Wildfire Risk Assessment

Classifying Remote Sensing Images for Wildfire Risk

Kori Thompson

In recent decades wildfires have become increasingly destructive natural disasters that have resulted in massive property damage, injuries, and fatalities. These disasters have also wreaked havoc on forest ecosystems. To mitigate these consequences, it is imperative to prevent wildfires. One key step in preventing wildfires is to identify wildfire hazards and risks as much as possible. This allows for the proper responses to be taken to reduce these risks. Fire risk assessments are time-consuming processes that require both the identification and evaluation of the potential risks. While methods using satellite data such as surface temperature, moisture levels, and other geoscientific data exist, these still require specialized knowledge. Introducing remote sensing imagery and satellite imagery into the field can help in mapping geographical features to these measurements and reduce the specialized knowledge bottleneck.

This analysis seeks to examine whether remote sensing imagery can be used to classify land areas into fire risk levels. The FireRisk dataset, a collection of remote sensing imagery of different classifications of wildfire risks, was used for the analysis. By utilizing computer vision techniques rather than other machine learning methods, it may be possible to create a simplified method for assessing fire risks. Due to the image resolutions and the amount of detail captured in the remote sensing imagery, deep learning models, in particular convolutional neural networks (CNN), can be of use. CNNs allow the model to learn features from raw pixel data unlike other machine learning models. These models can extract patterns from the pixel data which can be recognized anywhere in the image, allowing for greater ability to classify images. These models require less interpretation and less specialized knowledge than other machine learning models, making them more streamlined.

In this work, the FireRisk dataset was used to construct CNNS to classify images into 7 fire risk levels. The dataset itself was highly imbalanced, with a bias toward very low fire risk and non-burnable areas. Three baseline models were built and compared, with the best performing baseline model being advanced for additional fine tuning. The best performing baseline model used a ResNet50 backbone, pre-trained on the ImageNet dataset. After selecting the backbone, a new model was constructed to improve on the baseline model. Once the architecture was finalized, the model’s hyperparameters were tuned to further improve performance.

The analysis concludes that it may be possible to use remote sensing imagery to assess areas for fire risk. However, due to the imbalanced nature of the dataset, the model was not effective at detecting risks. Rather, it was highly effective at identifying non-burnable areas, water, or very low risk fire areas. Additional testing with a more balanced dataset may provide better results and increase the potential use of the model. The final model achieved a validation accuracy of 60.13%. After further examination of the F1 scores, this accuracy was found to be due to the high bias towards non-burnable, water, or very low fire risk classes. The model struggled to correctly identify low to very high-level fire risks. Overall, the use of remote sensing imagery for fire risk assessment may have some potential. However, relying only on remote sensing imagery may not be the most successful of methods on its own. Instead, it could help to augment other methods rather than stand on its own.

Data

The dataset used for this analysis is the FireRisk dataset from Shuchang Shen, Sachith Seneviratn, Xinye Wanyan, and Michael Kirley. It was proposed as a benchmark dataset for remote wildfire risk assessment. The dataset consisted of 91,872 labeled high-resolution remote sensing images collected using the National Agriculture Imagery Program (NAIP). Images were labeled using the fire risk classes supplied by the Wildfire Hazard Potential (WHP) raster dataset. The 7 fire risk classes are water, non-burnable, very low, low, moderate, high, and very high risk. The dataset was pre-split into training and validation datasets with 70,331 images in the training split and 21,541 images in the validation split. Each image was resized to 256 x 256 pixels to ensure all images were the same size for modeling.

The classes in the dataset were highly imbalanced with a bias toward lower fire risk levels. From the training dataset, 31% of all images were classified as very low fire risk and 25% of images were classified as non-burnable. Only 9% of the training images were classified as high fire risk and 4.6% were classified as very high. This is likely due to the lack of available images of moderate to very high fire risk areas and the abundance of images of very low fire risk areas. While it is generally desirable to have balanced classes, this does not reflect the reality of the situation in the real-world. There are generally fewer high and very high-risk areas compared to non-burnable or very low fire risk areas. Considering this, no efforts were made to balance the dataset such as over-sampling underrepresented classes or under-sampling the overrepresented classes.

Method

Baseline Models

Three models were constructed as baseline models. The first model used a pre-trained ResNet50 base. The ResNet base was chosen as the original paper also used a ResNet base for their baseline model. The second model used a pre-trained InceptionResNetV2 base, chosen as it was an improvement on the ResNet architecture. The third baseline model used a pretrained EfficientNetV2S base. All models used the pre-determined weights from training the models on the ImageNet dataset. The top layers of the ore-trained models were not used for the baseline models, only the convolutional layers. After the pre-trained base, a global average pooling layer was used to flatten the input to one dimension. This was followed by a dropout layer with a dropout rate of 0.2, a fully connected layer with 1,024 nodes, and another dropout layer. A fully connected layer with 7 nodes and a softmax activation was used as the output layer. The output produces an array of percentages for the image belonging to each class. The baseline models were each trained for 15 epochs and evaluated on both validation loss and accuracy.

