A Comprehensive Framework Incorporating Hybrid Deep Learning Model, Vi-Net, for Wildfire Spread Prediction and Optimized Safe Path Planning

by

Manavjit Singh Dhindsa

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2025

© Manavjit Singh Dhindsa 2025

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Forest fires are becoming more prevalent than ever, and their intensity and frequency are expected only to increase owing to climate change and environmental degradation. These fires severely threaten the economy, human lives, and infrastructure. Therefore, effective management of wildfires is of utmost importance, and accurately predicting the wildfire spread lies at the core of it. Reliable predictions of fire spread not only provide insights about the at-risk regions but also help in planning several mitigation activities including resource allocation and evacuation planning. This thesis introduces Vi-Net, an innovative hybrid deep learning model, which integrates the localized precision of U-Net with the global contextual awareness of Vision Transformers (ViT) to predict next-day wildfire spread with unprecedented accuracy.

This study utilizes an extensive multimodal dataset that accumulates data from different sources across the United States from 2012 to 2020 incorporating critical factors such as topographical, meteorological, anthropological(population density), and vegetation indices. These elements are vital for modeling the complex dynamics of wildfire spread. A significant challenge in this domain is the class imbalance as the "fire" points are generally quite less compared to "non-fire" points. The dataset used in this study had fire regions less than 5% of the total data. To address this issue, advanced loss functions, including Focal Tversky Loss(FTL), are employed, prioritizing accurate segmentation of fire-prone regions while minimizing false negatives. FTL modifies the focus towards hard-to-predict regions and crucial boundaries, thereby enhancing the model's predictive accuracy and reliability in practical scenarios.

Vi-Net addresses the complexities in modeling fire dynamics by synergizing the strengths of U-Net and ViT. U-Net excels in extracting fine-grained spatial details, crucial for segmenting localized fire areas accurately. In contrast, ViT captures broader spatial relationships through its self-attention mechanism, enabling the model to recognize patterns and dependencies over larger geographical extents. The integration of U-Net and ViT in Vi-Net allows for a comprehensive analysis that ensures high precision and recall, effectively balancing the sensitivity and specificity needed in wildfire predictions. This dual approach allows the model to process both detailed local information and extensive contextual data, making it exceptionally capable of identifying and predicting fire spread across diverse landscapes.

Experimental results highlight the superiority of Vi-Net over traditional models, achieving an F1 Score of 97.25% and an Intersection over Union(IoU) of 94.15% on the test dataset. These metrics highlight its capability to accurately capture localized fire patches

and long-range dependencies while avoiding overprediction. These advancements validate the model's potential to offer more nuanced predictions, capturing the interplay between micro and macro-level environmental dynamics.

In addition to predictive modeling, this research extends its practical applicability by integrating the predicted fire masks into an optimized A* algorithm for safe path planning. This step ensures actionable insights for emergency response teams, facilitating efficient evacuation routes and resource allocation while avoiding high-risk fire regions. Qualitative and quantitative analyses confirm the hybrid model's efficacy, with visualizations demonstrating Vi-Net's ability to preserve spatial detail while capturing broad environmental contexts, and path planning results illustrating the model's robustness and reliability.

This research not only sets a new benchmark for wildfire prediction models but also demonstrates the potential of hybrid deep learning systems in environmental science applications. By providing a robust framework for real-time wildfire management, Vi-Net could significantly influence future strategies in disaster response and resource allocation. Future enhancements could include integrating real-time data feeds to further improve the adaptability and predictive capabilities of the model, potentially revolutionizing wildfire management practices globally.

Acknowledgements

I would like to extend my deepest gratitude to my thesis supervisor, Dr. Kshirasagar Naik, for his unwavering support and patient mentorship throughout this research journey. His expertise has been essential in shaping both the direction and success of my work.

Special thanks to my committee members Dr. Marzia Zaman and Dr. Pin-Han Ho for their insightful feedback and constructive critiques.

My heartfelt thanks to my parents for their constant love, support, and encouragement. Their guidance and sacrifice have been crucial to my journey and have inspired me every step of the way.

I am also grateful to my research group, team members, and colleagues for their collaboration and helpful insights. Your cooperation has been vital to the success of this research. To everyone who played a role in this endeavor, no matter how small, I offer my sincere thanks.

Dedication

This thesis is dedicated to my parents, Narinder and Gurjot, who have always taught me the power of resilience and love, and to the guidance of the higher power.

Table of Contents

Author's Declaration ii

Abstract iii

Acknowledgements v

Dedication vi

List of Figures xi

List of Tables xiii

List of Symbols xiv

1 Introduction 1

1.1 Motivation 1

1.2 Objective 3

1.2.1 Research Questions 5

1.2.2 Contributions 6

1.3 Thesis Outline 8

2 Background & Literature Review **10**
2.1 Traditional Fire Spread Models . . . 10
2.1.1 Empirical Models . . . 11
2.1.2 Physics-Based Models . . . 11
2.1.3 Cellular Automata Models . . . 11
2.2 Forest Fires and Machine Learning . . . 12
2.2.1 Machine Learning Models . . . 12
2.2.2 Deep Learning Models . . . 13
2.2.3 Advanced Prediction Models . . . 14
2.2.4 Hybrid Architectures and Multimodal Data Integration . . . 15
2.3 Wildfire Datasets . . . 15
2.4 Path Planning Algorithms . . . 16
2.4.1 Safe Path Planning . . . 17
2.5 Imbalanced Handling and Performance Metric . . . 18
2.5.1 Imbalanced Handling with Loss Functions . . . 18
2.5.2 Performance Metric . . . 19

3 Problem Formulation **20**
3.1 Task Description . . . 20
3.2 Mathematical Formulation . . . 21

4 Dataset Description and Preparation **24**
4.1 Data Collection and Organization . . . 26
4.1.1 Data Collection . . . 26
4.1.2 Spatial and Temporal Alignment . . . 27
4.1.3 Region Selection . . . 27
4.1.4 Feature Processing and Transformation . . . 27
4.1.5 Data Storage and Format . . . 28

4.1.6 Structured Dataset . . . 28
4.2 Data Preprocessing . . . 28
4.2.1 Random Cropping . . . 28
4.2.2 Filtering Invalid Data . . . 29
4.2.3 Noise Augmentation . . . 29
4.2.4 Standardization . . . 29
4.3 Data Imbalance Handling . . . 30

5 Methodology **34**
5.1 Baseline Image Segmentation Model: UNet . . . 36
5.2 Baseline Image Segmentation Model: ViT . . . 39
5.3 Proposed Hybrid Approach: Vi-Net . . . 41
5.3.1 Vi-Net Architecture and Functionality . . . 41
5.3.2 Benefits of the Hybrid Vi-Net Approach . . . 44
5.3.3 Performance Metrics . . . 48
5.4 Safe Path Planning: Optimized A* Algorithm . . . 50
5.4.1 A* Search Algorithm . . . 50
5.4.2 Safe Path Planning Methodology . . . 53

6 Results & Experiments **57**
6.1 Experiments with Loss Functions . . . 57
6.1.1 Weigted Cross Entropy(WCE) with Logits . . . 58
6.1.2 Dice Loss . . . 58
6.1.3 Tversky Loss . . . 59
6.1.4 Focal Tversky Loss . . . 59
6.2 Wildfire Segmentation Results . . . 61
6.2.1 Quantitative Results . . . 61
6.2.2 Qualitative Results . . . 65

6.3 Safe Path Planning Results 69
6.3.1 Scenarios with Navigable Safe Routes 69
6.3.2 Scenarios with No Safe Routes 72

7 Conclusion & Future Work **74**
7.1 Conclusion 74
7.2 Future Work 76

References **78**

Glossary **85**

List of Figures

1.1 System Model 3

3.1 A Sample extraction of fire location and corresponding features. 22

3.2 Binary grid (extracted from the output of Wildfire Spread Prediction) converted into corresponding Fire Mask for application of Safe Path Planning. 23

4.1 Data Collection Framework 26

4.2 Example images from the dataset showing each of 12 features derived from various sources. 30

4.3 Imbalanced Data Ratio in Next-day Wildfire Dataset 32

5.1 High-Level Overview of Wildfire Spread Prediction and Safe Path Planning Framework. 35

5.2 Workflow Diagram showcasing the complete flow of working of prediction models 37

5.3 UNet Model Architecture. 38

5.4 ViT Model Architecture. 39

5.5 Model Architecture of the hybrid model, Vi-Net. 42

5.6 Overview of the Vi-Net Architecture. 45

6.1 The training and loss function plots for Vi-Net. 62

6.2 Results from three image segmentation models: U-Net, ViT, and Vi-Net. . 67

6.3 This results from the Vi-Net model compares fire boundary of ground truth(GT) and predictions results. 68

6.4 Application of OA* Algorithm on the Wildfire Spread Prediction results. . 69
6.5 Results of Successful Path Generated Based on Optimized A* Algorithm . 71
6.6 Results of No Path Exist Based on Optimized A* Algorithm 73

List of Tables

1.1 Contributions(C) and the corresponding addressed Research Questions(RQs) 7

4.1 Features in Next Day Wildfire Spread dataset chosen for prediction modeling. 25
4.2 Class Ratios Across Datasets 31

5.1 Summary of Prediction Outcomes. 48

6.1 Performance Metrics for Different Models and Loss Functions 60
6.2 Comparison of Performance Across The Models 64

List of Symbols

ΔA Change in fire area, where $\Delta A > 0$ indicates an increase in fire size 84

ϵ A small constant added to prevent division by zero during standardization or normalization 30, 84

$\mathbf{X}$ Input tensor representing the spatial-temporal data with dimensions $H \times W \times C$ 84

$\mathbf{X}_{crop}$ Cropped input tensor with dimensions $a \times a \times C$ 84

μ_{ch} The mean value of a feature channel ch 30, 84

σ_{ch} The standard deviation of a feature channel ch 30, 84

$f_{crop}(\mathbf{X}, a)$ A function that extracts a random square patch of size $a \times a$ from an input tensor $\mathbf{X}$ 84

$\mathbf{D_{NDWS}}$ Next-Day Wildfire Spread dataset used for supervised learning 84

$\mathbf{P_{spread}}$ Predicted spatial wildfire spread map 84

$\mathbf{F}$ Mapping function approximated by the deep learning model; $\mathbf{F} : \mathbf{X} \rightarrow \mathbf{Y}$ 84

$\mathbf{X}$ Input tensor representing environmental variables; $\mathbf{X} \in \mathbb{R}^{H \times W \times C}$ 84

$\mathbf{Y}$ Output tensor representing the predicted binary mask; $\mathbf{Y} \in \mathbb{R}^{H \times W}$ 84

Chapter 1

Introduction

1.1 Motivation

Forest fires, although an essential part of many ecosystems, have become an environmental and socio-economic threat in recent years. Earlier these fires played a crucial role in maintaining the ecological balance by removing dead vegetation and recycling the nutrients allowing regeneration of new flora. However, the increasing frequency of these fires, worsened by climate change, has raised serious concerns about their dangerous effects. For instance, the study in [25] postulates that the annual occurrences of very large fires (> 5000 hectares) will increase by 4 times between 2041–2070 as compared to 1971–2000. The year 2020 alone witnessed more than 4 million ha of land burnt in the United States including California's record-breaking largest individual fire to date, the "August Complex," which consumed over 400,000 ha [11]. The 2020 wildfires caused economic losses estimated at 19 billion and the destruction of approximately 11,116 structures in infrastructural losses [48]. The wildfires in 2020 released nearly 112 million metric tons of carbon and 1.2 million metric tons of $PM_{2.5}$. To put this value in context, this is more than 120 times the combined emissions from all Californian cars, trucks, and buses for the same year [48].

According to studies by Moritz et al. [41], the natural fire regime in many areas have been altered by decades of fire suppression policies, coupled with more recent climate change effects, resulting in more frequent and severe wildfires. This alteration affects biodiversity, water resources, carbon storage, and air quality, with fires becoming larger and more intense. For instance, the fires across the Amazon rainforest and the Australian bushfires of recent years have shown how extensive fires can become, leading to catastrophic biodiversity loss and substantial carbon emissions [9, 5].

From an economic perspective, the immediate costs of wildfires include destruction of property and natural resources, firefighting efforts, and post-fire rehabilitation. Long-term impacts, however, extend further to include loss of tourism revenue, reduction in property values, and increased insurance premiums. The economic burden also encompasses health care costs related to fire-induced air pollution, which affects millions of people globally. In the U.S. alone, the combined costs of fires can reach billions annually, as reflected in National Interagency Fire Center reports which document the escalating expenditures on fire suppression and management [7].

Such apocalyptic incidents make it crucial for scientists, researchers, real estate planners, and fire department agencies to improve upon the methods and practices utilized for: (a) precise and quicker estimation of fire spread prediction, and (b) rapid implementation of mitigation strategies. Accurate prediction of wildfire spread helps address both these objectives efficiently. The prediction of the trajectory of fire spread can help identify the potential areas of risk and its preciseness allows for planning mitigation activities, such as route planning for evacuation purposes or safe passage of first responders, within the affected regions with greater confidence.

Wildfire spread prediction involves modeling the complex relationships between various influential factors and fire danger. As noted by reference [32], the spread trajectory of forest fires is influenced by multiple meteorological factors, anthropogenic, and vegetation factors. These factors include wind speed,and direction, min. and max. temperature, elevation of the land, population density, humidity, and type of crop in the region. Furthermore, the interrelation between these factors significantly impacts the spread of wildfires. For example, fire spreads more rapidly downhill than uphill, contrasting with its behavior on flat terrain. If a fire advancing downhill is propelled by wind in the same direction, the likelihood and speed of fire spread increase, consequently reducing the response time available to first responders. Conversely, if the wind opposes the fire's direction, or if precipitation occurs, the probability of fire advancement through the area may significantly decrease or even cease entirely.

The integration of new technologies and data-driven approaches, such as remote sensing and deep learning (DL), is revolutionizing fire spread prediction. Machine learning (ML) presents a promising research area for addressing the challenges associated with forest fire spread prediction and management. Traditional methods have often fallen short due to the scale and complexity of forest fires. ML algorithms are capable of analyzing vast and intricate datasets and offer more accurate and timely predictions. However, with the advancement in ML technology, emerging algorithms like neural networks and deep learning (DL) techniques now make it possible to accurately predict fire spread patterns considering various meteorological, topographical, and environmental factors. Deep

learning (DL) based methods exhibit better and more significant accuracy in terms of fire spread prediction [4]. Considering the massive environmental and socio-economic losses, there is a pressing need to develop deep learning (DL) based approaches for wildfire spread prediction.

1.2 Objective

The primary objective of this thesis is to develop a sophisticated model for wildfire spread prediction using advanced machine learning techniques. Given the focus on supervised learning, this study aims to predict the precise spread of wildfire by formulating the problem as a semantic segmentation problem. This model is designed to leverage multimodal dataset: Next-Day Wildfire Spread dataset, [29] derived from satellite data, weather station data, and other environmental sources, integrating these into a comprehensive system that predicts the spatial-temporal dynamics of wildfire spread. The additional goal is to enhance the accuracy of wildfire predictions compared with state-of-the-art baseline models: U-Net and Vision Transformers(ViT). The study also aims to develop a reliable framework for generating safe evacuation routes, with a modified A* pathfinding algorithm, using the predictions generated by spread prediction algorithm. Figure 1.1 illustrates the integrated approach undertaken in this research.

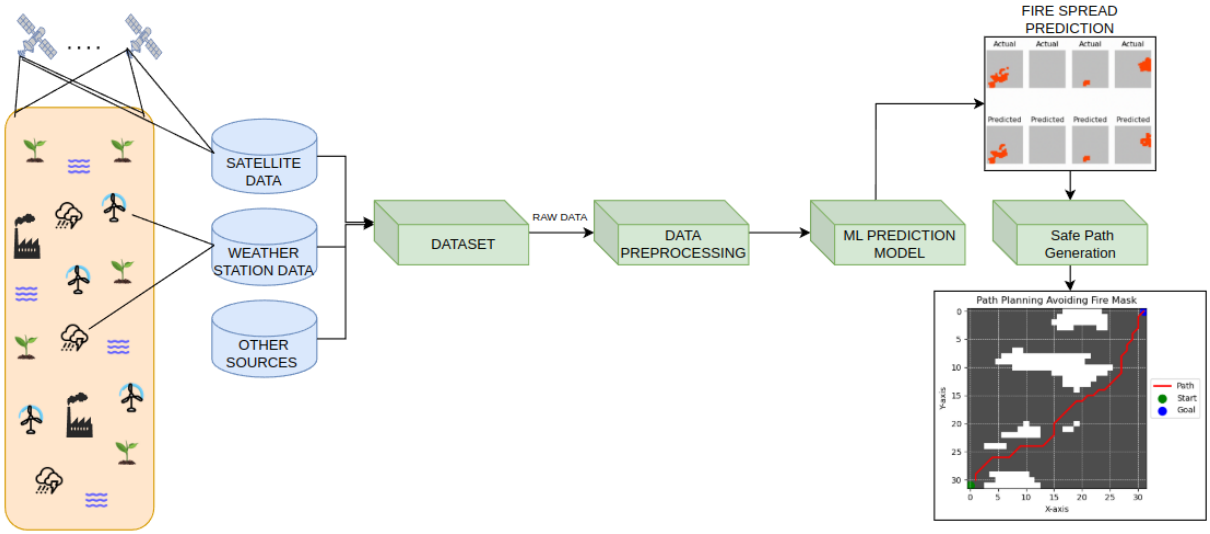


Figure 1.1: System Model

The Next-Day Wildfire Spread dataset is characterized by a significant imbalance, wherein the occurrence of fire events is considerably lower than that of non-fire events. The challenge associated with such an imbalanced dataset is that models trained on this data are prone to bias towards the majority class, consequently neglecting the minority class, which in this context is represented by fire events. Therefore, it is imperative to devise a methodology to address the imbalance present in the Forest Fire dataset.

Specific Objectives:

1. Data Integration and Preprocessing

- To implement robust preprocessing methods such as standardization, handling of missing values, and data augmentation to enhance model training and prediction accuracy.

2. Hybrid Modeling for Enhanced Prediction Accuracy

- To develop and validate a hybrid model, Vi-Net, which combines the localized precision of U-Net with the contextual awareness capabilities of Vision Transformers (ViT).
- To mathematically model the hybrid approach ensuring it leverages both local and global data features effectively. The model is represented as:

$$F_{\text{Vi-Net}} = \alpha \times f_{\text{U-Net}} + \beta \times f_{\text{ViT}} \tag{1.1}$$

  where $\mathbf{f}_{\text{U-Net}}$ and $\mathbf{f}_{\text{ViT}}$ are outputs from U-Net and ViT and α and β balance each model's contribution.

3. Handling Imbalanced Data

- To employ and compare advanced loss functions including Weighted Cross Entropy, Dice Loss, Tversky Loss, and Focal Tversky Loss, to address the class imbalance inherent in Next-Day Wildfire Spread dataset.

4. Route Planning Using Predictive Models

- To integrate the predictive outputs of the wildfire spread models into a route planning algorithm, specifically a modified A* algorithm, to enhance emergency response strategies.
- To mathematically formulate the route planning as a pathfinding problem in a grid where each cell's passability is determined by the predicted fire spread, employing a heuristic such as Euclidean distance to optimize the path.

5. Model Evaluation and Validation

- To rigorously evaluate the model using metrics such as Precision, Recall, F1 Score, and Intersection over Union (IoU) to assess its performance across different scenarios and datasets.
- To validate the model's generalization ability on unseen data and across diverse geographical regions, ensuring its applicability in real-world scenarios.

6. Theoretical and Practical Contributions to Wildfire Management

- To enhance the understanding of how accurately machine learning techniques can perform on real-world fire spread dataset, thereby contributing to the precision of wildfire spread predictions.
- To provide practical tools for wildfire risk assessment and emergency planning, thereby contributing to safer and more effective management practices.

1.2.1 Research Questions

The following are the research questions identified for our research work:

RQ1. How can machine learning methods be used to predict the spread of forest fires using computer vision?

RQ2. How can hybrid modeling techniques improve the accuracy of wildfire spread predictions?

RQ3. What are the limitations, if any, of integrating diverse remote-sensing data on the performance of wildfire spread prediction models?

RQ4. How do local and global contextual data contribute to the effectiveness of semantic segmentation in predicting wildfire-prone areas and what contribution does each of U-Net and ViT model provide?

RQ5. How do we deal with imbalanced data in multimodal datasets for wildfire predictions and how do various loss functions impact the performance of models trained on multimodal wildfire datasets?

RQ6. How effectively do predictive models generalize to new and unseen data and across different geographical regions?

RQ7. What metrics effectively demonstrate and compare the performance of the hybrid model over baseline models in wildfire spread prediction tasks?

RQ8. How can wildfire spread predictions be used in pathfinding to enhance emergency response strategies?

RQ9. What are the key considerations when designing algorithms for route generation in wildfire-affected areas?

1.2.2 Contributions

This research aims to bridge ga The contributions of this study can be summarized as follows:

- C1: We propose **Vi-Net**, a novel hybrid model combining the strengths of U-Net and Vision Transformer (ViT) to address the critical task of next-day wildfire spread prediction. By framing the problem as a semantic segmentation task, Vi-Net captures both local spatial details and global contextual relationships, enabling precise and interpretable predictions of fire-prone regions. This hybrid approach not only enables precise and interpretable predictions of fire-prone regions but also demonstrates superior performance in real-world applications. We rigorously tested Vi-Net on a multimodal wildfire dataset, comparing it against two state-of-the-art vision models, U-Net and ViT, to underscore its effectiveness in achieving a balance between fine-grained segmentation accuracy and global context capture.

- C2: To address the challenges of data imbalance and preprocessing in complex multimodal datasets, we incorporate tailored data handling strategies, including advanced

augmentation techniques and loss function adjustments, ensuring robust learning despite the inherent class imbalance and variability in the wildfire data. We also compare the performance of 4 different loss functions widely used for semantic segmentation tasks.

- C3: Experimental results and comprehensive analysis demonstrate that Vi-Net achieves superior performance over baseline models. This enhanced predictive capability establishes Vi-Net as a reliable and impactful tool for crisis management and resource allocation in wildfire-prone regions.

- C4: Extend the application of the Vi-Net model by developing a sophisticated pathfinding algorithm, which employs outputs from the wildfire spread prediction to compute optimal safe route(s). The process involves meticulous preparation of inputs incorporating fire spread scenarios, followed by the application of the pathfinding algorithm to ascertain routes that avoids exposure to active fire zones. The algorithm outputs a detailed representation of route generated by our pathfinding algorithm across the predictions of Vi-Net model, enhancing the utility of the model for real-time strategic decision-making in emergency response operations.

The experiments show that the F1 Score of our hybrid model, Vi-Net, is highest scoring at 97.2% surpassing state-of-the-art vision models: U-Net and ViT. The best performance of all the loss function among the four loss function compared: Weighted Cross Entropy, Dice Loss, Tversky Loss, and Focal Tversky Loss, was Focal Tversky Loss.

