Wildfire Spread Prediction in North America Using Satellite Imagery and Vision Transformer

Bronte Sihan Li, Ryan Rad
Northeastern University
Khoury College of Computer Science
{li.siha, r.rad}@northeastern.edu

***Abstract*—The recent surge in wildfire incidents, exacerbated by climate change, demands effective prediction and monitoring solutions. This study introduces a novel deep learning model, the Attention Swin U-net with Focal Modulation (ASUFM), designed to predict wildfire spread in North America using large-scale remote sensing data. The ASUFM model integrates spatial attention and focal modulation into the Swin U-net architecture, significantly enhancing predictive performance on the Next Day Wildfire Spread (NDWS) benchmark. Additionally, we have developed an expanded dataset that encompasses wildfire data across North America from 2012 to 2023, providing a more extensive basis for model training and validation. Our approach demonstrates state-of-the-art performance in wildfire spread prediction, offering a promising tool for risk mitigation and resource allocation in wildfire management. Related code is in our public repository https://github.com/bronteee/fire-asufm**

I. Introduction

Climate change has been correlated with a significant increase in the occurrence of wildfires as well as global forest loss, resulting in a devastating 42-44 percent loss in North America alone between 2001 and 2019 [1] and in 2022 alone over 6.6 million hectares of tree cover was lost [2]. Studies have found factors such as increased fuel acridity and flammability [3] and elevated levels of Persistent Positive Anomalies (PPAs) [4] to be contributing to fire ignitions with implications of a further upward trend as temperatures continue to rise globally. The detrimental impact is not limited to economic cost but also physical and mental health-related, disproportionately affecting elderly and low-income populations [5].

As a result of these often catastrophic events becoming more frequent, wildfire prediction and modeling have become crucial in risk mitigation and targeted resource allocation, providing key insights on fire behavior to allow for early planning efforts [6]. Simulation of wildfire spread from remote sensing data in particular offers an advantage in real-world applications due to its accessibility and extensibility, with potential utility in the timely mapping of wildfire activity to aid in emergency response and management [7].

Responding to the call for action by [8] to leverage machine learning in combating climate change challenges, a plethora of tools, datasets, and techniques have emerged, enhancing innovation in environmental science. For fire detection and prediction, deep learning is a promising approach due to the abundance of high-performing network architectures and efficiency in solving complex problems with large feature spaces such as the remote sensing data from satellite imagery. While deep learning models have generally achieved great performance on fire detection tasks with large-scale data, results from spread prediction efforts, however, have been highly inconsistent across dimensions including temporality, region, and size of the data used, performance metrics, and formulation of the problem [9]–[12]. This reflects the difficulty of the task as well as the need for a broader search for a generalized method capable of informing decisions across areas on a continental scale - we use vision transformers in a U-shaped formation combined with a more comprehensive data set to tackle this problem.

In this study, we tailor various state-of-the-art segmentation methodologies for large-scale wildfire spread modeling, utilizing remote sensing data. Our approach involves a comprehensive examination of input features, loss functions, deep network architectures, and mechanisms for attention and representation modulation. To evaluate the generalizability of our model, we have extended the scope of the Next Day Wildfire Spread (NDWS) dataset. Originally, NDWS encompassed data from 2012 to 2020 within the United States [7]. We have expanded this dataset by including fire incidents from 2012 to 2023 across North America. This expansion has nearly doubled the dataset's size. Building on our analyses, we introduce a wildfire prediction model based on a symmetrical encoder-decoder architecture, employing Swin-Unet with spatial attention and focal modulation [13]–[15]. This model is designed to predict fire locations for t+1 day.

In this research, we make the following contributions:

1) To our knowledge, this is the first study that enhances performance in large-scale, next-day wildfire prediction using remote sensing data. Our evaluation encompasses both the original benchmark dataset and our extended dataset.
2) We introduce and implement a transformer-based, UNet-like model, incorporating spatial attention and focal modulation. This model demonstrably surpasses the existing top-performing models on the NDWS dataset.
3) We present an expanded version of the NDWS dataset, which now includes data from 2012 to 2023 across North America, providing a more comprehensive resource for wildfire prediction.