The best performing baseline model used the ResNet50 backbone. It achieved both the highest validation accuracy of 58.02% and the lowest validation loss of 1.1128. The validation accuracy may be misleading as the classes are highly imbalanced. However, with the lowest validation loss, the ResNet backbone was deemed to be the best performer. While training the baseline model, one interesting trend of note was that both validation accuracy and validation loss tended to outperform their test counterparts. This may have been due to the validation accuracy and loss being calculated at the end of all batches while training accuracy and loss are averaged across the batches. Another reason for this trend may have been the higher concentration of very low-class instances in the validation set. If the model was defaulting to the most prevalent class and there was a higher concentration of the very low fire risk class in the validation set, it stands to reason that the accuracy would be higher and the loss lower.

Another possible explanation for the validation metrics outperforming the test metrics could also have been that the models were potentially under-capacity. This would have prevented the model training data from learning the data as well as it needed to. Similarly, all the baseline models had ragged trends in the validation accuracy and loss. This implies that none of the baseline models consistently generalized well to the data. The lack of consistent generalization by the models may suggest that the models were either over or underfitting the data, further supporting the idea that the models may have been under capacity.

Building the Final Model

A ResNet50 backbone was used for the final model as it performed best in the baseline testing. Additional fully connected layers and dropout layers were added to the final structure of the model. As previously noted, there was a possibility that the model was under-capacity due to the validation accuracy and loss outperforming the testing accuracy and loss. The additional fully connected layers were used to help reduce this possibility. While determining the final model architecture, the model was tested with two and three fully connected layers on top of the ResNet backbone. Dropout layers were added before and after each fully connected layer to help with regularization and to attempt to prevent overfitting.

Of these architectures, the model with two fully connected layers on top of the backbone performed the best. It achieved the highest validation accuracy, 59.7%, and lowest validation loss, 1.1064, compared to the baseline model and the three-layer model. As such, two fully connected layers were utilized in the final model architecture. An additional model was tested with two fully connected layers and no dropout layers. This proved to be the best performing model in terms of accuracy with a validation accuracy of 60.69%. However, the model appeared to quickly overfit with the lowest validation loss occurring in the first epoch. Similarly, with the highly imbalanced data, accuracy is less reliable and does not necessarily indicate a good overall performance. As such, the decision was made to keep the dropout layers in the final model architecture.

Fine-Tuning the Model

To fine-tune the final model, a hyperband search was used to find the best hyperparameters. Since the classes in the dataset are imbalanced, the model was optimized for validation loss rather than validation accuracy. Since loss measures the distance between the true value and the prediction, this was a better indicator of the model's overall performance than accuracy, which simply measures how often the prediction is correct. The dropout rate was tuned to change the amount of regularization used in the model. Each fully connected layer had its neurons tuned to find the best combination of neurons. The search space started at 512 and increased by 512 to 2,048 neurons. The idea was to find the number of neurons that minimized the loss of the model while preventing overfitting. Finally, the learning rate was tuned to assist the model in converging on optimal values for weights. By tuning the learning rate, the model may converge on the optimal solutions sooner and hopefully avoid overshooting the optimal solution or getting stuck at a sub-optimal solution. The tuned model had a dropout rate of 0.25, 512 neurons in both fully connected layers, and a learning rate of 0.00028948.

The tuned model achieved a validation accuracy of 0.6013, which was about four percentage points lower than that achieved by the authors of the FireRisk dataset with their ResNet model. This was expected as the final model architecture was selected based on minimizing the validation loss rather than maximizing validation accuracy. The authors did not appear to address the class imbalance in their assessment of the data either. While better than random chance, the accuracy of the model was misleading as shown by the F1 scores for each class. Images of water had the highest F1 score of 0.8545, likely because the images are distinctive from the other classes. The non-burnable class also had a high F1 score of 0.8253, while the very low risk class had an F1 score of 0.7020. This is likely due to the classes being overrepresented in the data. The high fire risk class had an F1 score of 0.3857, implying that the model did a better job of correctly identifying the class compared to the other fire risk levels. Low and very high fire risks have F1 scores of 0.1368 and 0.1247 respectively, while the moderate fire risk class had an F1 score of 0.0711. Based on these scores, the model struggles to identify low, moderate, and very high fire risk images. It did a moderately better job of identifying very high fire risk areas. Overall, the model was biased towards the classes more prevalent in the data and towards the highly distinctive classes like water.

