BC Wildfire Risk Prediciton Time-Series Dataset: 2002–2023

Zhengsen Xu[1], Lanying Wang[2], Yimin Zhu[1], Wentao Sun[3], Lincoln Linlin Xu[1*]

[1] Schulich School of Engineering, University of Calgary, T2N1N4 Calgary, Canada
(zhengsen.xu, yimin.zhu, lincoln.xu)@ucalgary.ca
[2] Department of Geography and Environmental Management, University of Waterloo, N2L3G1 Waterloo, Canada
lanying.wang@uwaterloo.ca
[3] Department of Systems Design Engineering, University of Waterloo, N2L3G1 Waterloo, Canada
wentao.sun@uwaterloo.ca

Keywords: Wildfire, British Columbia, Benchmark, Deep Learning.

Abstract

Wildfires are longstanding natural phenomena with significant impacts on ecosystems and communities. In recent years, Canada has experienced particularly severe wildfire effects, especially in British Columbia (BC), which has endured prolonged and impactful wildfire events. However, there is currently no specialized wildfire time-series dataset for BC that considers long-term temporal sequences and multiple driving factors, which are essential for data-driven approaches. To facilitate future research on data-driven wildfire risk and spread prediction, we have developed a dataset covering the entire BC province, encompassing 683 wildfire events from 2002-2023 at 500m resolution with daily observations. For each wildfire event, the dataset includes 20 driving factors, including vegetation status, meteorological factors, human activities, topographical features, and active fire detection. Based on this benchmark and similar datasets from other regions, we compared multiple deep learning models, including CNN-based, Transformer-based, and Mamba-based architectures, to explore the effectiveness of existing deep learning models in wildfire risk prediction. We found that model F1 scores were below 0.6, indicating that this new dataset presents a challenging non-linear modeling scenario that requires more advanced and tailor-designed deep learning models to improve wildfire risk prediction accuracy.

1. Introduction

In recent years, Canada has faced a significant increase in wildfire frequency, size, and intensity. Given the worsening fire-prone conditions due to climate change, the need for accurate wildfire risk prediction models has become more urgent (Ager et al., 2019). Wildfires in British Columbia (BC), Canada, have been particularly severe, with an increasing number and frequency of extreme wildfire events (Daniels et al., 2024; Baron et al., 2022; Hanes et al., 2019). From 2003 to 2022, BC experienced an average of 1,350 wildfires annually, burning approximately 284,000 hectares (Taylor et al., 2022; Daniels et al., 2024). Between 1980 and 2021, BC accounted for 37.9% of all wildfire evacuations in Canada, primarily due to the frequent overlap of people and fire-prone landscapes (Christianson et al., 2024).

During the extreme wildfire events of 2023, many regions in BC recorded unprecedented wildfire weather and fire behavior. Similar to other regions of Canada, BC's wildfire season was extended, with fires starting earlier and ending later (Daniels et al., 2024). Fires began in April and May in northern and subarctic forests, as well as in the temperate rainforests of the southern coastal regions. The peak of the fire season occurred in July and August, with 481 wildfires burning simultaneously in BC. New wildfires continued to ignite until October (Service, 2023). As of December, over 100 active wildfires were still burning (Daniels et al., 2024). The Donnie Creek Complex burned 619,073 hectares, setting a new record for the largest wildfire in BC's history (Service, 2023; Daniels et al., 2024). By the end of 2023, BC had experienced a total of 2,245 wildfires (Daniels et al., 2024), burning approximately 2.84 million hectares, accounting for 15.35% of the total area burned in Canada that year (Daniels et al., 2024). This season led to 208 evacuation orders, displacing 48,000 people from their homes. Additionally, 386 evacuation alerts were issued, affecting 137,000 people who were advised to leave if necessary (Service, 2023; Daniels et al., 2024). Insurance losses exceeded $720 million, with direct firefighting costs amounting to $817 million, ultimately expected to surpass $1 billion (Service, 2023).

Therefore, both industry and academia require accurate wildfire prediction tools. Deep learning tools, in particular, are receiving increasing attention in wildfire prediction research due to their nonlinear feature modeling capabilities and data-driven approach. Although data-driven methods have demonstrated significant advantages in various tasks, including wildfire prediction, they require large training datasets. Unfortunately, while wildfire research is abundant, few studies make their datasets publicly available (Gerard et al., 2023). Thus, despite the large volume and diversity of remote sensing data and the rapid progress of deep learning models, the lack of standardized and user-friendly datasets significantly hinders wildfire research for those unfamiliar with remote sensing data.

