OXFORD JOURNAL
OF STUDENT
SCHOLARSHIP

Deep Learning-Based Fire Segmentation Using U-Net

Ayushi Agarwal
techyayu2009@gmail.com

ABSTRACT

Wildfires are dangerous threats to ecological systems, infrastructure, livestock, and human life. Therefore, reliable and scalable fire detection methods are required to mitigate their impact. This study proposes a deep learning-based framework for fire segmentation using RGB images. A U-Net convolutional neural network architecture was implemented and trained on a dataset of approximately 11,100 images paired with corresponding pixel-wise binary segmentation masks to accurately segment fire-affected regions under diverse visual conditions. To evaluate generalization, the model's performance was analyzed across varying dataset sizes (100%, 50%, and 25%) and training durations (10 and 20 epochs). Evaluation was conducted using four segmentation metrics: Dice coefficient, Intersection over Union (IoU), precision and recall. The results demonstrate that the proposed model effectively identifies fire-affected regions, even in complex visual scenes. However, a noticeable decline in performance was observed as the dataset size decreased, emphasizing the importance of large training datasets in deep learning-based applications. Overall, these findings highlight the effectiveness of convolutional neural networks for fire segmentation. This research highlights the potential of models such as the one described above for future applications in wildfire monitoring and related computer vision tasks.

INTRODUCTION

Wildfires and Their Impact

Wildfires are any unplanned or uncontrolled fire affecting natural, cultural, industrial and residential landscapes. Wildfires represent a hazard that is primarily influenced by humans and thus to a degree can be predicted, controlled and in many cases, prevented (UNDRR, 2021). However, wildfires that are not controlled can result in adverse health conditions as a result of exposure to the gaseous air pollutants like carbon monoxide and nitrogen dioxide. Fires affect more than 350 million hectares every year worldwide (Musabyimana, 2022).

Fire Detection and Segmentation

Recent advances in sensing technologies, remote imaging, and computer vision have enabled new approaches for wildfire detection. However, many traditional methods still struggle with fire-like objects, elevated false alarm rates, difficulty detecting small fire regions, and high computational latency (Cui et al., 2020). As a result, deep learning has become an increasingly important tool for wildfire analysis because of its strong capability for automatic feature extraction and pattern recognition.

Deep Learning in Fire Analysis

Deep learning has been proven highly effective across a variety of computer vision tasks, from basic image classification and object detection to complex medical diagnostics and face recognition (Choudhary, 2022). In this specific field, these models have begun to outperform traditional methods by offering more precise detection and identification of fire-affected areas. Their real strength lies in their ability to learn spatial features, making them particularly effective for identifying complex fire patterns in images.

Existing Models and Research Gap

While the U-Net architecture has demonstrated excellent performance in medical image segmentation (Azad, 2022), it has also shown promise for fire segmentation tasks. Existing deep learning models for wildfire analysis, such as the FU-NetCast V2, often prioritize post-ignition spread prediction and have achieved high accuracy rates of 94.6% (Khennou, 2023). However, these models generally focus on forecasting rather than the critical need for immediate, accurate pixel-level segmentation of fire regions in image, which is essential for early response efforts.

Study Objective

The primary goal of this study is to develop a reliable deep learning pipeline for fire segmentation using RGB images. By using a U-Net architecture to perform pixel-level segmentation, this research seeks to accurately identify fire-affected regions across diverse visual scenes. Ultimately, the aim is to evaluate the effectiveness of U-Net for fire segmentation and provide a foundation for future research exploring its application to wildfire monitoring.

MATERIAL AND METHODS

This study was conducted by building and evaluating a deep learning model for fire detection using RGB imagery. The process involved collecting a dataset of labeled fire images, preparing the data for training, designing a U-Net architecture, and training the model to segment fire-affected regions. To measure how well the model performed, the following evaluation metrics were used: Dice coefficient, Intersection over Union (IoU), precision and recall. These metrics helped measure how accurately the model could identify fire-affected regions in new images.

2.1 Dataset

The dataset used in this study was obtained from the “Fire Segmentation Image Dataset” published on Kaggle by DiversisAI (DiversisAI, 2021). Images were categorized into two categories: Fire and Not Fire images, examples of such images are seen in Figure 1. Each image is paired with a corresponding binary segmentation mask that labels fire pixels. The dataset has 11,100 images and 11,100 segmentation masks.