979-8-3503-5409-6/24/$31.00 ©2024 IEEE
DOI 10.1109/CAI59869.2024.00278

II. Related Work

A. Modeling wildfire using remote-sensing data

Remote-sensing data is ideal for fire monitoring and early detection, as a result, there have been multiple efforts in developing computational models from satellite and geospatial data for both real-time burn identification and prediction [11], [16]–[18]. In general, these data-driven proposals select a relatively small group of fire samples from a specific region for training and testing, which have produced higher accuracy performance. [19] compares homogeneous vs. heterogeneous landscapes whose performance on heterogeneous was much worse, speaking to the complexity of wildfire spread prediction. The NDWS data set [7] is a large multivariate aggregation of 18545 fire events in the United States spanning over 10 years along with 11 other relevant features including elevation, wind direction and speed, minimum and maximum temperatures, humidity, precipitation, drought index, vegetation, population density, and Energy Release Component (ERC), at a spatial resolution of 1km per pixel. By combining 2D fire data with variables like topographical information and human activity representation which have been attributed to wildfires [20], it provides rich features that can be explored with deep learning approaches. Given the similarity in data complexity and imbalance, we take inspiration from medical segmentation techniques in devising a solution - the next section discusses the specifics.

B. Previous related work on the NDWS data set

In [10], the study uses ROC-AUC instead which is not an ideal evaluation metric given the overwhelming majority of sample pixels being negative, furthermore, shortening the time interval between previous fire down to 30 minutes from one day significantly reduces the future utility of the approach. [11] introduces a CNN-based method similar to that of the original data release, however, only a small subset of the data was used for training as well as testing. In the work of [21], variations of U-net with attention were analyzed and the study prioritized training efficiency.

C. Deep Learning approaches for wildfire spread prediction

[22] discussed the development of physical models for wildfire spread. Since growing advances in deep learning applications, a variety of neural network-based methods have been explored to simulate fire propagation including Irregular Graph Network (IGN) [12], CNN-based approaches [7] [23], and convolutional LSTM [24]. Semantic segmentation in image analysis involves assigning class probabilities to each pixel. For wildfire spread modeling from satellite imagery, this concept is adapted to predict fire presence at future coordinates (x, y), based on current ignition patterns, environmental data, and adjacent locations. Unlike standard RGB images, satellite data requires unique processing due to its distinct band relationships. The U-net architecture, as discussed in [25], offers an efficient U-shaped encoder-decoder structure ideal for segmentation, particularly in delineating fire perimeters. U-net's effectiveness in fire prediction has been well-documented [9], making it a solid baseline for initial experiments. The Swin-Unet model [26], an advancement over U-net, incorporates Swin Transformer blocks [27] for improved performance. Additionally, the integration of spatial attention, as explored in [14], further refines localization and context recognition, enhancing overall model accuracy.

III. Methods

Inspired by previous advances mentioned above, we present Attention Swin U-net with Focal Modulation (ASUFM) where each of the Transformer blocks consists of a normalization layer (LayerNorm), a focal modulation layer followed by a second normalization, and a shifted window multi-head self-attention, while the spatial attention travels along the skip connections between the encoder and decoder blocks to carry an emphasis on higher importance tokens. In combining these mechanisms of weighting, we ensure focus both across input bands and within an image layer such as the fire perimeter. Fig. 2 shows the overall architecture of this approach. In addition, including updated information on wildfire events is crucial to providing and evaluating an expansive solution, to this end, we perform additional GEE data extraction to obtain broader spatial and temporal coverage to deal with the complexity of wildfire spread modeling.

A. Spatial Attention

[17] demonstrates the effectiveness of applying spatially-aware attention around the fire perimeter from previous time steps. To this end, we bring forward spatial attention produced in the encoding pathways [14] to localize on the relevant tokens surrounding the fire, leveraging the skip connections in the U-shaped architecture. This provides further guidance on the model towards a correlation between previous day and next day combustion.