Furthermore, while many studies have proposed datasets, they often cover the Mediterranean region and the United States, with few datasets focusing on northern regions (Kondylatos et al., 2022; Huot et al., 2022; Gerard et al., 2023; Zhao et al., 2024). Northern regions, especially northern Canada, have vast forested areas that have been severely affected by frequent wildfires in recent years. While existing datasets can support deep learning-based wildfire prediction research, they cannot explore the ignition or propagation characteristics of wildfires in northern regions, nor are they conducive to establishing wildfire prediction models that align with the characteristics of wildfires in northern regions. Therefore, considering these two points, we propose a simple, user-friendly dataset covering the entire BC province with a long temporal resolution (2002-2023) and

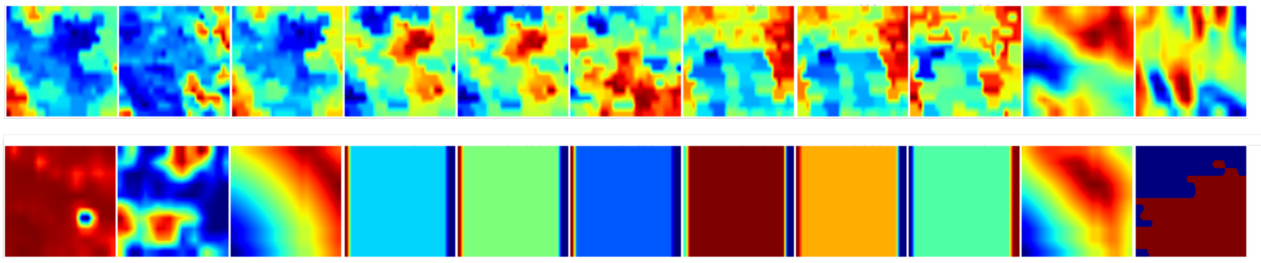

Figure 1. Illustration of selected driving factors in this dataset. From top left to bottom right: NDVI, LAI, EVI, Terra and Aqua's Bands 1, 2, and 7 reflectance products, DEM and DEM-based Aspect, slope, and hillshade, distance to the nearest water calculated from OpenStreetMap water distribution data, ERA-5's snow cover, V and U component wind speed, total precipitation, latent heat flux, 2m temperature, distance to the nearest infrastructure calculated from OpenStreetMap building and road distribution data, and MODIS active fire detection products.

multiple driving factors for wildfire prediction. We hope to promote future data-driven wildfire prediction research in northern regions. Additionally, based on this dataset and similar wildfire time-series datasets, we conduct next-day wildfire spread prediction experiments using various deep learning models, including CNN-based, Transformer-based, and Mamba-based architectures, to explore the effectiveness of existing advanced deep learning models in wildfire risk prediction tasks.

2. Dataset

2.1 Overview

Overall, the structure of this dataset is similar to WildfireSpreadTS and TS-satfire (Gerard et al., 2023; Zhao et al., 2024), that is, for each wildfire event, we construct datacubes with consistent temporal and spatial resolutions, with one datacube per day. Each datacube includes 20 types of drivers describing vegetation state, active fire detection, meteorological factors, human activity, and topographical factors, as described in Figure 1. The advantage of this architecture is that it has consistent spatiotemporal resolution, making it convenient for training and inference of various deep learning models, such as semantic segmentation models, time series prediction models, or video processing models.

Furthermore, considering the balance between availability, stability, and quality of long time series data, the datacube selects MODIS LAI products and Terra and Aqua's Bands 1, 2, and 7 to describe vegetation conditions; uses ERA5-Land's daily average 2m temperature, 10m U and V wind, total precipitation, and surface latent heat flux sum to represent daily meteorological conditions; topographical factors include SRTM's DEM data, from which slope, aspect, and hillshade are calculated, as well as distance to the nearest water body calculated from water distribution maps based on OpenStreetMap; for human activity factors, MODIS Land use products and distance to the nearest infrastructure calculated from distributions of roads, transportation, and railways based on OpenStreetMap are introduced; finally, since this dataset has a long time series, MODIS active fire detection products are used and wildfire event start and end times are obtained from the GlobFire dataset. Due to differences in temporal and spatial resolution and coordinate systems between various data sources, during production, the dataset was uniformly sampled to 1-day, 500m resolution products. The overall schematic of the dataset is shown in Figure 1.