Each image was first matched with its corresponding binary segmentation mask using consistent file identifiers provided in the original dataset of Kaggle. To prevent data leakage, the dataset was first partitioned into training and test sets at the image-mask pair level prior to any preprocessing steps. The split was performed using a fixed random seed of 42 to ensure reproducibility and the test set was kept completely unseen during all stages of training, hyperparameter tuning, and preprocessing decisions. After the split, all images were resized to 256 × 256 pixels to ensure uniform input dimensions for the neural network.

The dataset was then split into:

Table 1. Dataset Subsets (Training Set/Test Set)

Dataset Subset	Number of Images	Number of Segmentation Masks
Training Set	8,880	8,880
Test Set	2,220	2,220
Total	11,100	11,100

Figure 1. Depicts a sample of the Kaggle dataset showing representative “fire” and “not fire” images from the dataset.

2.2 Data Preparation

U-Net is a deep convolutional neural network that has been widely used for image segmentation tasks, particularly in medical imaging (Azad, 2022). Unlike many traditional deep learning models that require large amounts of training data, U-Net can achieve strong performance with relatively small datasets. The architecture consists of two stages: an encoder and a decoder. The encoder extracts increasingly complex features from the input image through a series of convolution and pooling operations, while the decoder reconstructs spatial information to produce a segmented output. The network contains nine blocks: four encoder blocks, four decoder blocks, and one shared bottleneck block. Each block consists of two 2D convolutional layers with 3×3 kernels and rectified linear unit (ReLU) activation functions, followed by 2D max-pooling layers in the encoder. The number of feature channels doubles at each down-sampling stage, allowing the model to learn increasingly detailed representations of image features. Finally, a 1×1 convolutional layer generates a binary segmentation mask, a pixel level classification map in which each pixel is assigned a value of 1 if it belongs to the target object and 0 if it belongs to the background. This mask enables the model to identify the location and boundaries of the object of interest within the image (Ronneberger, Fischer & Brox, 2015) .

2.3 U-Net Model Architecture

The model uses a set of input images and their corresponding binary segmentation masks. During training and based on the binary mask as the desired output, the model learns how to classify each pixel of the images into the different object labels. For this task, the model classifies each pixel into one of two classes: fire or background (non-fire).

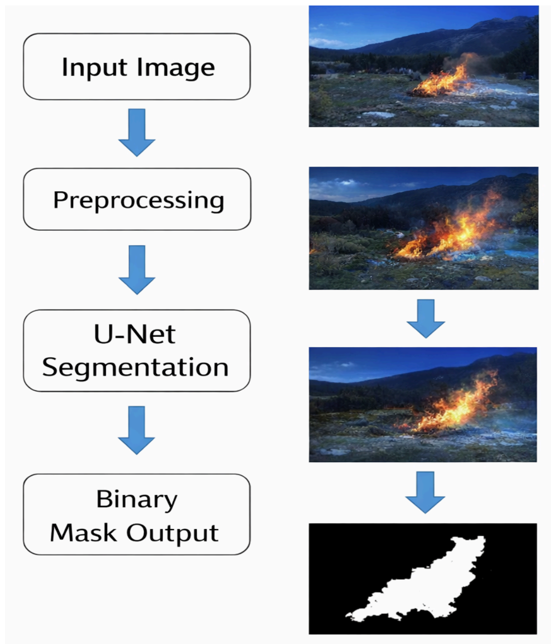

Figure 2. Proposed Architecture

2.4 Model Training

The U-Net model was implemented in Python using PyTorch and trained on an NVIDIA T4 GPU via Google Colab. To ensure reproducibility, a fixed random seed of 42 was set for all stochastic processes, including dataset shuffling and weight initialization. The input images were resized to 256×256 pixels and normalized using standard ImageNet mean [0.485,0.456,0.406] and standard deviation [0.229,0.224,0.225] to stabilize gradient descent.

The model was optimized using the Adam optimizer with a learning rate of 1×10^{-3} and a weight decay of 0 (Kingma & Ba, 2015). A composite loss function consisting of Binary Cross-Entropy (BCE) and Dice Loss was employed to balance pixel-level accuracy with spatial overlap, effectively addressing the class imbalance where fire pixels are rare relative to background pixels (Milletari et al., 2016). During inference, a threshold of 0.5 was applied to the model's output probabilities to convert continuous logits into binary segmentation masks. The model was trained with a batch size of 8 for 10 epochs for the first three tests. During each epoch, the model performed forward propagation on batches of images, calculated the composite loss, and updated network weights via backpropagation. To ensure the model learned to recognize fire patterns rather than memorizing the training data, the test set was held out entirely from the parameter-updating process.