In an attempt to make the model more sensitive towards the other less prevalent classes, class weights were added to the model. The class weights are used to weight the loss function for training, increasing the penalty for underrepresented classes. The weights and biases of the model would then be adjusted according to this weighted loss during backpropagation. The idea is that this will increase the sensitivity of the model to the underrepresented classes by penalizing the model for misclassifying instances belonging to those classes. In practice, the addition of the class weights did not significantly improve the performance of the model. The model did not reach as low of a validation loss as it did when not using class weights. Similarly, it achieved a lower validation accuracy than the model that did not use a weighted loss function. Using a weighted loss function did change the sensitivity of the model to certain classes. The F1 score for the high fire risk class decreased from 0.3857 to 0.2952. Similarly, the non-burnable class, very low class, and the water classes all saw reductions in their F1 scores. On the other hand, the model did become better at classifying very high-risk fire areas with its F1 score increasing to 0.3567 from 0.1247. The low class also saw a marked improvement in its F1 score, increasing from 0.1368 to 0.3495. The moderate class also saw a slight improvement in its F1 score as well. Overall, using the class weights did not improve the performance of the model, but did somewhat improve its ability to correctly classify low to very high-risk land areas.

Conclusion

This analysis sought to investigate whether it was possible to use remote sensing imagery to simplify the process of fire risk assessments. To do this, the FireRisk dataset was used to train a convolutional neural network to classify images into 7 fire risk levels. Three baseline models were constructed and trained to find the best base architecture for the model. From these models, the best base was found to be a ResNet50 base. Due to validation accuracy and loss consistently outperforming training loss and validation, it was assumed that the baseline model may have been undercapacity. An additional fully connected layer surrounded by dropout layers was added to the model architecture. This improved the performance, increasing validation accuracy and decreasing validation loss. The final model was then tuned for several hyperparameters using a hyperband search further improving performance. A weighted loss function was introduced to the tuned model to try and counteract the class imbalance in the dataset. While this did not improve the performance of the model in terms of loss or accuracy, it did slightly improve the model’s ability to correctly predict underrepresented fire risk classes.

Overall, the analysis found that the use of remote sensing imagery was not sufficient for automated fire risk assessment. Some promise was shown during the analysis and perhaps with a more balanced dataset the idea could become more feasible. The final tuned model performed modestly well with the highest validation accuracy of 60.13%. However, this accuracy was due to a high bias towards the non-burnable, water, and very low risk classes which were overrepresented in the data. It is likely that the model learned the patterns in these images the best and predicted these classes more frequently. The model struggled to correctly classify low to very high fire risk land areas. This essentially makes the model useless for performing a fire risk assessment as it could not reliably predict areas at risk for wildfires. Efforts to correct this bias were made by introducing a weighted loss function during model training. These efforts showed some promise as it did make the model slightly more sensitive to the underrepresented fire risk classes. However, the model’s performance suffered with validation accuracy falling to 52.15% and validation loss increasing. Even with the improvement in sensitivity towards the other fire risk classes, the model still struggled to identify areas at risk for wildfires and remained biased towards the overrepresented classes. As such, the model was deemed to not be of use for automating fire risk assessments.

Additional improvements to the model may change this and make the use of remote sensing imagery for fire risk assessment more viable. These changes include setting the initial biases of the output layer according to the prevalence of classes in the data. This provides the model with a better starting point for training and can help to mitigate class imbalance. This was not carried out for this analysis due to time constraints. Another improvement would be to either upsample the underrepresented classes using data augmentation or to downsample the overrepresented classes. By balancing the data, the idea is that the model will give equal weight to all. This could improve the ability of the model to correctly identify areas at risk for wildfires rather than finding areas not at risk for wildfires. Again, this was not considered for this analysis due to time and computing constraints.

While the use of remote sensing imagery alone does not provide satisfactory results, it could be enhanced using additional data. Some areas for future research may be the inclusion of satellite data such as air temperature or moisture levels in the model. This would provide additional context to the model that may help in deciphering the nuances between similar fire risk classes like high and very high risk or very low and low risk. Another area of research to consider is a time series of remote sensing imagery to see how the land has changed to help identify fire risk areas. Examining how the land has changed over time may reveal patterns that are not currently accessible with the given dataset. This may also be useful in identifying areas that are not currently at risk for wildfires but may become at risk.

Citations

@misc{shen2023fireriskremotesensingdataset,
title={FireRisk: A Remote Sensing Dataset for Fire Risk Assessment with Benchmarks Using Supervised and Self-supervised Learning},
author={Shuchang Shen and Sachith Seneviratne and Xinye Wanyan and Michael Kirley},
year={2023},
eprint={2303.07035},
archivePrefix={arXiv},
primaryClass={cs.CV},
url={https://arxiv.org/abs/2303.07035},
}