B. Focal Modulation

We apply focal modulation [15] to handle the complexity and heterogeneity of our input features and distinct relationship between bands that requires both spatial and band specificity, allowing for efficient context building around previous day fire coordinates across bands as well as extraction of information from the t day fire layer. From the results, we see that this helps achieve a balanced precision/recall compared to using spatial attention alone. After the tokenization operation q, focal modulation can be described as:

$$y_i = q(x_i) \odot h(\sum_{l=1}^{L+1} g_i^l \cdot z_i^l) \tag{1}$$

where each token x_i goes through a modulation process that produces gating value g_i^l and visual feature z_i^l through the linear layer h.

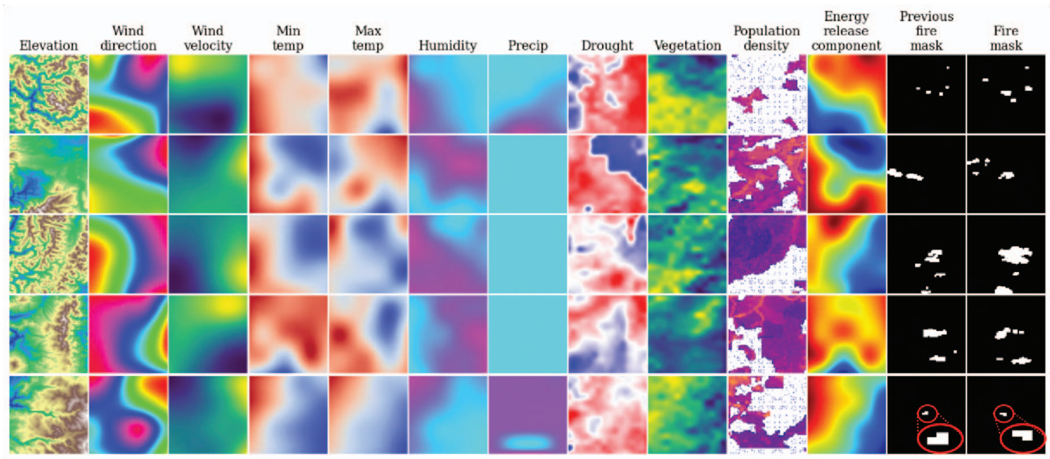

Fig. 1: Visualization of sample input bands and the target fire map.

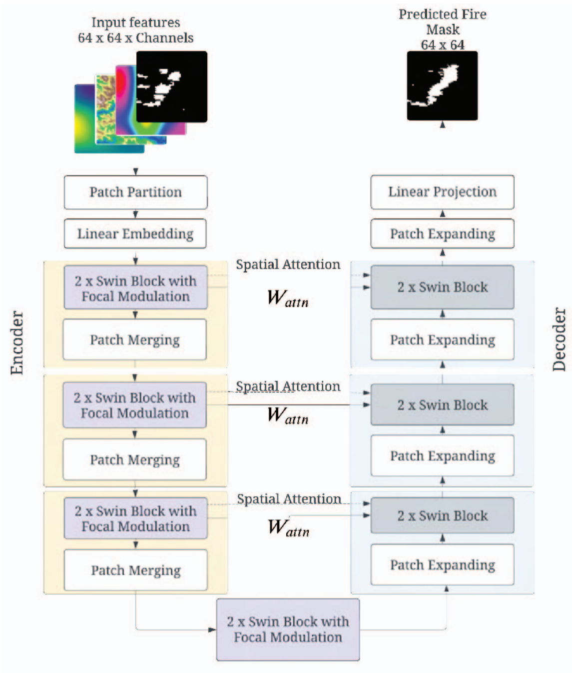

Fig. 2: Overall architecture of ASUFM.