2.5 Evaluation Metrics

To evaluate whether the model is operating as expected and how well it's trained, for unique problems, evaluation metrics are necessary (Akhyar, 2026). These metrics were found through extensive research on deep learning methodologies, specifically focusing on basic standards for binary image classification tasks such as wildfire detection (Cook, 2018).

Model performance was evaluated using four segmentation metrics:

● Dice Coefficient- It measures how well the predicted fire area matches the true fire area. It is calculated by the following formula

$$Dice = \frac{2TP}{2TP+FP+FN}$$

Where TP: True Positive, FP: False Positive and FN: False Negative.

● Intersection over Union- IoU measures the ratio of the intersection between predicted and ground truth masks to their union.

$$IoU = \frac{TP}{TP+FP+FN}$$

Where TP: True Positive, FP: False Positive and FN: False Negative.

● Recall- Recall measures the proportion of actual fire pixels that are correctly identified by the model.

$$Recall = \frac{TP}{TP+FN}$$

Where TP: True Positive and FN: False Negative.

● Precision- Precision measures the proportion of predicted fire pixels that are correct.

$$Precision = \frac{TP}{TP+FP}$$

Where TP: True Positive and FP: False Positive.

2.6 Implementation and Testing

To provide a consistent benchmark for comparing the experimental variations, a fixed test set of 2,220 images was utilized. This test set was held out from the training process and used to assess the generalization performance of all configurations (Baseline, 50% subset, 25% subset, and increased epochs).

Two primary variables were tested: the training dataset size and length of time influenced the model's performance. These factors determine how well the network captures the key visual features in the images and how reliably it can identify fire patterns in data it hasn't seen before.

In order to assess how robustly the model generalized to new data, Table 2 shows the experiments that were performed, each varying key training parameters. A total of three experimental variants were examined in addition to the baseline. First, the training dataset size was reduced to 50%, then further reduced to 25% of the original size. In both cases, the remaining training parameters remained unchanged. Machine learning algorithms typically perform best when they are provided with vast amounts of data. Reducing the sample size increases variability and creates bias in the results produced by the algorithm.

Second, the number of training epochs was increased from 10 to 20 while keeping the full training dataset. With more training iterations available (more epochs), there is a greater opportunity for the model to continue learning its features and improving its prediction accuracy. When additional training iterations allow for lower training errors and improved segmentation accuracy to occur, there is also a risk that excessive numbers of training epochs may result in overfitting occurring which means that the model begins to memorize the patterns of the training data instead of applying those learned features to predict values associated with new, unseen images (Bashir et al., 2020). Thus, assessing performance at various epoch levels is important for identifying a suitable compromise between not providing sufficient training to the model (under-training) and providing too much training to the model (over-training).

Table 2. Experimental Configurations

Test	**Description**	**# Training Images**	**# Test Images**	**Training %**	**Epochs**	**Batch Size**
Test 1	Base Model	8,880	2,220	80%	10	8
Test 2	Training size reduced to 50%	4,440	2,220	40%	10	8
Test 3	Training size reduced to 25%	2,220	2,220	20%	10	8
Test 4	Training epochs increased from 10 to 20	8,880	2,220	80%	20	8

RESULTS

Each graph in Figures 4a-4d corresponds to a different test condition, including variations in dataset size (100%, 50%, and 25%) and training duration (10 vs. 20 epochs). The x-axis represents training epochs, while the y-axis represents the value of each evaluation metric.

In the baseline configuration (Figure 4a), the model was trained on the full dataset (8,880 training images) for 10 epochs. The Dice score increased to 0.72 and IoU reached 0.58, while Precision and Recall remained above 0.73, indicating strong fire segmentation performance. When the training dataset was reduced to 50% (4,440 images) in Figure 4b, performance showed a slight decline, with Dice decreasing to 0.70 and IoU to 0.55. This indicates that reducing the amount of training data limits the model's ability to learn discriminatory fire features.

Further reduction of the dataset to 25% (2,220 images), as shown in Figure 4c led to a more noticeable performance drop, with Dice reaching 0.67 and IoU 0.53. This trend shows a consistent relationship between dataset size and segmentation quality: as training data decreases, the model becomes less effective at accurately segmenting fire regions due to the model seeing fewer examples of different wildfire shapes, sizes, and conditions.