Table 1.1: Contributions(C) and the corresponding addressed Research Questions(RQs)

Contribution	**Research Question**
C1	RQ1, RQ2, RQ4
C2	RQ3, RQ5
C3	RQ6, RQ7
C4	RQ8, RQ9

Table 1.1 organizes and outlines the connections between the key contributions of the thesis and the corresponding research questions they address. Contribution 1 (C1) addresses Research Questions RQ1, RQ2, and RQ4, offering a foundation for understanding the use of machine learning in forest fire spread predictions through hybrid models. Contribution 2 (C2) relates to RQ3 and RQ5, exploring the integration of diverse data and the handling of data imbalances in wildfire modeling. Contribution 3 (C3) focuses on RQ6 and RQ7, demonstrating the generalization and performance measurement of predictive models

across different environments. Finally, Contribution 4 (C4) tackles RQ8 and RQ9, which delve into the application of predictive models in emergency route planning and algorithm design for navigating through wildfire-affected areas. This structured approach ensures a comprehensive coverage of the various facets of wildfire prediction and management.

1.3 Thesis Outline

This thesis is organized to offer a comprehensive understanding of wildfire spread prediction using advanced machine learning models. It is structured as follows to provide a coherent flow of investigation from initial concept to final conclusions:

Chapter 2 offers a scholarly examination of the existing research in wildfire prediction. It surveys significant advancements and methodologies while identifying critical gaps in the field. This review sets the stage for the subsequent chapters by framing the current state of knowledge and technology in wildfire spread analysis and prediction.

Chapter 3 formally postulates the semantic segmentation task both intuitively and mathematically. This chapter outlines the complexity of wildfire dynamics, influenced by a network of environmental, meteorological, and anthropogenic factors. It frames the wildfire prediction challenge as a one-day-ahead forecasting problem, leveraging the temporal and spatial interactions of features.

Chapter 4 delves into the composition and preparation of the 'Next Day Wildfire Spread' dataset used for training the predictive models. It explains how data from various sources are integrated and aligned to form a robust dataset for model training. The chapter also discusses the spatial and temporal characteristics of the data, explaining how these aspects are crucial for accurate prediction modeling.

In Chapter 5, the focus shifts to the specific methodologies employed in the research. It begins with an overview of the modeling frameworks utilized, followed by detailed descriptions of the data collection, preprocessing, and the machine learning models deployed. Special emphasis is placed on the unique integration of U-Net and Vision Transformers to form the novel Vi-Net model, illustrating how this hybrid approach enhances prediction accuracy.

Chapter 6 presents detailed results from the experiments conducted are presented in this chapter. It compares the performance of the baseline models (U-Net and ViT) with the hybrid Vi-Net model across various metrics such as precision, recall, and F1-score. This chapter not only highlights the quantitative results of the models but also provides

qualitative insights through visual representations of the wildfire prediction maps generated by each model.

The final chapter synthesizes all findings from the research, discussing the implications and the significance of the results obtained. It evaluates the success of the hybrid Vi-Net model against established models and discusses the potential impact of this research on real-world wildfire management and prediction. Lastly, this chapter suggests future research directions and potential improvements in the field of wildfire spread prediction.

This structure ensures a logical progression of content, facilitating a detailed and systematic exploration of the research topic.

Chapter 2

Background & Literature Review

This chapter discusses the evolution of wildfire prediction models, from traditional empirical and physics-based models to advanced machine learning and deep learning approaches. Predicting and managing wildfires have dramatically improved with advances in data-driven approaches, utilizing modern machine learning (ML) and deep learning (DL) models. This evolution signifies a shift from empirical and physics-based models to more dynamic, data-informed strategies capable of addressing the complexities associated with predicting and managing wildfires. It reviews literature on the advancement in fire spread models, and their methodologies and showcases a transition from the use of simulation models towards integration of multimodal data in learning models for enhanced predictive accuracy.

This section covers various existing wildfire datasets that can be used for various tasks like fire detection, fire spread/growth prediction, burned area analysis, and much more. A highly common problem of class imbalance and their evaluation metrics were also discussed in this literature review. In this section, we also discuss various categories of path-planning algorithms. The review was focused on graph-based algorithms and further provided an emphasized discussion on route navigation in disaster scenarios, such as wildfires, or the presence of moving obstacles in a dynamic environment, leading to establishing what constitutes as safe path planning.

2.1 Traditional Fire Spread Models

Researchers have always tried to predict the nature of fire spread accurately to effectively plan for wildfire management. Before the advent of robust ML and DL models, several

models were developed to capture the fire spread dynamics. The prediction results serve as the basis for various mitigation strategies, from real estate planning to resource allocation, especially in fire-prone areas.

2.1.1 Empirical Models

Empirical models rely heavily on accumulated data from historical wildfire events to predict fire behavior under similar future conditions. The groundbreaking work by Rothermel established a quantitative approach for predicting the rate of spread and intensity of wildfires, utilizing fundamental fuel and environmental parameters. This model has laid the groundwork for numerous subsequent studies and applications in wildfire prediction [3]

2.1.2 Physics-Based Models

Mathematical models provide a framework for simulating wildfire spread by incorporating equations that represent physical processes. The BEHAVE system, as detailed by Andrews (1986), exemplifies this approach, offering a comprehensive suite of predictive tools that are essential for both planned burns and real-time firefighting operations [2]. Finney's development of FARSITE illustrates the application of these principles, integrating spatial and temporal factors to predict fire growth across landscapes [17]. FlamMap [18] have utilized semi-empirical methods, blending observed fire behavior with theoretical frameworks to enhance landscape-scale prediction capabilities.

2.1.3 Cellular Automata Models

Among the innovative approaches in wildfire modeling, cellular automata provide a unique way to simulate the discrete and stochastic nature of fire spread. Karafyllidis and Thanailakis (1997) introduced a model that uses cellular automata to predict fire dynamics, incorporating complex interactions between cells that represent different parts of the forest [33]. Similarly, Hernández Encinas et al. (2007) proposed a hexagonal cellular automaton model that enhances the spatial resolution of fire front simulation, offering more accurate predictions of fire spread [28].

2.2 Forest Fires and Machine Learning

In recent years, machine learning techniques and algorithms have vastly improved and shown great potential in transforming the field of wildfire spread prediction. This development has led to more nuanced Machine Learning and Deep Learning models that can capture the complex dynamics and other intricacies of wildfire propagation with high accuracy and precision. This literature review aims to examine the existing body of research on the advancement of different Machine Algorithms employed by different research works, identifying key challenges and opportunities in the wildfire spread domain.

2.2.1 Machine Learning Models

Recent advancements in the robustness of machine learning models have gained significant interest from researchers working in the forest fire domain. [32] highlights understanding and application of ML in addressing the complex challenges associated with wildland fires. In particular, [32] includes a section on methods of fire behavior prediction including fire spread and growth highlighting the development and advancement of fire spread simulators and models over time.

Researchers have explored a wide range of ML models for wildfire spread prediction including popular models like Random Forest(RF), K-Nearest Neighbours(KNN) and Bayesian Networks(BN). Markuzon et al. did a study [40] wherein they experimented with RFs, BN, and KNN to predict the spread of fire after 1 or 2 days after the observation was made and found each of these ML models performed similarly, and with Random Forest accurately classifying over 75% cases. [30] presented the task of wildfire spread prediction as an image segmnetation problem and compared the performance of Logistic Regression and RF with a neural network-based model called Convolutional Autoencoder and noted that the neural network-based model was better at capturing spatial information in the image dataset than the ML models.

SVMs have been utilized to classify areas based on their susceptibility to fire, drawing on historical fire occurrence data to train predictive models [12]. Gradient Boosting are powerful ensemble learning models known for their predictive accuracy in wildfire risk assessment by sequentially correcting errors of weak learners, thus strengthening the model with each iteration [63].

2.2.2 Deep Learning Models

The integration of deep learning methods has transformed wildfire prediction, introducing models that efficiently process large datasets to identify complex patterns. With the increased robustness of these models in recognizing patterns and processing images, researchers are moving in the direction of forecasting the fire spread by formulating it as an image-processing problem. Convolutional Neural Networks (CNNs), known for their effectiveness in image processing, have been adapted for analyzing high-resolution satellite imagery to predict wildfire perimeters, achieving significant performance improvements over traditional models. For instance, a CNN-based model outperformed the FARSITE model by approximately 20% in accuracy [24][44]. [38] implemented a multi-temporal CNN to utilize and model data available at different temporal resolutions to increase predictive accuracy. Notably, deep learning techniques, specifically CNN-based variants like U-Net [34] and DeepLab [8], have been pivotal in advancing fire detection and segmentation, offering significant improvements over traditional image processing techniques.

Fire spread is a spatio-temporal dynamics that weigh equally on temporal dynamics as much as on spatial features. Therefore, the ability to model the spread of fire with respect to time is a critical avenue. While CNN and its variant are great at modeling the spatial dynamics, they have limited abilities in terms of capturing temporal dependencies. Recurrent Neural Networks (RNNs) and their popular variant, Long Short-Term Memory Networks (LSTMs), have been critical in modeling the temporal dynamics of wildfires. [44] employs LSTM for temporal predictions of wildfire spread on geospatial data like weather and temperature.

Another important technique employed is the use of the "attention" mechanism. Attention-based architectures integrate spatial and channel attention to enhance focus on critical features. [64] used attention modules with a CNN model architecture to enhance feature representation at both spatial and channel levels. The performance of the attention-based models is validated through a comparative analysis with benchmark systems, including a more sophisticated autoencoder architecture and the FARSITE semi-empirical fire behavior model, showcasing the efficacy of attention mechanism in this domain. [19] introduced a custom attention-based U-Net architecture, demonstrating state-of-the-art performance. Interestingly, this architecture also improved training time by a significant measure while utilizing fewer computational resources which is critical for faster and accurate results in real-time scenarios.

2.2.3 Advanced Prediction Models

In recent years, advanced deep-learning(DL) models have significantly improved the accuracy and efficiency of wildfire spread prediction. These models utilize complex neural network architectures to process multimodal data and capture the intricate spatiotemporal dynamics of fire propagation. Another important DL architecture that has gauged the interest of research community is the Transformers. They are particularly used in areas where contextual or "memory" is required to learn and remember the previous tokens for new predictions. These models have taken the NLP domain by storm by showing prowess in capturing long-range dependencies and enabling parallel processing significantly improving training and inference speed. Transformers can effectively process larger input sequences compared to traditional models like RNNs, which tend to forget earlier inputs in long sentences.

Recent advancements in transformers have seen the adaptation of Vision Transformers (ViTs) for environmental and disaster management applications, including wildfire prediction. The studies cited in [22] highlight how ViTs and related architectures like Swin Transformers and SegFormer are tailored to improve upon traditional CNN models by better handling the global context within visual data. This capability is particularly advantageous in scenarios like wildfire spread where the spatial extent and temporal dynamics are critical for accurate predictions. As wildfires are influenced by complex interactions between local conditions and global climate patterns, models that incorporate teleconnections offer a promising avenue for predictions at sub-seasonal to seasonal scales. The work presented in [43] explores the use of teleconnection-driven transformers that integrate global climatic factors to predict wildfire activity. This approach acknowledges the interconnectedness of global climate systems and their impact on local wildfire occurrences, providing a more holistic view of fire dynamics. This approach has been supported by research indicating the predictive power of global climatic indices on wildfire occurrences [61, 52]. Chen et al. [61] use simple autoregressive statistical methods to combine volume pressure deficit values with oceanic indicators. Yu et al. [52] use a statistical pre-processing to identify the most prominent oceanic indicators modulating burned areas in Africa and then use the findings to select input features for tree-based machine learning models.

Emerging models like Earthformer and Climax use advanced computational techniques to simulate Earth system dynamics, providing a framework that captures both the complexity and scale necessary for effective wildfire prediction [20, 42]. Earthformer [20] combines spatio-temporal attention with learnable global vectors that are meant to summarize the dynamics of a system. Climax [42] is proposed as a transformer-based foundational model for weather and climate at coarse resolutions, working with global input or on a very

large spatial domain. GraphCast [36] uses graph neural networks that capture long-range interactions by defining graphs at several resolutions.

2.2.4 Hybrid Architectures and Multimodal Data Integration

Hybrid architectures, developed in recent years, combining different neural network architectures have emerged as powerful tools to to handle wildfire spread prediction. Often, the combination of two or more models helps combat the limitations of individual models. Task-specific hybrid models in particular prove to be a great asset, as in the case of fire spread prediction, as the the combination of the model can be developed specifically to the expected output of the task. This review examines several key hybrid architectures and their contributions to the field of fire spread prediction.

Marjani, et al. introduces a hybrid model, CNN-BiLSTM [39], which combines Convolutional Neural Networks (CNNs) with Bidirectional Long Short-Term Memory (BiLSTM) networks. This approach utilizes the CNNs to extract spatial features and and BiLSTM to capture temporal dependencies in the wildfire spread patterns. [50] uses the MA-Net model, originally developed for tumor segmentation, for predicting wildfire spread direction and speed. The MA-Net model takes advantage of a multi-scale attention network to handle complex spatial relationships making it particularly suitable for analyzing the multifaceted nature of wildfire propagation.

Alliare et al.[1] introduced a hybrid architecture design of a Deep Neural Network(DNN) that handles both 2D fields and scalar inputs. This model is trained on a large multimodal dataset of simulated fire scenarios, and demonstrated excellent performance, explaining 94% of the variance in the output. Xu et al. integrated GIS data with ML-based cellular automata hybrid architecture to simulate fire dynamics, achieving improved accuracy in modeling fire spread across varying terrains. This hybrid model incorporated localized transformation rules derived from least squares support vector machines (LSSVM) for better precision [59].

2.3 Wildfire Datasets

Extensive research efforts have been dedicated to finding and compiling the data from different sources into one place under a common data type so that the fire spread model can capture the exact interplay of various influencing factors like wind speed, direction, min. and max. temperature, humidity, vegetation indices, etc. mirroring the real-world complex

dynamics of wildfire spreading. The development and refinement of such datasets are crucial for advancing the capabilities of prediction models that aid in wildfire management. There is no set benchmark dataset in the field of fire prediction or fire spread prediction. However, we researched and reviewed wildfire datasets that exist in the domain of forest fire or wildfire spread prediction while assessing each for our specific task of wildfire spread prediction.

WildfireDB, [51], is a comprehensive dataset integrating historical fire data with weather, vegetation, and topographical factors across the continental USA, aiming to support data-driven wildfire spread models. WildfireSpreadTS provides a multi-modal, multi-temporal dataset that captures the dynamics of active wildfires at a high temporal resolution, facilitating the development of predictive models for active fire spread [21].

Mesogeos, [35], focuses on the Mediterranean region, offering a dataset that includes meteorological conditions, vegetation characteristics, and human activities to support various machine learning applications in wildfire modeling. Next Day Wildfire Spread, [30], combines 2D fire data with multiple explanatory variables such as elevation, vegetation, and wind speed, and direction, specifically curated to predict next-day wildfire spread.

These datasets are instrumental in training advanced machine learning models capable of predicting the spread of wildfires. However, these datasets have challenges of their own such as resolution constraints, class imbalance, missing data, and label noise, which can significantly impact the prediction results of the models. After thorough review and assessment, [30] was chosen for the task of wildfire spread prediction. This dataset compiles data from various sources and provides an excellent basis for a nuanced understanding of the complex dynamics of fire spread.

2.4 Path Planning Algorithms

Path planning is a crucial problem in various science fields ranging from route navigation and development of travel and delivery apps to robotics, autonomous vehicles, and artificial intelligence. The core challenge is to determine an optimal path from a starting position to a target destination while considering constraints such as obstacles, dynamic environments, and computational efficiency. Over the years, researchers have developed a variety of algorithms like graph-based, sampling-based, optimization-based algorithms, and many more to address this challenge. This literature review explores key graph-based path-planning algorithms and then delves into usage of such algorithms in the field of safe path planning in dynamic environments such as forest fires.

Dijkstra, in 1959, gave one of the most fundamental and widely used path planning algorithm of all time, later named as Dijkstra's Algorithm [13]. This algorithm computes the shortest path from a start node to all other nodes in a weighted graph. This algorithm always gives an optimal route. However, Dijkstra's algorithm is computationally expensive for large graphs, as it explores all nodes regardless of their proximity to the target.

A* algorithm, another hugely popular path planning algorithm, is built upon Dijkstra's algorithm and incorporates heuristics to guide the direction of the path towards a defined target or destination node. This significantly reduced computational overhead, which was the only concern in the Dijkstra's algorithm. A* also achieves optimal path given the condition that the heuristic function is admissible and consistent [27].

2.4.1 Safe Path Planning

Safe path planning, in the context of scenarios such as forest fires or the presence of moving obstacles, presents a critical challenge. The route estimation gets tricky here as the dynamic nature of such environments prevents pre-planning of the paths and thus, the logic to avoid moving obstacles or generate a safe path in a dynamic environment must ensure adaptability and robustness against such scenarios. The algorithm must be optimal for navigation while minimizing the risk exposure. This review discusses the literature on safe path planning, particularly for first responders in scenarios involving dynamic hazards such as fires, toxic plumes, or floods, and the risk profile of available equipment.

Wang et al. [56] addressed the challenge of safe path planning in an environment containing moving hazards with uncertain boundaries. They proposed a spatial-temporal data model to represent the nature of obstacles and introduced a modified A* algorithm for safe route navigation. This approach accounts for varying risk levels and integrates constraints such as responder profiles and equipment limitations. The system efficiently generates safe routes by categorizing roads into "safe," "partially safe," and "unsafe" based on risk modeling, demonstrating applicability in real-time dynamic environments.

A modified Dijkstra algorithm was proposed by [55] that incorporates the dynamic properties of obstacles (e.g., plume concentration or flood depth) and defines safety criteria based on the risk tolerances of first responders. Routes are optimized for minimizing risk and travel time, even when the destination lies within the hazard zone calculated based on risk profile of the equipment. This approach successfully navigates complex scenarios. In the context of fire emergencies, [59] proposed an improved A* algorithm that incorporates traffic efficiency and timeliness in road-free environments.

2.5 Imbalanced Handling and Performance Metric

During the analysis of our research, a significant research gap was found regarding the class imbalance in the dataset in the forest fire domain. The prevailing approach for data collection does not inherently cater to this issue of class imbalance. The general trend is that "fire" points tend to be a lot less in number compared to "non-fire" fire points. This in turn leads to the model not being able to correctly measure true points and simple performance metrics like accuracy reflect a false number of accuracy of the model which is not a reflection of the true scenario [57].

For instance, a scenario with fire to non-fire points in the ratio of 1:10 shows an imbalance in the dataset. However, if our model predicts all the points as non-fire then performance metrics, such as overall accuracy(OA) would suggest our model has a 90% of accuracy which is not a true reflection of the nature of our data. To avoid this issue, several techniques have been employed such as oversampling, undersampling, and choosing an equal number randomly from the major class to match the number of minor classes as shown in studies [37].

2.5.1 Imbalanced Handling with Loss Functions

None of these techniques work without generating or removing data points which takes away the "real-world" reflection of the true nature of areas affected by forest fire. Loss functions play a pivotal role in such scenarios. Loss functions not only help better train the model but also put more emphasis on minor class and penalize the model more if they incorrectly classify the minor class data points which increases the model's performance and handles the imbalance while reflecting the true nature of forest fires [19].

The study done in [50] proposed an integration of two loss functions, Dice Loss and Focal Loss, to handle data imbalance. The combined loss function approach promoted greater diversity in model predictions and enhanced the model's ability to accurately predict large fires. [19] employed and compared the Weighted Binary Cross-Entropy (BCE) and Focal Loss concerning performance as well as computation efficiency. Focal Loss dynamically adjusts the loss contribution of each example, focusing on misclassified or uncertain samples while Weighted BCE applies fixed weight to minority class. It was also noted that Focal Loss dramatically reduced training times compared to Weighted BCE, making it a more suitable choice for large, imbalanced datasets.

2.5.2 Performance Metric

Evaluating the effectiveness of machine learning models on imbalanced datasets requires careful consideration of performance metrics, as highlighted in the work of researchers such as [57], [44], and [60]. Within the domain of forest fire prediction, certain metrics, particularly recall, are of critical importance. Sensitivity assesses the model's ability to accurately identify fire events in relation to the total number of actual fire occurrences. Achieving a high precision and recall score is essential for decision-makers, as it provides valuable insights into resource allocation and budgeting for fire mitigation strategies. Equally important is the ability to pinpoint areas where fires are unlikely, aiding in more efficient planning and resource deployment. These metrics quantify the performance of the model and the scores help in determining the confidence on the results of the model.

The study conducted in [45] highlights different metrics for segmentation. The discussion on evaluation metrics for semantic segmentation is well-supported in the literature. Both F1 Score and Intersection over Union (IoU) are widely used due to their complementary strengths in assessing segmentation performance.

F1 Score

The F1 Score is defined as the harmonic mean of precision and recall. It ensures a balanced evaluation of a model's performance. Precision measures the proportion of correctly predicted pixels among all predicted pixels, while recall quantifies the proportion of correctly identified pixels among all actual pixels. The F1 Score is useful in datasets with class imbalances, like most of wildfire datasets, where certain objects occupy significantly smaller areas in an image. By considering both false positives and false negatives, the F1 Score mitigates biases toward high precision or high recall, providing a more reliable measure of segmentation quality.

Intersection over Union (IoU)

IoU is also referred to as the Jaccard Index. It is a widely used metric in semantic segmentation, defined as the ratio of the intersection of the predicted segmentation mask and the ground truth mask to their union. This metric provides a direct spatial accuracy check, ensuring that the predicted regions align well with actual object boundaries.

Chapter 3

Problem Formulation

This chapter describes the problem of predicting wildfire spread in a structured manner, integrating various environmental, meteorological, and anthropogenic factors into a predictive model. The complexity of modeling wildfire dynamics is articulated, with a focus on the mathematical formulations that support the predictive tasks. Path Planning is a vital task, especially in the context of wildfires ranging from planning evacuation routes and resource allocation to ensuring the safety of first responders. This chapter also describes the task of safe path planning in fire-affected areas and the integration of wildfire spread prediction with a safe path planning module for automated safe path planning. This sets a clear framework for the subsequent chapters, which delve into dataset preparation and model development.

3.1 Task Description

Wildfire spread prediction is a vital task, critical for safeguarding lives, property, and ecosystems. The temporal and spatial dynamics of wildfires are shaped by interdependent environmental variables, meteorological, and anthropogenic factors. Addressing the challenge of accurately forecasting wildfire spread requires understanding how these factors interact over time and space. Wildfires evolve dynamically, necessitating predictive models capable of handling high-dimensional, multivariate dataset, and temporal datasets.

Time-series data are inherently temporal, containing sequential observations where current and past states influence future outcomes. In the context of wildfires, time-series modeling reflects how environmental variables like temperature, wind speed, vegetation,

and humidity evolve and interact to drive the fire's spatial propagation. These variables form a causative chain, where changes in one factor (e.g., wind speed) directly affect others, influencing the fire's spread rate and direction.

Causality is central to the predictive task, as the dynamics of wildfires are governed by cause-effect relationships across multiple modalities. For instance, a decrease in humidity increases vegetation dryness, raising fire susceptibility. Similarly, strong winds can cause rapid fire spread by carrying embers over long distances. To accurately model these phenomena, it is essential to capture both temporal dependencies (e.g., today's weather influencing tomorrow's fire behavior) and spatial correlations (e.g., wind patterns affecting nearby regions).

