A Multimodal Machine Learning Framework for Enhanced Wildfire Prediction in Canada

Yanwen, Guan Qiang, *Student Member, IEEE*, and Fang Fang, *Senior Member, IEEE*

***Abstract*—Wildfires pose an increasing threat to Canada's ecosystems, economy, and public safety as climate change drives more volatile environmental conditions. Existing wildfire monitoring systems, relying on limited data sources, often lack the accuracy and timeliness required for effective response. This study presents a comprehensive multimodal machine learning framework that combines LSTM and Vision Transformer (ViT) architectures to integrate visual map data with numerical weather variables for improved wildfire occurrence prediction in Canada. By using both spatial and temporal information, this multimodal approach demonstrates significant improvements over traditional single source prediction methods, offering a robust foundation for addressing Canada's growing wildfire challenges.**

***Index Terms*—Forest Fire Weather Index (FWI), Wildfire Prediction, Modality,**

I. Introduction

Wildfires have become an increasingly serious threat to Canada's ecosystems, economy, and public safety. Each year, wildfires burn millions of acres across the country, particularly in provinces such as British Columbia and Alberta. These events result in significant ecological disruption, economic losses, and risks to human health and infrastructure. Notable incidents, such as the 2016 Fort McMurray wildfire, highlight the growing scale and impact of these disasters. Climate change has further intensified wildfire activity, contributing to longer fire seasons, higher temperatures, and drier conditions that increase the likelihood and severity of wildfires [1].

Traditional wildfire monitoring systems are heavily relying on meteorological data, satellite imagery, and expert judgment. Although effective to some extent, these methods often struggle with timely and accurate predictions due to the complex and dynamic nature of wildfire behavior. Machine learning (ML) offers a promising solution by enabling the analysis of large, multidimensional datasets to detect patterns and make data-driven forecasts. Integrating ML with conventional systems can significantly improve Canada's wildfire prediction capabilities, supporting earlier warnings, better resource allocation, and more effective fire management strategies [2].

Wildfires pose an increasing threat to Canada's diverse ecosystems, critical infrastructure, and public safety, risks that are intensifying under the influence of climate change and increasing environmental volatility. Current wildfire prediction and early warning systems often rely on limited data sources and may not fully capture the complex region-specific drivers of wildfire occurrence, particularly in remote and northern areas.

To address these challenges, we propose developing a multimodal machine learning framework for wildfire occurrence prediction across Canada. Our approach integrates two complementary data sources: visual map data, including fire danger and hotspot maps from satellite imagery, and numerical meteorological data that encompass temperature, humidity, wind speed, and precipitation measurements. By combining these diverse information sources, our framework aims to provide a more comprehensive and accurate assessment of the risk of next day wildfires throughout Canada.

A. Literature Review

Recent advancements in UAV-based wildfire monitoring have focused on improving early detection accuracy using deep learning techniques. One study proposes a novel FT-ResNet50 model that leverages transfer learning to enhance forest fire recognition from UAV imagery, which is often limited by small labeled datasets and varying visual perspectives during image acquisition. The model adapts a pretrained ResNet50 network by fine-tuning three convolutional blocks using the Adam optimizer and the Mish activation function while incorporating a focal loss function to better handle class imbalance. Experimental results demonstrate that FT-ResNet50 achieves a recognition accuracy of 79.48%, outperforming the baseline ResNet50 by 3.87% and VGG16 by 6.22%. These findings highlight the potential of combining transfer learning with optimized loss functions and activation strategies to improve wildfire detection performance in UAV systems. The study underscores the growing importance of data-efficient deep learning models in real-time, remote wildfire monitoring applications. [3]

Another advancements in drone-based wildfire detection have leveraged deep learning to enhance detection accuracy and operational efficiency in complex forest environments. Zhu et al. [4] propose an improved YOLOv8-based model that introduces several architectural innovations to address challenges such as false positives due to vegetation, terrain variability, and lighting changes. The model replaces full convolution with local convolution in the C2F module and integrates an EMA module to improve channel interaction and contextual awareness while reducing computational complexity. Additionally, the AgentAttention module is incorporated into the backbone to enhance feature extraction robustness

[0]Guan Qiang and Fang Fang are with the Department of Electrical and Computer Engineering, and Fang Fang is also with the Department of Computer Science, Western University, London, ON N6A 3K7, Canada. (email: {gqiang, fang.fang}@uwo.ca)

by combining Softmax and linear attention mechanisms. To further improve detection across varying scales and angles, the BiFormer module is introduced for adaptive fusion of global and local features. Experimental results demonstrate that the proposed model achieves a precision of 93.57% and recall of 88.51%, outperforming the baseline YOLOv8 by 5.05% and 2.72%, respectively. The model also achieves substantial reductions in FPS, GFLOPs, and parameter count, indicating improved real-time performance and deployment efficiency. This research offers a powerful solution for early wildfire warning and response, contributing to forest resource protection and public safety.

Agarwal et al. [5] investigate the use of Vision Transformers (ViTs) for wildfire detection from LandSat-8 satellite imagery, comparing their performance against traditional Convolutional Neural Networks (CNNs). While CNNs have demonstrated high accuracy in prior studies, they are computationally intensive and limited to local feature extraction. In contrast, ViTs offer efficient training and the ability to capture both local and global context. The study finds that a ViT model slightly outperforms a baseline CNN by 0.92%; however, the authors' own implementation of a CNN-based UNet architecture outperforms all other models with an Intersection over Union (IoU) of 93.58%, surpassing the baseline UNet by 4.58%. These results suggest that while ViTs are promising alternatives for wildfire detection, well-tuned CNN architectures such as UNet remain state-of-the-art for satellite-based image segmentation tasks. The study highlights the complementary strengths of ViTs and CNNs, contributing to the evolving landscape of remote sensing-based wildfire monitoring systems.