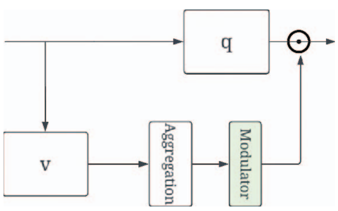

Fig. 3: Focal modulation operations

C. Swin Transformer Block with Focal Modulation

The Swin transformer backbone [28] exchanges the multi-head self-attention (MSA) with a shifted window module to achieve increased granularity in vision tasks, introducing hierarchical and local context building. Evident in Fig 5, Table I and Table. IV, the vanilla Swin configurations tend to over-index on the previous day's fire masks thus sacrificing precision. To address this imbalance, we apply early focal modulation [15] layers as an intermediary between the original input z^{l-1} and subsequent attention layers. This results in the following formulation:

$$\hat{z}_m^{l-1} = (LN(FM(z^{l-1})) \tag{2}$$
$$\hat{z}^l = W\text{-}MSA(LN(\hat{z}_m^{l-1})) + z^{l-1}, \tag{3}$$
$$z^l = MLP(LN(\hat{z}^l)) + \hat{z}^l, \tag{4}$$
$$z_m^l = (LN(FM(z^l)), \tag{5}$$
$$\hat{z}^{l+1} = SW\text{-}MSA(LN(z_m^l)) + z^l, \tag{6}$$
$$z^{l+1} = MLP(LN(\hat{z}^{l+1})) + \hat{z}^{l+1} \tag{7}$$

where $\hat{z}_m^{l-1}$ denotes the output after focal modulation at block $l-1$, which is passed into the $W\text{-}MSA$ and $SW\text{-}MSA$ modules after normalization with LayerNorm(LN). Fig. 4 presents a diagram visualizing a couple of aforementioned blocks.

IV. Experiments

A. Experimental Setup

1) Data Set: The NDWS data set is comprised of 18454 samples containing data sources compiled using Google Earth Engine (GEE) tools, each sample contains 12 layers of 64×64 remote sensing images as input including previous day fire mask, and one layer of the target next day fire mask. We perform data pre-processing using clipping and normalization using values provided from the original data release [7] to reduce the impact of extreme values. In addition, we also experiment with downsizing and random cropping of the

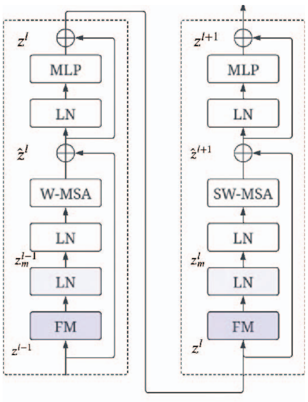

Fig. 4: Two connecting Swin Transformer with focal modulation blocks

features into 32×32 pixels. To expand on the original data, by extracting fires from 2012 to 2023 in coordinates covering North America we reach a total sample size of 51388, a detailed comparison is in Table III and Table II describes each input variable.

2) Implementation and Training Details: Due to the imbalance of the positive (i.e., fire) and negative (i.e., no fire) classes in this data set where most pixels do not contain fire, we discovered that the loss function commonly used in segmentation approaches did not produce desirable results. Taking into account the distinction between this problem and typical segmentation, we explore combinations of loss functions to handle the class imbalance and concluded that weighted binary cross entropy (Eq. 8) with a positive weight of 3.0 produced the best results, where N is the total number of fire samples, y_i is the true label of the i-th sample, $\hat{y}_{ij}$ is the predicted probability of the j-th pixel being in a positive class. w_1 and w_0 are the weights associated with the positive and negative classes, respectively.

$$\text{WBCE} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{P}[w_1 \cdot y_{ij} \cdot \log(\hat{y}_{ij}) + w_0 \cdot (1-y_{ij}) \cdot \log(1-\hat{y}_{ij})] \tag{8}$$

We train each of the models from scratch with an ADAM optimizer with a weight decay of 0.05 on a single NVIDIA A100-80G or a single RTX A6000 GPU. For managing the learning rate, we use the cosine annealing learning rate scheduler with the warm-up, and the batch size is 16. All input features are clipped and normalized, and the number of input features is also explored and selected based on the ablation study presented in [7], which we show the results in the later section.