However, increasing the number of training epochs from 10 to 20 while keeping the full dataset (Figure 4d) resulted in improved performance, with Dice increasing to 0.73 and IoU reaching 0.59. This suggests that additional training iterations allow the model to better understand and improve segmentation precision. However, the improvement is relatively small compared to the gains from increasing dataset size, indicating that data availability has a stronger impact on performance than training duration.

Overall, the results show two clear trends: increasing dataset size consistently improves model performance, while increasing epochs provides marginal gains after sufficient training.

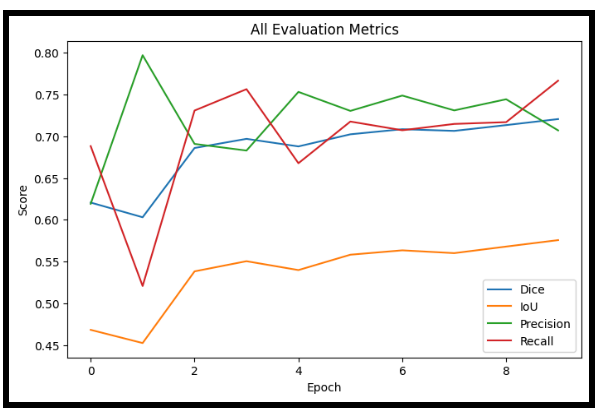

Figure 4a. Baseline Model Testing with all Evaluation Metrics

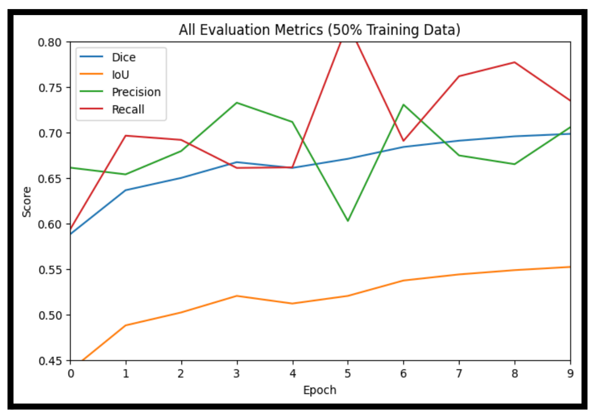

Figure 4b. Training size reduced to 50%

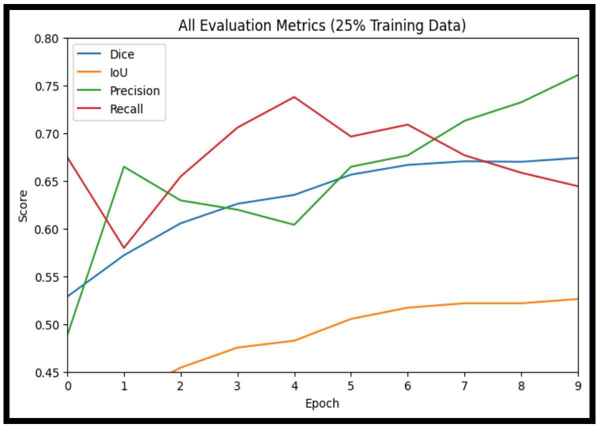

Figure 4c. Training size reduced to 25%

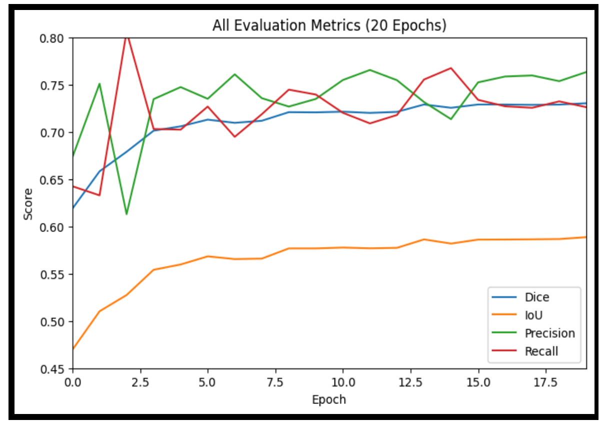

Figure 4d. Training epochs increased from 10 to 20

Table 3 presents the final performance evaluation metrics achieved by the model on the test set. While Figures 4a-4d illustrate the evolution of these metrics across training epochs, the values in Table 3 represent the model's performance at the conclusion of the training process.