2.2 Driving Factors

2.2.1 Vegetation Status Vegetation serves as fuel for wildfires, and the state of this fuel is crucial for predicting wildfire ignition and spread. Given the rapid dynamic changes of wildfires, stable products that can provide long time series are very scarce. Therefore, in addition to selecting the LAI product with 500m 4-day resolution and 500m daily NDVI and EVi products, we also selected MODIS Terra and Aqua's 500m daily resolution Band 1, 2, and 7 surface reflectance products. The combination of these bands can effectively distinguish healthy vegetation from burned areas and is relatively sensitive to clouds and smoke (Gerard et al., 2023). In addition to the above products, time series of vegetation indices such as NDVI and EVI also have long time spans, but they cannot provide complete temporal coverage, so this study did not incorporate these data into the dataset.

2.2.2 Meteorological Factors Weather and climate factors are major drivers of wildfire activity, influencing the occurrence, burned area, and fire behavior. Therefore, they are essential factors in various wildfire prediction models. High temperatures, low relative humidity, insufficient precipitation, and strong winds are key short-term factors that influence the ignition and spread of fires. In this study, we use the daily composite 1° resolution ERA5-Land product. When selecting meteorological factors, in addition to essential factors that directly affect fuel dryness, wind transport, and fire spread—such as 2m temperature, Eastward component of the 10m wind, Northward component of the 10m wind, total precipitation sum, and surface latent heat flux sum—we also introduced Snow coverage, which is characteristic of northern regions. Snow coverage not only affects surface temperature and humidity but can also significantly delay the exposure and drying process of fuels in winter and spring, thereby having an important impact on the probability of wildfire ignition and the initial spread rate.

2.2.3 Topographical Factors Topographical factors also play an important role in the ignition and spread of wildfires. Steep slopes (slope) help fire to rise quickly uphill, increasing the rate of spread; aspect affects sunlight exposure and dryness, thereby influencing the flammability of fuels; and hillshade reflects, to some extent, the shading of terrain on solar radiation, indirectly affecting local microclimate and burning conditions. Therefore, we selected high-resolution SRTM DEM data and calculated slope, aspect, and hillshade based on it. Additionally, water bodies often serve as natural barriers that can significantly reduce the probability of fire spread, so we also calculated the distance to the nearest water body using OpenStreetMap's water distribution map, further enhancing the model's ability to represent fire behavior.

2.2.4 Human Activity Factors Human activities are an important source of wildfire ignition. Considering data availability,

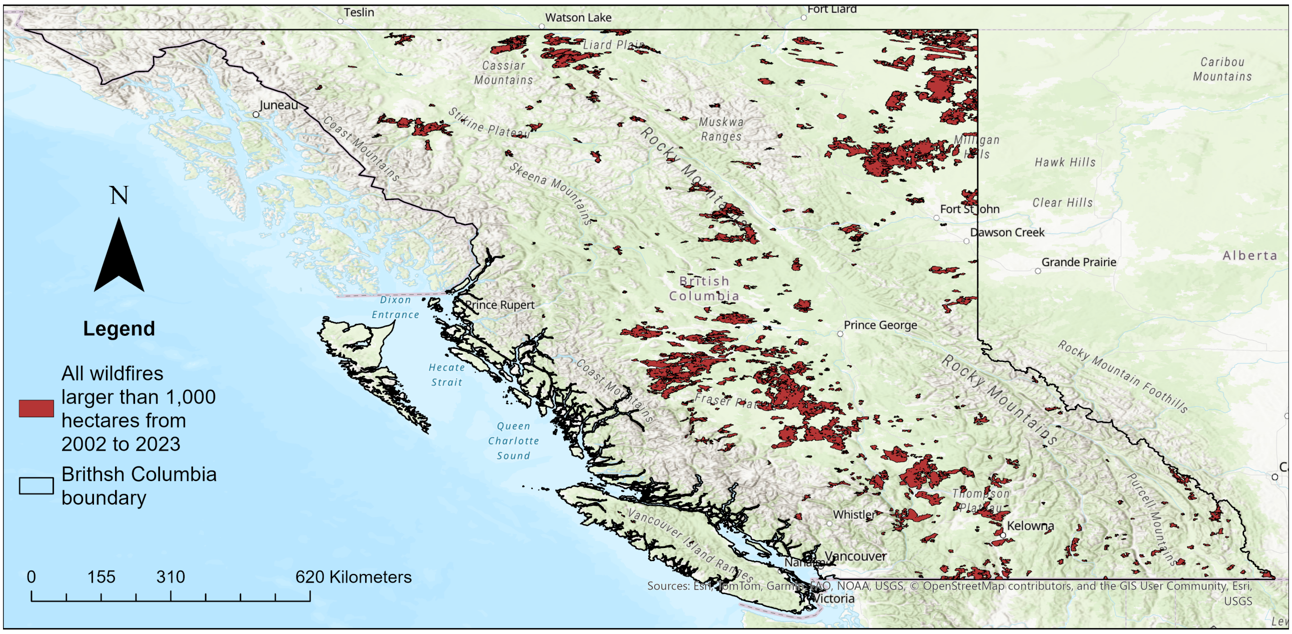

Figure 2. Spatial distribution of wildfire events in British Columbia from 2002 to 2023