Natekar et al. [6] explore the use of Long Short-Term Memory (LSTM) networks, a type of Recurrent Neural Network (RNN), for time-series-based wildfire occurrence prediction in India. The study leverages historical meteorological data to train the LSTM model, aiming to forecast potential wildfire events with high temporal accuracy. The model achieved an impressive prediction accuracy of 94.77%, demonstrating its effectiveness in capturing temporal dependencies related to wildfire ignition patterns. This approach underscores the utility of deep learning in proactive forest management and climate resilience strategies. By enabling early warning systems and informed resource allocation, the model offers a valuable tool for governmental and environmental organizations focused on conserving sensitive forest ecosystems under increasing climate stress.

Wildfires pose a significant threat to ecological systems and human infrastructure, particularly in forest-rich regions like Alberta, Canada. Liang et al. [7] propose a predictive model to estimate the scale of wildfires—defined by the duration and area burned—using meteorological data from the Canada National Fire Database (CNFDB). The study evaluates three neural network architectures: Backpropagation Neural Network (BPNN), Recurrent Neural Network (RNN), and Long Short-Term Memory (LSTM). After preprocessing steps such as multicollinearity testing and feature normalization, the models were trained using historical wildfire and weather data. Among the tested models, the LSTM network achieved the highest classification accuracy of 90.9%, demonstrating its superior capacity to model temporal dependencies in meteorological variables. The study confirms the feasibility of early stage wildfire scale prediction, offering a valuable tool for emergency response teams to implement timely and proportionate mitigation strategies based on the predicted severity of the fire.

More advancements in machine learning (ML) and deep learning (DL) have significantly enhanced the predictive capabilities for early wildfire detection and severity assessment. Krishna et al. [8] proposed a novel Radial Basis Function Neural Network with Adaptive Variability Functionality (RBFNN-AVF) to predict and classify wildfires using image-based data and ecological variables. The study aimed to build a robust early warning system capable of identifying forest fires and assessing their potential severity before significant damage occurs. The model was trained and validated using a large dataset, achieving a high classification accuracy of 98.5% by correctly predicting 21,375 out of 21,697 wildfire instances. The researchers also introduced a five-level severity scale ranging from Level 1 (normal) to Level 5 (extremely severe), allowing detailed categorization of hazards. Compared to six other approaches, RBFNN-AVF outperformed all in predictive precision, highlighting its effectiveness in real-time image classification and ecological data analysis. The study demonstrates the feasibility of integrating ML-based image recognition systems with aerial surveillance technologies such as drones, offering a scalable and responsive solution for wildfire monitoring in densely forested regions of India. These findings contribute to the growing body of research supporting AI-driven environmental risk management tools.

B. *Motivation*

The motivation for this work comes from the opportunity to improve wildfire prediction using a dual-modality approach that combines Vision Transformers and LSTMs to jointly model spatial imagery and temporal weather data.

Our literature review reveals remarkable advancements in computer vision methodologies for wildfire prediction, which have improved ability to detect fires using satellite and unmanned aerial vehicle (UAV) imagery. Recent research has demonstrated impressive performance in identifying fire patterns and signatures within individual images, with these approaches proving valuable for real-time fire detection and monitoring applications.

However, we have found a fundamental limitation in existing methodologies: their exclusive focus on spatial features extracted from individual image frames. This approach overlooks a critical dimension of wildfire behavior, the temporal evolution of environmental conditions. Wildfires are inherently dynamic events that develop and progress in response to complex interactions among various environmental factors, including wind patterns, humidity levels, fuel availability, and topographic characteristics. We argue that effective prediction of wildfire occurrence and progression requires understanding not only the instantaneous conditions visible in a single image, but also how these conditions evolve and interact over time.

Therefore, the temporal dimension becomes particularly crucial when considering that many wildfire precursors manifest as subtle changes in environmental conditions that accumulate over days or weeks before ignition occurs.

To address this critical gap, we propose a novel hybrid deep learning architecture that synergistically combines vision transformers (ViT) with long-short-term memory (LSTM) networks. Our approach takes advantage of the complementary strengths of these two architectures to capture both spatial and temporal aspects of wildfire prediction. The Vision Transformer component excels at capturing complex spatial relationships within satellite imagery by treating each image as a sequence of patches and employing self-attention mechanisms to learn global dependencies across the entire image. This capability makes ViT particularly effective at identifying subtle visual cues and environmental patterns that may serve as precursors to fire events. Unlike traditional CNNs that process images through localized convolutions, the self-attention mechanism in ViT enables the model to consider long-range spatial dependencies, potentially identifying fire-conducive conditions across large geographic areas. We complement this spatial analysis capability with LSTM networks specifically designed to model temporal dependencies across sequences of daily satellite observations. The LSTM architecture is well-suited for capturing long-term temporal patterns and dependencies in sequential data, allowing our model to learn how environmental conditions evolve over time and identify temporal signatures that precede wildfire events.

By integrating these two powerful architectures, our hybrid model simultaneously learns the spatial structure of environmental conditions on any given day and tracks how these conditions change and interact over extended time periods.

C. Contribution

One of the key contributions of our work lies in developing a robust and automated data collection pipeline that supports wildfire prediction across Canada through multi-modal data integration. We recognized that accurate and timely forecasting requires comprehensive datasets, which led us to curate an extensive collection spanning from 2017 to 2023, sampled at daily intervals and covering Canada's complete geographic extent. Our dataset integrates two primary data modalities: satellite imagery and meteorological variables. We systematically collected geospatial fire-related maps from the Canadian Wildland Fire Information System (CWFIS), focusing on two main categories. First, we gathered FireM3 maps, which include daily hotspots, season-to-date hotspots, and combined fire weather and hotspot overlays. Second, we obtained Fire Weather Index (FWI) maps encompassing fire danger ratings, drought codes, fine fuel moisture codes, and other relevant environmental indices. We accessed each map through structured URL patterns based on map type and date, which allowed us to achieve scalable and automated retrieval of historical data. The detailed image's metedata are shown in Fig. 1. The data pipeline code is available on GitHub at https://github.com/YanwenWang1125/Wildfire-Predication-in-Canada.