B. Experiment Results

We present analysis on the Attention Swin-Unet with focal modulation compared to other models in Table I, demonstrating significant improvement in various metrics compared to previous metrics [7] [21]. Keeping in mind the heavy imbalance of positive (i.e., fire) and negative (i.e., no fire) classes, we evaluate the performance of each method with a Dice score (Eq. 9) and PR-AUC (Eq. 10). We are able to increase the state-of-the-art by 31% using the introduced network, furthermore, we verify model generalizability and balanced performance on our extended data set, and the results are in Table IV.

$$DSC(X,Y) = \frac{2 \times |X \cap Y|}{|X| + |Y|} \tag{9}$$

$$\text{PR-AUC} = \int_0^1 \text{Precision}(\text{Recall})\, d(\text{Recall}) \tag{10}$$

TABLE I: Performance comparison of the proposed ASUFM with state-of-the-art techniques on the NDWS dataset

Methods	PR-AUC	Precision	Recall	Dice Score
U-net [25]	0.3081	0.3767	0.4026	0.3824
Attention U-net [29]	0.3217	0.3897	0.4184	0.3973
SU [30]	0.3132	0.3653	0.4230	0.3862
ASU [14]	**0.3522**	**0.3807**	**0.4660**	**0.4132**
FocalU-net [15]	0.2830	0.3645	0.3107	0.3303
RevCol [31]	0.1277	0.1772	**0.4584**	0.2499
ASUFM	**0.3728**	**0.4301**	0.4050	**0.4109**

TABLE II: Input variables in the augmented NDWS dataset

Variable	Min	Max	Mean	Description
elevation	-9.00	3900.00	854.48	Elevation in meters
th	-505875.44	37670.47	75.78	Wind direction in degrees clockwise
vs	-98.15	87.71	3.88	Wind speed in m/s
tmmn	0.0	433.54	281.81	Min temperature in Kelvin
tmmx	0.0	320.75	296.52	Max temperature in Kelvin
sph	-0.01	0.05	0.01	Specific humidity
pr	-137.11	138.29	0.25	Precipitation in mm
pdsi	-152.91	31.36	-1.23	Pressure
NDVI	-32768.0	32767.0	5220.83	Normalized Difference Vegetation Index (NDVI)
population	0.0	221.11	0.21	Estimated number of people in residence
erc	-655.15	762.68	54.59	NFDRS fire danger index energy release component
PrevFireMask	-1.0	1.0	-0.02	Previous day fire mask

TABLE III: The original NDWS vs. our extended NDWS

Data set	Temporal Coverage	Geographical Coverage	Total Size
Original NDWS	2012-2020	United States	18545
Extended NDWS (ours)	2012-2023	North America	31760

TABLE IV: Comparison of different models when tested on our extended NDWS (ours) data set

Methods	Params	Dice Score	Precision	Recall	PR-AUC
U-net [25]	17M	0.3493	0.2945	0.3805	0.2945
Attention U-net [29]	8M	0.3650	0.3721	0.3759	0.3157
SU [30]	27M	0.3943	0.3669	**0.4573**	0.3555
ASU [14]	27M	**0.4052**	**0.3915**	**0.4533**	**0.3760**
FocalU-net [15]	25M	0.3460	0.3809	0.3319	0.3227
RevCol [31]	31M	0.2122	0.2170	0.3471	0.1176
ASUFM	35M	**0.4066**	**0.4345**	0.4096	**0.3974**