Table 3. Experimental Results of U-Net

Model	Dice Coefficient	IoU	Precision	Recall
U-Net	0.72	0.58	0.71	0.77

The evaluation results demonstrate that the U-Net model achieves robust segmentation performance, with a final Dice coefficient of 0.72 and an IoU of 0.58 on the test set. These metrics confirm the model's ability to identify fire-affected regions in unseen imagery, providing a reliable baseline for future research in wildfire monitoring.

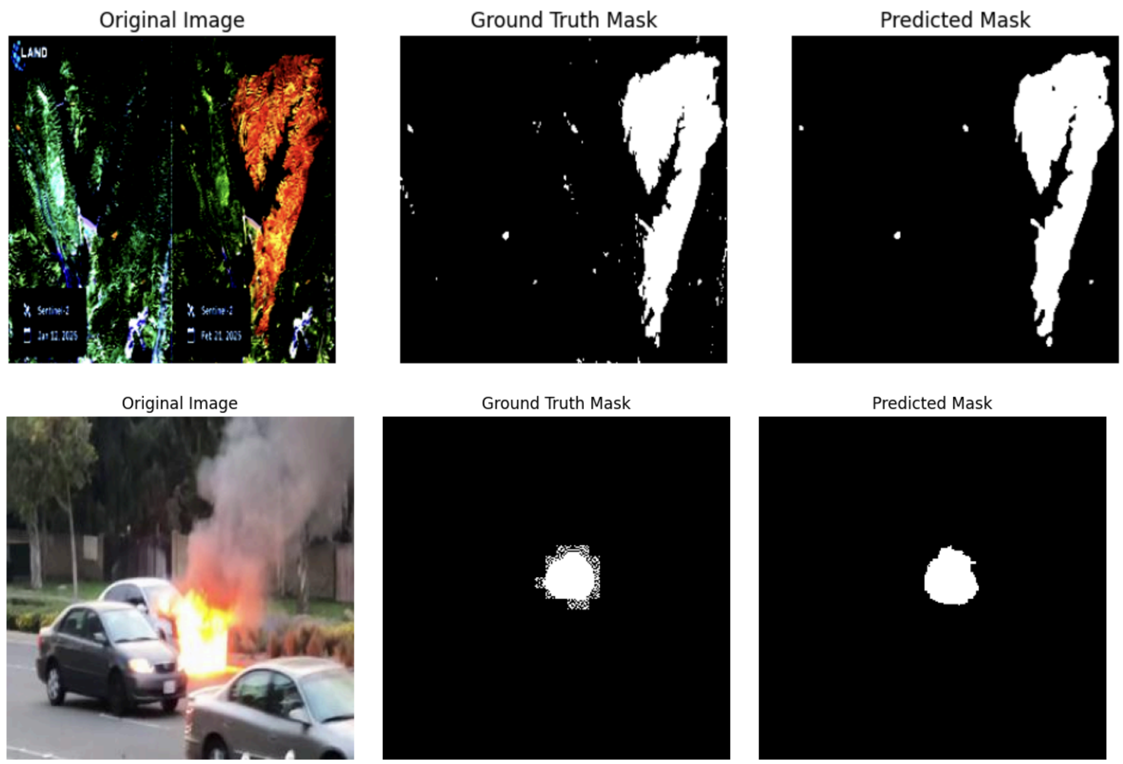

Figure 5. U-Net Results

CONCLUSION

The proposed U-Net model demonstrates robust performance in fire segmentation from RGB imagery, achieving a final Dice coefficient of 0.72, an IoU of 0.58, and a recall of 0.77 on the test set. These results confirm the model's ability to identify fire-affected areas across diverse visual scenes. The analysis presented in Figures 4a-4d highlights the influence of two primary factors: dataset size and training duration on the model's segmentation capability.

Experiments using the full (100%) dataset consistently yielded the highest Dice and IoU scores. Conversely, models trained on reduced 50% and 25% datasets exhibited a decline in performance. This trend indicates that larger, more diverse datasets allow the model to better generalize by learning discriminatory features across a wider variety of fire shapes and environmental conditions.

The comparison between training durations also shows that increasing epochs from 10 to 20 improved segmentation metrics. While this suggests that additional iterations allow the model to refine its learned features, the marginal gains observed indicate that data availability, rather than training duration, is the more significant driver of model performance. This finding aligns with the risk of overfitting, where excessive training time could lead the model to memorize training data rather than improving its predictive accuracy on unseen images (Roelofs, 2019).