To complement the imagery, we extracted corresponding meteorological labels for each day, including temperature, wind speed, rainfall, and precise timestamp information. We ensured temporal consistency across modalities by aligning these labels with the imagery through careful matching of embedded dates in filenames. Therefore, We developed a comprehensive Python-based pipeline to automate the entire data collection workflow, including URL generation, image downloading, label extraction, and integrity validation. Our pipeline incorporates built-in mechanisms to handle common data challenges such as missing files, naming inconsistencies, and corrupted downloads, thereby ensuring high data quality and completeness throughout the collection process.

Finally, we organized the resulting dataset using a structured directory format with standardized naming conventions and comprehensive metadata files, making it readily accessible for machine learning applications. To promote reproducibility and support the broader research community, we plan to publicly release the complete dataset—including both aligned images and meteorological labels—alongside detailed documentation and the complete code for our data collection pipeline.

This data collection effort establishes a solid foundation for training and evaluating multi-modal models in our wildfire prediction contexts.

II. Dataset Description: The Canadian Forest Fire Weather Index (FWI) System

The Canadian Forest Fire Weather Index (FWI) System is a foundational tool for wildfire risk assessment and management across Canada, widely adopted by governmental agencies, researchers, and operational wildfire managers to monitor and predict forest fire danger in a variety of ecological contexts [9], [10]. The FWI System integrates meteorological data with empirically derived models of fuel moisture and fire behavior, producing a set of indices that collectively describe the potential for fire ignition, spread, and intensity under prevailing environmental conditions.

A. Structure and Components

The FWI System comprises six principal components, divided into two categories: fuel moisture codes and fire behavior indices.

1) Fuel Moisture Codes: Fuel moisture codes are quantitative ratings of the moisture content in different layers of the forest floor and dead organic matter. These codes are crucial, as fuel moisture is a primary determinant of ignition probability and fire spread severity [11].

- **Fine Fuel Moisture Code (FFMC):** Reflects the moisture content of surface litter and other fine fuels. Sensitive to short-term weather changes, it indicates the ease of ignition and flammability of fine materials [9].
- **Duff Moisture Code (DMC):** Estimates the average moisture content in loosely compacted organic layers

FireM3 Maps	
Map Name	**Maptype**
Daily Hotspots	tri
Season-to-date Hotspots	apt
Daily hotspots & FWI	fwih

FWI Maps	
Map Name	**Maptype**
Fire Danger	fdr
Fire Weather Index	fwi
Fine Fuel Moisture Code	ffmc
Duff Moisture Code	dmc
Drought Code	dc
Initial Spread Index	isi
Buildup Index	bui
Daily Severity Rating	dsr

Fig. 1. Overview of the FireM3 and Fire Weather Index (FWI) map types used in the dataset. Each map type corresponds to a specific variable or indicator relevant to wildfire monitoring and prediction

of moderate depth, indicative of fuel consumption in medium-sized woody material and duff layers [12].

- **Drought Code (DC):** Measures the moisture content of deep, compact organic layers, representing cumulative seasonal drought effects and the likelihood of persistent smoldering in deep duff and large logs [11].

2) Fire Behavior Indices: The following indices are derived from the fuel moisture codes and meteorological variables, providing operational metrics for fire management:

- **Initial Spread Index (ISI):** A composite index estimating the expected rate of fire spread, primarily as a function of wind speed and FFMC. ISI is effective for signaling periods of heightened spread potential [9].
- **Buildup Index (BUI):** Synthesizes information from the DMC and DC to estimate the total amount of fuel available for combustion, reflecting the cumulative drying of organic layers [10].
- **Fire Weather Index (FWI):** Integrates the ISI and BUI to provide a general index of fire intensity, widely used as a daily operational indicator of overall fire danger [13].

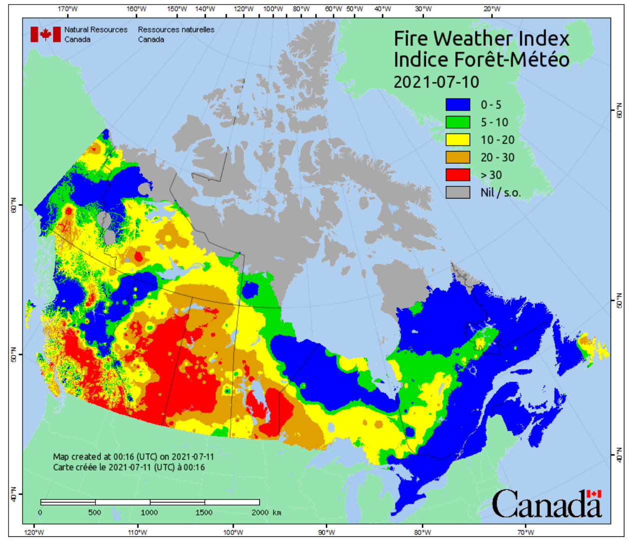

Fig. 2. The Fire Weather Index is a component of the Canadian Forest Fire Weather Index (FWI) System. It is a numeric rating of fire intensity. It is based on the ISI and the BUI, and is used as a general index of fire danger throughout the forested areas of Canada.

3) Daily Severity Rating (DSR): The Daily Severity Rating (DSR) is an additional component that provides a numeric estimate of the anticipated difficulty of controlling fires, based on the FWI. The DSR reflects the expected operational effort required for fire suppression and supports tactical decision-making [14].