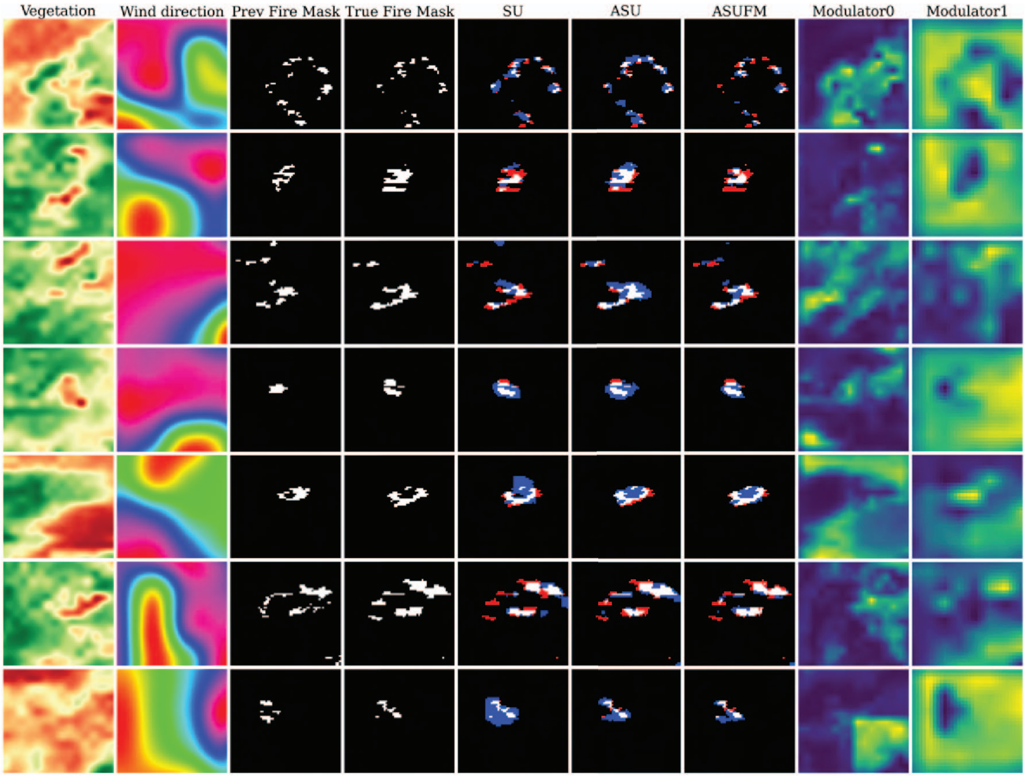

Fig. 5: Qualitative comparison of the ASUFM and ASU [14] and Swin U-net [30] models. Here, white represents True Positives (TP), ***red*** *represents False Negatives (FN), and* ***blue*** *represents False Positives (FP). The last column depicts an upsampled modulator at encoder layers* 0 *and* 1*, corresponding to the green block in Fig 3.*

TABLE V: Comparison of different loss functions

Loss Function	Dice Score	Precision	Recall	PR-AUC
BCE	**0.415**	0.360	**0.502**	**0.350**
BCE + Focal Tversky	0.340	**0.450**	0.340	0.278

TABLE VI: ASUFM results with different input features, the 6 *features are Wind direction, precipitation, specific humidity, vegetation, elevation, and previous day fire.*

Input features	Dice Score	Precision	Recall	PR-AUC
6 features	**0.4109**	0.4301	**0.4050**	**0.3728**
12 features	0.3559	**0.4778**	0.2926	0.3556

C. Discussion

1) Limitations: Our work is an initial exploration of tailoring current state-of-the-art segmentation techniques on wildfire spread modeling with multi-spectral data. The challenging nature of this problem, as highlighted in [7], compounded with ambiguity from the MODIS data leaves this an open research problem to be explored. Early experiments on various input bands suggest that some features may have confounding effects on the final prediction accuracy (see Table VI. for example) rather than improvement. While this study focuses on the input bands most likely to contribute to fire ignition and propagation [32] [7], further examination of feature selection could be useful. On the other hand, due to the high variability and limited extensibility in available data sets for wildfire spread, comparison across multiple sources becomes difficult. Other factors not explored include finer temporal resolution, other spatial resolution, and larger multi-modal models, which may offer additional capacity for extracting diverse information. Notably, [17] used FT (Focal Tversky) loss [33] for training their CNN-based models to counter the majority of pixels being background. However, our experiments in both training solely on FT loss and pre-training with weight BCE followed by FT loss, did not produce desirable results - details can be found in Table V.

2) Potential applications: Many previous studies have depended on specialized data that is challenging to acquire in real-time, consequently limiting their practical applicability in real-world scenarios. This work leverages GEE data which can be easily extended to apply to a real-time prediction map. In addition, we provide the pre-trained model weights that can be fine-tuned on region-specific data.

3) Future work: [34] outlines additional publicly available satellite based wildfire datasets and highlights the importance of baseline models and datasets. In future work we aim to look at incorporating input dimension from additional sources such as LANDFIRE [35] which provides in-depth fuel and land information that may unlock better performance on spread prediction. Meanwhile, datasets that are continuous temporally remain scarce and we hope to continue developing models that can make real-time predictions leveraging live remote-sensing imagery.