Framing wildfire prediction as a one-day-ahead forecasting problem leverages the causality embedded in time-series data. This approach harnesses historical and current observations to estimate the fire's spatial extent for the next day. The sequential nature of the task aligns with time-series principles, emphasizing the predictive relationship between temporal inputs and future outcomes.

This research work also aims to integrate wildfire spread prediction with safe path planning to demonstrate the application of the results achieved for mitigation activities. This task is done by integrating the wildfire spread prediction module with the safe path planning module wherein the output of the spread prediction module is the input for the safe path planning module.

3.2 Mathematical Formulation

We represent the wildfire prediction problem as a spatial-temporal task by structuring the input data into a three-dimensional tensor $\mathbf{X} \in \mathbb{R}^{H \times W \times C}$. Here, H and W define the spatial dimensions of the region of interest (height and width), while C denotes the number of feature channels corresponding to environmental variables like wind speed, temperature, and vegetation indices. Mathematically, the input tensor can be expressed as:

$$\mathbf{X} \in \mathbb{R}^{H \times W \times C} \tag{3.1}$$

where each element $x_{(h,w,f)} \in \mathbf{X}$ represents the value of feature f at spatial location (h, w).

The objective is to predict a binary segmentation mask $\mathbf{Y} \in \mathbb{R}^{H \times W}$, where each element $y_{(h,w)}$ indicates the presence (1) or absence (0) of fire at location (h, w) for the following day.

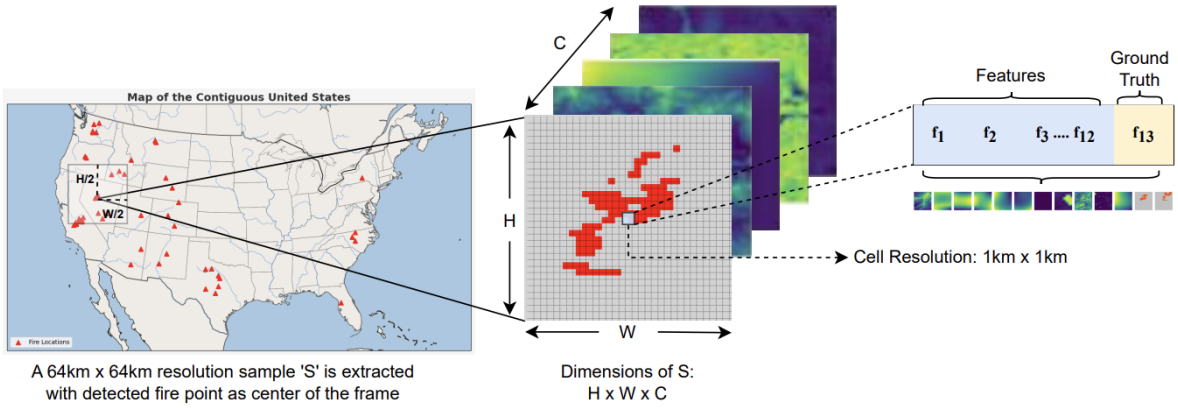


Figure 3.1: A Sample extraction of fire location and corresponding features.

This task encapsulates the spatial-temporal dynamics of wildfire propagation, enabling informed decision-making for emergency response and resource allocation:

$$\mathbf{Y} \in \mathbb{R}^{H \times W} \tag{3.2}$$

The relationship between $\mathbf{X}$ and $\mathbf{Y}$ is defined through a mapping function $\mathbf{F}$, parameterized by a deep learning model. This mapping captures the complex interdependencies and temporal patterns within the input features:

$$\mathbf{F} : \mathbf{X} \rightarrow \mathbf{Y} \tag{3.3}$$

As illustrated in Fig. 3.1, each sample corresponds to a 64 km $\times$ 64 km grid, extracted at 1 km $\times$ 1 km resolution. These grids are centered on a detected fire point and enriched with features like weather indices, vegetation metrics, and topographic data, serving as inputs to the model. The safe path planning task comprises converting the fire predictions into a binary mask and applying the optimized A* algorithm(OA*) on the binary grid representing a fire mask. Fig 3.2 shows an example where a complex spiral-shaped fire mask is converted into a binary grid, which acts as a platform for the application of the OA* algorithm and converted to a corresponding fire mask with pixel values as 0 or 1 representing "fire" or "no fire".

To address the unique challenges of wildfire spread prediction, we propose a hybrid model, Vi-Net, integrating the strengths of U-Net and ViT. U-Net excels at capturing

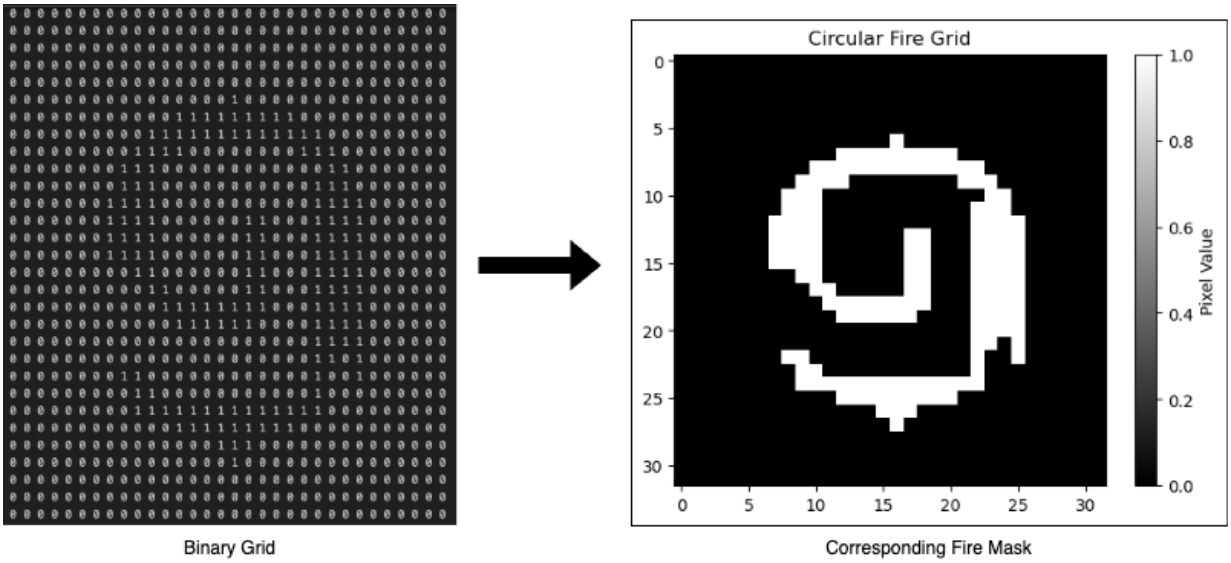


Figure 3.2: Binary grid (extracted from the output of Wildfire Spread Prediction) converted into corresponding Fire Mask for application of Safe Path Planning.

local spatial details, essential for precise boundary segmentation, while ViT provides global contextual understanding, crucial for identifying broader fire propagation patterns. This combination ensures a robust balance between fine-grained accuracy and global spatial coherence.

By grounding our formulation in time-series data principles and leveraging causality, the proposed framework aims to enhance the interpretability and reliability of wildfire predictions, paving the way for actionable insights in fire management. In subsequent sections, we detail the data preprocessing methods employed to construct the input tensor $\mathbf{X}$, the strategies used to mitigate class imbalance, and the design of the hybrid Vi-Net model for predicting $\mathbf{Y}$. This formulation underpins our broader goal of enhancing the reliability of wildfire spread predictions.

Chapter 4

Dataset Description and Preparation

This chapter provides a detailed account of the dataset used in this study, including the sources, collection methods, and preparation processes. It discusses the challenges and solutions associated with handling imbalanced data and ensuring the quality and integrity of the dataset for training robust predictive models. This chapter is crucial for understanding the empirical basis upon which the predictive models are built.

Recent growth in remote sensing and machine learning (ML) techniques has accelerated the development of sophisticated models for predicting wildfire spread. We take advantage of these advancements to train our state-of-the-art models for the prediction of wildfire spread.

The Next Day Wildfire Spread (NDWS) dataset collects data from various locations across the United States during the period of active wildfires and is designed to train models to predict wildfire spread with high precision. The dataset covers a time frame from 2012 to 2020 with 18,545 fire events split into training, validation, and testing subsets using an 8:1:1 ratio, incorporating multiple data sources aligned through Google Earth Engine (GEE). It was observed that fire increased in size in 58% of the samples (10,798 samples), while it decreased in 39

The dataset combines 13 essential features that influence wildfire behavior, collected from several reputable remote sensing and environmental databases. These features provide a multimodal perspective on wildfire spread and include both static and dynamic variables, as detailed in Table 4.1.

Table 4.1: Features in Next Day Wildfire Spread dataset chosen for prediction modeling.

Nos	**Features**	**Impacts**	**Temporal Resolution**	**Spatial Resolution**	**Unit**
1	Elevation	Influences fire behavior due to changes in air density and moisture at different altitudes.	Constant	30 m	m
2	Wind Direction	Affects the direction and speed of fire spread, influencing containment strategies.	Hourly	$0.5^\circ \times 0.625^\circ$	degrees
3	Wind Speed	Influences the rate of fire spread, especially in areas with abundant fuel.	Hourly	$0.5^\circ \times 0.625^\circ$	m/s
4	Minimum Temperature	Can slow down fire spread overnight.	Daily	0.5°	°C
5	Maximum Temperature	Increases fire intensity and rate of spread.	Daily	0.5°	°C
6	Humidity	Can reduce or exacerbate fire intensity.	Daily	0.5°	%
7	Precipitation	Reduces wildfire risks by moistening fuels.	Daily	0.5°	mm
8	Drought Index	Indicates potential for severe conditions due to prolonged dry periods.	Variable	0.5°	Index
9	Vegetation Cover	Determines fuel availability and type, affecting fire intensity and spread.	Constant	30 m	%
10	Population Density	Impacts evacuation strategies and risk assessment in fire-prone areas.	Every 5 years	1 km	people/km^2
11	Energy Release Component	Measures potential energy available for combustion in wildfires.	Daily	0.5°	Index
12	Previous Fire Mask	Tracks and predicts wildfire movement on day 't'.	Daily	0.5°	Binary
13	Fire Mask	Tracks and predicts wildfire movement on day 't+1'.	Daily	0.5°	Binary

4.1 Data Collection and Organization

The dataset development involved several key steps to ensure spatial and temporal consistency across features and regions, a framework for which is shown in Fig. 4.1. These steps are described in detail below.

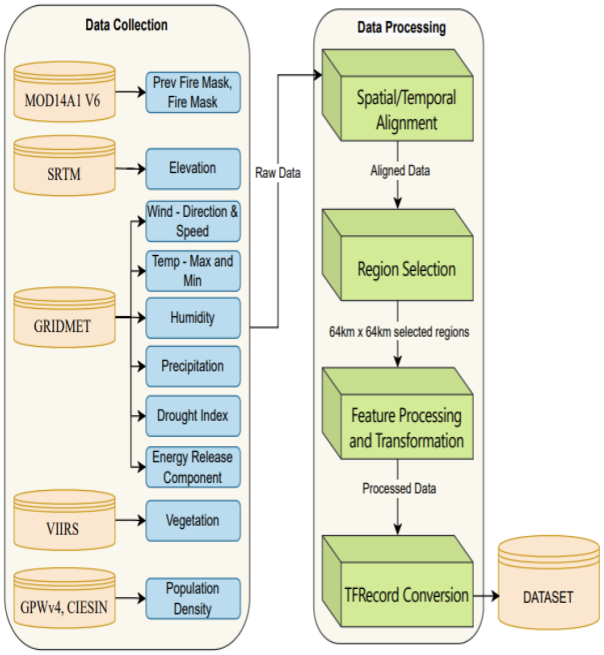


Figure 4.1: Data Collection Framework

4.1.1 Data Collection

Remote sensing Integration: Using Google Earth Engine (GEE), researchers aggregated the data in [30] from multiple sources. Each source provided distinct information relevant to wildfire behavior:

- **MOD14A1 V6:** Provided daily fire mask data at 1 km resolution.
- **SRTM:** Supplied high-resolution topographical data at 30 m, later downsampled.
- **GRIDMET:** Supplied weather (temperature, wind, precipitation, humidity) and drought data at 4 km resolution.

- **VIIRS:** Provided vegetation indices at 0.5 km resolution.
- **GPWv4:** Delivered population density at 1 km resolution.

4.1.2 Spatial and Temporal Alignment

Resolution Adjustment: To ensure consistent resolution, each feature was resampled to 1 km using appropriate techniques:

- Downsampling for high-resolution data (e.g., Shuttle Radar Topography Mission (SRTM) at 30 m).
- Upsampling for lower-resolution data (e.g., GRIDMET at 4 km and VIIRS at 0.5 km).

Temporal Aggregation: Weather data, recorded every 6 hours, was averaged to produce a single daily value. This process allowed for stable daily metrics, reducing intra-day variability.

4.1.3 Region Selection

Active Fire Detection: Only areas with detected fire activity on day t were included, with each region centered on a 1 km $\times$ 1 km cell where fire was observed. This selection method focuses on relevant data for predicting fire spread.

Region Size: A 64 km $\times$ 64 km region was captured around each detected fire point, as shown in Fig. 3.1. This region size is designed to provide adequate spatial context for understanding fire spread while maintaining computational manageability.

4.1.4 Feature Processing and Transformation

Normalization and Clipping: Each feature was clipped to avoid extreme outliers, using physically meaningful limits or data percentiles (0.1% and 99.9%). After clipping, features were normalized by subtracting the mean and dividing by the standard deviation.

Conversion to Image Data: Each feature was structured as a layer in a 64 $\times$ 64 grid, creating a multi-channel image format. The fire mask from day $t + 1$ served as the target label.

4.1.5 Data Storage and Format

TFRecord Conversion: The final dataset was saved in TFRecord format, optimized for TensorFlow. This format supports efficient data access and loading, critical for large-scale machine learning (ML) tasks.

4.1.6 Structured Dataset

Each sample in the dataset represents a 64 km × 64 km grid with 1 km × 1 km cells, capturing spatial features around detected wildfires. Each feature (e.g., temperature, vegetation, population density) is represented as a distinct channel within the 2D image format, as represented in Fig. 4.2. This multi-channel format is ideal for convolutional neural networks, which can leverage spatial and contextual information across features.

4.2 Data Preprocessing

In this study, we use a variety of data preprocessing methods to enhance the data quality and consistency for our models to better generalization capabilities. The methods that are used include standardization, noise augmentation, filtering of invalid data, and random cropping. Below is a detailed discussion of each technique.

4.2.1 Random Cropping

Random cropping is applied to introduce variability and prevent the model from overfitting on specific sections of the data. Initially, the input images are of the dimensions 64 km x 64 km regions and the fire mask are centered within each region. Thus, the model learns to predict the fire spread in the center of the images(predictions) when trained on uncropped data.

To avoid this, we augment the images in the data input pipeline by cropping 32 km x 32 km regions from the original dimensions(see (4.1)). Randomly selecting patches within the input images allows the model to generalize better and focus on different regions of the input data.

From (3.1), we have $\mathbf{X} \in \mathbb{R}^{H \times W \times C}$. Random cropping is defined as a function $f_{crop}(\mathbf{X}, a)$ that extracts a square patch of size $a \times a$ from $\mathbf{X}$ at a randomly chosen location, producing:

$$\mathbf{X}_{crop} = f_{crop}(\mathbf{X}, a), \quad \mathbf{X}_{crop} \in \mathbb{R}^{a \times a \times C} \tag{4.1}$$

In this study, we use a crop size of $a = 32$ to match the model's input dimensions.

4.2.2 Filtering Invalid Data

After cropping, we apply a filtering process to ignore data points that contain uncertain labels in their "previous fire mask" channel. The fire information data processed from Google Earth Engine (GEE) contained 3 classes: fire, no fire, and an additional class for uncertain labels. These labels accounted for cloud coverage or other unprocessed data in fire masks over each region. Data points with label values of -1 represent uncertain labels and are excluded from further processing. This step ensures that only valid, complete data is used for training the model and in further loss and performance calculations. This prevents the model from learning incorrect associations due to missing or undefined data points.

4.2.3 Noise Augmentation

Noise augmentation is introduced to enhance the model's robustness to small variations in feature values, simulating real-world fluctuations. Small random fluctuations were added to specific feature channels based on their statistical properties such as mean and standard deviation of the dataset. This ensures that the model learns to generalize better by encountering slightly altered versions of input data rather than overfitting to fixed patterns. By adjusting feature values within a realistic range, this process helps the model generalization better to unseen data and mitigates overfitting by encouraging it to learn from varied inputs. This manipulation allows the model to develop robustness by learning from varied versions of the input data.

4.2.4 Standardization

Standardization is a feature scaling method that centers each feature around the mean and scales the resultant distribution to unit standard deviation. The features in our dataset exhibit a wide variety of ranges, leading to disparities in their influence on the model. As a result, the model tends to favor features that have larger values, potentially skewing its predictions and overall performance.

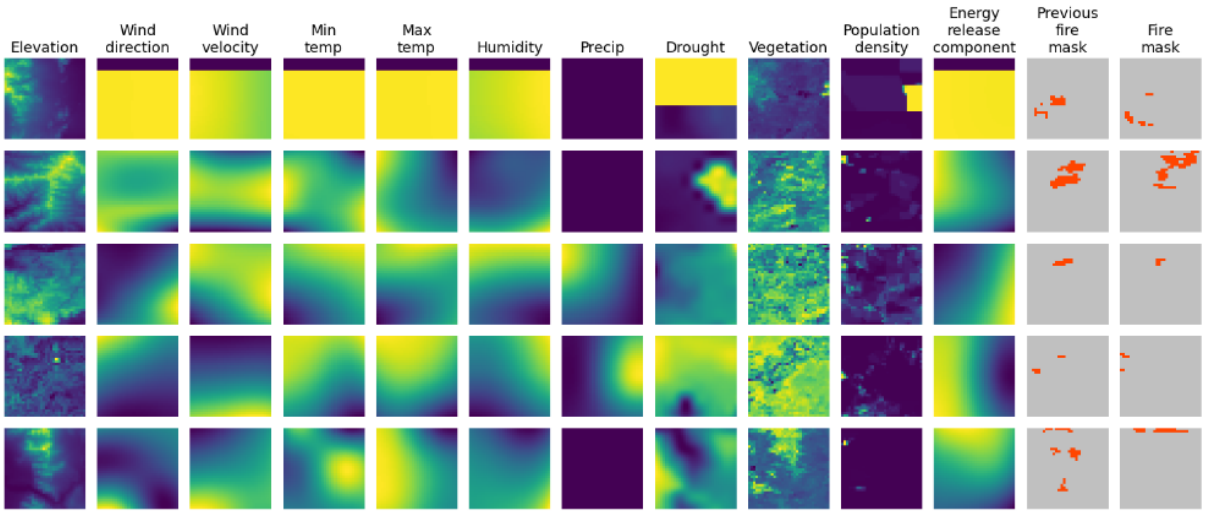


Figure 4.2: Example images from the dataset showing each of 12 features derived from various sources.

To avoid this issue and to stabilize the training process, we perform Z-score standardization for each feature, according to (4.2). For a given feature $\mathbf{X}_{ch}$, standardization is performed and the standardized feature $\mathbf{X}_{ch}^{std}$ is then given by:

$$\mathbf{X}_{ch}^{std} = \frac{\mathbf{X}_{ch} - \mu_{ch}}{\sigma_{ch} + \epsilon} \tag{4.2}$$

where μ_{ch} is the mean, σ_{ch} is the standard deviation of the variable and ϵ is a small constant added to prevent division by zero.

Overall, the above-discussed preprocessing techniques enhance the data quality, consistency, and robustness, improving the performance and generalization ability of our models.

4.3 Data Imbalance Handling

In the context of wildfire spread prediction, the effective handling of class imbalance is crucial. Previous data preprocessing strategies, such as random cropping and filtering invalid data, serve as vital preliminary steps to ensure the quality of input data; however, these methods do not inherently address the issue of imbalance within the dataset. Therefore, the loss function is key in mitigating the skewness of class distribution, thus enabling more balanced learning from both majority and minority classes.

Class imbalance is quantitatively assessed by calculating the ratio of the number of instances in each class to the total number of instances across the dataset. Specifically, for each class label, the ratio is determined by:

$$\text{Class Ratio} = \frac{\text{Number of Instances in Class}}{\text{Total Number of Instances}} \tag{4.3}$$

Table 4.2 shows the fire class and no fire class ratios present in the training, validation, and test set. Such unequal ratios highlight the urgent need for effective handling of this issue so that the model learns accurately over both the classes and not just the majority ("no fire") class.

Table 4.2: Class Ratios Across Datasets

Datasets	**Fire (%)**	**No Fire (%)**
Training Set	3.60	96.40
Validation Set	4.17	95.83
Test Set	3.74	96.26

Fig. 4.3 depicts the distribution of "fire" and "no fire" classes across three datasets: training, validation, and testing in the Next-day wildfire spread dataset. The plot illustrates a significant imbalance in the class distribution, also discussed in the table 4.2. The massive difference between the classes raises concerns about the model potentially developing a bias in its predictions. In this scenario, if the class imbalance is not handled, even if the model predicts all pixels as "False" and is checked for simple metrics, say accuracy, the model would achieve an accuracy upwards of 90%. However, this high accuracy is misleading as the model would not have "learned" anything from the minority classes. Therefore, it is critical that the imbalance is handled efficiently and that more nuanced performance metrics are used for truly reflecting the model's performance such as F1 Score and IoU.

The issue of class imbalance is central to wildfire datasets, as the "Fire" class generally comprises a small fraction of the dataset compared to the "No Fire" class as shown in Table 4.2. Such imbalance can lead to models favoring the majority "No Fire" class, undermining the predictive accuracy for critical fire regions. The core of addressing class imbalance in our model is the implementation of an appropriate loss function. Specifically, the Focal Tversky Loss (FTL), is a loss function that is particularly suited for imbalanced datasets and hard-to-predict regions. This approach ensures balanced sensitivity to both classes while prioritizing regions critical to fire spread prediction.