B. Data Availability and Update Frequency

FWI System data are updated daily throughout the year, offering near-real-time assessments of fire danger at national, regional, and local scales. The Canadian Wildland Fire Information System (CWFIS) disseminates these indices through interactive maps and downloadable datasets, ensuring broad accessibility for both research and operational use [10].

C. Scientific and Operational Relevance

The FWI System has been validated and adapted for use in different forest types and climatic regimes, both within Canada . Its modular structure allows integration with advanced predictive models, remote sensing, and machine learning frameworks, enhancing the precision and utility of wildfire risk forecasts in the context of climate change and increasing fire activity [15], [16].

III. Dataset Description: Fire M3 Satellite-Derived Hotspots

Remote sensing has become a pivotal tool in wildfire detection, monitoring, and management, especially over large

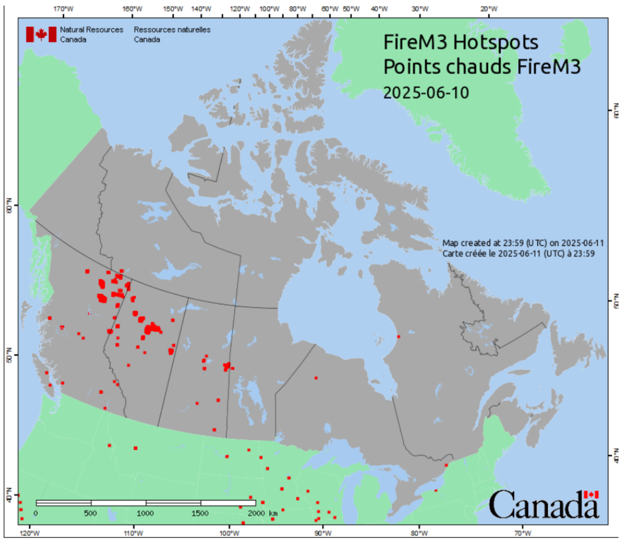

Fig. 3. A hotspot is a satellite image pixel with high infrared intensity, indicating a heat source. Hotspots from known industrial sources are removed; the remaining hotspots represent vegetation fires, which can be in forest, grass, cropland, or logging debris. A hotspot may represent one fire or be one of several hotspots representing a larger fire. Not all fires can be identified from satellite imagery, either because the fires are too small or because cloud cover obscures the satellite's view of the ground.

and inaccessible areas [17], [18]. The Fire Monitoring, Mapping, and Modeling (Fire M3) system utilizes satellite-derived thermal anomaly data—commonly referred to as *hotspots*—to provide near-real-time information on active vegetation fires across Canada and adjacent regions. This dataset underpins both operational fire management and scientific research on wildfire occurrence, behavior, and impacts.

A **hotspot** is defined as a satellite image pixel exhibiting abnormally high infrared intensity, indicative of a surface heat source. These thermal anomalies are primarily associated with active combustion processes. To ensure the reliability and specificity of the dataset for vegetation fires, hotspots originating from known industrial sources (e.g., gas flares, power plants) are systematically excluded through spatial masking and ancillary data analysis [19]. The remaining hotspots are attributed to vegetation fires, which may occur in forests, grasslands, croplands, or logging debris.

It is important to note that a single hotspot may represent an individual fire event or may constitute one of several contiguous hotspots collectively delineating a larger fire perimeter. The spatial and temporal resolution of hotspot detection is inherently limited by the sensor characteristics and overpass frequency. Additionally, not all active fires are detectable via satellite: small fires may fall below the sensor's detection threshold, and cloud cover or dense smoke can obscure the satellite's view, resulting in underreporting [20].

Fire M3 hotspots are derived from multiple complementary satellite platforms, each contributing unique spatial and temporal characteristics to the dataset:

- **Advanced Very High Resolution Radiometer (AVHRR):** AVHRR data are provided by the U.S. National Oceanic and Atmospheric Administration (NOAA) National Environmental Satellite, Data and Information Service (NESDIS). AVHRR sensors, aboard NOAA's polar-orbiting satellites, offer moderate spatial resolution (approximately 1.1 km at nadir) and have been foundational in global fire monitoring since the 1980s [21].
- **Moderate Resolution Imaging Spectroradiometer (MODIS):** MODIS sensors, onboard NASA's Terra and Aqua satellites, provide higher spatial (1 km for thermal bands) and temporal resolution (up to four observations daily at mid-latitudes). Fire M3 incorporates MODIS active fire products from NASA LANCE FIRMS and the USDA Forest Service Active Fire Mapping Program, Remote Sensing Applications Center (RSAC) [22]. MODIS active fire algorithms are widely validated and provide robust detection of medium to large fires [20], [23].
- **Visible Infrared Imaging Radiometer Suite (VIIRS):** VIIRS, launched aboard the Suomi NPP and NOAA-20 satellites, offers improved spatial resolution (375 m for fire detection bands) and enhanced sensitivity to smaller or lower-intensity fires compared to MODIS [18]. VIIRS active fire data are sourced from NASA LANCE FIRMS, University of Maryland, and USDA Forest Service RSAC.

Fire M3 maps and reports are updated daily during the Canadian wildfire season, typically from May through September. This high-frequency update cycle supports near-real-time fire monitoring and response, as well as retrospective analysis of fire occurrence patterns.

While the Fire M3 hotspot dataset provides comprehensive coverage and valuable temporal granularity, several limitations must be acknowledged:

- *Detection Limitations:* Small, low-intensity, or short-lived fires may not be detected, particularly under cloud cover or dense smoke.
- *Spatial Uncertainty:* The geolocation accuracy is constrained by the sensor's spatial resolution and georeferencing algorithms.
- *Source Attribution:* Despite rigorous filtering, some non-vegetation heat sources may be misclassified as wildfires, and conversely, some vegetation fires may be excluded.
- *Temporal Gaps:* The revisit frequency of polar-orbiting satellites may result in temporal gaps, especially in rapidly evolving fire events.