we introduced MODIS Land Use products to reflect the potential impact of land use types on wildfire occurrence, as there are significant differences between agricultural areas and forest lands in terms of fuel types and human intervention. At the same time, based on the distribution maps of roads, transportation facilities, and railways from OpenStreetMap, we calculated the distance to the nearest infrastructure. Human activities often cluster along infrastructure, potentially bringing fire sources that trigger ignition events; meanwhile, roads and railways can also act as fire breaks to some extent, affecting the path of fire spread.

2.2.5 Wildfire Data Finally, due to the long time span covered by this dataset, we adopted MODIS active fire detection products (MOD14A1 and MYD14A1) to identify daily fire point information, and combined with the GlobFire dataset to obtain the start and end times of corresponding wildfire events, thereby constructing a relatively complete time series of wildfire events. MODIS fire point products have advantages such as global coverage, daily updates, and long time series, which can stably capture the location and intensity information of fire activities, especially suitable for large-scale, long-term trend analysis. The GlobFire dataset further aggregates fire points at the event level, providing the duration and spatial range of each wildfire, which helps us to model and analyze from single fire point to event level. The combination of the two not only enhances the completeness and accuracy of time-series data but also provides reliable target output labels for subsequent model training.

2.3 Data Preprocessing

First, following the processing workflow of the WildfireSpreadTS dataset, we obtained the GlobFire dataset, which provides the spatial range and time span of wildfire events, while MODIS active fire products only provide isolated daily fire point detection results. After obtaining this data, we first filtered out small wildfires less than 1000ha based on the spatial extent of wildfire events. Then, based on the polygon of each fire in the GlobFire dataset, we established a bounding rectangle to mask all driving features and MODIS active fire products. It is worth noting that due to the large differences in coordinate systems and spatial resolutions of various driving data, before masking, we first unified the data coordinates to the EPSG: 3005 coordinate system to maintain accurate area systems and used the nearest neighbor method to resample all data to 500m. Moreover, to ensure that wildfire events are continuously captured and to fully consider the importance of temporal factors, we also added a 6-day buffer outside the start and end time range of wildfires. That is, data within 6 days before the start and 6 days after the end will also be used to establish the dataset. Additionally, to account for the diversity of data time series, for non-daily resolution data, when constructing datacubes, the data closest to the time of that datacube was incorporated into the datacube.

2.4 Dataset Statistics

As shown in Figure 2, from 2002 to 2023, the spatial distribution of wildfire events in British Columbia was primarily concentrated in the southern and central interior regions, including the Okanagan, Cariboo, and Thompson-Nicola regions. These areas have dry forests and frequent fire sources, leading to frequent fires. Clusters of fire events also appeared in northeastern British Columbia, especially around the Peace River region, while coastal areas, northwestern British Columbia, and rainforest areas such as Vancouver Island and Haida Gwaii had extremely low fire occurrence rates.

Combined with Figure 3, it is evident that both the total burned area and total duration of wildfires in British Columbia (BC) show a significant upward trend. In 2002, the total burned area across the province was only 7,392 hectares, with a total duration of 51 days, while by 2023, these two values had increased to 2.191 million hectares and 3,719 days, representing increases of approximately 296 times and 73 times, respectively. Among these, 2023 not only set a record for the largest burned area but was also the year with the longest duration. The interannual variations between the two are highly correlated, with a Pearson correlation coefficient of 0.96, indicating that the larger the burned area, the longer the duration of the fire.