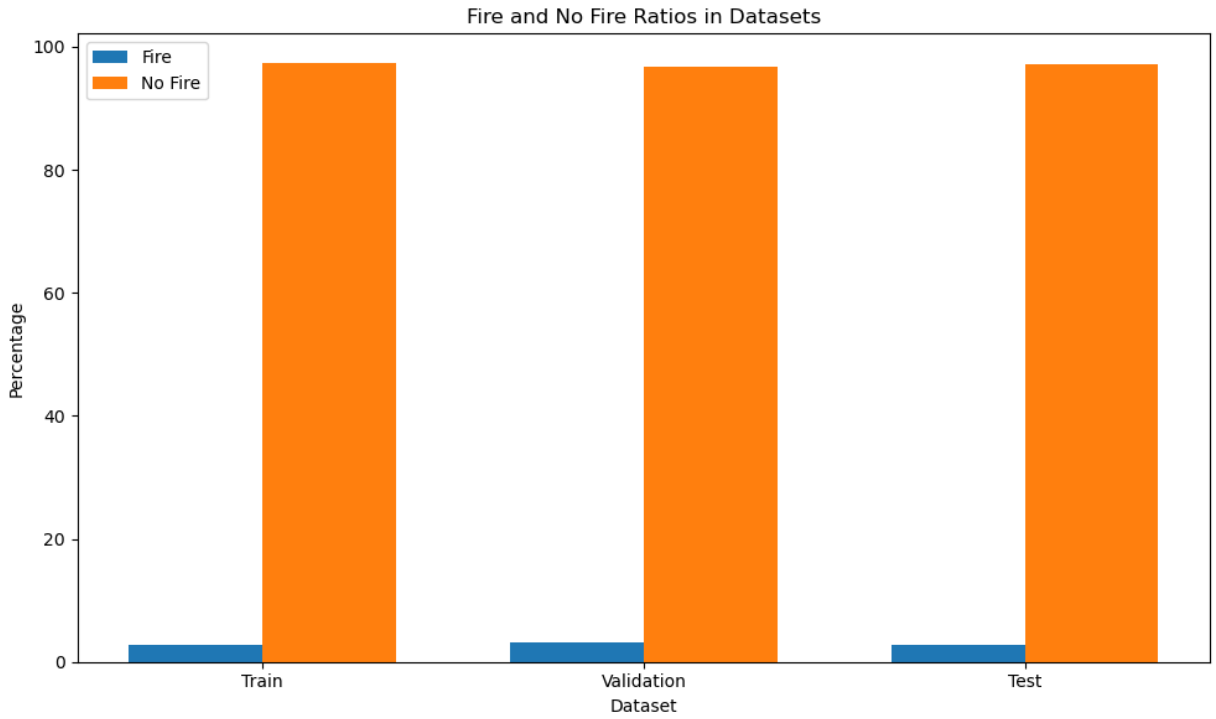


Figure 4.3: Imbalanced Data Ratio in Next-day Wildfire Dataset

This loss function mitigates this risk by assigning a higher penalty to false negatives, ensuring the model is more sensitive to detecting fire regions. At the same time, it controls false positives to prevent overprediction. What makes Focal Tversky Loss especially effective is its ability to focus on hard-to-predict areas, such as small or sparse fire regions and boundary edges, which are often overlooked by simpler loss functions. These challenging areas are given higher importance during training, enabling the model to learn fine-grained distinctions and make confident predictions in critical scenarios. This is particularly valuable for wildfire spread prediction, where both localized precision and a broad contextual understanding are essential for accurate segmentation.

Aside from the Focal Tversky Loss, we have also implemented and experimented with other loss functions such as Dice Coefficient Loss, Tversky Loss, and Weighted Cross-Entropy(WCE) Loss within our code. These functions are known for their effectiveness in segmentation tasks by focusing on the overlap between the predicted and actual labels, which is particularly useful for spatial data like ours. A detailed analysis of how each of the models performs across each loss function is explained in section 6.

The strategic selection of the Focal Tversky Loss function directly addresses the disproportionate class representation, ensuring that the predictive performance across all classes is balanced and reliable.

Chapter 5

Methodology

This chapter presents the methodological framework utilized for next-day wildfire spread prediction, formulated as a semantic segmentation problem. This section is organized into 3 subsections: two baseline models, U-Net and Vision Transformer (ViT), and a novel hybrid model, Vi-Net. Both the chosen baseline models are state-of-the-art in the field of image segmentation. U-Net, a modified CNN with encoder-decoder architecture, is employed for its ability to capture precise spatial details essential for accurate segmentation. Vision Transformer (ViT), on the other hand, leverages its global attention mechanism to model long-range dependencies across environmental features. Finally, in our proposed Vi-Net hybrid model we combined the localized precision of U-Net with the global contextual awareness of ViT, to provide an enhanced and robust solution for wildfire spread prediction. To demonstrate the applications of the hybrid model, we integrate the fire spread predictions obtained from the Vi-Net model with an optimized A*(OA*) A* algorithm to find paths that avoid fire masks while maintaining optimality. This work simulates a near real-time safe path planning framework. Fig 5.1 shows a high-level overview of the framework employed for wildfire spread prediction and its integration with the safe path planning(OA*) module to produce safe navigable routes from the generated spread predictions in near real-time.

semantic segmentation is a computer vision problem that involves classifying the pixels from the input image into one of several categories. These categories could be an object or a region in the image. Essentially, we want every pixel assigned to a label based on its semantic meaning. The output of a semantic segmentation task is typically a segmentation mask, which is a matrix or an image, the same as the size of the input image, with each pixel assigned a label. In the context of wildfire spread prediction given our dataset, we are classifying each pixel of a geographic area based on the likelihood of fire spread and current

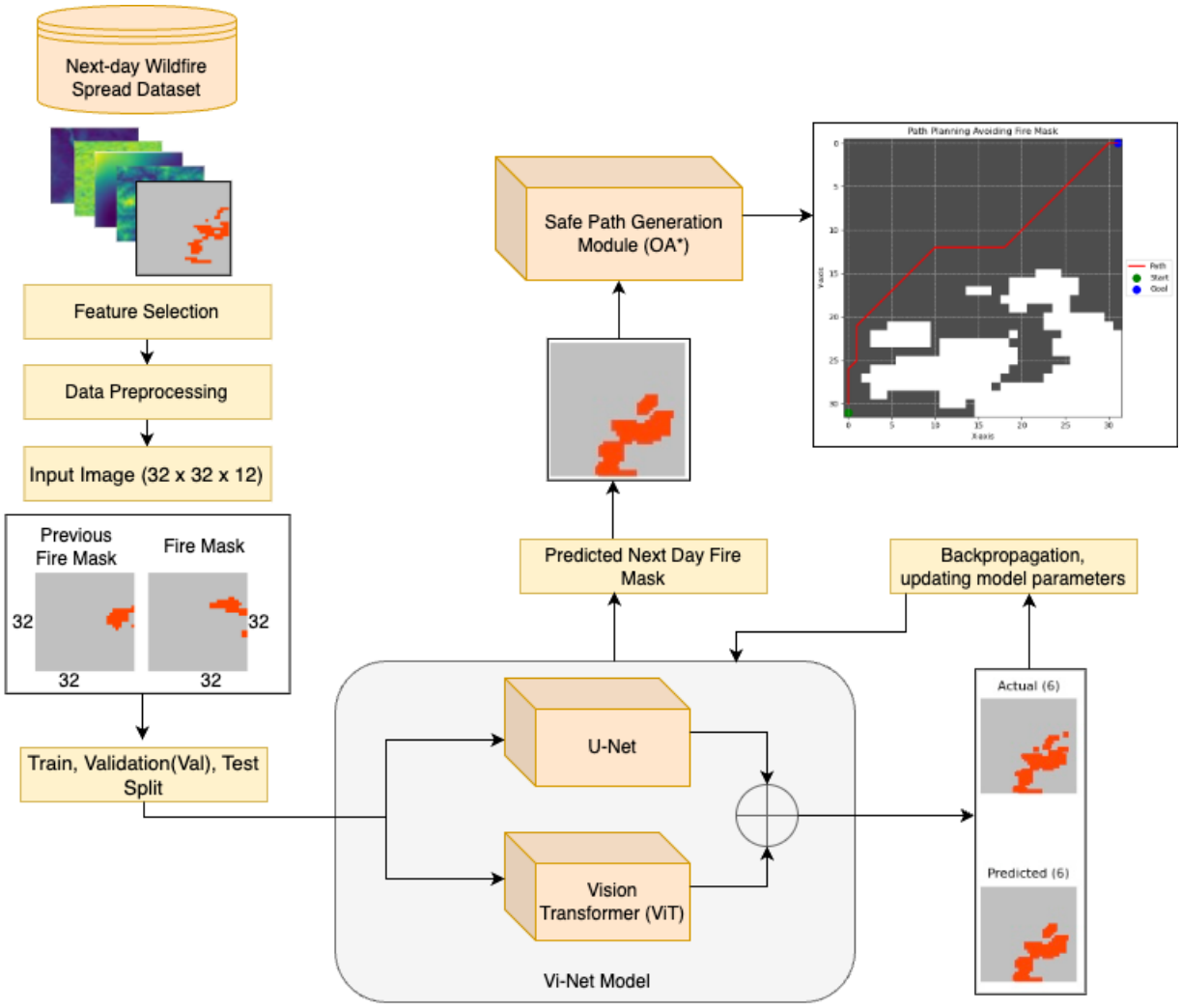


Figure 5.1: High-Level Overview of Wildfire Spread Prediction and Safe Path Planning Framework.

environmental and topographical factors. The goal is to predict and map the spread of fire by analyzing different environmental features at a pixel level and classifying each pixel as "fire" or "no fire" based on the probability score of each pixel of the likelihood of it being considered as a fire pixel and then comparing it with our threshold. The threshold is a probability score that we set to classify our pixels as "fire" or "no fire", the pixels that have a probability score more than the threshold after taking into account the features get classified as "fire" and the others as "no fire" pixels. We keep this parameter the same across all three models. A complete workflow of the 3 models, U-Net, ViT, and the proposed hybrid model, Vi-Net, is demonstrated in Fig.5.2. The following subsections detail each model's architecture, functionality, and unique contributions in addressing the challenges inherent in wildfire prediction tasks.

5.1 Baseline Image Segmentation Model: UNet

The U-Net model was originally proposed by Ronneberger et al. [46], and has been a state-of-art model for image segmentation tasks because of its effectiveness in locating boundaries and objects in an image. They showed remarkable improvements in performance compared to previous existing methods. U-Nets are a special type of convolutional neural network with an encoder-decoder-style architecture. Using the U-Net model, we can make accurate pixel-level predictions for determining areas at risk for our next-day fire spread prediction task.

Our U-Net model consists of an encoder (contraction path), a bottleneck, a decoder (expansion path), and an output layer as seen in Fig. 5.3. The U-Net model takes a tensor representing an input image, $\mathbf{X} \in \mathbb{R}^{H \times W \times C}$ and feeds it to the encoder. In the encoder, the image is downsampled over each layer to decrease the spatial dimensions, which helps the model to focus on larger and more abstract features and progressively capture the feature hierarchies to extract fine-grained, local features. Each block of the encoder has two convolutional layers with kernel size $= 3 \times 3$, followed by Batch Normalization and LeakyReLU activation layer:

$$f_{enc,i} = \sigma(BN(Conv(f_{i-1}) * W_i + b_i)) \tag{5.1}$$

where $f_{enc,i}$ represents the feature map at each layer i, W_i and b_i are the weights and biases, and σ is the LeakyReLU activation function. At the end of each encoder block, Max-Pooling, of window size 2×2 is added, it halves the spatial dimension while retaining the essential features.

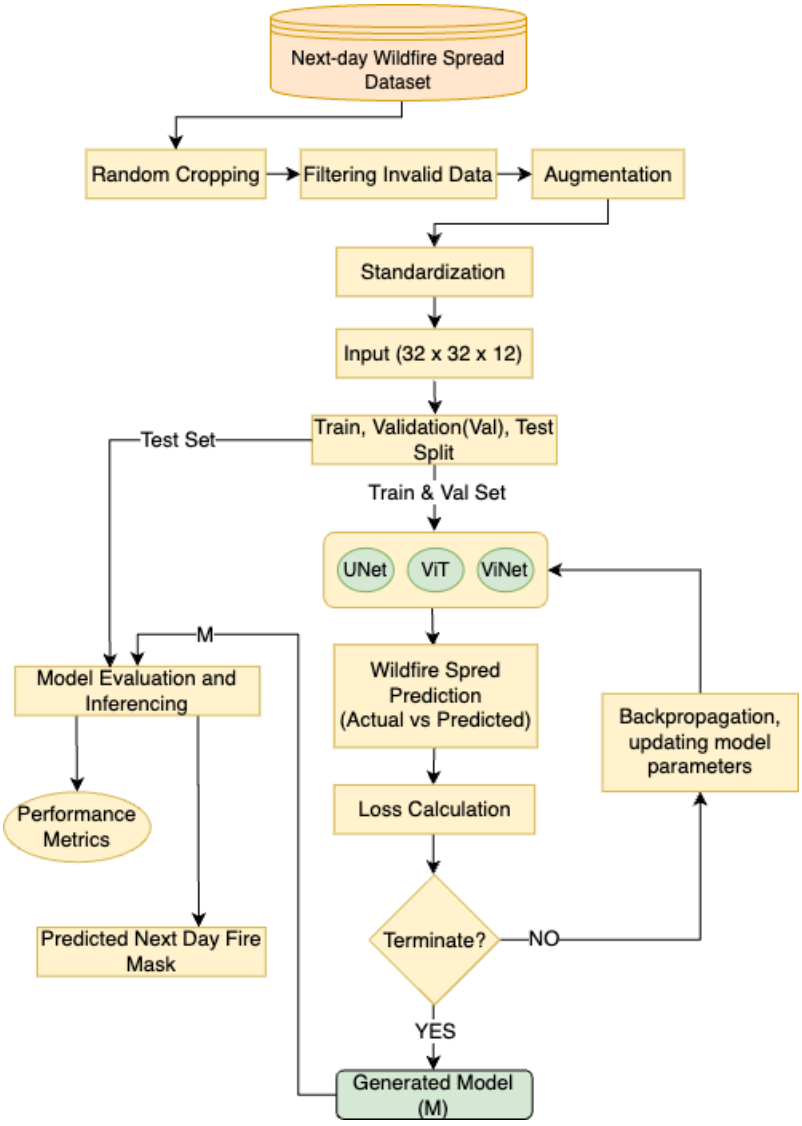


Figure 5.2: Workflow Diagram showcasing the complete flow of working of prediction models

The Bottleneck block serves as a transition between encoder and decoder blocks. It operates at the lowest spatial resolution and captures most abstract representations of the input features. This is where the model aggregates all the learning from the encoder blocks enabling it to form a comprehensive, cohesive contextual understanding of the input,

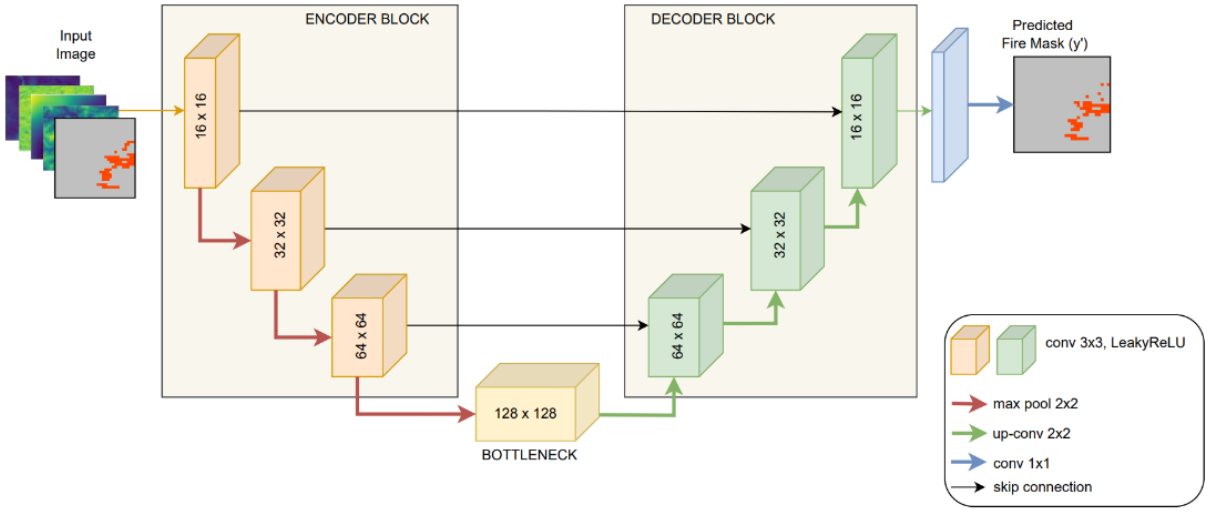


Figure 5.3: UNet Model Architecture.

such as identifying which areas are at most risk of fire considering the environmental and topographical data. It can be represented as:

$$f_b = \sigma(BN(Conv(f_L) * W_b + b_b)) \tag{5.2}$$

where f_b is the feature map in bottleneck and L is the deepest layer.

The decoder block works on reconstructing the dimensions and fine-tuning the output. It upsamples the feature maps to reconstruct the original input dimensions and uses the skip connections to merge feature maps from the corresponding encoder layer.

$$f_{dec,i} = \sigma(BN(Conv(Concat(Up(f_{dec,i+1}), f_{enc,i})))) \tag{5.3}$$

where $f_{dec,i}$ is the decoder feature map at i.

After the input image is upsampled back to the original dimensions, it passes through an output layer, which is the final layer in our model. It is essential for our segmentation task as here, each pixel will be classified either "fire" or "no fire". The output layer is a 1×1 convolution with a sigmoid activation, outputting a segmentation map, $\hat{y} \in R^{H \times W}$, where each pixel indicates the probability of fire spread:

$$\hat{y} = \sigma(\text{Conv}_{1\times1}(f_{\text{dec},1})) \tag{5.4}$$

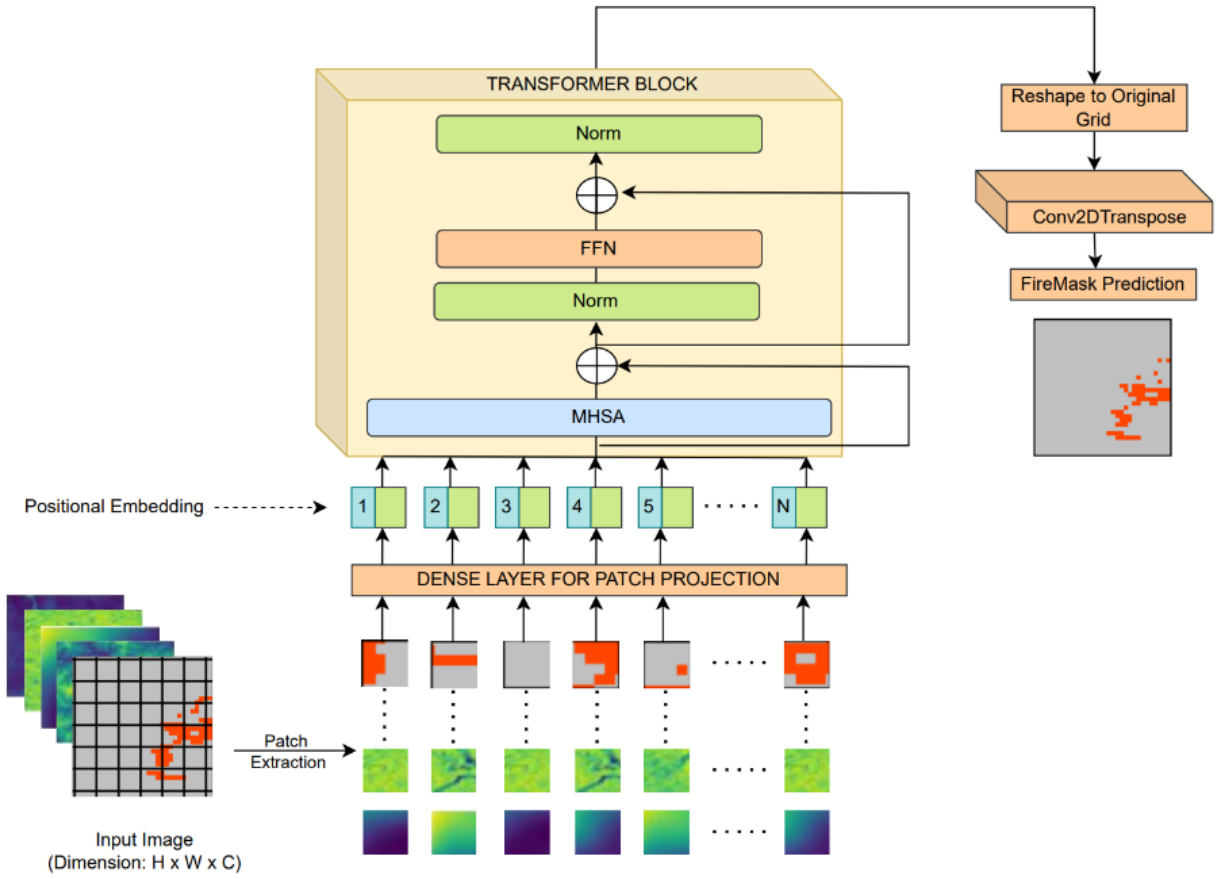


Figure 5.4: ViT Model Architecture.

5.2 Baseline Image Segmentation Model: ViT

Vision Transformers (ViT), first introduced by Dosovitskiy et al. [15], brings the powerful transformers model, primarily used for NLP tasks, to the realm of computer vision. It introduces a paradigm shift in the image analysis domain by parallelly processing images as opposed to the sequential processing in convolutional network models, which traditionally dominated this domain.

Fig. 5.4 shows the architecture of our ViT model. ViT processes an input image by dividing it into smaller, non-overlapping patches. From (3.1) we have $\mathbf{X} \in \mathbb{R}^{H \times W \times C}$, this image is divided into patches and each patch, $\mathbf{x}$, can be defined as:

$$\mathbf{x} \in \mathbb{R}^{p \times p \times C} \tag{5.5}$$

where p is the patch size. For our input image of size 32×32, we define the patch size of 4×4.

Each of these patches is flattened into a one-dimensional vector, $\mathbf{x}' \in \mathbb{R}^{p^2 \cdot C}$. These vectors are then projected to a higher dimensional embedding space, z_i:

$$z_i = W_e \mathbf{x}'_\mathbf{i} + b_e, \quad i = 1, \ldots, N \tag{5.6}$$

where $W_e \in \mathbb{R}^{D \times (p^2 \cdot C)}$ is the learnable weight matrix, b_e is the bias, D is the dimension of embedded space and N is the number of patches. This embedding helps the model learn more abstract features than just pixel values.

Since ViTs do not have inherent spatial understanding like convolutional networks, to address this issue, positional encodings PE_i are added to each embedded patch to retain spatial structure:

$$z_i^{\text{pos}} = z_i + \text{PE}_i \tag{5.7}$$

These encodings allow the model to understand the relative positions of the patches, which is essential for tasks like segmentations where spatial relationships affect predictions.

The patch embeddings with the positional encoding are fed to transformer blocks. Each transformer block is a self-contained transformer layer consisting of a Multi-Head Self-Attention layer and a Feed-Forward Network. In our ViT model, we deploy 4 transformer blocks with 8 self-attention heads.

- **Multi-Head Self-Attention(MHSA) layer:** Self Attention is the process by which the model learns to weigh the importance of different parts of input data relative to each other. The model calculates "attention scores" for each patch. According to Vaswani et al. [54], given query Q, key K, and value V vector from the patch embedding, Attention can be defined as:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{5.8}$$

  where d_k is the dimension of the key vector. The model uses several such attention "heads" in parallel, each potentially focusing on different input relationships. MHSA concatenates outputs from all these multiple attention heads:

$$MHSA(z) = (head_1 + head_2 + \ldots + head_n)W_o \tag{5.9}$$

  with $W_o \in \mathbb{R}^{D \times D}$ is the weight matrix for output projection.