The Fire M3 hotspot dataset is integral to wildfire risk assessment, fire behavior modeling, emissions estimation, and ecosystem impact studies [24], [25]. Its integration with other geospatial and meteorological data supports operational decision-making and long-term research into wildfire dynamics under changing climatic conditions.

IV. Methodology

Our proposed multimodal neural network architecture combines the complementary strengths of Vision Transformers

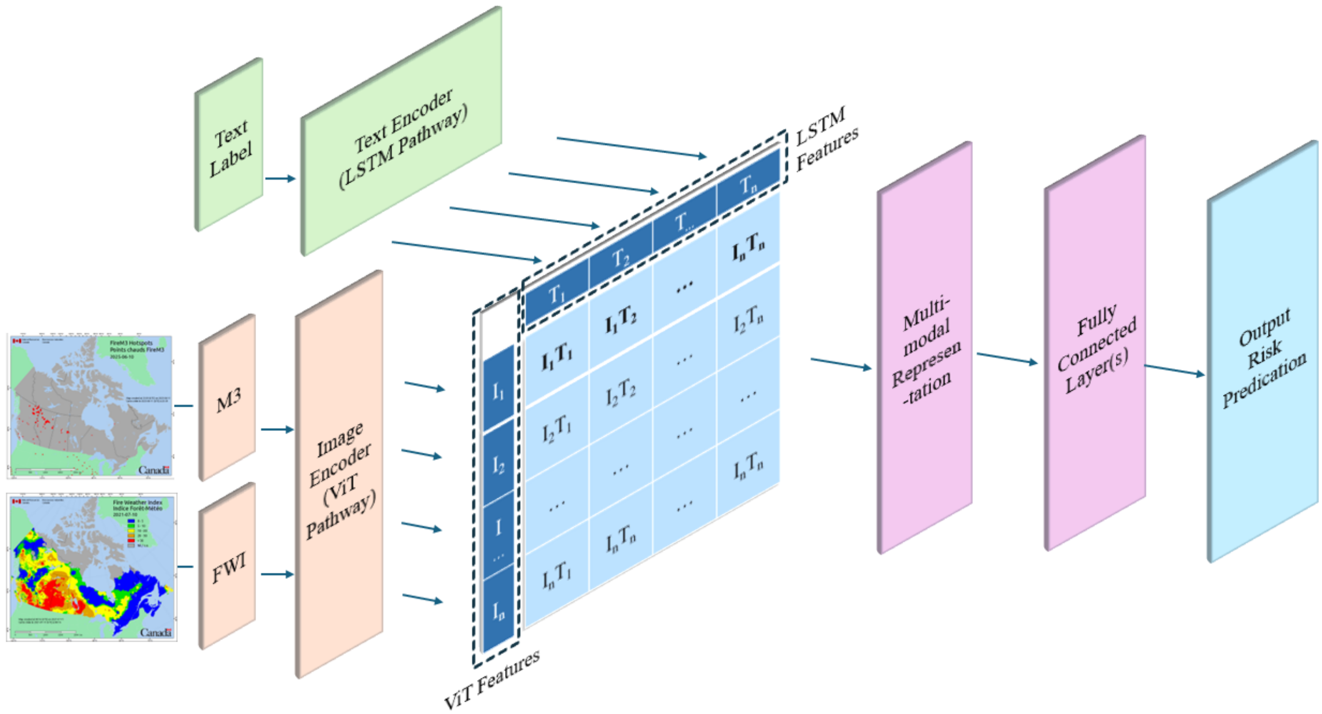

Fig. 4. Workflow of the proposed multimodal neural network integrating spatial (ViT) and temporal (LSTM) pathways for wildfire risk prediction.

(ViT) and Long Short-Term Memory (LSTM) networks to tackle the challenging problem of wildfire risk prediction. The architecture operates through two parallel processing pathways that handle spatial and temporal information separately before integration. The spatial pathway processes visual data such as Fire Weather Index (FWI) maps and hotspot imagery. Following standard ViT methodology, each input image is partitioned into fixed-size patches that are flattened and transformed into embedding vectors through linear projection. To maintain spatial context, these embeddings are enhanced with positional encodings before being fed into stacked transformer encoder blocks. The self-attention mechanisms within these blocks effectively capture long-range spatial dependencies and complex patterns across the imagery, ultimately producing high-level spatial features through either the final classification token or pooled output. In parallel, the temporal pathway models sequential patterns in meteorological data. Time series data including temperature, humidity, and wind speed measurements are processed through stacked LSTM layers, which excel at capturing temporal evolution and contextual relationships within these sequential measurements. The LSTM network distills this temporal information into its final hidden state or pooled output vector. The integration phase combines spatial features from the ViT pathway with temporal features from the LSTM pathway through concatenation, creating a unified multimodal representation. This fused vector passes through fully connected layers with non-linear activation functions, allowing the model to learn complex interactions between spatial and temporal modalities. The final output layer generates wildfire risk predictions using either linear activation for regression or sigmoid/softmax activation for classification tasks. This integrated architecture enables the model to simultaneously reason over spatial fire patterns and temporal meteorological dynamics, leading to more robust and accurate wildfire risk assessments than single-modality approaches.

A. Data Sources and Preprocessing:

Spatial data consists of Fire Weather Index (FWI) maps and M3 satellite-derived hotspot images. The FWI system provides a quantitative rating of fire danger based on weather conditions and fuel moisture [9]. M3 images, generated from satellite observations such as MODIS or VIIRS, indicate active fire locations and intensities [26]. All images are resized to 224×224 pixels and normalized to match the requirements of the ViT backbone. Data augmentation, including random cropping, horizontal flipping, and color jittering, is applied during training to improve generalizability.