V. Conclusion

This research presents a significant advancement in the field of wildfire spread prediction using deep learning and remote sensing data. By developing the ASUFM model, which innovatively combines spatial attention and focal modulation with the Swin U-net architecture, we have established a new benchmark in predictive accuracy on the NDWS dataset. Our expanded dataset, covering North America up to 2023, provides an enriched foundation for further research and model development. The ASUFM model not only demonstrates superior performance in predicting wildfire spread but also offers a scalable and adaptable framework for real-time wildfire monitoring. Looking ahead, we aim to enhance our model by incorporating recursive inference for longer-term predictions, expanding the dataset with additional feature bands, and developing risk assessment tools for vulnerable populations.

References

[1] A. Tyukavina, P. V. Potapov, M. C. Hansen, A. H. Pickens, S. Stehman, S. Turubanova, D. Parker, V. Zalles, A. Lima, I. Kommareddy, X. Song, L. Wang, and N. L. Harris, "Global trends of forest loss due to fire from 2001 to 2019," in *Frontiers in Remote Sensing*, 2022.

[2] M. C. Hansen, P. V. Potapov, R. Moore, M. Hancher, S. A. Turubanova, A. Tyukavina, D. Thau, S. V. Stehman, S. J. Goetz, T. R. Loveland, A. Kommareddy, A. Egorov, L. Chini, C. O. Justice, and J. R. G. Townshend, "High-resolution global maps of 21st century forest cover change," 2013.

[3] J. T. Abatzoglou and A. P. Williams, "Impact of anthropogenic climate change on wildfire across western us forests," *Proceedings of the National Academy of Sciences*, vol. 113, pp. 11 770 – 11 775, 2016.

[4] A. R. Sharma, P. Jain, J. T. Abatzoglou, and M. D. Flannigan, "Persistent positive anomalies in geopotential heights promote wildfires in western north america," *Journal of Climate*, 2022.

[5] S. Masri, E. Scaduto, Y. Jin, and J. Wu, "Disproportionate impacts of wildfires among elderly and low-income communities in california from 2000–2020," *International Journal of Environmental Research and Public Health*, vol. 18, 2021.

[6] J. P. Minas, J. W. Hearne, and J. W. Handmer, "A review of operations research methods applicable to wildfire management," *International Journal of Wildland Fire*, vol. 21, no. 3, pp. 189–196, 2012.

[7] F. Huot, R. L. Hu, M. Ihme, Q. Wang, J. Burge, T. Lu, J. Hickey, Y.-F. Chen, and J. R. Anderson, "Deep learning models for predicting wildfires from historical remote-sensing data," *ArXiv*, vol. abs/2010.07445, 2020.

[8] D. Rolnick, P. L. Donti, L. H. Kaack, K. Kochanski, A. Lacoste, K. Sankaran, A. S. Ross, N. Milojevic-Dupont, N. Jaques, A. Waldman-Brown, A. S. Luccioni, T. Maharaj, E. D. Sherwin, S. K. Mukkavilli, K. P. Kording, C. P. Gomes, A. Y. Ng, D. Hassabis, J. C. Platt, F. Creutzig, J. T. Chayes, and Y. Bengio, "Tackling climate change with machine learning," *ACM Computing Surveys*, vol. 55, pp. 1 – 96, 2019.

[9] R. Ghali and M. A. Akhloufi, "Deep learning approaches for wildland fires using satellite remote sensing data: Detection, mapping, and prediction," *Fire*, 2023.

[10] K. Shah and M. Pantoja, "Wildfire spread prediction using attention mechanisms in u-net," *2023 3rd International Conference on Electrical, Computer, Communications and Mechatronics Engineering*, pp. 1–6, 2023.

[11] M. Marjani and M. S. Mesgari, "The large-scale wildfire spread prediction using a multi-kernel convolutional neural network," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2023.

[12] W. Jiang, F. Wang, G. Su, X. Li, G. Wang, X. Zheng, T. Wang, and Q. Meng, "Modeling wildfire spread with an irregular graph network," *Fire*, 2022.

[13] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, and M. Wang, "Swin-unet: Unet-like pure transformer for medical image segmentation," in *ECCV Workshops*, 2021.

