience
e. Nyah - plan to take a predictive analytics course at my university to deepen my understanding of topics learned during this program

Sources

- https://www.tensorflow.org/hub/tutorials/tf2_image_retraining
- https://www.kaggle.com/code/abdelazizmakhlouf/fire-classification-model
- https://towardsdatascience.com/types-of-biases-in-data-cafc4f2634fb
- https://cdnsciencepub.com/doi/pdf/10.1139/er-2020-0019?download=true
- https://arxiv.org/pdf/2112.02447.pdf
- https://firststreet.org/our-mission