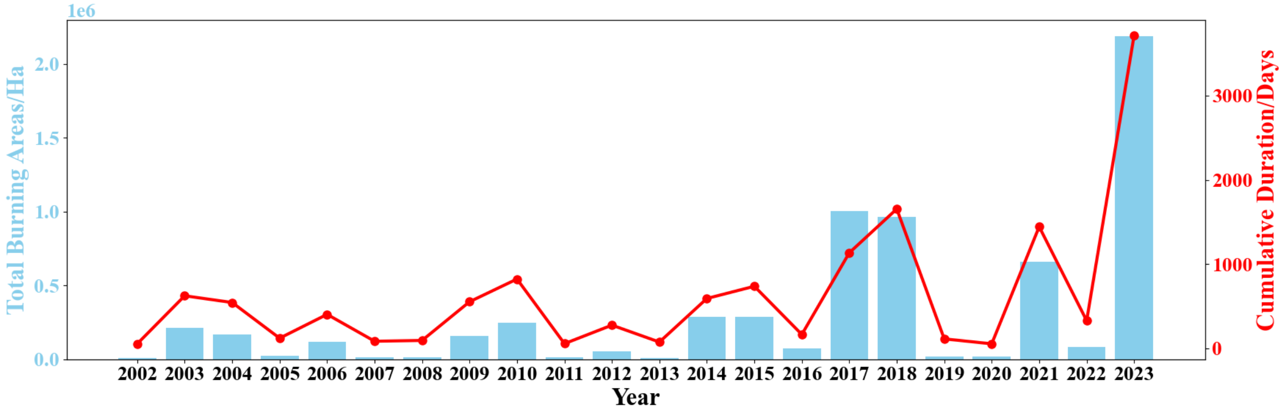

Figure 3. Annual trends in total burned area and total duration of wildfires in British Columbia from 2002 to 2023

2.5 Comparison with Existing Datasets

Currently, publicly available wildfire spread prediction datasets remain relatively scarce, with representative ones including the WildfireSpreadTS dataset and the Next Day Wildfire Spread dataset. WildfireSpreadTS is a multimodal time-series wildfire spread prediction dataset constructed by Gerard et al. (2023), covering 607 fire events in the United States from 2018 to 2021 . The significant advantage of this dataset lies in its consistent daily temporal resolution and 375m spatial resolution, integrating 23 input channels covering vegetation status, meteorological factors, topographical features, and daily fire point distribution. Its structural design is particularly suitable for multi-temporal modeling, helping to evaluate the performance of deep learning models in fire spread prediction. However, this dataset only covers the western United States, making it difficult to support generalization research for Canada or high-latitude regions.

Another dataset, Next Day Wildfire Spread, was developed by Huot et al. (2022), covering 18,545 fire samples in the United States from 2012 to 2020. This dataset focuses on the next-day fire spread prediction task based on single-day observations, with rich remote sensing variables (such as terrain, meteorology, drought indices, population density, etc.), a spatial resolution of 1 km, emphasizing the pixel-level classification task of whether a fire spreads to surrounding areas. Its main feature is the spatial alignment processing of multi-source observation variables and fire labels, very suitable for image segmentation-type deep learning methods. However, this dataset similarly does not cover Canada or other high-latitude regions where forest fires frequently occur, and lacks cross-day time-series information.

In comparison, the BC Wildfire Event Dataset constructed in this research fills the gaps in geographical region and time span of the above datasets. This dataset covers 683 wildfire events in British Columbia from 2002 to 2023, with a longer time series, higher latitude representation, and integrates 20 driving factors (including vegetation indices, active fire points, meteorology, human activities, and terrain), using uniform 500m spatial resolution and daily temporal resolution. Moreover, the structure of this dataset is compatible with various modeling methods, such as time series prediction, video modeling, and semantic segmentation, possessing good scalability and universality.

3. Comparative Experiments

To explore the effectiveness of advanced deep learning methods in wildfire risk prediction and ensure the robustness of experiments, we conducted comparisons of CNN-based, Transformer-based, and Mamba-based deep learning networks on three datasets, including our dataset, the WildfireSpreadTS dataset, and the Next Day Wildfire Spread dataset.

3.1 Wildfire Risk Prediction Methods Comparison

Specifically, CNN-based models mainly include UNet (Ronneberger et al., 2015), DeepLabV3+ (Chen et al., 2018), and ConvLSTM (SHI et al., 2015), which are classic and widely applied structures in the field of semantic segmentation. UNet adopts an encoder-decoder architecture and establishes fusion between high-resolution features and deep semantic information through skip connections, effectively enhancing boundary detection capabilities. DeepLabV3+ further introduces dilated convolution and Atrous Spatial Pyramid Pooling (ASPP) structures, enhancing the model’s perception ability of multi-scale features, suitable for fire prediction tasks with complex spatial distributions. Besides static image segmentation models, we also introduced ConvLSTM networks to further model temporal dimension information. ConvLSTM embeds convolution operations into Long Short-Term Memory networks (LSTM), maintaining spatial structure while having the ability to model temporal dynamic features, suitable for handling spatiotemporal coupling relationships in the daily evolution process of wildfires. For Transformer-based networks, we selected UNETR, which combines UNet’s encoding-decoding framework with Transformer’s global modeling capability, able to better capture long-distance dependencies between wildfire spatiotemporal features (Hatamizadeh et al., 2022). Finally, Mamba is a new type of state space model architecture proposed in recent years, with good long sequence modeling capability and computational efficiency. In our experiments, we adopted Video Vision Mamba (Vivim), which effectively processes video sequence information through a sliding state update mechanism, suitable for multi-temporal wildfire spread modeling tasks (Yang et al., 2024).