- **Feed-Forward Network(FFN) layer:** The output of the MHSA is passed through an FFN, consisting of two linear transformation layers with ReLU activation in between:

$$FFN(\mathbf{m}) = ReLU(\mathbf{m}W_1 + b_1)W_2 + b_2 \tag{5.10}$$

where $\mathbf{m}$ is the output from MHSA, W_1 and W_2 are the weights matrices of linear layers, b_1 and b_2 are the corresponding biases. Typically, $W_1 \in \mathbb{R}^{D \times D_f}$ and $W_2 \in \mathbb{R}^{D_f \times D}$; D_f is of higher dimensional order and larger than D, which is the dimensionality of the $\mathbf{m}$. The FFN layer initially increases the representation size and then scales it back to its original size during the second step.

The final output from the last transformer layer is reshaped back into a $2D$ grid to match the original dimensions. An upsampling layer is used to increase the resolution to the original($32km \times 32km$), this ensures each pixel location is covered as in the original image. Following this step, a final 1×1 convolution with sigmoid activation is applied to generate the segmentation mask:

$$\hat{y} = \sigma(Conv2D(Reshape(Z_{out}))) \tag{5.11}$$

where $\hat{y} \in \mathbb{R}^{H \times W}$ represents the predicted segmentation mask, with each pixel containing a probability value indicating the likelihood of fire on that location the next day.

5.3 Proposed Hybrid Approach: Vi-Net

The proposed Vi-Net architecture combines the strengths of U-Net and Vision Transformer (ViT) to enhance next-day wildfire spread prediction. This hybrid model captures local, fine-grained spatial information through U-Net and long-range global dependencies through ViT, making it suitable for tasks requiring a nuanced understanding of spatial relationships across environmental data.

5.3.1 Vi-Net Architecture and Functionality

As shown in Fig 5.5, our hybrid approach consists of two parallel paths, for U-Net and ViT, that independently process the input image $\mathbf{X} \in \mathbb{R}^{H \times W \times C}$. Each path captures distinct features before combining their outputs in a concatenation layer.

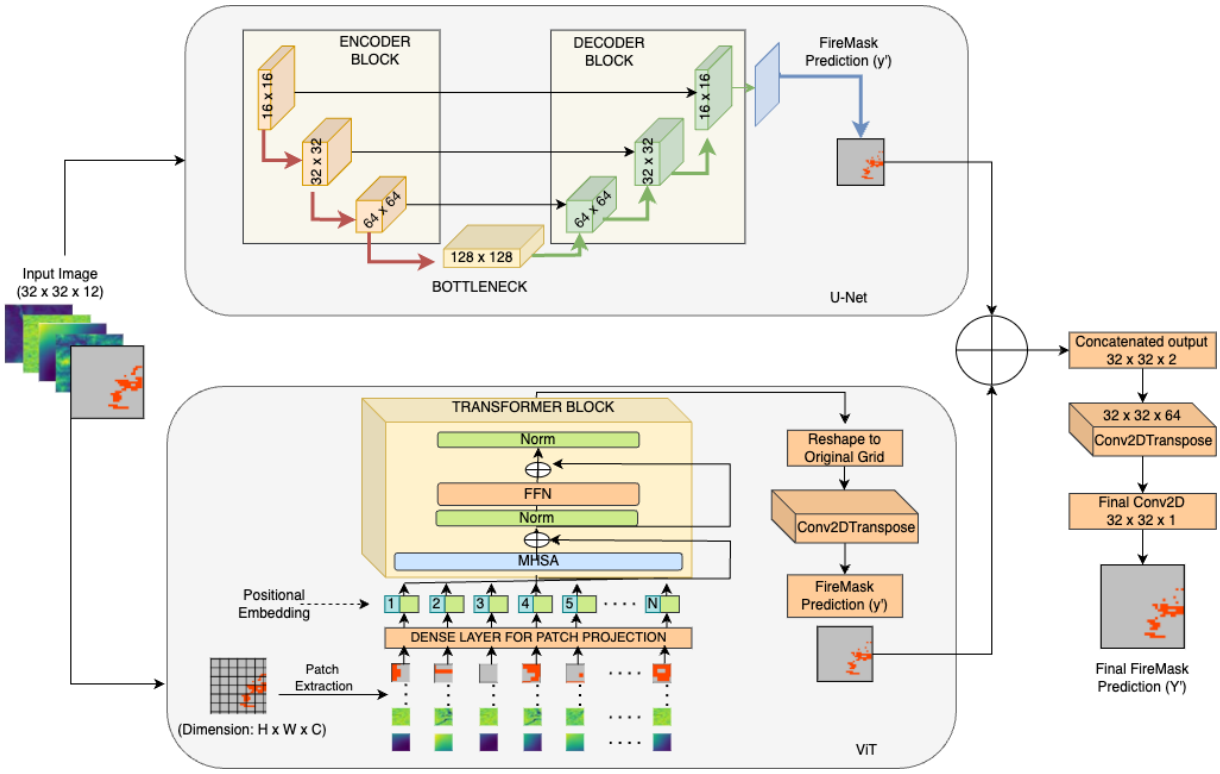


Figure 5.5: Model Architecture of the hybrid model, Vi-Net.

Unet Path (Local Feature Extraction) The U-Net component of Vi-Net captures localized, hierarchical features essential for pixel-level precision. The encoder in U-Net progressively down-samples the image to produce multi-scale feature maps as shown in (5.1). Each encoding layer follows:

$$Encoder_i = f_{\text{enc},i} \tag{5.12}$$

for all i-th layer feature maps. The lowest spatial resolution is reached at the bottleneck, capturing abstract representations of the input image (5.2). In the decoder (5.3), U-Net concatenates upsampled feature maps from the previous deeper decoder layer and corresponding encoder layers through skip connections:

$$Decoder_i = f_{dec,i} = \oplus_i(f_{dec,i+1}, Encoder_i) \tag{5.13}$$

Skip connections ensure the fine-grained information is preserved and enable precise reconstruction of spatial details which is crucial for accurately delineating regions at risk for fire spread. The final U-Net output $\mathbf{f}_{\text{U-Net}}$ is generated through a 1×1 convolution and sigmoid activation (5.4). Therefore, we can define $\mathbf{f}_{\text{U-Net}}$ as:

$$f_{\text{U-Net}} = \bigoplus_i f_{\text{dec},i} \tag{5.14}$$

ViT Path (Global Contextual Awareness) In parallel, the ViT works on extracting global dependencies from the input features. ViT divides the input image into non-overlapping patches($\mathbf{x}'$), (5.5), embedding each patch into a higher-dimensional space (5.6) to create an abstract representation:

$$\text{Embed}(\text{Patch}_i) = W_{\text{patch}} \cdot \text{Patch}_i + b_{\text{patch}} \tag{5.15}$$

$\text{Embed}(\text{Patch}_i)$ represents the i-th patch embedding. Positional encodings PE_i are added to each embedded patch:

$$\text{Patch}_i^{\text{pos}} = \text{Embed}(\text{Patch}_i) + \text{PE}_i. \tag{5.16}$$

The encoded patches are processed through multi-head self-attention (MHSA) and feed-forward layers, modeling global dependencies across patches, as defined in (5.8), (5.9) and (5.10). The final output from ViT is reshaped back to a $2D$ grid and up-sampled to match U-Net's spatial dimensions (5.11), yielding the globally contextualized output $\mathbf{f}_{\text{ViT}}$:

$$f_{\text{ViT}} = \text{Reshape}(\text{FFN}(\text{MHSA}(\text{Embed}(\mathbf{x}') + \text{PE}))) \tag{5.17}$$

Concatenation and Output Layer The outputs from U-Net and ViT, $\mathbf{f}_{\text{U-Net}}$ and $\mathbf{f}_{\text{ViT}}$, are concatenated in a weighted summation:

$$f_{\text{combined}} = \alpha \times f_{\text{U-Net}} + \beta \times f_{\text{ViT}} \tag{5.18}$$

where α and β balance each model's contribution. This introduces greater flexibility in factoring in the contributions from a particular model and can be adjusted as per the requirements of a task. For equal contribution from both models, we set α and β at 0.5. This combined representation then passes through a final 1×1 convolution and sigmoid activation:

$$\hat{Y} = \sigma(\text{Conv}_{1\times1}(f_{\text{combined}})) \tag{5.19}$$

producing the segmentation mask $\hat{Y} \in \mathbb{R}^{H\times W}$, indicating fire spread probabilities for each pixel. A high-level overview of the working of Vi-Net is shown in Fig. 5.6.

Algorithm 1 demonstrates the pseudocode for our hybrid model, Vi-Net, which integrates both U-Net and Vision Transformer (ViT) architectures to enhance image segmentation performance for wildfire spread prediction. This algorithm outlines the process from inputting an image dataset to producing a fire segmentation mask, systematically explaining each step involved in the model. It highlights how both the models work parallelly on the input set and provide results as per their individual strengths and then combine the outputs in a way to enhance the prediction capabilities, illustrating our method's dual approach to handling image data for segmentation.

5.3.2 Benefits of the Hybrid Vi-Net Approach

Vi-Net achieves a comprehensive representation by combining U-Net's localized precision with ViT's global context. The U-Net path preserves high-resolution details, essential for accurate delineation, while ViT captures broader relationships that impact fire spread, such as environmental factors across distant regions.

U-Net's limited receptive field restricts its capacity for capturing long-range dependencies. ViT's self-attention mechanism overcomes this by allowing the model to learn relationships between distant patches, a critical factor in wildfire prediction. ViT's patch-based approach may sacrifice fine-grained spatial detail. U-Net compensates with skip connections that maintain pixel-level detail, ensuring both macro and micro-level segmentation accuracy.

The Concatenation of $f_{\text{U-Net}}$ and f_{ViT} balances both models' strengths. U-Net's multi-scale, locally enriched features (5.4) complement ViT's global contextualization (5.11),

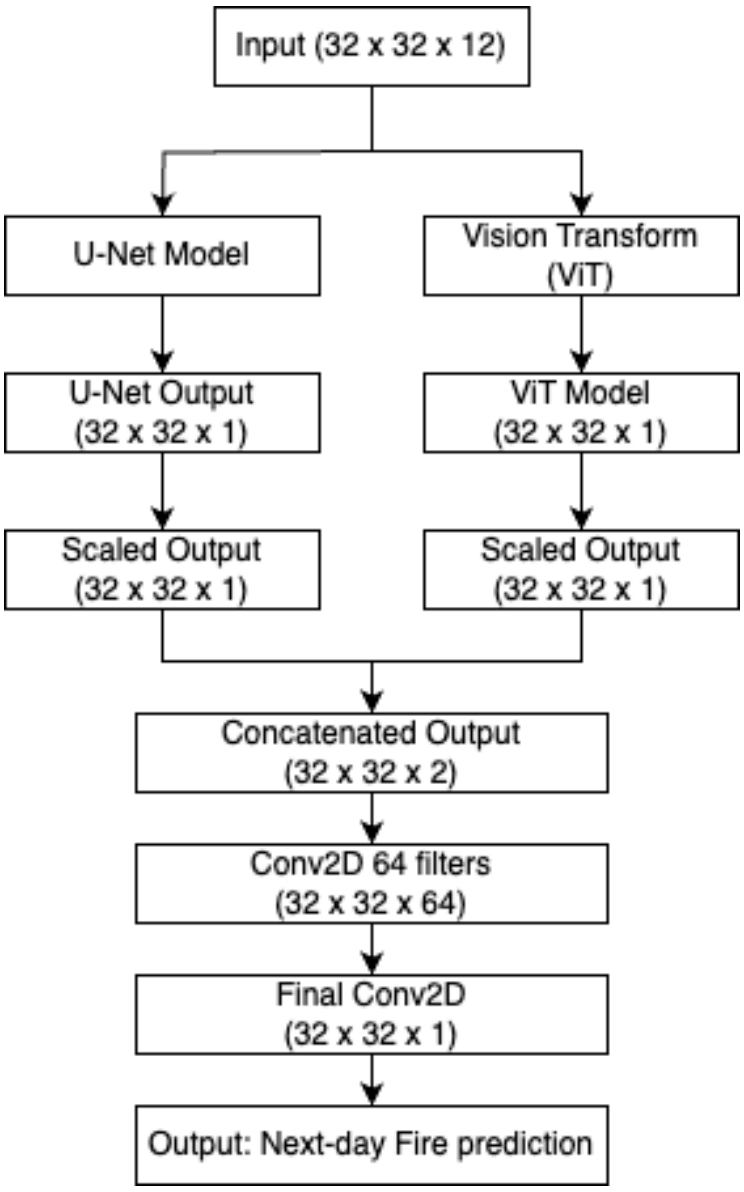


Figure 5.6: Overview of the Vi-Net Architecture.

resulting in a unified segmentation mask that provides a nuanced, accurate prediction as given by (5.18) and (5.19).

The hybrid Vi-Net architecture successfully combines U-Net's spatial detail preservation with ViT's global contextual awareness, providing a robust approach for next-day wildfire spread prediction. By addressing the individual limitations of U-Net and ViT, Vi-Net offers a significant improvement in accurately segmenting areas at risk of fire, advancing predictive capabilities in complex environmental applications.

Algorithm 1 Hybrid Vi-Net Model for Image Segmentation

Input: Image $X \in \mathbb{R}^{H \times W \times C}$, U-Net and ViT parameters
Output: Segmentation mask $\hat{Y}$
U-Net Encoder:
for $i = 1$ to L **do**
    $c_i \leftarrow \text{Conv2D}(X, F_i) \rightarrow \text{BN} \rightarrow \text{LeakyReLU}$
    if $i < L$ **then** $p_i \leftarrow \text{MaxPooling2D}(c_i)$
    end if
end for
Save encoder outputs c_i for skip connections
U-Net Decoder:
for $i = L$ to 1 **do**
    Retrieve c_i and concatenate with upsampled previous layer
    Conv2D $\rightarrow$ BN $\rightarrow$ LeakyReLU $\rightarrow$ Upsample (if $i > 1$)
end for
U-Net Output: segmentation_output
ViT Path:
Patch Extraction:
Divide X into patches of size $p \times p$
for each patch x_i **do**
    $z_i \leftarrow \text{Dense}(\text{Flatten}(x_i)) + \text{PE}_i$
end for
Transformer Layers:
for each block in ViT **do**
    Apply Multi-Head Self-Attention $\rightarrow$ Feed-Forward
end for
ViT Output: ViT_output
Scale outputs:
    scaled_segmentation_output $\leftarrow 0.5 \times$ segmentation_output
    scaled_ViT_output $\leftarrow 0.5 \times$ ViT_output
Concatenate(C):
final_output $\leftarrow$ C(scaled_segmentation_output, scaled_ViT_output)
$x \leftarrow \text{Conv2D}(\text{final_output}, 64, (3, 3)) \rightarrow \text{ReLU}$
$\hat{Y} \leftarrow \sigma(\text{Conv2D}(x, 1, (1, 1)))$
Return: $\hat{Y}$

5.3.3 Performance Metrics

To evaluate the performance of our model in wildfire spread prediction, we analyze its predictions in terms of True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). These terms provide the foundation for understanding how well the model identifies fire regions versus non-fire regions and are integral to calculating key performance metrics.

In machine learning, predictions are categorized into four key outcomes: True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). TP refers to instances correctly identified as positive outcome, which in the context of wildfire spread prediction corresponds to regions correctly predicted as fire. TN represents instances correctly identified as negative outcomes, meaning regions accurately predicted as non-fire. FP are cases incorrectly predicted as positive outcome, which occur when the model predicts fire in regions that are not actually on fire, leading to leading to overprediction and potential misallocation of resources. FN are cases incorrectly predicted as negative outcome, referring to missed fire regions that are actually on fire, posing a significant risk of overlooking hazardous zones(underprediction). These outcomes provide the foundation for evaluating the model's performance. Table 5.1 summarizes the prediction outcomes with respect to outcome probabilities in our task of wildfire spread predicition.

Term	**Definition**
True Positives (TP)	Regions correctly predicted as fire by the model that are actually on fire in the ground truth.
True Negatives (TN)	Regions correctly predicted as non-fire when they are indeed not on fire in the ground truth.
False Positives (FP)	Regions incorrectly predicted as fire by the model that are not on fire in the ground truth, leading to overprediction.
False Negatives (FN)	Regions incorrectly predicted as non-fire by the model that are actually on fire in the ground truth, leading to underprediction.

Table 5.1: Summary of Prediction Outcomes.

These fundamental quantities provide the basis for deriving critical performance metrics. Each metric offers unique insights into the model's behavior, particularly in handling

the trade-offs between correctly identifying fire regions and minimizing false alarms, which is especially relevant in the context of our imbalanced dataset. Below, we define these metrics in detail.

Precision

Precision measures the accuracy of positive predictions. It reflects the proportion of positive identifications that were actually correct and is particularly valuable in scenarios where the cost of a false positive is high.

$$\text{Precision} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Positives (FP)}} \tag{5.20}$$

Recall

Recall, also known as sensitivity or true positive rate, measures the model's ability to identify all relevant instances (true positives) within the actual positives available during testing. It is critical in situations where missing a positive instance is costly.

$$\text{Recall} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Negatives (FN)}} \tag{5.21}$$

F1 Score

The F1 Score is the harmonic mean of precision and recall. It is used to balance the trade-offs between precision and recall in a single metric, which is especially useful when dealing with imbalanced datasets.

$$\text{F1 Score} = 2 \cdot \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{5.22}$$

Jaccard Index(IoU)

The Jaccard index, also called Intersection over Union(IoU), is a measure of the similarity between two sets. For segmentation tasks, it can be used to evaluate the overlap between

regions of predicted and ground truth sets. It is defined as the ratio of the intersection area to the union area of the predicted and ground truth regions.

$$\text{IoU} = \frac{\text{Area of Intersection}}{\text{Area of Union}} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \tag{5.23}$$

Both IoU and F1 Score provide complementary perspectives on segmentation performance. IoU effectively quantifies spatial accuracy, making it a robust metric for evaluating the alignment of predicted segmentation masks with ground truth labels. Meanwhile, the F1 Score ensures a balance between segmentation completeness and correctness, reducing the risk of over-segmentation or under-segmentation. The combination of these two metrics allows for a more holistic evaluation of segmentation models, facilitating fair comparisons across different architectures and datasets.

5.4 Safe Path Planning: Optimized A* Algorithm

The objective of this part of our research is to develop an enhanced path planning algorithm based on A* (A-star) that can dynamically avoid fire areas in the predicted fire mask given by the Vi-Net model, as represented by binary grid maps. The output of the fire masks from the wildfire spread prediction acts as input to our path-planning algorithm. This algorithm is designed to identify safe paths in scenarios where both standard movement and movement close to fire zones are considered. Our safe path planning methodology capitalizes on the strengths of the traditional A* algorithm while optimizing it for efficient and less risky route generation in fire-affected environments.

5.4.1 A* Search Algorithm

A* search is an algorithm that has been popular in the pathfinding research community. This algorithm extends on Dijkstra's algorithm and uses best-first search making it efficient in finding the shortest path between a start and a goal node. It is the faster and more optimal approach for finding the shortest paths in graphs in most cases.

Like Dijkstra, A* works by making a lowest-cost path tree from the start node to the goal node. However, the A* algorithm introduces a heuristic into a regular graph-searching algorithm, essentially planning at each step so a more optimal decision is made. The heuristic function allows the algorithm to estimate the cost of the cheapest path from the current node to the goal node, prioritizing paths that are better expected to lead to the

goal faster. A* is particularly favored for its performance and accuracy as it can efficiently process graphs by strategically exploring the most promising nodes first, based on a cost function given by $f(n)$:

$$f(n) = g(n) + h(n) \tag{5.24}$$

where $g(n)$ is the cost from the start node to the current node n. $h(n)$ is a heuristic function that estimates the cost of the cheapest path from n to the goal. The heuristic function is chosen specific to the problem and when chosen appropriately, it ensures both efficiency and optimality of the algorithm. Algorithm 2 shows the working of a traditional A* algorithm.

Heuristic Function in A* Algorithm

A* is known as an informed search algorithm as it uses heuristics to estimate the cost of the cheapest path from a node to the goal. This heuristic works like a "smart, informed guess" that helps estimate the cheapest cost from the current node in the grid to the goal node. The heuristic is problem-specific and can be used to control A* behavior. The $h(n)$ part of the Eq. 5.24 must be admissible, that is the heuristic, $h(n)$, must never overestimate the cost to reach the goal node from the current node($h(n) \leq TrueCost$).

Heuristics can be classified into 2 categories: (i)exact heuristics, and (ii) approximate heuristics. Exact heuristics precomputes the exact value of $h(n)$. It is an ideal case but is very time-consuming and computationally expensive making it often an impractical way of calculating heuristics. Approximate heuristics are much faster and can be used to approximate the distance on a graph. The heuristic function for grid maps uses approximate distance heuristics that match the allowed movements. Some of the well-known heuristic functions are :

- Manhattan Distance: This heuristic is used in grid-based pathfinding scenarios where the movement is restricted only to horizontal and vertical steps.

  $$h(n) = |x_{\text{goal}} - x_n| + |y_{\text{goal}} - y_n| \tag{5.25}$$

  This heuristic assumes a uniform cost for horizontal and vertical movements and does not account for diagonal traversal.

Algorithm 2 A* Search Algorithm

```
1: Input: Start node s, Goal node g, Graph G
2: Output: Shortest path from s to g
3: Initialize open list O with start node s
4: Initialize closed list C as empty
5: g(s) ← 0                                   ▷ Cost of start node initialized 0 s
6: f(s) ← h(s)                                ▷ Estimated cost from s to g
7: while O is not empty do
8:     n ← node in O with the lowest f(n) value
9:     if n = g then
10:        return ReconstructPath(s, g)
11:    end if
12:    Move n from O to C
13:    for each neighbor m of n do
14:        if m ∈ C then
15:            continue
16:        end if
17:        tentative_g ← g(n) + cost(n, m)
18:        if m ∉ O then
19:            Add m to O
20:        else if tentative_g ≥ g(m) then
21:            continue
22:        end if
23:        g(m) ← tentative_g
24:        f(m) ← g(m) + h(m)
25:        parent(m) ← n
26:    end for
27: end while
28: return failure
```

- Euclidean Distance: This heuristic is used when movement in any direction is allowed. It allows movement in diagonal directions as well and gives the shortest path between 2 points.

$$h(n) = \sqrt{(x_{\text{goal}} - x_n)^2 + (y_{\text{goal}} - y_n)^2} \tag{5.26}$$

  This heuristic assumes a uniform cost for all movements.

In contrast to the Manhattan distance, Euclidean distance measures the straight-line distance between two points. However, it gives a better approximation of the true shortest path in environments that permit free movement in any direction.

5.4.2 Safe Path Planning Methodology

In the context of navigating through fire-affected regions, we modify the traditional A* algorithm to ensure that we get the shortest possible path that avoids the risk of running into a hazardous area. Therefore, the aim of this methodology is not only to provide the shortest route between two points in a fire-affected area but also the safest path between those points. The modified A* algorithm implemented in this research work specifically addresses this by integrating additional logic to effectively avoid fire regions altogether.

The change to the traditional A* algorithm involves incorporating a fire avoidance heuristic. The predicted fire mask($\hat{Y}$) is converted to a binary grid with the nodes under the fire mask denoted by '1' and other nodes as '0', indicating a safe, free-from-fire region. The algorithm treats any cell in the binary grid marked 1 (indicating fire presence) as impassable. This direct approach ensures that the pathfinding logic completely avoids areas marked by fire, enhancing safety and reliability. During the exploration of paths, each node considers only those neighboring nodes that are not within fire-affected regions. Euclidean Distance is employed as the heuristic function for this traversal. This filtering is achieved by checking the state of each neighbor before adding it to the set of nodes to be explored. If a node falls within a fire region, it is excluded from further consideration, effectively pruning paths that would lead through hazardous areas.