Traditional fire detection methods such as color-based approaches rely on manually engineered features and fixed thresholds. For example, the method proposed by Jamali et al. uses saliency detection, HSV color filtering, and texture analysis (LBP-TOP) combined with an SVM classifier to detect fire regions (Jamali et al., 2019). While this approach can perform well under controlled conditions, it depends heavily on hand crafted rules and multiple processing stages. In contrast, the proposed U-Net model performs end-to-end learning and automatically extracts hierarchical features directly from data for pixel-level segmentation. This allows the model to better generalize across different fire shapes, lighting conditions, and backgrounds without relying on manually tuned thresholds or separate feature extraction steps.

One limitation of this study is that the dataset used does not consist exclusively of satellite imagery. Instead, it contains a mixture of fire and non-fire RGB images, meaning the model was evaluated for fire segmentation rather than satellite-based wildfire detection. Thus, the results should not be interpreted as demonstrating performance on satellite imagery or real-time wildfire monitoring. In addition, the dataset does not provide information about atmospheric conditions, smoke density, weather factors, or sensor characteristics that may influence model performance in real-world environments. Future work could evaluate the proposed model on satellite imagery and real-time wildfire datasets to assess its applicability to wildfire monitoring. The model could also be expanded to classify different levels of fire intensity.

REFERENCES

Akhyar, A., Lee, J., Shahar, N., et al. (2025). Automated forest fire detection in ecological monitoring using enhanced deep learning networks. Sci Rep 16, 2039. https://doi.org/10.1038/s41598-025-31707-6

Azad, R., Aghdam, E. K., Rauland, A., Jia, Y., Avval, A. H., Bozorgpour, A., Karimijafarbigloo, S., Cohen, J. P., Adeli, E., & Merhof, D. (2022). Medical Image Segmentation Review: *The success of U-Net. arXiv*. https://arxiv.org/abs/2211.14830

Bashir, D., Montañez, G. D., Sehra, S., Segura, P. S., & Lauw, J. (2020). *An information-theoretic perspective on overfitting and underfitting*. arXiv preprint arXiv:2010.06076. https://doi.org/10.48550/arXiv.2010.06076

Choudhary, K., DeCost, B., Chen, C., Jain, A., Tavazza, F., Cohn, R., Park, C. W., Choudhary, A., Agrawal, A., Billinge, S. J. L., Holm, E., Ong, S. P., & Wolverton, C. (2022). Recent advances and applications of deep learning methods in materials science. *npj Computational Materials*, *8*(1). https://doi.org/10.1038/s41524-022-00734-6

Cook, J. D. (2018). *Accuracy, precision, and recall*. John D. Cook. https://www.johndcook.com/blog/2018/09/06/accuracy-precision-and-recall/

Cui, F. (2020). Deployment and integration of smart sensors with IoT devices detecting fire disasters in huge forest environments. *Computer Communications*, *150*, 818–827. https://doi.org/10.1016/j.comcom.2019.11.051

DIVERSISai. (2022). *Fire Segmentation Image Dataset*. Kaggle.com. https://www.kaggle.com/datasets/diversisai/fire-segmentation-image-dataset

Jamali, M., Karimi, N., & Samavi, S. (2019). Saliency-based fire detection using texture and color features. *arXiv preprint arXiv:1912.10059*. https://arxiv.org/abs/1912.10059

Khennou, F., & Akhloufi, M. (2023). Improving wildland fire spread prediction using deep U-Nets. *Science of Remote Sensing*, *8*, 100101–100101. https://doi.org/10.1016/j.srs.2023.100101

Kingma, D. P., & Ba, J. (2017). Adam: *A method for stochastic optimization*. arXiv preprint arXiv:1412.6980. https://doi.org/10.48550/arXiv.1412.6980

Milletari, F., Navab, N., & Ahmadi, S.-A. (2016). V-Net: *Fully convolutional neural networks for volumetric medical image segmentation*. arXiv preprint arXiv:1606.04797. https://doi.org/10.48550/arXiv.1606.04797

Musabyimana, J. D. (2022). *T.I.P.F.F.-Life*. In MIT Solve. https://solve.mit.edu/solutions/59182

PreventionWeb. (2021). Wildfire. https://www.preventionweb.net/hazards/wildfire

Roelofs, R., Fridovich-Keil, S., Miller, J., Shankar, V., Hardt, M., Recht, B., & Schmidt, L. (2019). *A Meta-Analysis of Overfitting in Machine Learning*. NeurIPS 2019. https://papers.neurips.cc/paper_files/paper/2019/file/ee39e503b6bedf0c98c388b7e8589aca-Paper.pdf

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. https://arxiv.org/abs/1505.04597