3.2 Implementation Details

In the specific experimental process, we used all driving factors at time point t as model input, with active fire detection results at time point $t+1$ as the supervision signal (ground truth) for training, aiming to achieve prediction of next-day fire distribution based on current environmental drivers. In dataset partitioning, we selected years covering approximately 70%, 20%, and 10% of the data to construct training, validation, and test sets, respectively, to maximize the exploration of the model’s ability to predict unknown wildfire risks.

Table 1. Performance comparison of different models on three wildfire datasets

Model	Next Day			WildfireSpreadTS			BC Wildfire Events		
	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
UNet	19.05	21.02	19.99	56.67	53.03	54.79	20.49	25.03	22.53
DeepLabV3+	18.85	39.48	25.52	47.58	60.27	53.18	23.04	41.02	29.51
ConvLSTM	21.64	28.96	24.77	53.94	59.72	56.68	24.67	29.21	26.75
UNETR	30.04	26.79	28.31	55.28	50.62	52.84	35.21	28.47	31.48
ViViM	19.61	33.96	24.87	42.31	66.24	51.64	21.85	38.12	27.78

During training, we set the batch size to 64 and the total number of training epochs to 2000. Considering that wildfire distribution has severe class imbalance characteristics, i.e., fire pixels usually only account for a very small proportion spatially, we adopted a combination loss function of Binary Cross Entropy (BCE) and Focal Loss to enhance the model's learning ability for minority classes (fire points). In terms of optimizers, we used AdamW with an initial learning rate of 1e-5 to enhance training stability and generalization capabilities. To accurately evaluate the performance of various models, we used precision, recall, and F1-score as evaluation metrics.

3.3 Baseline Results and Discussion

First, in the Next Day Wildfire Spread dataset, all models performed poorly, with F1 scores ranging from 19.99 to 28.31, as shown in Table 1. The Transformer-based UNETR and CNN-based DeepLabV3+ performed best, with F1 scores of 28.31 and 25.52, respectively. ViViM and ConvLSTM performed next best, with F1 scores of 24.87 and 24.77; while UNet performed worst, with an F1 score of 19.99, approximately 8.31 lower than the best-performing UNETR. Additionally, in this dataset, all models except UNETR had higher Recall than Precision.

Second, in the WildfireSpreadTS dataset, various models performed significantly better than in the Next Day Wildfire Spread dataset. ConvLSTM had the highest F1 score at 56.68, while the Mamba-based ViViM performed worst at 51.64, about 5.04 lower. The F1 scores of the remaining models did not differ significantly, with UNet, DeepLabV3+, and UNETR scoring 54.79, 53.18, and 52.84, respectively. Notably, similar to the Next Day Wildfire Spread dataset, in this dataset, most models—UNet, DeepLabV3+, and ViViM—had higher Recall metrics than Precision metrics.

Finally, in our BC Wildfire Events dataset, overall model performance was close to that of the Next Day Wildfire Spread dataset but slightly higher. Specifically, the Transformer-based UNETR performed best with an F1 score of 31.48, about 3.17 higher than UNETR's 28.31 in the Next Day dataset; followed by DeepLabV3+ and ViViM, with F1 scores of 29.51 and 27.78, respectively, very close in performance, and slightly improved compared to their F1 scores in the Next Day dataset (25.52 and 24.87, respectively). ConvLSTM's F1 score in this dataset was 26.75, ranking fourth, showing a slight improvement compared to the Next Day dataset. UNet still performed worst, with an F1 score of 22.53, but higher than its 19.99 in the Next Day dataset. Additionally, in this dataset, models' Recall was still generally higher than Precision, such as DeepLabV3+ with a Recall of 41.02, higher than its Precision (23.04), similar to the performance in the previous two datasets.