Beyond just avoiding the fire cells, this algorithm also implements an adjustable buffer, which is essentially a safe distance from the closest fire cell to avoid risk. This ensures the algorithm does not plan a path too close to the fire cells and is implemented by checking the surrounding cells within a specified buffer radius. If any cell within this buffer is marked as 'fire' or $'1'$, the cell under consideration is marked unsafe for passage. This mechanism ensures a safe distance from fire-affected areas, which is critical in situations where fire

might spread or proximity to fire can be hazardous even without direct contact (e.g., due to heat or smoke).

Algorithm 3 demonstrates essential utility functions to support the A* search mechanism optimized for safe path planning. The heuristic Function calculates the Euclidean distance between two nodes. With this function, traversal in all directions is made possible to better capture the possible trajectory of a path in cases of wildfires. '$IsSafe'$ function evaluates the safety of a node by inspecting the surrounding area within a specified buffer for fire regions. By using this function, the algorithm will only consider nodes that are at a safe distance defined by a buffer from the fire areas. '$ReconstructPath'$ function constructs the route from the goal node back to the start node by tracing through a map of predecessors. It maps out the final path once the destination is reached based on the safest and shortest route identified during the path search.

Algorithm 4 showcases the core search algorithm of the optimized search process. The predicted fire mask, output of the Vi-Net model, acts as input to the optimized A* search algorithm. Other inputs to the algorithm are start and goal nodes and a safety buffer. This core algorithm makes use of functions described in algorithm 3 to compute the Euclidean-based heuristic function, the evaluation of the safe nodes for traversal, and for reconstruction of the optimal path. if found, based on the set criteria. First, the open and closed sets are initialized to manage nodes during the search, along with maps to keep track of actual and estimated costs to reach each node. The algorithm processes nodes from the open set, prioritizing by lowest estimated total cost and ensuring each node's safety through the 'Is_Safe' function'. It evaluates adjacent nodes in eight possible movement directions, updating paths and costs based on a combination of actual traversed distance($g(n)$) and heuristic estimations($h(n)$). If a path to the goal is successfully found, it is returned; otherwise, the algorithm indicates that no safe path could be established.

The optimized A* algorithm ensures that the pathfinding process is not only focused on finding the shortest path but also on ensuring the safety and possibility of the path in dynamic and potentially hazardous conditions. This algorithm is particularly useful in scenarios such as emergency evacuations, allocation of firefighting personnel, or autonomous vehicle navigation in environments where safety is paramount.

This method revamps the classic A* algorithm improving its applicability to more complex real-world scenarios where risk factors must be managed strategically. It leverages the core strengths of the A* algorithm to find the least-cost path in a computationally effective way. The choice of heuristic ensures faster and optimal results.

Algorithm 3 Utility Functions for Optimized A* Path Planning

Input: Fire mask $\hat{Y}$, start node *start*, goal node *goal*, buffer buf
function HEURISTIC($node1, node2$)
    return $\sqrt{(node1_x - node2_x)^2 + (node1_y - node2_y)^2}$
end function
function ISSAFE($grid, node, buf$)
    for $dx = -buf$ to buf **do**
        for $dy = -buf$ to buf **do**
            $nx \leftarrow node_x + dx$
            $ny \leftarrow node_y + dy$
            if $nx < 0$ or $nx \geq$ length of $grid$ or $ny < 0$ or $ny \geq$ length of $grid[0]$ or $grid[nx][ny] = 1$ **then**
                return false
            end if
        end for
    end for
    return true
end function
function RECONSTRUCTPATH($cameFrom, current$)
    $path \leftarrow$ empty list
    while $current$ in $cameFrom$ **do**
        Prepend $current$ to $path$
        $current \leftarrow cameFrom[current]$
    end while
    Prepend *start* node to $path$
    return $path$
end function

Algorithm 4 Optimized A* Search Algorithm Using Predicted Fire Mask

Input: Fire mask $\hat{Y}$, start node *start*, goal node *goal*, buffer buf
Output: Path from *start* to *goal* avoiding fire-affected areas
procedure ASTAR($grid, start, goal, buf$)
    $dirs \leftarrow$ set of 8 directions for movement (including diagonals)
    $openSet \leftarrow$ priority queue containing (HEURISTIC($start, goal$), $start$)
    $cameFrom \leftarrow$ empty map
    $gScore \leftarrow$ map with default value of ∞, $gScore[start] \leftarrow 0$
    $fScore \leftarrow$ map with default value of ∞, $fScore[start] \leftarrow$ HEURISTIC($start, goal$)
    $closedSet \leftarrow$ empty set
    while *openSet* is not empty **do**
        $current \leftarrow$ node in *openSet* with the lowest $fScore$ value
        if $current = goal$ **then**
            return RECONSTRUCTPATH($cameFrom, current$)
        end if
        Remove *current* from *openSet*
        Add *current* to *closedSet*
        for each dir in $dirs$ **do**
            $neigh \leftarrow (current_x + dir_x, current_y + dir_y)$
            if not ISSAFE($grid, neigh, buf$) **then**
                continue
            end if
            $tempG \leftarrow gScore[current] +$ distance between *current* and *neigh*
            if $neighinclosedSet$ and $tempG \geq gScore[neigh]$ **then**
                continue
            end if
            if $tempG < gScore[neigh]$ or $neigh \notin openSet$ **then**
                $cameFrom[neigh] \leftarrow current$
                $gScore[neigh] \leftarrow tempG$
                $fScore[neigh] \leftarrow tempG +$ HEURISTIC($neigh, goal$)
                if $neigh \notin openSet$ **then**
                    Add *neigh* to *openSet*
                end if
            end if
        end for
    end while
    return false ▷ Path not found
end procedure

Chapter 6

Results & Experiments

This chapter presents the setup of the experiment, the evaluation metrics used, and the results achieved in wildfire spread prediction and safe path planning. It provides a comprehensive analysis of the comparison between baseline models and the proposed hybrid model, Vi-Net, the performance of various appropriate loss functions, and the working of safe path generation within given constraints. The analysis gives deep insights gained from the study while highlighting areas for improvement. This chapter plays a key role in demonstrating the efficacy of the proposed model and its practical implications for wildfire management.

All experiments were conducted on the dataset as described in the section 4, which contains various meteorological and topographical features like windspeed, vegetation index, fire mask at day t along with fire mask at day $t+1$ as ground truth. The input images were resized to 32×32 and preprocessed for modeling. For wildfire spread prediction, we used TensorFlow and trained the models on a machine with an NVIDIA Geforce RTX 3060Ti GPU. Each model was trained with a batch size of 32, and Adam as an optimizer. The dataset was split as follows: 14979 for the training set, 1877 for the validation set, and 1689 for the test set.

6.1 Experiments with Loss Functions

Choosing an appropriate loss function is crucial for tasks like wildfire spread prediction. Wildfire datasets often exhibit a stark imbalance in class representation, as the number of non-fire areas is significantly greater than the actively burning areas over a captured

geographical location. The choice of the loss function directly impacts the model's learning capabilities as it influences how the gradients are calculated during training. It enhances the model's robustness and generalization ability across unseen data.

While training the model, we experimented with different loss functions and noted the performance of these loss functions across the 3 models as seen in Table 6.1. The 4 loss functions chosen for the experiments are described below:

6.1.1 Weigted Cross Entropy(WCE) with Logits

In segmentation tasks, the cross entropy loss attempts to measure the difference between the probabilities of the target and predicted labels. WCE addresses the class imbalance by applying different weights to the positive and negative classes, effectively giving more priority to the minority class.

$$L(y, \hat{y}) = -\sum_{i=1}^{N} w_i \cdot [y_i \cdot \log(\hat{y}_i) + (1 - y_i) \cdot \log(1 - \hat{y}_i)] \tag{6.1}$$

where y is the true labels, $\hat{y}$ is the predicted probability, N is the number of samples, w_i is the weight, adjusted inversely to class frequencies. This enhances sensitivity to less frequent wildfire spread patterns and improves prediction accuracy.

6.1.2 Dice Loss

Dice Loss measures the similarity between predicted and target segmentation masks. It is based on the Dice Coefficient and offers a direct measure of the model's ability to capture the area of overlap between the predicted and actual masks

$$\text{Dice Loss} = 1 - \frac{2|Y \cap T|}{|Y| + |T|} \tag{6.2}$$

where Y represents the predicted segmentation and T represents the actual segmentation. This function focuses on the correctly identified pixel to the combined total of target and predicted pixels and promotes accurate delineation of the fire-affected areas.

6.1.3 Tversky Loss

Tversky Loss is an extension of Dice Loss wherein it adds flexibility to control relative weights of false positives(FP) and false negatives(FN). This is particularly useful in imbalanced datasets.

$$\text{Tversky Loss} = 1 - \frac{|Y \cap T|}{|Y \cap T| + \alpha|Y \setminus T| + \beta|T \setminus Y|} \tag{6.3}$$

where Y represents the predicted segmentation and T represents the actual segmentation, α and β are weights that control the relative importance of FP and FN. Customizable α and β give us more flexibility over sensitivity reducing the overfitting over non-fire pixels.

6.1.4 Focal Tversky Loss

Focal Tversky Loss further enhances the Tversky Loss by introducing a "focus" parameter to give more importance to the difficult samples, thus enhancing the performance in advanced segmentation tasks.

$$\text{Focal Tversky Loss} = (\text{Tversky Loss})^{\gamma} \tag{6.4}$$

where γ is the focusing parameter, enhancing the importance of correcting misclassification. It enhances overall prediction accuracy and robustness in diverse environments by paying more attention to complex fire spread patterns.

We compare the results of all the experiments conducted for 4 loss functions across the 3 models in table 6.1.

The table 6.1 presents the performance metrics of U-Net, Vision Transformer (ViT), and the proposed hybrid model, Vi-Net, for predicting the spread of wildfires on the following day, evaluated using four distinct loss functions: Weighted Cross-Entropy (WCE), Dice Loss, Tversky Loss, and Focal Tversky Loss. The dataset exhibits significant imbalance, with "Fire" regions making up only 3.6%–4.17% of the total. To address this, predictions were implemented with a threshold of 0.85, prioritizing Precision by classifying only high-confidence regions as fire. This approach reduces false positives but also lowers Recall by failing to identify some fire regions.

U-Net excels in capturing the fine details of fire regions due to its convolutional architecture and skip connections. It achieves its best F1 Score of 0.8403 with Focal Tversky Loss, which effectively manages class imbalance and regions that are challenging to predict.

Table 6.1: Performance Metrics for Different Models and Loss Functions

Models	**Loss Functions**	**Precision**	**Recall**	**F1 Score**
U-Net	WCE	0.9433	0.7018	0.8048
	Dice Loss	0.7491	0.8712	0.8056
	Tversky Loss	0.9957	0.7050	0.8255
	Focal Tversky Loss	0.9542	0.7506	**0.8403**
ViT	WCE	0.8492	0.5671	0.6801
	Dice Loss	0.9815	0.5635	0.7159
	Tversky Loss	0.9939	0.5559	0.7130
	Focal Tversky Loss	0.9873	0.5941	**0.7418**
Vi-Net	WCE	0.9250	0.9322	0.9286
	Dice Loss	0.9656	0.8999	0.9316
	Tversky Loss	0.9626	0.9418	0.9521
	Focal Tversky Loss	0.9834	0.9619	**0.9725**

However, its Recall is limited, as U-Net struggles with detecting fire regions influenced by long-range factors such as wind patterns, a limitation stemming from its narrow receptive field.

ViT achieves high Precision, peaking at 0.9939 with Tversky Loss, enabled by its ability to model global relationships across the image. However, its Recall remains low (maximum 0.5941 with Focal Tversky Loss), as its patch-based processing sacrifices the detection of local details, making it difficult to identify small or sparse fire regions. This limitation is exacerbated by the high threshold setting, which filters out uncertain predictions.

Vi-Net, the hybrid model, combines the strengths of U-Net and ViT to deliver the best overall performance. It achieves an F1 Score of 0.9725 with Focal Tversky Loss, along with a Precision of 0.9834 and a Recall of 0.9619. Vi-Net leverages local details through U-Net's skip connections and long-range dependencies via ViT's self-attention mechanism. This dual approach effectively addresses U-Net's limitations in global context and ViT's challenges with spatial detail.

Focal Tversky Loss emerges as the most suitable loss function, adept at balancing false negatives ($\alpha = 0.7$) and focusing on hard-to-predict regions ($\gamma = 0.75$). This configuration, combined with Vi-Net, ensures high-confidence predictions and robust performance, even under strict thresholds.

In conclusion, Vi-Net outperforms U-Net and ViT individually by leveraging their combined strengths. It provides precise, reliable predictions for the spread of wildfires, positioning it as an invaluable tool for managing complex environmental challenges.

6.2 Wildfire Segmentation Results

6.2.1 Quantitative Results

Given the input image data of $64 \times 64 \times 12$ dimensions, the Vi-Net model can predict the spread of fire over 24 hours by classifying each pixel as "fire" or "no fire" in the next time step. The training and validation loss calculated over 100 epochs, with early stopping, for predicting wildfire spread across Vi-Net is shown in Fig. 6.1.

The left plot of Fig. 6.1(Loss Over Epochs) illustrates the model's training and validation loss over epochs. Initially, both training and validation losses exhibited a steep decline, indicating rapid learning. The training loss continued to decrease steadily and maintained a low value close to zero throughout the remaining epochs. In contrast, the validation loss showed greater fluctuation but generally remained low after the initial sharp decline, suggesting the model's robustness against overfitting. The right plot of Fig. 6.1(Accuracy Over Epochs) displays the F1 score for both training and validation sets. The training F1 score approached 1.0, indicating excellent model performance on the training set. The validation F1 score also improved significantly in the initial epochs and closely approached the training F1 score, achieving its peak near 0.95. The validation F1 score exhibited minor variations but remained high, demonstrating the model's effectiveness in generalizing to new data. Overall, the depicted metrics illustrate that the hybrid U-Net and ViT model performed exceptionally well on the task of wildfire spread prediction. The convergence of training metrics and the sustained high validation metrics indicate both strong learning and generalization capabilities.

Table 6.2 highlights the performance of U-Net, Vision Transformer (ViT), and the proposed hybrid model, Vi-Net, in predicting wildfire spread over training, validation, and test datasets. We compare the models on the metrics described in section 5.3.3, Intersection over Union (IoU), F1 Score, Precision, and Recall, offering insights into the strengths and



Figure 6.1: The training and loss function plots for Vi-Net.

weaknesses of each model in capturing fire regions while minimizing false positives (FP) and false negatives (FN).

U-Net performs consistently across datasets and demonstrates a strong ability to avoid false positives, as it achieves a high precision of 0.9981 on the training set and 0.9542 on the test set. However, its recall is comparatively lower, indicating a significant number of false negatives, where actual fire regions were missed. This aligns with U-Net's dependence on convolutional operations and skip-connections, which excel in capturing fine-grained spatial details but struggle with long-range dependencies. Consequently, U-Net performs well on samples with smaller or localized fire regions but underperforms over samples with fire spread involving broader environmental factors such as wind or vegetation continuity.

ViT achieves the highest IoU (0.9130) on the training set, reflecting its ability to model global dependencies effectively. ViT excels in precision (0.9873 on the test set), but its recall remains consistently lower (maximum 0.5941), which highlights its ability to correctly classify high-confidence fire regions but difficulty in detecting actual fire regions. This can be attributed to its patch-based processing, which sacrifices fine spatial details for global contextual awareness. While ViT effectively captures the influence of large-scale environmental factors, such as wind patterns and drought indices, its inability to detect smaller or sparse fire regions leads to a higher number of false negatives. This limitation affects its overall F1 Score (0.7418 on the test set), as the task of wildfire spread prediction over the period of 24 hours has more instances of samples that showcase the need to capture spatial details than global patterns.

Vi-Net achieves a balanced and superior performance across all datasets, with the highest F1 Score (0.9725), precision (0.9979), and recall (0.9619) on the test set. The model's recall highlights its effectiveness in minimizing false negatives, demonstrating a high level of precision in identifying both small and large fire regions. This success is attributed to Vi-Net's integration of U-Net's localized feature preservation and ViT's global contextual modeling. By combining U-Net's ability to retain spatial precision through skip connections with ViT's capacity for long-range dependency modeling via self-attention, Vi-Net effectively captures both micro-level and macro-level patterns of wildfire spread. The weighted fusion of U-Net and ViT outputs ensures that the model leverages their complementary strengths, reducing both false positives and false negatives.

High precision in wildfire spread prediction is of paramount importance to minimize false positives, ensuring that resources are directed to genuinely at-risk areas. The high threshold of 0.85 used in prediction limits the inclusion of uncertain predictions as fire regions. While this approach inherently reduces false positives, it often risks lowering recall by excluding some true positives. However, Vi-Net's hybrid architecture mitigates

this trade-off effectively, maintaining high recall while benefiting from increased precision. This balanced performance highlights Vi-Net's robustness in accurately segmenting fire regions under high confidence prediction thresholds, making it particularly well-suited for risk-sensitive applications like wildfire management.

Vi-Net demonstrates a significant advantage over U-Net and ViT by combining their individual strengths, resulting in a robust solution for wildfire spread prediction. Its ability to achieve high IoU and F1 Scores while maintaining a balanced Precision and Recall highlights its effectiveness in addressing both localized and global dependencies. These results establish Vi-Net as a better model for managing complex environmental challenges, providing accurate predictions while minimizing false positives and negatives.

Table 6.2: Comparison of Performance Across The Models

Dataset	**U-Net**				**ViT**				**Vi-Net**			
	IoU	F1	P.	R.	IoU	F1	P.	R.	IoU	F1	P.	R.
Training set	0.7879	0.8813	0.9981	0.7890	0.5625	0.7200	0.9130	0.5943	0.9803	0.9900	0.9979	0.9822
Validation set	0.7409	0.8511	0.9727	0.7566	0.5261	0.6894	0.9527	0.5402	0.9902	0.9880	0.9951	0.9809
Test set	0.7325	0.8403	0.9542	0.7506	0.5829	0.7418	0.9873	0.5941	**0.9415**	**0.9725**	0.9834	0.9619

6.2.2 Qualitative Results

This section presents visual results for each prediction model. Each model's performance is illustrated with actual and predicted fire spread images, as shown in Fig. 6.2. These visualizations highlight the strengths and challenges of each model in capturing fire boundaries and affected areas. All three models accurately predict the spread of wildfire whereas the proposed model Vi-Net showcases superior predictions as compared to the baseline models. The visual results of the three models highlight their distinct strengths and weaknesses.

The predictions by U-Net in Fig. 6.2a demonstrate precise localization capabilities, effectively preserving fine-grained information in fire spread mapping. For smaller and well-defined fire regions, the predictions overlap significantly with the ground truth. However, for next-day wildfire spread dataset, occasionally struggles to detect fires accurately when they are spatially distant from one another, resulting in missed regions or incomplete predictions.

In the predictions by ViT (Fig. 6.2b), we notice that the model captures larger fire regions or spread-out patches more effectively. The ViT excels in extracting global dependencies, capturing fires and it's spread in multiple regions across the input image frame even when they are far apart. Nevertheless, it misses delineating the boundaries of the fires, leading to less precise localization.

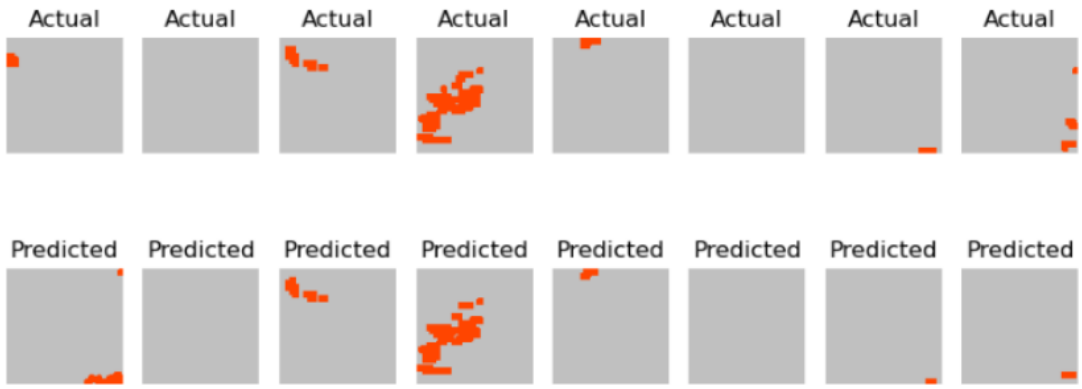
The predictions by Vi-Net (Fig. 6.2c) show the most balanced and accurate results among the three models. Vi-Net integrates the strengths of both local and global contextual information, achieving a superior balance. It produces highly precise and comprehensive predictions of fire spread, accurately capturing both fine details and spatially distributed patterns, showcasing its advanced segmentation capabilities.

Fig. 6.3 provides boundary analysis of the Actual(ground truth) vs Prediction results of the output by the Vi-Net model. Blue outlines in the first row are outlines of Prediction engraved on GT and Red outlines in the second row are outlines of GT engraved on prediction results. Both red and blue colored outlines resemble each other very closely demonstrating that our Vi-Net model does an excellent job in capturing the complex intricacies of wildfire spread dynamics. Integration of UNet and ViT has enabled the Vi-Net model to accurately capture global dependencies as well as precisely capture the spatial features which results in predictions having the same fire mask shape as the actual shape.

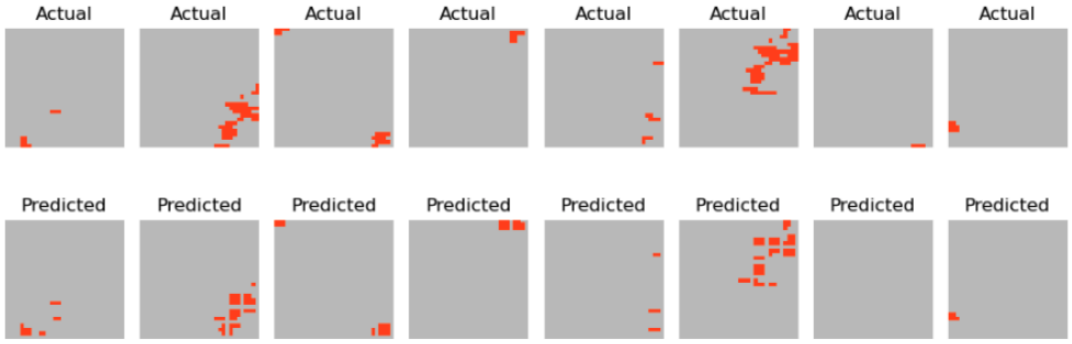
For instance, in the first and sixth predictions, fire masks are small and spread out but the Vi-Net model does a great job of locating those fire masks as they are actually present showcasing the model's prowess in capturing long-range fire masks. Conversely, in the fifth and seventh predictions, it is clear whether it is a single small fire mask(fifth) present in

the area or a complex structured fire mask(seventh) present towards a corner, our model shows excellent capability in recognizing the spatial pattern producing closely similar fire mask shape.