Meteorological variables include temperature, relative humidity, wind speed, and dew point, collected at regular intervals from meteorological stations or reanalysis datasets. Each variable is normalized to zero mean and unit variance. For each prediction instance, a fixed-length sequence of historical meteorological measurements is extracted, forming a matrix of dimensions$T \times 4$, where T denotes the temporal window length.

B. Training Configuration

The model training employs different loss functions depending on the problem formulation: binary cross-entropy (BCE) for binary risk classification and categorical cross-entropy for multiclass scenarios. Specifically, the loss for a single sample in the binary case is defined as

$$\mathcal{L}_{\text{BCE}} = -[y \log(\hat{y}) + (1 - y) \log(1 - \hat{y})], \tag{1}$$

where y is the ground truth label and $\hat{y}$ is the predicted probability. We use the Adam optimizer for efficient weight updates, with the learning rate determined through grid search within the typical range of $[10^{-4}, 10^{-3}]$ to ensure optimal convergence. Several regularization techniques prevent overfitting during training. Dropout regularization is applied after dense layers with rates between 0.2 and 0.5, while early stopping monitors validation loss and halts training when no improvement occurs over a predetermined number of epochs. Batch sizes are selected based on available GPU memory constraints, typically ranging from 16 to 64 samples per batch. To leverage existing visual knowledge, the Vision Transformer pathway is initialized with weights pretrained on large-scale datasets such as ImageNet. The model then undergoes fine-tuning on wildfire-specific data, allowing it to adapt pretrained visual representations to the unique characteristics of wildfire imagery while maintaining the benefits of transfer learning.

C. Evaluation Methodology

We partition the dataset into training, validation, and test sets using a 70/15/15 split while maintaining both temporal and spatial independence between subsets. This partitioning strategy prevents information leakage and ensures the model cannot exploit correlations across time or geographic regions, providing a more realistic assessment of generalization capability for operational wildfire risk prediction. Our evaluation framework employs multiple metrics to comprehensively assess model performance. Standard classification metrics include accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC-ROC). To address potential class imbalance common in wildfire datasets, we also report the area under the precision-recall curve (AUC-PR) which provide more robust performance measures for rare wildfire events. For a more robust evaluation in the presence of class imbalance, the Matthews correlation coefficient (MCC) is also computed:

$$\text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}, \tag{2}$$

where TN denotes true negatives. Additionally, Cohen's kappa statistic,

$$\kappa = \frac{p_o - p_e}{1 - p_e}, \tag{3}$$

is used to quantify agreement between predicted and true labels, where p_o is the observed agreement and p_e is the expected agreement by chance.

For enhanced interpretability, we visualize confusion matrices to reveal class-specific prediction patterns and identify potential sources of misclassification. An ablation study systematically compares our proposed multimodal architecture (ViT+LSTM) against unimodal baselines (ViT-only and LSTM-only), demonstrating the individual contribution of each modality to overall predictive performance. This comprehensive evaluation framework ensures rigorous and transparent assessment of the model's effectiveness in wildfire risk classification.

V. Simulation Result

We evaluated the performance of our proposed multimodal ViT+LSTM architecture against unimodal baselines using a comprehensive suite of metrics. Table V presents the quantitative results across all evaluation metrics on the test set.

The multimodal ViT+LSTM model demonstrated superior performance across all measures, achieving an accuracy of 0.87, precision of 0.83, recall of 0.81, and F1-score of 0.82. The model's discrimination capability was particularly strong, with AUC-ROC and AUC-PR scores of 0.91 and 0.79, respectively, indicating excellent ability to distinguish between wildfire and non-wildfire events despite class imbalance. The robust MCC value of 0.71 and Cohen's kappa of 0.77 confirm strong agreement between predictions and ground truth, substantially outperforming both unimodal baselines. Analysis of confusion matrices Fig.V, Fig.V, Fig.V reveals significant improvements in the multimodal model's classification accuracy. The ViT+LSTM architecture misclassified only 9% of actual wildfire events as non-events, compared to 18% for ViT-only and 21% for LSTM-only models. This reduction in false negatives, coupled with improved precision scores, demonstrates the model's enhanced capability to detect genuine wildfire risks while minimizing false alarms—a critical consideration for operational wildfire management systems.

Additionally, the calibration curves in Figure.V demonstrate that the ViT+LSTM model is the best calibrated, closely following the diagonal line of perfect calibration across most probability ranges. In contrast, the ViT-only model tends to be slightly overconfident at higher predicted probabilities, while the LSTM-only model is generally underconfident, particularly at the upper end of the probability scale. These results indicate that combining the ViT and LSTM architectures improves the reliability of the predicted probabilities of the occurrence of wildfires.

The ablation study shows the complementary nature of spatial and temporal information processing. While the ViT-only model excelled with spatially rich samples and the LSTM-only model captured temporal patterns effectively, only the integrated multimodal approach achieved consistently high performance across all evaluation metrics. This finding underscores the importance of combining both modalities for comprehensive wildfire risk assessment. Our comprehensive evaluation confirms that the proposed multimodal ViT+LSTM architecture delivers substantial improvements in both predictive accuracy and reliability. The robust performance across

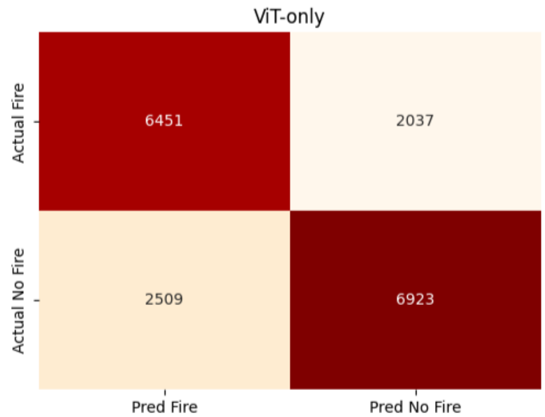

Fig. 5. Confusion matrix illustrating the classification performance of the ViT-only model.