In a comprehensive comparison, we found that all models performed worse on the Next Day Wildfire Spread dataset compared to the other two datasets, followed by our BC Wildfire Events dataset. This may be because both of these datasets use MODIS products with insufficient wildfire detection capability as ground truth. MODIS products have insufficient detection capability for small and low-temperature wildfires compared to VIIRS, leading to unclear boundaries and possible omissions. However, the BC Wildfire Events dataset showed a significant improvement in model performance compared to the Next Day Wildfire Spread dataset, possibly due to the introduction of more driving factors. Additionally, we found that in all datasets, most models had higher recall rates than their corresponding precision, suggesting that the models are better at identifying possible wildfires than obtaining accurate predictions from these possibilities. In other words, the models may struggle to find the true paths from possible wildfire spread routes.

In terms of model performance comparison, since the purpose of this experiment was next-day wildfire spread prediction, which did not involve long time series information, time series prediction-oriented models like UNETR, ConvLSTM, and ViViM did not show significant advantages in this comparative experiment. Furthermore, due to the randomness of deep learning models and the uneven distribution of wildfire prediction datasets, i.e., significant differences in the spatiotemporal distribution of wildfires in different years, an n-fold model training and testing strategy might be necessary to maximize the fairness of experiments.

Overall, all model performances were not very high, with F1 scores below 0.6, indicating that the randomness of wildfire occurrence or spread and the weak nonlinear interactions between various features are too strong for simple deep learning models used for semantic segmentation to capture. Additionally, the extremely low proportion of wildfire pixels among all label pixels presents a class imbalance problem, which is a major challenge for wildfire prediction based on segmentation models. Furthermore, the inaccuracy of wildfire detection datasets also greatly hinders models from learning correct wildfire spread patterns.

4. Conclusion and Limitations

British Columbia is a region with frequent wildfires that requires focused attention, but simple and standardized wildfire event datasets targeting this region are still relatively scarce. Based on this gap, we proposed a dataset covering 683 wildfire events with areas over 1000ha in BC from 2002 to 2023. In this dataset, for each wildfire event, we established datacubes with daily resolution, consisting of 20 driving factors considering fuel state, meteorology, topography, and human activities, with unified spatiotemporal resolution, making it convenient for researchers in both remote sensing and non-remote sensing fields, such as computer vision, to use directly. Additionally, based on this dataset and two other similar wildfire prediction datasets, Next Day Wildfire Spread and WildfireSpreadTS, we tested advanced deep learning semantic segmentation networks and found that existing deep learning networks perform poorly on these datasets, with F1 scores below 0.6. This indicates that even existing deep learning models struggle with wildfire prediction tasks that have strong uncertainty and are severely affected by pixel class imbalance.

Establishing high-quality remote sensing datasets is a very challenging task. Although we have taken the first step in establishing a wildfire prediction dataset for BC, we must acknowledge that the existing dataset still has many deficiencies. First, BC's cloudy conditions cause missing passive optical remote sensing data, which not only interferes with the accurate and timely description of wildfire driving factors, such as fuel state, but cloud and smoke cover also interfere with MODIS's accurate fire point detection, posing challenges for establishing accurate ground truth. Furthermore, the inaccuracy of MODIS products, especially the insufficient ability to detect small wildfires, led us to remove wildfire events with spread ranges less than 1000ha when establishing the dataset. In subsequent updated datasets, we will combine higher spatial resolution VIIRS products to supplement MODIS's deficiency in detecting small wildfire events.

Based on this, the label data provided by this dataset can be considered to a large extent as inaccurate or incomplete labels. Therefore, we recommend that users consider using inaccurate or incomplete supervision training strategies or model structures when establishing data-driven methods using this dataset. To address this issue, future dataset updates will consider adding synthetic aperture radar products or consider translating SAR into multispectral products to address data gaps.

On the other hand, when establishing the dataset, priority was given to data availability, so fuel state only considered a few vegetation indices and MODIS reflectance products, which is not fine-grained enough to describe fuel conditions; the temporal resolution of land cover type products is annual, making it difficult to accurately reflect changes in land cover types, especially those caused by wildfires. When updating future datasets, we will consider more data sources and datasets with better capabilities to capture spatiotemporal dynamic changes.

Third, the purpose of establishing the dataset is to facilitate deep learning-based, especially semantic segmentation model-based, wildfire spread or wildfire risk prediction for near-term time frames, and did not consider climate factors such as El Niño or the Southern Oscillation that have been proven to have significant correlations with wildfire ignition risk, nor long-term drought factors such as the Palmer Drought Severity Index. Future dataset updates will consider introducing more climate factors and long-term meteorological data to facilitate long-term wildfire risk research.