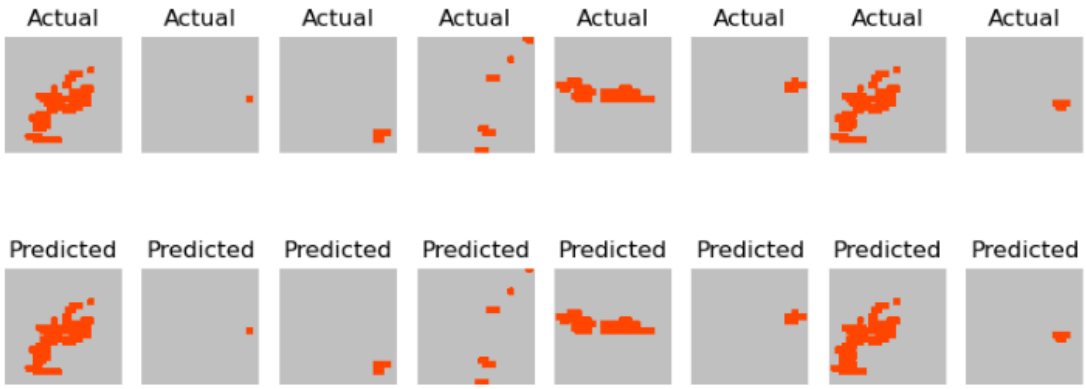
Both the quantitative and qualitative results demonstrate that Vi-Net outperforms the U-Net and ViT in all evaluation metrics, achieving an F1 score of 0.9725 and an IoU of 0.9415. The Vi-Net employs the localized spatial feature extraction ability of U-Net and the global context understanding ability of ViT which gives the advantage to the hybrid approach over the U-Net and ViT considered individually.



(a) U-Net Results.



(b) ViT Results.



(c) Vi-Net Results.

Figure 6.2: Results from three image segmentation models: U-Net, ViT, and Vi-Net.

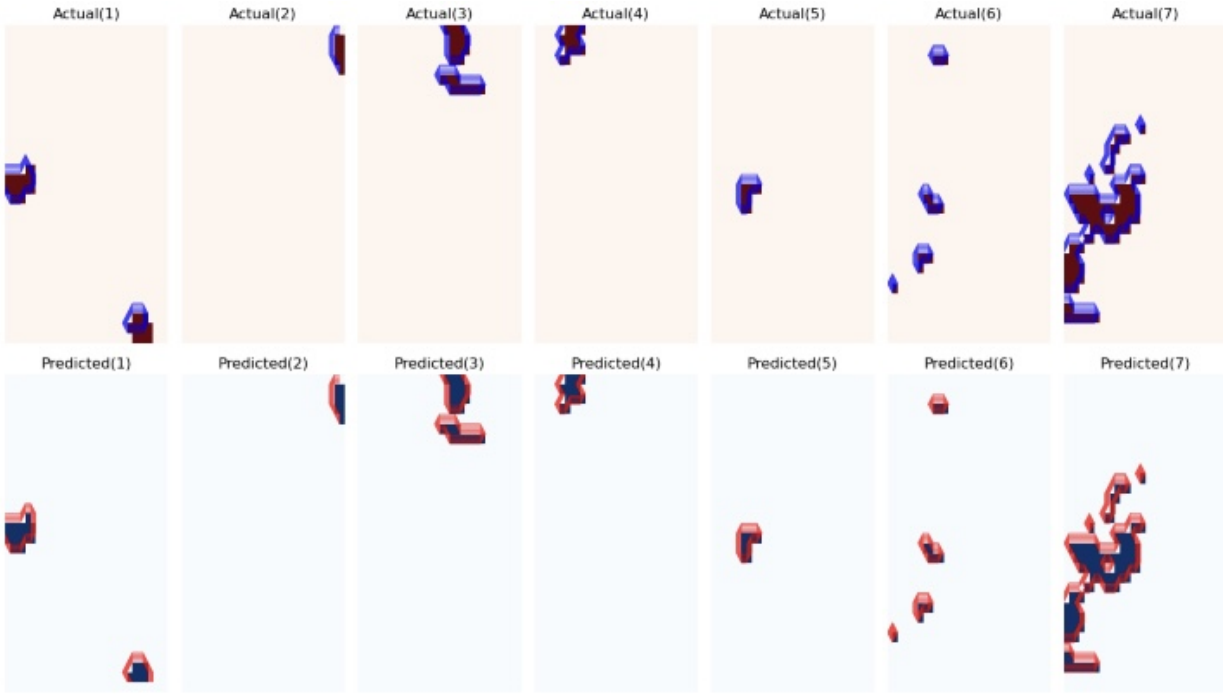


Figure 6.3: This results from the Vi-Net model compares fire boundary of ground truth(GT) and predictions results.

6.3 Safe Path Planning Results

This section presents the results of the safe path planning achieved through the application of the modified A* algorithm, which leverages the fire spread predictions obtained from the Vi-Net model. The objective of this section is to evaluate the performance of the optimized A*(OA*) algorithm in finding traversable paths avoiding the fire-affected regions. The OA* algorithm achieves this by converting the received fire mask from the Vi-Net model into a binary grid of 0 and 1 representing non-fire and fire respectively, and finds the safe routes considering the specified safety constraints. The analysis is divided into two parts: section 6.3.1 discusses scenarios where a feasible and safe path exists and section 6.3.2 highlights cases where no feasible path can be found due to environmental constraints or configuration challenges. In both cases, the algorithm's ability to adapt to various complexities and ensure safety is critically assessed, providing valuable insights into its suitability for real-world applications. Fig 6.4 extends on Fig. 3.2 and demonstrates the application and final result of the OA* algorithm on the input binary grid. The received fire masks(outputs of the spread prediction) are converted into the binary grid to lay the foundation for our OA* algorithm to work on, converted to a corresponding fire map for ease of representation. In the final step, the red line between the green and blue markers shows the safe path between the start and goal node considering all the safety constraints.

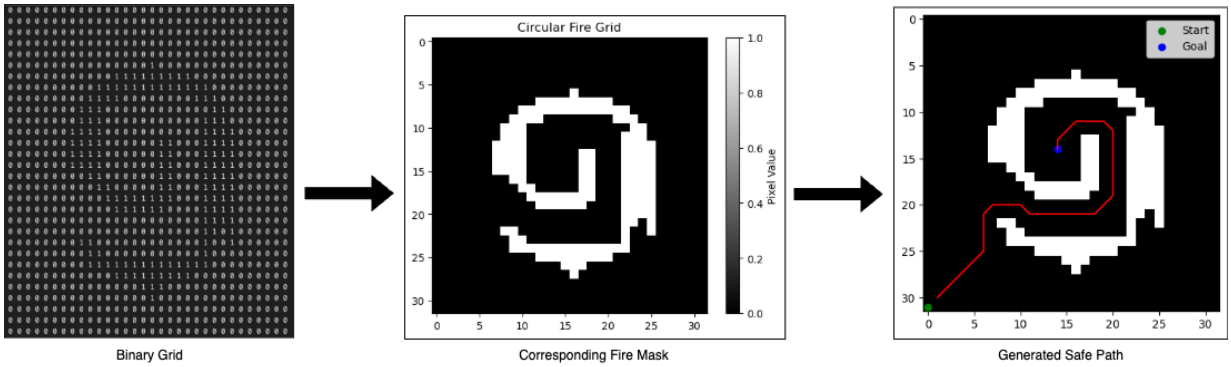


Figure 6.4: Application of OA* Algorithm on the Wildfire Spread Prediction results.

6.3.1 Scenarios with Navigable Safe Routes

The results in Fig 6.5 show the application of the OA* algorithm for path planning, utilizing the wildfire spread prediction results achieved as the output of the Vi-Net model. The figure

shows the path from the start node (green marker) to the goal node (blue marker). The red lines represent the calculated paths, while the binary grid highlights fire-affected areas (white cells) and safe zones (gray cells).

- Scenario 1 (Fig. 6.5a): The path traverses a relatively sparse and disconnected region. The algorithm selects a direct route while maintaining a safe buffer distance from fire areas.
- Scenario 2 (Fig. 6.5b): Here, fire regions display denser and more complex, with multiple obstructions separating the start and goal points. The path demonstrates the algorithm's ability to navigate through narrow corridors while maintaining a safe buffer distance.
- Scenario 3 (Fig. 6.5c): In this scenario, a large continuous fire zone dominates the environment, causing the algorithm to identify a longer but safer path, considering the buffer, around the affected area. The generated path reflects the algorithm's capacity to balance shortest route planning with safety, effectively avoiding large clusters.
- Scenario 4 (Fig. 6.5d): This scenario presents a highly complex and obstructive scenario with a spiral distribution of fire-affected regions. The algorithm successfully navigates through a highly constrained environment, showcasing its robustness in complex spatial arrangements.

In all the above scenarios, not only does the modified A* algorithm avoid fire-affected regions, but it also maintains a safe distance buffer to ensure safe passage. This demonstrates the effectiveness of the buffer mechanism implemented in the is_safe function, which checks surrounding cells of potential path nodes to exclude unsafe options. The results highlight the efficiency and adaptability of the algorithm in different settings, from sparse distributions (as in scenario 1) to highly complex arrangements such as spiral fire masks (as in scenario 4), the paths remain optimal and feasible, indicating that the algorithm can handle varying degrees of environmental complexity. The Euclidean distance heuristic employed in the algorithm contributes to generating direct paths, and minimizing detours while the safe distance buffer ensures the safety of the generated paths between the start and the goal.

The results obtained validate that the modified A* algorithm is suitable for applications requiring safe and efficient navigation in fire-prone environments. The integration of this algorithm with the ViNet (hybrid) model for fire spread prediction demonstrates a powerful

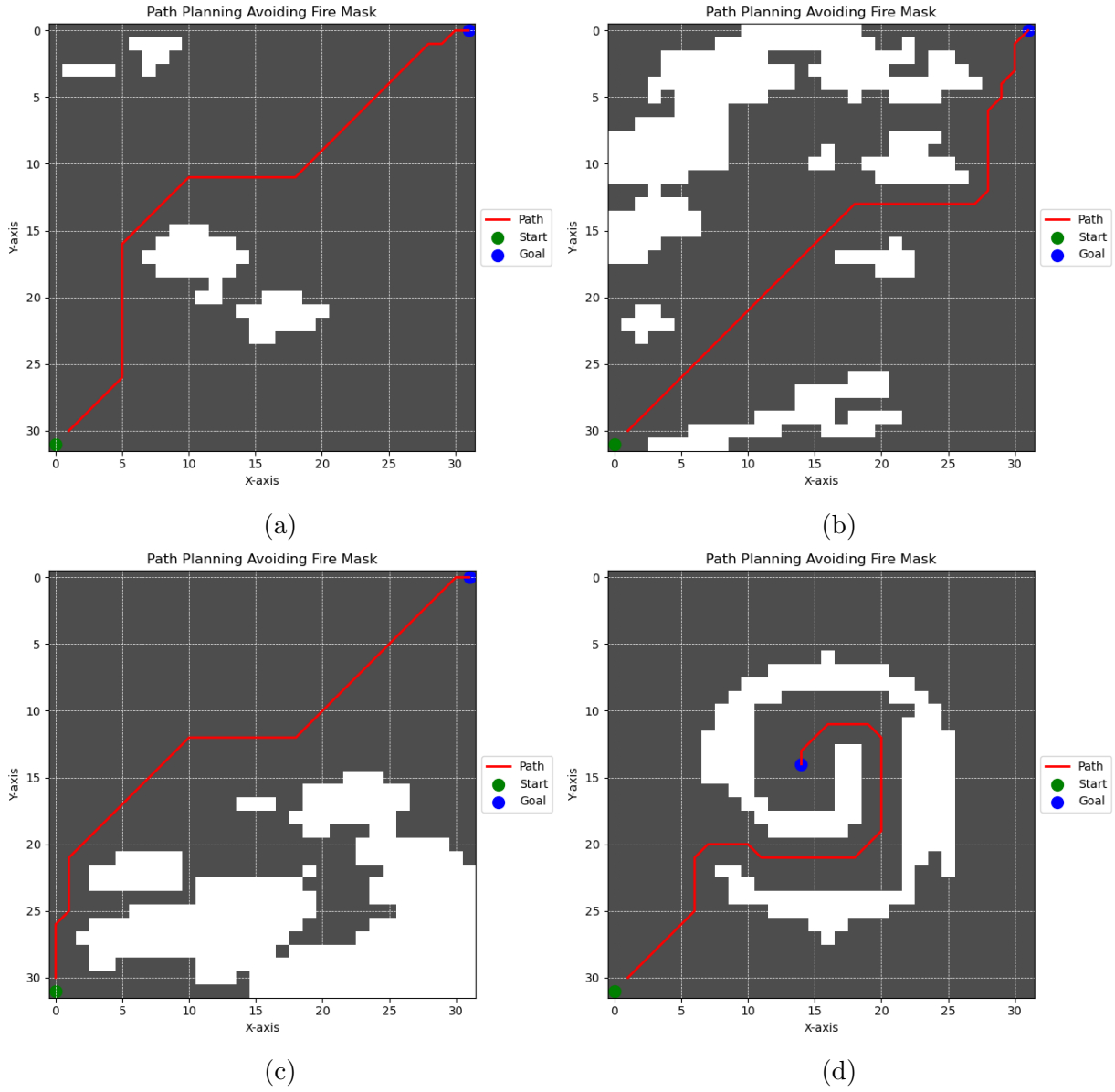


Figure 6.5: Results of Successful Path Generated Based on Optimized A* Algorithm

synergy between data-driven fire spread modeling and algorithmic safe path planning, providing a robust framework for navigation in hazardous environments.

6.3.2 Scenarios with No Safe Routes

The results obtained in figure 6.6 show cases where the algorithm does not generate a path between given start and goal points due to specific constraints or challenges. The optimized A*(OA*) algorithm ensures that a route is not suggested between the start and goal node that does not satisfy the given safety constraints at all steps. These cases highlight the accuracy of the algorithm in identifying when pathfinding is not feasible, thus ensuring the integrity of its output by avoiding unsafe or impractical solutions. The absence of a red path indicates that the algorithm successfully determined that no safe route exists from the start node (green marker) to the goal node (blue marker).

- Scenarios 1 & 2 (Fig. 6.6a & 6.6b): These cases demonstrate situations where no path is possible due to complete blockages of any safe passages caused by areas under fire and the applied safe distance. As fire is a dynamic obstacle that is constantly moving, the safe distance buffer ensures that no such routes are generated that compromise safety, leaving no navigable route between the start and the goal nodes. The inability of the OA* algorithm to generate a path shows that it adheres strictly to safety protocols, enforced by the safe buffer mechanism.
- Scenarios 3 & 4 (Fig. 6.6c & 6.6d): These situations occur when the fire mask directly overlaps or is too close to the specified start or goal nodes. As a result, the OA* algorithm correctly identifies that a path cannot be planned, ensuring accuracy and compliance with safety constraints. This behavior ensures the integrity of the algorithm, as it avoids compromising safety standards by initiating or concluding paths in unsafe areas. It is important for safe path planning that the OA* algorithm covers edge cases like these and only suggests a path when the complete path adheres to the enforced safety protocols. Such cases also highlight the importance of proper placement of start and goal nodes in practical implementations.

These no-path scenarios validate the robustness and reliability of the modified A* algorithm in handling edge cases and upholding the safety protocols as specified by the users. By accurately capturing high-risk regions and enforcing safety margins, the algorithm ensures that paths are only suggested when they are truly safe and feasible. This level of precision is critical for real-world applications, such as evacuation planning or safe traversal of firefighting personnel, where suggesting an unsafe route could have catastrophic consequences.

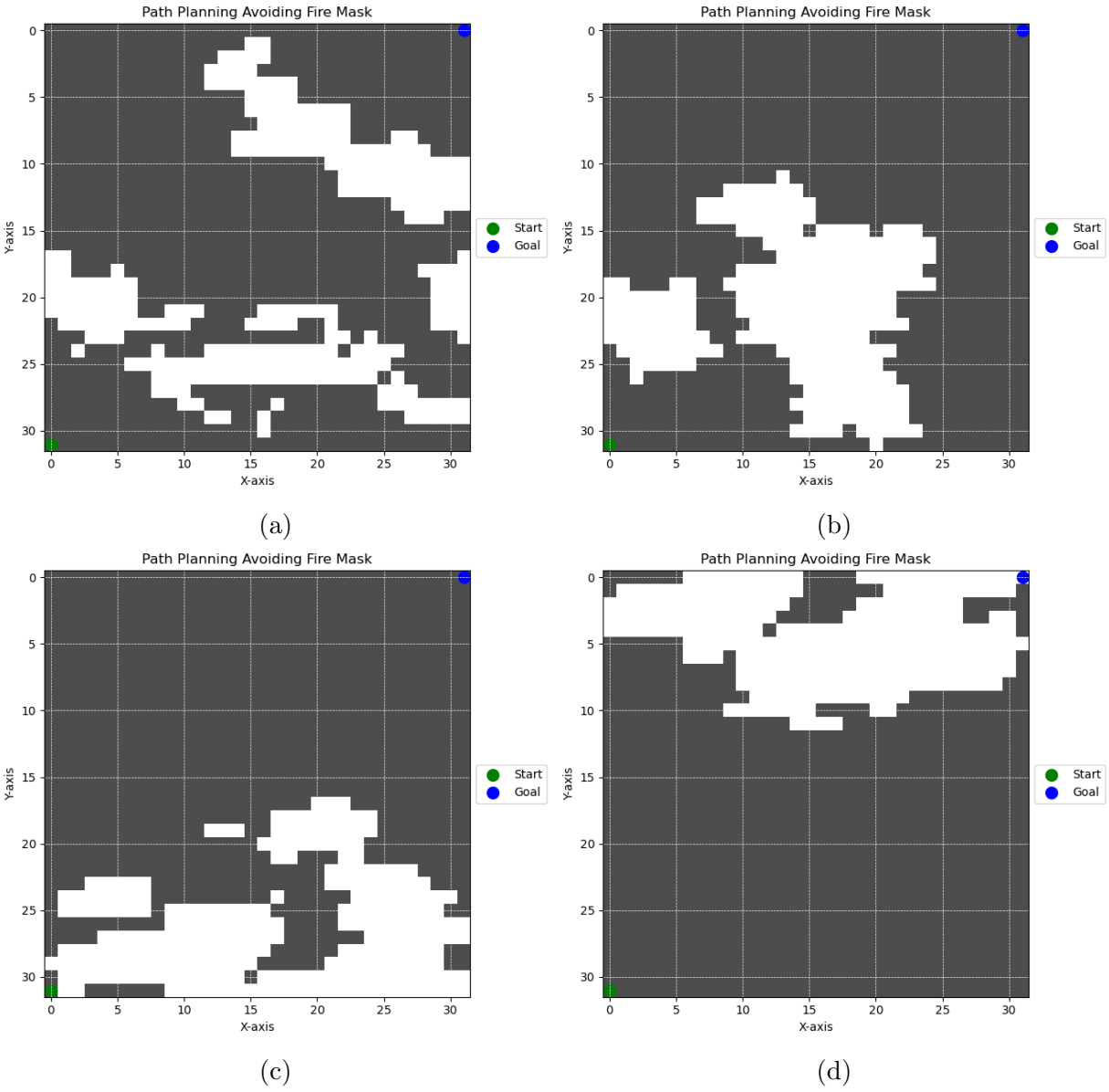


Figure 6.6: Results of No Path Exist Based on Optimized A* Algorithm

Chapter 7

Conclusion & Future Work

This chapter summarizes the findings of the thesis, discusses the results of the experiments conducted, and outlines potential areas of for future work. It reflects on the study's contributions to the field of wildfire spread prediction and broader impacts of fire mitigation strategies, such as safe path planning. The chapter concludes with suggestions and potential areas of advancement in enhancing the fire spread model, extending the dataset, increasing the robustness of the path planning algorithms, and ideas for better modeling strategies.

7.1 Conclusion

This research attempted to address a challenging yet critical problem in the domain of wildfire spread prediction, aiming to develop an advanced deep learning (DL) model for predicting fire spread on a real-world fire spread dataset including various topographical, meteorological, anthropological, and vegetation factors. This study also aims to leverage wildfire spread prediction for mitigation efforts, with a specific emphasis on safe path planning which is essential for planning evacuation routes and resource allocation. Wildfire disasters pose significant threats to infrastructure, human lives, and the economy, necessitating the development of accurate and reliable predictive models to facilitate proactive disaster management strategies. The dynamic nature of wildfire behavior, compounded by challenges such as imbalanced datasets, limited high-resolution environmental data, and the need to model complex spatial and temporal interactions, has historically hindered the development of robust prediction systems. This thesis addresses these challenges by examining state-of-the-art ML models in the field of semantic segmentation over images

and their architectural designs to enhance the predictive accuracy of wildfire spread modeling systems. As a result of extensive experimentation and rigorous evaluation, this study delivers a hybrid model that overcomes individual shortcomings of the baseline models. This study also delivers a methodology to use the predictions for safe path-finding using an modified A* algorithm.

The U-Net model demonstrated high efficiency in localized precision of fire regions, making use of its encoder-decoder architecture with skip connections to preserve pixel-level spatial details. However, its limited receptive field constrained its ability to understand complex scenarios where context from distant parts of the image is critical for wildfire dynamics. Conversely, the ViT model effectively captured these global dependencies by utilizing a self-attention mechanism, that helps capture long-range dependencies within an image. Despite its strength in contextual modeling, ViT faced challenges in capturing fine details, particularly for small or sparse fire regions because unlike U-Nets, ViTs do not have built-in inductive biases such as translation invariance and locality.

The hybrid Vi-Net model addressed these limitations by a strategic fusion of the strengths of U-Net and ViT. A unique intersection has been recognized where two baseline state-of-the-art models display opposing strengths and limitations. Strategic integration of these models could potentially alleviate their individual weaknesses and improve overall efficacy. This integration resulted in superior predictive performance across all metrics, achieving a high F1-score of 97.25% and an IoU of 0.9415 on the test set. Vi-Net's ability to maintain spatial detail while leveraging global context highlights its robustness and adaptability in handling the complex interactions that govern wildfire spread.

A core focus of this study was the evaluation of loss functions to address the challenge of class imbalance inherent in wildfire prediction datasets. Four widely-used loss functions, Weighted Cross Entropy, Dice Loss, Tversky Loss, and Focal Tversky Loss, were rigorously tested across all models. The experimentation revealed Focal Tversky Loss as the most effective, particularly for Vi-Net, due to its ability to balance false negatives(FNs). The use of Focal Tversky Loss further enhanced the Vi-Net's performance, effectively addressing class imbalance and emphasizing the accurate prediction of challenging regions, such as fire boundaries and isolated patches.

The predictive models were also integrated into an modified A* algorithm for safe path planning. This application provided a practical extension of the research, enabling the generation of safe routes that avoided fire-prone areas. The system's ability to detect scenarios with no viable paths demonstrated its reliability and robustness in operational contexts. By combining accurate prediction with actionable insights, this work bridges the gap between predictive analytics and real-world disaster mitigation.