[14] E. K. Aghdam, R. Azad, M. Zarvani, and D. Merhof, "Attention swin u-net: Cross-contextual attention mechanism for skin lesion segmentation," 2022.

[15] J. Yang, C. Li, X. Dai, and J. Gao, "Focal modulation networks," 2022.

[16] Y. Chen, S. Hantson, N. Andela, S. R. Coffield, C. A. Graff, D. C. Morton, L. E. Ott, E. Foufoula-Georgiou, P. Smyth, M. L. Goulden, and J. T. Randerson, "California wildfire spread derived using viirs satellite observations and an object-based tracking system," *Scientific Data*, vol. 9, 2022.

[17] Y. Zou, M. Sadeghi, Y. Liu, A. Puchko, S. Le, Y. Chen, N. Andela, and P. Gentine, "Attention-based wildland fire spread modeling using fire-tracking satellite observations," *Fire*, 2023.

[18] M. Marjani, S. A. Ahmadi, and M. Mahdianpari, "Firepred: A hybrid multi-temporal convolutional neural network model for wildfire spread prediction," *Ecological Informatics*, 2023.

[19] J. L. Hodges and B. Y. Lattimer, "Wildland fire spread modeling using convolutional neural networks," *Fire Technology*, pp. 1–28, 2019.

[20] J. K. Balch, B. A. Bradley, J. T. Abatzoglou, R. C. Nagy, E. J. Fusco, and A. L. Mahood, "Human-started wildfires expand the fire niche across the united states," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 114, pp. 2946 – 2951, 2017.

[21] J. Fitzgerald, E. Seefried, J. E. Yost, S. L. Pallickara, and N. Blanchard, "Paying attention to wildfire: Using u-net with attention blocks on multimodal data for next day prediction," *Proceedings of the 25th International Conference on Multimodal Interaction*, 2023.

[22] A. L. Sullivan, "A review of wildland fire spread modelling, 1990-present, 1: Physical and quasi-physical models," *arXiv: Geophysics*, 2007.

[23] D. Radke, A. Hessler, and D. Ellsworth, "Firecast: Leveraging deep learning to predict wildfire spread," in *International Joint Conference on Artificial Intelligence*, 2019.

[24] J. Burge, M. Bonanni, M. Ihme, and L. Hu, "Convolutional lstm neural networks for modeling wildland fire dynamics," *ArXiv*, vol. abs/2012.06679, 2020.

[25] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *ArXiv*, vol. abs/1505.04597, 2015.

[26] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, and M. Wang, "Swin-unet: Unet-like pure transformer for medical image segmentation," 2021.

[27] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," 2021.

[28] Z. Liu, H. Hu, Y. Lin, Z. Yao, Z. Xie, Y. Wei, J. Ning, Y. Cao, Z. Zhang, L. Dong, F. Wei, and B. Guo, "Swin transformer v2: Scaling up capacity and resolution," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 11 999–12 009.

[29] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert, "Attention u-net: Learning where to look for the pancreas," in *Medical Imaging with Deep Learning*, 2018.

[30] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, and M. Wang, "Swin-unet: Unet-like pure transformer for medical image segmentation," in *Proceedings of the European Conference on Computer Vision Workshops*, 2022.

[31] Y. Cai, Y. Zhou, Q. Han, J. Sun, X. Kong, J. Li, and X. Zhang, "Reversible column networks," in *International Conference on Learning Representations*, 2023.

[32] J. Hodges and B. Lattimer, "Wildland fire spread modeling using convolutional neural networks," *Fire Technology*, vol. 55, 03 2019.

[33] N. Abraham and N. M. Khan, "A novel focal tversky loss function with improved attention u-net for lesion segmentation," *2019 IEEE 16th International Symposium on Biomedical Imaging*, pp. 683–687, 2018.

[34] R. Ghali and M. A. Akhloufi, "Deep learning approaches for wildland fires using satellite remote sensing data: Detection, mapping, and prediction," *Fire*, vol. 6, no. 5, 2023.

[35] "U.s. department of the interior, geological survey. landfire 2.0.0 database." 2023.