Finally, it is worth noting that in this experiment, wildfire spread prediction was treated as a simple deep learning semantic segmentation task, and classic pixel-level segmentation accuracy evaluation metrics—precision, recall, and their weighted average, F1—were selected to evaluate model performance. However, wildfire prediction is not a simple semantic segmentation task because wildfire prediction research tends to focus on models' ability to evaluate wildfire risk in different areas, rather than simple segmentation probabilities. Additionally, the emphasis on wildfire risk and reducing wildfire-caused losses means that evaluation metrics should focus more on predicting all possible wildfire spread trends, so there should be greater tolerance for false positive predictions. Based on these considerations, future research should establish reasonable wildfire risk or wildfire spread evaluation metrics for evaluating pixel-level prediction performance based on deep learning or other data-driven methods.

References

Ager, A. A., Palaiologou, P., Evers, C. R., Day, M. A., Ringo, C., Short, K., 2019. Wildfire exposure to the wildland urban interface in the western US. *Applied geography*, 111, 102059.

Baron, J. N., Gergel, S. E., Hessburg, P. F., Daniels, L. D., 2022. A century of transformation: fire regime transitions from 1919 to 2019 in southeastern British Columbia, Canada. *Landscape Ecology*, 37(10), 2707–2727.

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation. *Proceedings of the European Conference on Computer Vision (ECCV)*.

Christianson, A. C., Johnston, L. M., Oliver, J. A., Watson, D., Young, D., MacDonald, H., Little, J., Macnab, B., Bautista, N. G., 2024. Wildland fire evacuations in Canada from 1980 to 2021. *International Journal of Wildland Fire*, 33(7).

Daniels, L. D., Dickson-Hoyle, S., Baron, J. N., Copes-Gerbitz, K., Flannigan, M. D., Castellanos Acuna, D., Hoffman, K. M., Bourbonnais, M., Wilkinson, S. L., Roeser, D. et al., 2024. The 2023 wildfires in British Columbia, Canada: Impacts, drivers, and transformations to coexist with wildfire. *Canadian Journal of Forest Research*.

Gerard, S., Zhao, Y., Sullivan, J., 2023. WildfireSpreadTS: A dataset of multi-modal time series for wildfire spread prediction. *Advances in Neural Information Processing Systems*, 36, 74515–74529.

Hanes, C. C., Wang, X., Jain, P., Parisien, M.-A., Little, J. M., Flannigan, M. D., 2019. Fire-regime changes in Canada over the last half century. *Canadian Journal of Forest Research*, 49(3), 256–269.

Hatamizadeh, A., Tang, Y., Nath, V., Yang, D., Myronenko, A., Landman, B., Roth, H. R., Xu, D., 2022. Unetr: Transformers for 3d medical image segmentation. *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 574–584.

Huot, F., Hu, R. L., Goyal, N., Sankar, T., Ihme, M., Chen, Y.-F., 2022. Next day wildfire spread: A machine learning dataset to predict wildfire spreading from remote-sensing data. *IEEE Transactions on Geoscience and Remote Sensing*, 60, 1–13.

Kondylatos, S., Prapas, I., Ronco, M., Papoutsis, I., Camps-Valls, G., Piles, M., Fernández-Torres, M.-Á., Carvalhais, N., 2022. Wildfire danger prediction and understanding with deep learning. *Geophysical Research Letters*, 49(17), e2022GL099368.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. N. Navab, J. Hornegger, W. M. Wells, A. F. Frangi (eds), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, Springer International Publishing, Cham, 234–241.

Service, B. W., 2023. 2003 wildfire season summary. Technical report, BC Ministry of Forests, Victoria, BC.

SHI, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-k., WOO, W.-c., 2015. Convolutional lstm network: A machine learning approach for precipitation nowcasting. C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, R. Garnett (eds), *Advances in Neural Information Processing Systems*, 28, Curran Associates, Inc.

Taylor, S., Daniels, L., Copes-Gerbitz, K., Forbes, K., 2022. *1.5 Wildfires, in Resilient Pathways Report: Co-creating new Knowledge for Understanding Risk and Resilience in BC*. Open File 8910, 102–129.

Yang, Y., Xing, Z., Yu, L., Huang, C., Fu, H., Zhu, L., 2024. Vivim: a video vision mamba for medical video segmentation.

Zhao, Y., Gerard, S., Ban, Y., 2024. Ts-satfire: A multi-task satellite image time-series dataset for wildfire detection and prediction.