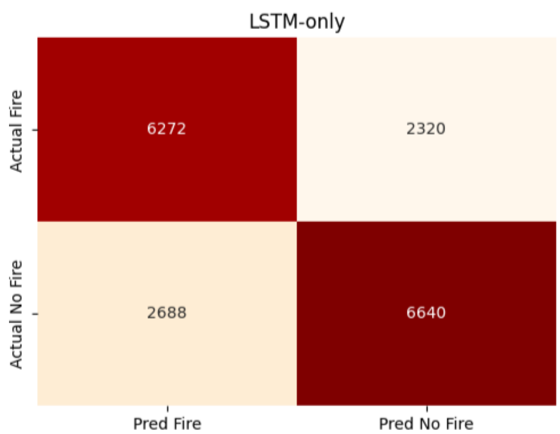

Fig. 6. Confusion matrix illustrating the classification performance of the LSTM model.

standard metrics and specialized class-imbalance measures supports its potential deployment in real-world wildfire risk management applications, where both precision and recall are essential for effective decision-making.

VI. Conclusion

Wildfires pose an escalating threat to Canada's ecosystems, economy, and public safety, with climate change intensifying both their frequency and severity. Traditional monitoring and early warning systems, while useful, often struggle to provide timely and accurate predictions due to the complex interactions among environmental factors. This project addresses these limitations by developing a comprehensive multimodal machine learning framework that integrates visual map data with numerical weather variables for improved wildfire occurrence prediction across Canada. Our experimental findings demonstrate that data fusion through the ViT+LSTM model substantially enhances predictive performance compared to single-modality approaches. The superior accuracy, precision, recall, F1-score, and AUC metrics achieved by the multimodal architecture highlight the critical value of combining

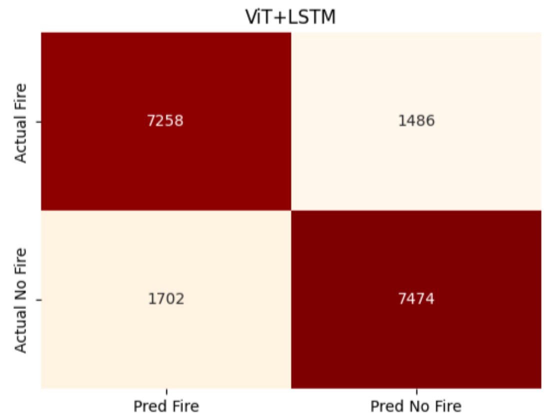

Fig. 7. Confusion matrix illustrating the classification performance of the combined LSTM and ViT model.

TABLE I
Performance comparison of ViT-only, LSTM-only, and ViT+LSTM models.

Model	Acc	Pres	Rec	F1	AUC-ROC	AUC-PR	MCC	Cohen's kappa
ViT-only	0.81	0.76	0.72	0.74	0.85	0.68	0.59	0.62
LSTM-only	0.79	0.73	0.70	0.71	0.83	0.65	0.54	0.58
ViT+LSTM	**0.87**	**0.83**	**0.81**	**0.82**	**0.91**	**0.79**	**0.71**	**0.77**

spatial and temporal information for wildfire forecasting. The framework's ability to generate high-resolution spatiotemporal risk maps provides emergency managers, policymakers, and at-risk communities with a powerful decision-support tool for more effective early warnings, resource allocation, and response planning. The integration of advanced machine learning with data fusion techniques offers considerable promise for strengthening Canada's wildfire prediction capabilities. Continued development and operational deployment of such multimodal systems will be essential for building resilience against the intensifying wildfire threat under changing climatic conditions, ultimately supporting more proactive and informed wildfire management strategies.

VII. Future Work

We plan to expand the multimodal framework by incorporating additional data sources, including real-time satellite imagery and vegetation indices, which could provide richer environmental context for risk assessment. Improving model generalization across diverse Canadian regions remains a priority, particularly given the varied climatic and geographical conditions that influence wildfire behavior. We also intend to explore more advanced architectures beyond the current ViT+LSTM approach. State-of-the-art models such as Swin Transformers, ConvNeXt, and multimodal foundation models offer potential improvements in both accuracy and efficiency. Additionally, integrating large language models (LLMs) could enable the system to incorporate textual weather reports, historical fire documentation, and expert knowledge, creating

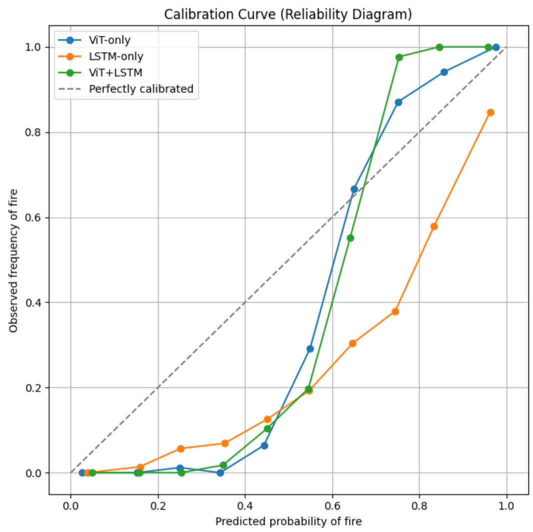

Fig. 8. Calibration curves (reliability diagrams) for ViT-only, LSTM-only, and ViT+LSTM models on wildfire prediction.

a more comprehensive understanding of wildfire risk factors through natural language processing.

References

[1] Natural Resources Canada, "The state of canada's forests: Annual report 2022." https://www.nrcan.gc.ca/, 2022. Accessed: 2025-05-12.

[2] P. Jain, S. C. P. Coogan, S. G. Subramanian, M. Crowley, S. Taylor, and M. D. Flannigan, "A review of machine learning applications in wildfire science and management," *Environmental Reviews*, vol. 28, no. 4, pp. 478–505, 2020.