In conclusion, this research advances wildfire spread prediction by leveraging cutting-edge machine learning methodologies. The hybrid Vi-Net model, supported by its innovative architecture and tailored loss functions, sets a new benchmark in predictive accuracy and reliability. Its integration into path planning demonstrates the practical implications of the approach, providing a comprehensive framework for risk assessment and mitigation. Future work could explore real-time data integration, further enhancing the system's scalability and responsiveness in dynamic wildfire scenarios. This study underscores the importance of combining technical innovation with practical applications to address the pressing challenges of wildfire management, ultimately contributing to the safety of ecosystems and communities alike.

7.2 Future Work

While this research serves as a strong foundation, there are several promising avenues to explore for future research direction:

1. **Extending the Dataset:** While the current dataset offers substantial insights, there are 2 major categories this dataset can extend into: (a) including temporal resolution > 1 day and (b) extending geographical limitations. Future research could focus on including additional regions and countries. Incorporating new variables, such as metrics related to ignition points, latitudes, and longitudes could further enhance the model's capacity for generalization and improve its wildfire spread prediction accuracy.

2. **Prediction Algorithm Approaches:** The Vi-Net model showed robust performances, however, the increase in complications in the dataset will call for more nuanced models. Federated Learning(FL) techniques open up an innovative avenue for collaboration in training the model. Graph-based datasets/models also provide an interesting area to be explored if the prediction results are intended to be used for path-planning applications.

3. **Path Planning Algorithm Approaches:** While the A* gives a huge advantage in calculating path using considerably fewer computational resources and guaranteeing optimal path, there is a scope for innovation by employing ML techniques. Reinforcement Learning(RL) is another avenue for the self-training model to find a path given the constraints. This approach is particularly useful in navigation for unmanned and autonomous vehicles sector.

4. **Integration of Real-Time Data Streams:** Incorporating real-time data from satellite imagery, weather sensors, and ground-based monitoring stations instead of historical data can improve the timeliness and accuracy of predictions. Dynamic updates to the model could better reflect rapidly changing environmental conditions.

5. **Emergency Response Systems Enhancement:** The present integration of the OA* algorithm gives an innovative approach to utilize fire spread prediction in near real-time for various mitigation activities in an assumed road-free environment. Integrating the same results with publically available map APIs such as OpenStreetMap and Google Maps API provides a very valuable area to explore in terms of navigating routes in a real-world scenario. These efforts would provide much better results for mitigation activities, such as evacuation planning in WUI areas. Also, integrating predictive models with real-time decision-making tools can improve emergency response.

6. **Interdisciplinary Collaboration:** To make the model more practical and impactful, collaboration with government bodies, ecologists, and fire management professionals could be explored. This interdisciplinary strategy would help bridge the gap between technological innovation and its application in developing effective forest management policies.

7. **Explainability and Interpretability:** Improving the interpretability of model predictions by integrating explainable AI (XAI) techniques can aid in understanding why a model predicts a specific fire spread pattern. This assists researchers in identifying primary causes of fire spread in a region and helps stakeholders build trust in prediction results and decision-making.

References

[1] Frédéric Allaire, Vivien Mallet, and Jean-Baptiste Filippi. Emulation of wildland fire spread simulation using deep learning. *Neural networks*, 141:184–198, 2021.

[2] Patricia L. Andrews. *BEHAVE: Fire behavior prediction and fuel modeling system-BURN Subsystem*. US Department of Agriculture, Forest Service, Intermountain Research Station, 1986.

[3] Patricia L. Andrews. The rothermel surface fire spread model and associated developments: A comprehensive explanation. 2018.

[4] Henintsoa S Andrianarivony and Moulay A Akhloufi. Machine learning and deep learning for wildfire spread prediction: A review. *Fire*, 7(12):482, 2024.

[5] David M Bowman et al. Fire in the earth system. *Science*, 324(5926):481–484, 2020.

[6] Canadian Climate Institute. Fact sheet: Climate change and wildfires, July 2024. Accessed: 2024-12-30.

[7] National Interagency Fire Center. Annual wildfire summary report, 2021. Report.

[8] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4):834–848, 2017.

[9] Mark A. Cochrane. *Tropical Fire Ecology*. Springer, 2021.

[10] Jonathan D Coop, Sean A Parks, Camille S Stevens-Rumann, Scott M Ritter, and Chad M Hoffman. Extreme fire spread events and area burned under recent and future climate in the western usa. *Global Ecology and Biogeography*, 31(10):1949–1959, 2022.

[11] Jonathan D. Coop, Sean A. Parks, Camille S. Stevens-Rumann, Scott M. Ritter, and Chad M. Hoffman. Extreme fire spread events and area burned under recent and future climate in the western usa. *Global Ecology and Biogeography*, 31(10):1949–1959, April 2022.

[12] Paulo Cortez and Anabela Morais. Using svm for forest fire prediction. *Journal of Environmental Management*, 2019.

[13] E.W. DIJKSTRA. A note on two problems in connexion with graphs. *Numerische Mathematik*, 1:269–271, 1959.

[14] Alexey Dosovitskiy. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

[15] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale, 2021.

[16] Mark Finney. An overview of flammap fire modeling capabilities. 2006.

[17] Mark A. Finney. *FARSITE: Fire Area Simulator-model development and evaluation.* 1998.

[18] Mark A Finney. An overview of flammap fire modeling capabilities. In *In: Andrews, Patricia L.; Butler, Bret W., comps. 2006. Fuels Management-How to Measure Success: Conference Proceedings. 28-30 March 2006; Portland, OR. Proceedings RMRS-P-41. Fort Collins, CO: US Department of Agriculture, Forest Service, Rocky Mountain Research Station. p. 213-220*, volume 41, 2006.

[19] Jack Fitzgerald, Ethan Seefried, James E Yost, Sangmi Pallickara, and Nathaniel Blanchard. Paying attention to wildfire: Using u-net with attention blocks on multimodal data for next day prediction. In *Proceedings of the 25th International Conference on Multimodal Interaction*, ICMI '23, page 470–480, New York, NY, USA, 2023. Association for Computing Machinery.

[20] Zhihan Gao, Xingjian Shi, Hao Wang, Yi Zhu, Yuyang Bernie Wang, Mu Li, and Dit-Yan Yeung. Earthformer: Exploring space-time transformers for earth system forecasting. *Advances in Neural Information Processing Systems*, 35:25390–25403, 2022.

[21] Sebastian Gerard, Yu Zhao, and Josephine Sullivan. Wildfirespreadts: A dataset of multi-modal time series for wildfire spread prediction. *Advances in Neural Information Processing Systems*, 36:74515–74529, 2023.

[22] Rafik Ghali, Moulay A Akhloufi, Marwa Jmal, Wided Souidene Mseddi, and Rabah Attia. Wildfire segmentation using deep vision transformers. *Remote Sensing*, 13(17):3527, 2021.

[23] Paolo Grasso and Mauro S Innocente. Physics-based model of wildfire propagation towards faster-than-real-time simulations. *Computers & Mathematics with Applications*, 80(5):790–808, 2020.

[24] Maxfield E Green, Karl Kaiser, and Nat Shenton. Modeling wildfire perimeter evolution using deep neural networks. *arXiv preprint arXiv:2009.03977*, 2020.

[25] Jessica Halofsky, David Peterson, and Brian Harvey. Changing wildfire, changing forests: the effects of climate change on fire regimes and vegetation in the pacific northwest, usa. *Fire Ecology*, 16:4, 12 2020.

[26] Jessica E Halofsky, David L Peterson, and Brian J Harvey. Changing wildfire, changing forests: the effects of climate change on fire regimes and vegetation in the pacific northwest, usa. *Fire Ecology*, 16(1):1–26, 2020.

[27] Peter E Hart, Nils J Nilsson, and Bertram Raphael. A formal basis for the heuristic determination of minimum cost paths. *IEEE transactions on Systems Science and Cybernetics*, 4(2):100–107, 1968.

[28] Luis Hernández Encinas, Sara Hoya White, Angel Martín del Rey, and Gerardo Rodríguez Sánchez. Modelling forest fire spread using hexagonal cellular automata. *Applied Mathematical Modelling*, 31:1213–1227, 2007.

[29] Fantine Huot, R. Lily Hu, Nita Goyal, Tharun Sankar, Matthias Ihme, and Yi-Fan Chen. 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, 2022.

[30] Fantine Huot, R Lily Hu, Nita Goyal, Tharun Sankar, Matthias Ihme, and Yi-Fan Chen. 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, 2022.

[31] Fantine Huot, R Lily Hu, Matthias Ihme, Qing Wang, John Burge, Tianjian Lu, Jason Hickey, Yi-Fan Chen, and John Anderson. Deep learning models for predicting wildfires from historical remote-sensing data. *arXiv preprint arXiv:2010.07445*, 2020.

[32] Piyush Jain, Sean C. P. Coogan, Sriram Ganapathi Subramanian, Mark Crowley, Steve Taylor, and Mike D. Flannigan. A review of machine learning applications in wildfire science and management. *CoRR*, abs/2003.00646, 2020.

[33] Ioannis Karafyllidis and Adonios Thanailakis. A model for predicting forest fire spreading using cellular automata. *Ecological Modelling*, 99:87–97, 1997.

[34] Fadoua Khennou and Moulay A Akhloufi. Improving wildland fire spread prediction using deep u-nets. *Science of Remote Sensing*, 8:100101, 2023.

[35] Spyridon Kondylatos, Ioannis Prapas, Gustau Camps-Valls, and Ioannis Papoutsis. Mesogeos: A multi-purpose dataset for data-driven wildfire modeling in the mediterranean. *Advances in Neural Information Processing Systems*, 36, 2024.

[36] Remi Lam, Alvaro Sanchez-Gonzalez, Matthew Willson, Peter Wirnsberger, Meire Fortunato, Ferran Alet, Suman Ravuri, Timo Ewalds, Zach Eaton-Rosen, Weihua Hu, et al. Learning skillful medium-range global weather forecasting. *Science*, 382(6677):1416–1421, 2023.

[37] Yudong Li, Zhongke Feng, Shilin Chen, Ziyu Zhao, and Fengge Wang. Application of the artificial neural network and support vector machines in forest fire prediction in the guangxi autonomous region, china. *Discrete Dynamics in Nature and Society*, 2020(1):5612650, 2020.

[38] Mohammad Marjani, Seyed Ali Ahmadi, and Masoud Mahdianpari. Firepred: A hybrid multi-temporal convolutional neural network model for wildfire spread prediction. *Ecological Informatics*, 78:102282, 2023.

[39] Mohammad Marjani, Masoud Mahdianpari, and Fariba Mohammadimanesh. Cnn-bilstm: A novel deep learning model for near-real-time daily wildfire spread prediction. *Remote Sensing*, 16(8):1467, 2024.

[40] Natasha Markuzon and Stephan Kolitz. Data driven approach to estimating fire danger from satellite images and weather information. In *2009 IEEE Applied Imagery Pattern Recognition Workshop (AIPR 2009)*, pages 1–7, 2009.

[41] Max A. Moritz et al. Learning to coexist with wildfire. *Nature*, 515:58–66, 2014.

[42] Tung Nguyen, Johannes Brandstetter, Ashish Kapoor, Jayesh K Gupta, and Aditya Grover. Climax: A foundation model for weather and climate. *arXiv preprint arXiv:2301.10343*, 2023.

[43] Ioannis Prapas, Nikolaos-Ioannis Bountos, Spyros Kondylatos, Dimitrios Michail, Gustau Camps-Valls, and Ioannis Papoutsis. Televit: Teleconnection-driven transformers improve subseasonal to seasonal wildfire forecasting. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3754–3759, 2023.

[44] David Radke, Anna Hessler, and Dan Ellsworth. Firecast: Leveraging deep learning to predict wildfire spread. In *IJCAI*, pages 4575–4581, 2019.

[45] TS Rajalakshmi and R Senthilnathan. Dataset and performance metrics towards semantic segmentation. *International Journal of Engineering and Management Research*, 13(1):40–49, 2023.

[46] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation, 2015.

[47] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pages 234–241. Springer, 2015.

[48] Hugh D. Safford, Alison K. Paulson, Zachary L. Steel, Derek J. N. Young, and Rebecca B. Wayman. The 2020 california fire season: A year like no other, a return to the past or a harbinger of the future? *Global Ecology and Biogeography*, 31(10):2005–2025, April 2022.

[49] Hugh D Safford, Alison K Paulson, Zachary L Steel, Derek JN Young, and Rebecca B Wayman. The 2020 california fire season: A year like no other, a return to the past or a harbinger of the future? *Global Ecology and Biogeography*, 31(10):2005–2025, 2022.

[50] Dmitrii Shadrin, Svetlana Illarionova, Fedor Gubanov, Ksenia Evteeva, Maksim Mironenko, Ivan Levchunets, Roman Belousov, and Evgeny Burnaev. Wildfire spreading prediction using multimodal data and deep neural network approach. *Scientific Reports*, 14(1):2606, 2024.

[51] Samriddhi Singla, Ayan Mukhopadhyay, Michael Wilbur, Tina Diao, Vinayak Gajjewar, Ahmed Eldawy, Mykel Kochenderfer, Ross Shachter, and Abhishek Dubey.

Wildfiredb: An open-source dataset connecting wildfire occurrence with relevant determinants. In *NeurIPS Thirty-fifth Annual Conference on Neural Information Processing Systems*, 2021.

[52] Qi Tong and Thomas Gernay. Mapping wildfire ignition probability and predictor sensitivity with ensemble-based machine learning. *Natural Hazards*, 119(3):1551–1582, 2023.

[53] A Vaswani. Attention is all you need. *Advances in Neural Information Processing Systems*, 2017.

[54] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need, 2023.

[55] Zhiyong Wang and Sisi Zlatanova. Safe route determination for first responders in the presence of moving obstacles. *IEEE transactions on intelligent transportation systems*, 21(3):1044–1053, 2019.

[56] Zhiyong Wang, Sisi Zlatanova, and Peter van Oosterom. Path planning for first responders in the presence of moving obstacles with uncertain boundaries. *IEEE Transactions on Intelligent Transportation Systems*, 18(8):2163–2173, 2017.

[57] Zili Wang, Binbin He, and Xiaoying Lai. Balanced random forest model is more suitable for wildfire risk assessment. In *IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium*, pages 3596–3599. IEEE, 2022.

[58] Qing Xu, Shisheng Feng, Qun Sun, Xinming Zhu, Ruoxu Chen, Xiaoyang Lihua, and Beibei Wu. A method of planning disaster emergency rescue paths in road-free environment. *Computational Intelligence and Neuroscience*, 2022(1):2987852, 2022.

[59] Yiqing Xu, Dianjing Li, Hao Ma, Rong Lin, and Fuquan Zhang. Modeling forest fire spread using machine learning-based cellular automata in a gis environment. *Forests*, 13(12):1974, 2022.

[60] Suwei Yang, Massimo Lupascu, and Kuldeep S Meel. Predicting forest fire using remote sensing data and machine learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 14983–14990, 2021.

[61] Qing Zhu, Fa Li, William J Riley, Li Xu, Lei Zhao, Kunxiaojia Yuan, Huayi Wu, Jianya Gong, and James Randerson. Building a machine learning surrogate model for wildfire activities within a global earth system model. *Geoscientific Model Development*, 15(5):1899–1911, 2022.

[62] Katelyn Zigner, Leila MV Carvalho, Seth Peterson, Francis Fujioka, Gert-Jan Duine, Charles Jones, Dar Roberts, and Max Moritz. Evaluating the ability of farsite to simulate wildfires influenced by extreme, downslope winds in santa barbara, california. *Fire*, 3(3):29, 2020.

[63] Yufei Zou et al. Attention-based wildland fire spread modeling using fire-tracking satellite observations. *Fire*, 6:289, 2023.

[64] Yufei Zou, Mojtaba Sadeghi, Yaling Liu, Alexandra Puchko, Son Le, Yang Chen, Niels Andela, and Pierre Gentine. Attention-based wildland fire spread modeling using fire-tracking satellite observations. *Fire*, 6(8):289, 2023.

Glossary

$PM_{2.5}$ Fine particulate matter with a diameter less than 2.5 micrometers, which is harmful to human health and the environment 1, 84

A* algorithm A* (A-star) is a pathfinding and graph traversal algorithm that finds the shortest path by combining the actual cost to reach a node and the estimated cost to the goal. 34, 50, 51, 53, 54, 84

August Complex The largest individual wildfire in California's history, which burned over 400,000 hectares in 2020 1, 84

biodiversity loss The decline or extinction of various species, often due to habitat destruction caused by events such as large-scale wildfires 1, 84

causality The relationship between cause and effect, critical for understanding how environmental and meteorological factors influence wildfire dynamics 21, 23, 84

class distribution The proportion of instances in each class relative to the total dataset 30, 84

class imbalance A situation in machine learning where one class is significantly overrepresented compared to others, leading to biased predictions 30, 31, 84

climate change A long-term alteration in global or regional climate patterns, often attributed to anthropogenic activities, influencing wildfire frequency and severity 1, 84

computer A programmable machine that receives input data, stores and manipulates the data, and provides formatted output 84

crop size The dimensions of the randomly selected patch extracted from the original image during random cropping 84

deep learning (DL) 2, 3, 74, 84

Dice loss A loss function for image segmentation tasks that maximizes the Dice coefficient, emphasizing overlap between predicted and actual regions 84

environmental variables Factors like wind speed, temperature, and vegetation indices that influence the spread of wildfires 20, 21, 84

False Negatives (FN) A prediction error where the model fails to predict a positive instance, classifying it as negative 48, 84

False Positives (FP) A prediction error where the model incorrectly predicts a negative instance as positive 48, 84

filtering invalid data The process of excluding data points with uncertain or missing labels to ensure model training on reliable and complete data 30, 84

fire mask A dataset layer that identifies areas with active fire, no fire, or uncertain regions, often derived from remote sensing data 28, 29, 84

fire regime The pattern, frequency, and intensity of fires in a particular ecosystem, which can be influenced by natural and human factors 1, 84

fire spread prediction The process of forecasting the trajectory and behavior of wildfires using environmental, meteorological, and topographical factors 2, 84

fire suppression Human interventions aimed at extinguishing or controlling wildfires, often resulting in ecological imbalances over time 1, 84

Focal Tversky loss A variant of Tversky loss that focuses on hard-to-classify samples by further penalizing misclassified regions 84

generalization The ability of a machine learning model to perform well on unseen data by avoiding overfitting on the training dataset 28, 29, 84

Google Earth Engine (GEE) A cloud-based geospatial analysis platform used for processing large-scale environmental data 24, 26, 29, 84

GPWv4 Gridded Population of the World version 4, a dataset providing population density estimates globally 84

GRIDMET A gridded surface meteorological dataset providing weather and drought data at high resolution 27, 84

heuristic A heuristic is a problem-solving approach that uses practical shortcuts, approximations, or intuitive judgments to find solutions efficiently when exact methods are impractical. 50, 51, 53, 54, 84

input tensor A multi-dimensional array representing input data with dimensions for height, width, and feature channels, typically denoted as $\mathbf{X} \in \mathbb{R}^{H \times W \times C}$ 84

loss function A mathematical function used in machine learning to measure the difference between predicted and actual labels 57–59, 61, 84

machine learning (ML) A subset of artificial intelligence that uses statistical methods to enable machines to improve tasks through experience 2, 24, 28, 84

meteorological factors Weather-related variables such as wind speed, temperature, and humidity that influence wildfire behavior 2, 84

modified A* algorithm A heuristic-based pathfinding algorithm modified to account for wildfire spread predictions to generate safe evacuation routes 69, 70, 72, 75, 84

multimodal dataset A dataset that integrates data from multiple sources such as satellite imagery, weather station data, and environmental sensors 84

multivariate dataset A dataset comprising multiple variables or features, often used to model complex systems with interdependent factors 20, 84

Next Day Wildfire Spread (NDWS) A dataset designed to train models to predict the spread of wildfires using multimodal features 24, 84

noise augmentation A data augmentation technique where random noise is added to input data to improve model robustness and generalization 28, 29, 84

overfitting A scenario in which a machine learning model learns patterns specific to the training data but performs poorly on new, unseen data 28, 29, 84

random cropping A data augmentation technique where a random portion of the input image is extracted to introduce variability and prevent overfitting 28, 30, 84

random cropping function A function $f_{crop}(\mathbf{X}, a)$ that extracts a square patch of size $a \times a$ from an input tensor $\mathbf{X}$ at a random location 84

remote sensing The acquisition of information about an object or phenomenon without making physical contact, often using satellite or aerial imagery 2, 24, 26, 84

robustness The ability of a machine learning model to maintain performance despite small variations or noise in the input data 29, 84

segmentation A process in computer vision to divide an image into meaningful parts or regions, such as binary segmentation for predicting fire presence or absence 21, 84

semantic segmentation A computer vision task that classifies each pixel in an image into predefined categories, commonly used in spatial data analysis like wildfire spread prediction 34, 84

Shuttle Radar Topography Mission (SRTM) A mission that provided high-resolution digital elevation models, often used for topographical analysis 27, 84

spatial-temporal dynamics The interaction of spatial and temporal factors that govern the behavior and spread of wildfires over time 84

spatial-temporal dynamics The interplay between spatial and temporal factors in a system, such as wildfire spread across regions over time 22, 84

standardization A feature scaling method that centers data around its mean and scales it to unit standard deviation, often used to stabilize machine learning model training 28, 29, 84

tensor A multi-dimensional array used for data representation in machine learning and computer vision tasks. 84

TensorFlow An open-source machine learning framework for developing and training complex models 28, 84

TFRecord A binary file format for storing large datasets efficiently, optimized for TensorFlow models 28, 84

time-series data Data consisting of sequential observations where past states influence future outcomes, used in temporal modeling tasks 20, 21, 23, 84

True Negatives (TN) A correct prediction where a model identifies a negative instance as negative. 48, 84

True Positives (TP) A correct prediction where a model identifies a positive instance as positive. 48, 84

Tversky loss A loss function for imbalanced segmentation tasks, extending Dice loss by controlling the penalty for false positives and false negatives 84

U-Net A convolutional neural network model designed for biomedical image segmentation, excelling at capturing spatial details 22, 34, 36, 41, 43, 44, 46, 47, 59–61, 63–66, 75, 84

very large fires Wildfires exceeding 5000 hectares in size, often associated with significant ecological and economic impacts 1, 84

Vi-Net A hybrid model combining U-Net and Vision Transformers to enhance wildfire spread prediction by leveraging both local and global spatial context xi, 22, 23, 34, 36, 41–46, 50, 54, 57, 59–61, 63–66, 69, 75, 76, 84

VIIRS Visible Infrared Imaging Radiometer Suite, an instrument that provides high-resolution vegetation and surface observations 27, 84

ViT Vision Transformer, a deep learning model that processes image patches using transformer architecture, excelling at global contextual understanding 22, 23, 34, 36, 39–41, 43, 44, 46, 47, 59–61, 63–66, 75, 84

weighted cross-entropy loss A loss function that assigns higher weights to the minority class, mitigating class imbalance by penalizing the model more for misclassifications of the minority class 84

wind speed A meteorological factor representing the speed of wind, which significantly influences the spread and direction of wildfires 2, 84

Z-score standardization A method of standardization where each data point is transformed by subtracting the mean and dividing by the standard deviation 30, 84