[3] L. Zhang, M. Wang, Y. Fu, and Y. Ding, "A forest fire recognition method using uav images based on transfer learning," *Forests*, vol. 13, no. 7, p. 975, 2022.

[4] W. Zhu, S. Niu, J. Yue, *et al.*, "Multiscale wildfire and smoke detection in complex drone forest environments based on yolov8," *Scientific Reports*, vol. 15, p. 2399, 2025.

[5] A. Agarwal, J. Gearon, R. Rank, and E. Chenevert, "Fighting fires from space: Leveraging vision transformers for enhanced wildfire detection and characterization," 2025.

[6] S. Natekar, S. Patil, A. Nair, and S. Roychowdhury, "Forest fire prediction using lstm," in *2021 2nd International Conference for Emerging Technology (INCET)*, pp. 1–5, 2021.

[7] H. Liang, M. Zhang, and H. Wang, "A neural network model for wildfire scale prediction using meteorological factors," *IEEE Access*, vol. 7, pp. 176746–176755, 2019.

[8] T. V. S. Krishna, V. Bhargavi, S. Jyothsna, V. Rakesh, D. Bikshapathi, and D. Saritha, "Leveraging radial basis function neural networks with adaptive variability functionality to predict wild fires," in *2023 4th International Conference on Smart Electronics and Communication (ICOSEC)*, pp. 1583–1588, 2023.

[9] C. E. Van Wagner, "Development and structure of the canadian forest fire weather index system," Tech. Rep. Forestry Technical Report 35, Canadian Forestry Service, 1987.

[10] Canadian Wildland Fire Information System (CWFIS), "Fire weather index (fwi) system." https://cwfis.cfs.nrcan.gc.ca/, 2024. Natural Resources Canada.

[11] B. M. Wotton, "Interpreting and using outputs from the canadian forest fire danger rating system in research applications," *Environmental and Ecological Statistics*, vol. 16, no. 2, pp. 107–131, 2009.

[12] B. D. Lawson and O. B. Armitage, *Weather guide for the Canadian Forest Fire Danger Rating System*. Natural Resources Canada, 2008.

[13] B. J. Stocks, M. E. Alexander, B. M. Wotton, C. N. Stefner, M. D. Flannigan, S. W. Taylor, N. Lavoie, J. A. Mason, G. R. Hartley, M. E. Maffey, and G. N. Dalrymple, "Canadian forest fire danger rating system: An overview," *Forestry Chronicle*, vol. 65, no. 6, pp. 450–457, 1989.

[14] A. J. Dowdy, G. A. Mills, K. Finkele, and W. J. de Groot, "Evaluation of the canadian forest fire weather index system as a predictor of fire activity in the australian landscape," *International Journal of Wildland Fire*, vol. 18, no. 8, pp. 881–889, 2009.

[15] X. Wang, D. K. Thompson, G. A. Marshall, C. Tymstra, R. Carr, and M. D. Flannigan, "Increasing frequency of extreme fire weather in canada with climate change," *Climatic Change*, vol. 130, no. 4, pp. 573–586, 2017.

[16] P. Jain, S. C. P. Coogan, S. G. Subramanian, M. Crowley, S. Taylor, and M. D. Flannigan, "A review of machine learning applications in wildfire science and management," *Environmental Reviews*, vol. 30, no. 2, pp. 165–186, 2022.

[17] L. Giglio, W. Schroeder, and C. O. Justice, "The collection 6 modis active fire detection algorithm and fire products," *Remote Sensing of Environment*, vol. 178, pp. 31–41, 2016.

[18] W. Schroeder, P. Oliva, L. Giglio, and I. A. Csiszar, "The new viirs 375 m active fire detection data product: Algorithm description and initial assessment," *Remote Sensing of Environment*, vol. 143, pp. 85–96, 2014.

[19] D. P. Roy, L. Boschetti, and A. Smith, "Satellite remote sensing of fires: Progress and challenges," *Current Opinion in Environmental Sustainability*, vol. 36, pp. 48–57, 2019.

[20] L. Giglio, J. Descloitres, C. O. Justice, and Y. J. Kaufman, "An enhanced contextual fire detection algorithm for modis," *Remote Sensing of Environment*, vol. 87, no. 2-3, pp. 273–282, 2003.

[21] Z. Li, S. Nadon, J. Cihlar, and B. Stocks, "Satellite mapping of canadian boreal forest fires: evaluation and comparison of algorithms," *International Journal of Remote Sensing*, vol. 21, no. 16, pp. 3071–3082, 2000.

[22] USDA Forest Service, "Usda forest service active fire mapping program," 2024. https://fsapps.nwcg.gov/afm/.

[23] C. O. Justice, L. Giglio, S. Korontzi, J. Owens, J. T. Morisette, D. Roy, J. Descloitres, S. Alleaume, F. Petitcolin, and Y. Kaufman, "The modis fire products," *Remote Sensing of Environment*, vol. 83, no. 1-2, pp. 244–262, 2002.

[24] B. M. Wotton, C. A. Nock, and M. D. Flannigan, "Forest fire occurrence and climate change in canada," *International Journal of Wildland Fire*, vol. 26, no. 7, pp. 985–994, 2017.

[25] F. H. Johnston, S. B. Henderson, Y. Chen, J. T. Randerson, M. Marlier, R. S. DeFries, P. Kinney, D. M. J. S. Bowman, and M. Brauer, "Estimated global mortality attributable to smoke from landscape fires," *Environmental Health Perspectives*, vol. 120, no. 5, pp. 695–701, 2012.

[26] L. Giglio, W. Schroeder, and C. O. Justice, "The collection 6 modis active fire detection algorithm and fire products," *Remote Sensing of Environment*, vol. 178, pp. 31–41, 2016.